Structural Studies of the Anti-HIV Human Protein APOBEC3G Catalytic Domain: A Dissertation

University of Massachusetts Medical School

Structural studies of the anti-HIV human protein APOBEC3G catalytic domain

A Dissertation Presented

By

Shivender Mohan Datt Shandilya

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences,
Worcester
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Biochemistry & Molecular Pharmacology

August 12, 2011

Structural studies of the anti-HIV human protein APOBEC3G catalytic domain

A Dissertation Presented

By

Shivender Mohan Datt Shandilya

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

---

Celia A. Schiffer, Ph.D., Thesis Advisor

Mohan Somasundaran, Ph.D., Thesis Advisor

William E. Royer, Ph.D., Member of Committee

David G. Lambright, Ph.D., Member of Committee

Dan Bolon, Ph.D., Member of Committee

Hiroshi Matsuo, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

---

Mary Munson, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

---

Anthony Carruthers, Ph.D.
Dean of the Graduate School of Biomedical Sciences

Biochemistry & Molecular Pharmacology

August 12, 2011

Table of Contents

List of Figures & Tables

List of Abbreviations

A3G	APOBEC3G
AFM	Atomic Force Microscopy
AID	Activation Induced Deaminase
AIDS	Acquired Immunodeficiency Syndrome
APOBEC	Apolipoprotein-B mRNA Editing Catalytic polypeptide
APOBEC3G	Apolipoprotein-B mRNA Editing Catalytic polypeptide #3 subfamily G
CA	Capsid
CBFβ	Core Binding Factor β subunit
CCR5	C-C chemokine Receptor type 5
CD4	Cluster of Differentiation 4
cDNA	Complementary DNA
CRL	Cullin RING Ligase
CTD	C-terminal Domain
CXCR4	C-X-C chemokine Receptor type 4
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ds-	double stranded–
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
GFP	Green Fluorescent Protein
HAART	Highly Active Anti-retroviral Therapy
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
HMM	High Molecular Mass
HSQC	Heteronuclear Single Quantum Coherence
HTS	High-throughput Screening
IC_{50}	Half maximal inhibitory concentration
IN	Integrase

LINE	Long INterspersed Element
LMM	Low Molecular Mass
LOPAC	Library of Pharmacologically Active Compounds
LTR	Long Terminal Repeat
NC	Nucleocapsid
NMR	Nuclear Magnetic Resonance
NNRTI	Non-nucleoside Reverse Transcriptase Inhibitor
NOE	Nuclear Overhauser Effect
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NTD	N-terminal Domain
PDB	Protein Data Bank
PI	Protease Inhibitor
PIC	Pre-integration Complex
PR	Protease
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
RT	Reverse Transcriptase
RTC	Reverse Transcription Complex
ss-	single stranded–
TLR	Toll-like Receptor
TRIM5α	Tripartite Motif containing Protein 5α
UDG	Uracil DNA Glycosylase
Vif	Virion Infectivity Factor

जीवन अस्थिर अनजाने ही
हो जाता पथ पर मेल कहीं
सीमित पग-डग, लम्बी मंज़िल
तय कर लेना कुछ खेल नहीं
दाएँ-बाएँ सुख-दुख चलते
सम्मुख चलता पथ का प्रमाद
जिस जिससे पथ पर स्नेह मिला
उस उस राही को धन्यवाद ।

साँसों पर अवलम्बित काया
जब चलते-चलते चूर हुई
दो स्नेह-शब्द मिल गए, मिली
नव स्फूर्ति थकावट दूर हुई
पथ के पहचाने छूट गए
पर साथ-साथ चल रही याद
जिस जिससे पथ पर स्नेह मिला
उस उस राही को धन्यवाद ।

जो साथ न मेरा दे पाए
उनसे कब सूनी हुई डगर
मैं भी न चलूँ यदि तो भी क्या
राही मर लेकिन राह अमर
इस पथ पर वे ही चलते हैं
जो चलने का पा गए स्वाद
जिस जिससे पथ पर स्नेह मिला
उस उस राही को धन्यवाद ।

कैसे चल पाता यदि न मिला
होता मुझको आकुल-अन्तर
कैसे चल पाता यदि मिलते
चिर-तृप्त अमरता-पूर्ण प्रहर
आभारी हूँ मैं उन सबका
दे गए व्यथा का जो प्रसाद
जिस जिससे पथ पर स्नेह मिला
उस उस राही को धन्यवाद ।

– शिव मंगल सिंह "सुमन"

Your only obligation
in any lifetime
is to be true to yourself
...

— Illusions : The Adventures of a Reluctant Messiah
[Richard Bach, 1977]

Lovingly dedicated to
Mama and Papa*ji*, for their unyielding support,
patience, love and countless sacrifices;
Ambika, the best sister any brother could ask for;
and Kamal, my dear wife,
for all her love and support
and the joy she brings into my life
every single day.

Acknowledgements

An effort of such magnitude as a doctoral thesis, though extraneously an individual endeavor, requires the help, guidance and support of numerous individuals and is thus, not just an individual accomplishment. I have been extremely lucky to have received the magnanimous support from many, many people to reach this far and sincerely thank them all.

First and most importantly, my Parents, whose love, encouragement and support throughout the years is central to all that I have ever accomplished. I thank them for their many, many sacrifices and their untiring efforts for my education. I salute them for their patience and thank them for their support as I have made way through one field of study to the next, leaving one perfectly good career for another, over the years. Thank you, first for the freedom and then for all the support, that has allowed me to make these bold decisions over the years and for that I am truly grateful.

With equal importance, I thank my wife, Kamalpreet, who I first met at UMMS, for her love and support through the good and the bad, the easy and the tough times and everything in between. Her untiring efforts, love and trust have made this work possible and are the source of my strength. I am thankful, every day, for having you by my side and for all you have done, and still do, to make life full of promise and joy.

A very special thanks to my sister, Ambika, for being there whenever I was in need of support, inspiration, guidance and at times, the right amount of pushing and prodding! Thank you, for being not just a sister, but an excellent friend over all these years and I would never have reached this far without your help and best wishes.

I am grateful for the many faceted assistance from my maternal uncle, Raghunath Kumar, who helped me every step of the way to consider, plan for and attain higher education in the U.S. It is only becasue of his constant help and support that I am here, and for all he has done I thank him sincerely.

I have been truly blessed to have many wonderful mentors over the years and the foremost is Dr. Celia Schiffer, my mentor during this Ph.D. program. Words may never be enough to describe my respect for her or my gratitude for the help, support and guidance she provided during this journey. Thank you for the opportunity, patience, encouragement and believing in me at times when I had trouble doing so. You are better than the best mentor I could have ever wished for and for that I am truly grateful.

I am indebted to Dr. Mary Munson, my TRAC and DEC chair, for the opportunity to "foster" in her lab as I started my research project, for teaching me the basics of laboratory bench work and most importantly, for inculcating a strong and rigorous work ethic in doing science. All of which has proved invaluable throughout the challenges of my project, and I am confident, will be indispensable throughout my career.

A sincere and heartfelt thanks to Dr. Mohan Somasundaran, my co-mentor, for his guidance, assistance and kind words of wisdom. I thank you for providing an "outsider's" viewpoint that is critical to the success of any project, especially one so specialized and focused at this level of research.

I am thankful to Dr. Bill Royer for his ever so patient guidance in decrypting the mysteries of X-ray diffraction crystallography techniques, theory and methods. His help, efforts and generosity in sharing his time and resources have made possible the data collection and analysis needed to solve the various crystal structures required for my project.

I would like to thank Dr. David Lambright and Dr. Dan Bolon for their mentorship and guidance as members of my TRAC and DEC committees. I am thankful for the many inputs over the years which proved essential to succeed at a decidedly demanding project.

I would like to acknowledge and thank all my collaborators for their help and assistance, especially Dr. Reuben Harris and Dr. Hiroshi Matsuo and their respective lab members at the University of Minnesota. Without their kind assistance this project would not

have been conceived or executed in its present form.

I am grateful to my past mentor, Dr. Richard Deth at Northeastern University, for introducing me to the world of structural biology and thus helping me discover my passion. I am very thankful to Dr. Michael Czech for convincing me to apply to UMMS when I was deciding to join a Ph.D. program and for his initial help and guidance as I started the program.

I am very grateful to all my colleagues in the Schiffer lab, past and present members, for their help and support over the years. A special thanks to Ellen Nalivaika for keeping the lab "functional", her guidance and help in tough cloning jobs and the wonderful collegial environment she makes possible by her unique mix of uncompromising civility and unyielding discipline. Sincere thanks to Dr. Madhavi Nalam for teaching me "hands-on" crystallography and the various methods and techniques essential to doing this work and successfully navigating the sea of crystallography knowledge. I am grateful for the help provided by Dr. Akbar Ali and Dr. Hong Cao during the many stages of this project.

Thank you, my fellow lablings, you know who you are. I thank you for sharing the ups and downs during my time in the lab, your selfless help, camaraderie and friendship that made each day in the lab a lot better than it would otherwise have been. And most of all, for the coffees, lunches, parties, teas and all the fun!

I have been lucky to be blessed with great friends and I am grateful for their help, support and the gift of friendship both at UMMS and outside. Thank you: Vijay, Amol, Vidya, Sushanth, Mandeep, Neeraj, Jainendra, Sriram, Nishant, Harish, Manian, Rajintha, Shamik, Jaspreet, Ramesh, Rajarshi, Smriti, Veena, Thilinie, Shuvasree and many, many others for everything over the years.

Thank you!

ABSTRACT

HIV/AIDS is a disease of grave global importance with over 33 million people infected world-wide and nearly 2 million deaths each year. The rapid emergence of drug resistance, due to viral mutation, renders anti-retroviral drug candidates ineffective with alarming speed and regularity. Instead of targeting mutation prone viral proteins, an alternative approach is to target host proteins that interact with viral proteins and are critical for the HIV life-cycle. APOBEC3G is a host anti-HIV restriction factor that can exert tremendous negative pressure by hypermutating the viral genome and has the potential to be a promising candidate for anti-retroviral therapeutic research.

The work presented in this thesis is focused on investigating the A3G catalytic domain structure and implications of various observed structural features for biological function. High-resolution crystal structures of the A3G catalytic domain were solved using data from macromolecular X-ray crystallographic experiments, revealing a novel intermolecular zinc coordinating motif unique to A3G. Major intermolecular interfaces observed in the crystal structure were investigated for relevance to biochemical activity and biological function.

Co-crystallization with a small-molecule A3G inhibitor, discovered using high-throughput screening assays, revealed a cysteine residue near the active site that is critical for inhibition of catalytic activity by catechol moieties. The serendipitous discovery of covalent interactions between this inhibitor and a surface cysteine residue led to further biochemical experiments that revealed the other cysteine, near the active site, to be critical for inhibition.

Computational modeling was used to propose a steric-hinderance based mechanism of action that was supported by mutational experiments. Structures of other human APOBEC3 homologs were modeled using *in-silico* methods examined for similarities and

differences with A3G catalytic domain crystal structures. Comparisons based on these homology models suggest putative structural features that may endow substrate specificity and other characteristics to the APOBEC3 family members.

Chapter I

Introduction

Chapter I
Introduction

1.1 HIV

1.1.1 AIDS and Global impact

Human Immunodeficiency Virus (HIV), discovered in 1983 as the causative agent of Acquired Immunodeficiency Syndrome (AIDS) (1–3), caused 1.8 million deaths due to AIDS in 2009 and currently infects an estimated 33.3 million people globally with over 7000 new infections per day or 2.8 million every year (4). The negative socioeconomic impact of this epidemic has hit hardest the developing and underdeveloped countries, particularly in Sub-Saharan Africa and Asia, extracting a substantial human price.

According to recent statistics, the incidence of new infections and fatalities from AIDS is on a downward trend in countries with reasonable access to preventive and therapeutic measures (4). Despite multi-billion dollar, multi-national efforts on education, prevention, therapeutic management and research, numerous challenges remain. The most significant being an absence of curative therapy due to rapid emergence of resistance to drugs that target HIV proteins (5–9). Drug resistance arises primarily from viral mutation patterns that lead to co-evolution of target viral proteins and their corresponding viral interaction partners or substrates, while disrupting binding to drugs. Coupled with persistence of latent HIV infection in host reservoir tissues, even in patients that may not progress to AIDS, drug resistance leads to failure of anti-HIV therapy. Novel therapeutic approaches, in order to remain effective, must be directed at targets that HIV depends on for its life-cycle but, being host proteins, are incapable of mutating and thus contributing to drug resistance (8, 10–13).

1.1.2 Viral life-cycle

HIV, a lentivirus of the retroviridae family, is an enveloped, positive-sense, single-stranded RNA virus (14, 15). The viral life-cycle may be broadly divided into six stages (Figure 1.1):

(i) **Fusion and Entry into the host cell:** HIV targets the host $CD4^+$ T-cells, macrophages and microglial cells. Entry is primarily mediated through interaction between the CD4 receptor and the gp120 subunit of the viral envelope glycoprotein (*env* or gp160). The α-chemokine receptor CXCR4 and the β-chemokine receptor CCR5, individually or together, act as co-receptors mediating viral entry(16). gp120 binds to the CD4 receptor with high affinity, subsequent binding with CXCR5/CCR5 co-receptor(s) trigger conformational changes that expose the gp41 subunit of *env* that contains a "fusion peptide" which is inserted into the host cell membrane. This insertion triggers a refolding of gp41 into an extended coiled-coil structure that assembles into a thermodynamically stable six-helix bundle. The formation energy of this highly stable structure forces the viral and host-cell membranes into close proximity, triggering a fusion of the two membranes and through a series of intermediate states, leads to formation of a "fusion pore" that allows the viral capsid to enter the host cell cytoplasm.

(ii) **Uncoating and Reverse transcription of the viral genomic RNA:** The viral capsid or core is a conical macromolecular assembly, 100-120 nM long and 50-60 nM wide at the broadest end. Nearly 1500 capsid monomeric subunits (*CA*) assemble into a fullerene cone formed by hexameric and pentameric structural sub-assemblies that arise from N-terminus-to-N-terminus (NTD-NTD) and C-terminus-to-C-terminus (CTD-CTD) interactions. In addition to the viral genomic RNA, it is comprised of *CA* (capsid) and *NC* (nucleocapsid), *RT* (reverse transcriptase), *IN* (integrase), *vpr* (viral protein R) and numerous host proteins packaged during viral assembly, such as APOBEC3G and Cyclophilin A. Reverse transcription of the single-stranded viral genomic RNA to double-stranded DNA, a prerequisite for integration

with the host genome, happens in the RTC (reverse transcription complex) composed of the capsid proteins, RT and viral genomic RNA and/or newly reverse transcribing DNA. The capsid is critical for delivering the RTC, also referred to as the PIC (pre-integration complex) following uncoating and reverse transcription, to the nuclear pore. Successful delivery of the uncoated, reverse transcribed viral genome to the nuclear pore is essential for integration (11).

(iii) **Nuclear import, Integration and Transcription:** Prior to nuclear transport and integration into the host genome, the reverse transcribed viral cDNA is associated with a tetramer of HIV integrase and other host proteins in the PIC. Other host proteins in the PIC include factors that prevent auto-integration (BAF), compact viral cDNA (HMGA1) and aid HIV integrase (LEDGF, hSNF5). The viral integrase is associated with LTR ends of the viral cDNA and drives the ends into close juxtaposition with host DNA followed by integration via strand transfer reactions (17). The elongation of RNA transcribed from the integrated viral genome is aided by binding of the HIV transcription transactivator protein *Tat* to the transactivating response region (*TAR*) in the newly transcribed RNA (18).

(iv) **Nuclear export and Translation:** The transcribed RNA may be exported from the host nucleus as unspliced (Gag, GagPol and genomic RNA), singly spliced (Env, Vif, Vpr, Vpu) or doubly spliced (Nef, Rev, Tat). The HIV protein, Rev, plays the a critical role in the export of unspliced RNA from the nucleus as it binds to the Rev response element (*RRE*) region in the RNA and prevents splicing. Once exported, the spliced RNA is translated by the host cellular machinery to viral proteins. The Gag and GagPol poly-proteins are precursors to viral structural proteins and enzymes respectively, all essential to assembly, budding and maturation of viral particles.

(v) **Assembly, Budding and Maturation:** The Gag poly-protein associates with the host cell membrane and initiates assembly of viral components, including genomic RNA and the ESCRT family of host proteins that are required for budding (TSG101, AIP1/ALIX) (15).

The HIV structural proteins matrix (*MA*), capsid (*CA*) and nucleocapsid (*NC*) are derived from the Gag poly-protein after cleavage by HIV protease (*PR*) following budding. This cleavage of Gag and GagPol poly-proteins is essential for viral particle maturation and subsequent host cell infection.

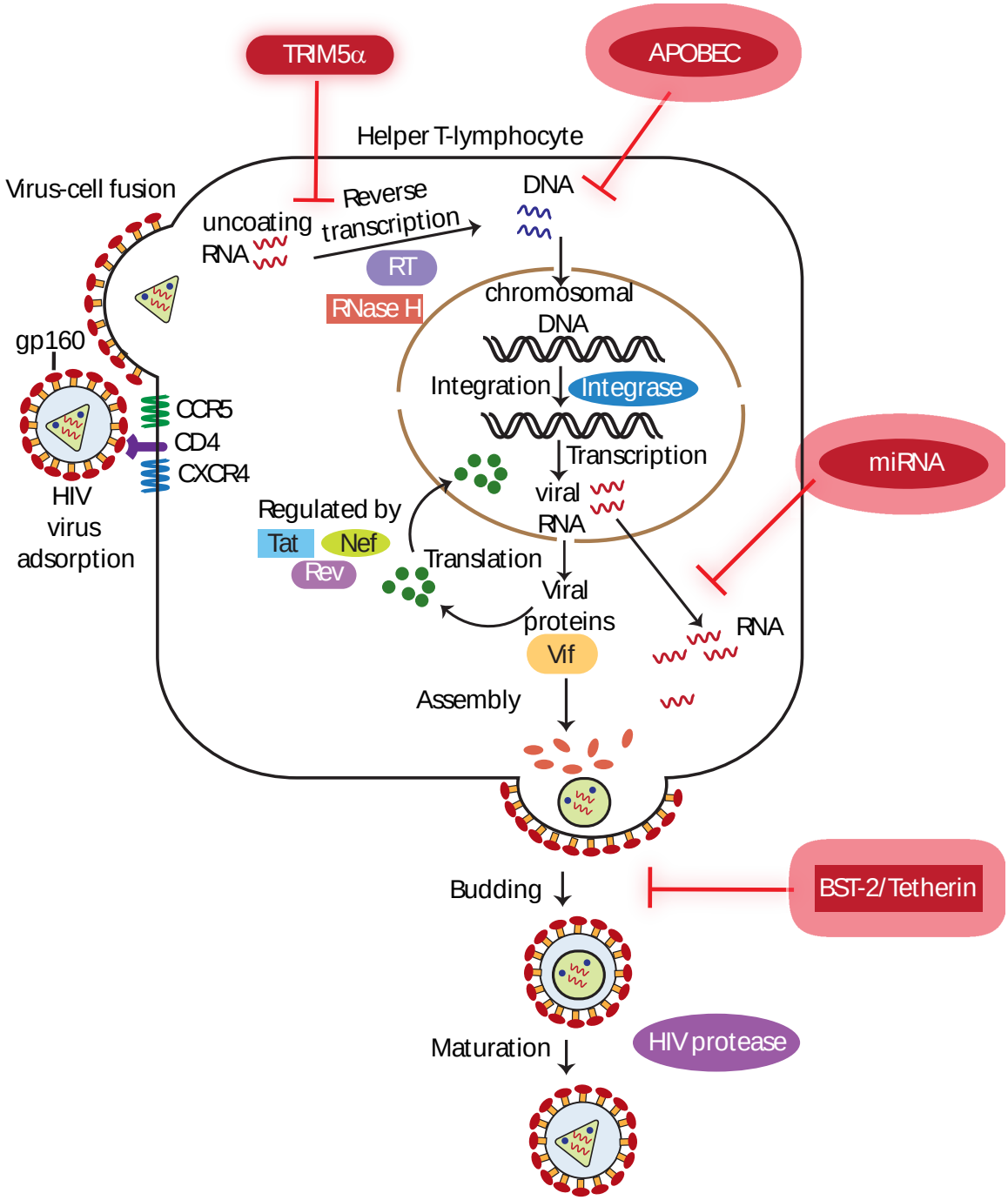

TRIM5α
APOBEC
Helper T-lymphocyte
Virus-cell fusion
uncoating
RNA
Reverse
transcription
RT
RNase H
DNA
chromosomal
DNA
Integration
Integrase
Transcription
viral
RNA
gp160
CCR5
CD4
CXCR4
HIV
virus
adsorption
Regulated by
Tat
Nef
Rev
Translation
Viral
proteins
Vif
RNA
miRNA
Assembly
Budding
BST-2/Tetherin
Maturation
HIV protease

Figure 1.1

Figure 1.1 : HIV life-cycle and host anti-HIV factors (red boxes). Detailed discussion under Subsection 1.1.2 : Viral life-cycle and Section 1.2 : Host anti-HIV factors. Based on : ref. (19).

1.1.3 Pathogenesis and disease

HIV infects the target hosts through contact of viral inoculum with host mucosal surfaces such as during unprotected sexual intercourse, by transfusion of infected blood (or derivatives such as plasma) and by mother-to-child transmission (*MTCT*) during pregnancy, parturition or breast feeding. $CD4^+$ T-cells are the primary viral hosts in the majority of infections whereas memory T-cells are long-term reservoirs of transcriptionally silent provirus. Initial acute HIV infection (acute retroviral syndrome) is characterized by a flu-like viremia within 4 to 10 days with symptoms including fever, night sweats, arthralgia/myalgia, lymphadenopathy, nausea, vomiting, diarrhea and unexplained weight-loss. During this acute phase, viral load in the peripheral blood may be as high as 1×10^6/mL to 10×10^6/mL. Most of the symptoms subside over 4 to 8 weeks as the viral infection becomes clinically latent. Seroconversion is indicated by the detection of anti-HIV antibodies by ELISA as the HIV p24 antigen drops to undetectable levels. The $CD4^+$ T-cell counts rebound but never approach pre-infection levels. On average, over the course of 8 to 10 years, the constant deterioration of the immune system ($CD4^+$, $CD8^+$ and other cells) due to HIV replication, results in AIDS. The diagnostic criteria for AIDS include a deteriorating $CD4^+$ count below 200/µL, emergence of opportunistic infections and neoplasms and the appearance of syncytia-forming (SI) phenotypes of HIV (20).

1.1.4 Therapeutic approaches and Limitations

Therapeutic agents have adverse effects

The earliest class of therapeutic agents used for treating HIV infections were Nucleoside Reverse Transcriptase Inhibitors (*NRTIs*: Zidovudine/Azidothymidine, Didanosine, Zalcitabine etc.) and Acyclic Nucleoside Phosphonates (Adefovir, Tenofovir, Cidofovir etc.). These drugs were characterized by intolerance and severe adverse reactions due to mitochondrial toxicity. Later, Non-Nucleoside Reverse Transcriptase Inhibitors (*NNR-*

TIs: Nevirapine, Efavirenz etc.), that bind non-competitively to RT were developed. These drugs had fewer side-effects than NRTIs but caused complications such as skin rashes and hepatotoxicity. Protease inhbitors (*PIs*: Saquinavir, Ritonavir, Nelfinavir, Darunavir etc.), target HIV protease and disrupt viral maturation. PIs are potent inhibitors of the hepatic cytochrome P450/CYP3A4 enzymes and can cause severe drug interaction issues. Other complications include PI associated lipodystrophy, hepatic transaminase elevation and sensory neuropathy (20).

Major limitations of therapeutic agents

Three classes of drugs mentioned above, NRTIs, NNRTIs and PIs, have been the backbone of Highly Active Anti-retroviral Therapy (HAART) in combination with the newer integrase and entry inhibitors discussed below. However, in addition to adverse effects, the greatest challenge to their efficacy and use arises from drug-resistance. Drug resistance, a change in molecular recognition, is caused by mutations in the HIV genome that occur as a result of error-prone reverse transcription by RT and possibly by the action of APOBEC3G, a cellular cytidine deaminase causing viral genomic G-to-A hypermutations(21). These mutations cause structural changes in the drug targets, such as HIV protease, that lead to a reduction or loss of binding/affinity to the drug molecule(s) without significant changes to the target's biological activity (22–25). Mutations in the drug target (protease) may be accompanied by mutations in the viral substrates (Gag/GagPol), a phenomenon referred to as "co-evolution" that derails drug binding while allowing normal or near normal target/substrate interaction (26, 27).

Recent anti-HIV drugs are targeted at HIV integrase/IN (Raltegravir), HIV gp41 (Enfuvirtide), HIV Gag protein (Bevirimat) or the host cell CCR5 receptor (Maraviroc) with the aim to minimize adverse effects and mitigate drug resistance (7, 28, 29). However, drug resistance has emerged rapidly in almost all cases, sometimes even during early phases of

clinical trials. This rapid emergence of drug resistant virus populations quickly renders ineffective anti-retroviral drugs, resulting in an absence of a curative treatment.

1.2 Host anti-HIV factors

During the course of viral life-cycle, HIV interacts with numerous host proteins to subvert the cellular machinery and pathways to its advantage (28). Some of these host proteins exert a deleterious effect on HIV and have evolved as "restriction factors" against retroviruses. The virus attempts to evade or degrade these proteins during replication using viral and other host proteins. Given the rapid emergence of drug resistance amongst viral proteins, these host factors are crucial therapeutic targets that may be leveraged by enhancing their activation or reducing their HIV driven degradation, in principle, without risking emergence of drug resistance. The principal host targets in this regard are the tripartite motif protein TRIM5α, interferon induced protein BST-2/Tetherin, cellular miRNAs and the cytidine deaminases APOBEC3G/3F discussed below.

TRIM5α

TRipartite Interaction Motif-containing 5 isoform alpha is a human cytoplasmic protein that as a part of the anti-retroviral innate immune response inhibits HIV infection (30–32). TRIM5α has been shown to act by sensing the hexameric viral capsid lattice, much like a pattern recognition receptor (PRR), binding to it and causing its accelerated disintegration during viral entry (33, 34). This rapid uncoating of the viral genome disrupts RT activity effectively preventing generation of proviral DNA (35).

BST-2/Tetherin

Bone marrow STromal cell antigen 2, a membrane targeted innate immune response protein, was recently discovered to adversely impact virion budding, an effect that was demon-

strated to be negated by the viral vpu protein (36). The budding virions appear unable to detach from the host cell membrane and/or each-other due to the "simple and direct" effect of BST-2, also known as "Tetherin" as it tethers the virions to the infected host cell (37, 38).

Cellular miRNAs

MicroRNA molecules (miRNAs) are 19-25 nucleotide long, single stranded RNA molecules that play a critical role in the regulation of cellular protein expression and function (39, 40). Recent evidence shows they may play important roles in regulating HIV infection directly by acting against HIV nef (miR-29a) or indirectly by repressing cellular Cyclin T1 (miR-198) to decrease viral infectivity (41–45).

1.2.1 APOBEC proteins

The APOBEC proteins are named after the first member of the family APOBEC1 or "apolipoprotein B mRNA editing enzyme, catalytic polypeptide #1", a cytidine deaminase responsible for post-transcriptional modification of the apolipoprotein B mRNA changing a CAA codon into a stop codon (46, 47). Other members of this family in humans include APOBEC2, APOBEC3 (seven subfamily members: A, B, C, D/E, F, G and H) and APOBEC4, all characterized as cytidine deaminases (48–50).

A short note on cytidine deaminases

Deaminases are a class of enzymes found in fungi, archea and metazoa that act on free nucleosides or nucleotides (cytosine/cytidine) or on polynucleotide sequences containing adenines: such as ADAR (Adenosine Deaminase RNA-specific) that acts on double stranded RNA or ADAT (Adenosine Deaminase t-RNA specific) that acts on tRNA. Similar enzymes acting on cytidines in polynucleotides are AID (Activation Induced Deminase) that acts

on DNA and APOBECs (Apolipoprotein-B mRNA-Editing Catalytic Polypeptide) that act on RNA/DNA. The defining characteristic of these enzymes is the presence of one or more [H/C]XE and PCXXC motifs that coordinate a zinc atom at the enzyme active site (51), although exceptions exist (52). Antibody class switch recombination and IgG differentiation result from the action of Activation Induced Deaminase (AID) (53, 54) with recent evidence also suggesting a role in DNA demethylation and epigenetic control (55). APOBEC3 proteins help defend against retroviruses such as HIV (A3G) and DNA viruses like AAV (A3A) (56, 57). Adenosine Deaminases Acting on RNA or ADARs convert adenosine to inosine in host and virus encoded double stranded RNA molecules that is read as guanosine and thus modulate protein expression and activity 58. ADATs (Adenosine Deaminase Acting on tRNA), on the other hand are known to act on the wobble position of $tRNA^{Ala}$:34 and may be required for cell cycle regulation (59).

Cytidine deaminases have been the focus of various drug development efforts over the years. Human cytidine deaminase (hCDA) was one of the earliest targets to be investigated because it was shown to degrade the anticancer drugs cytosine arabinoside and 5-azacytidine (60–63). Non-mammalian cytosine deaminases, in combination with small molecules such as 5-fluorocytosine, have been investigated as suicide inhibitors for use in anti-cancer gene therapy efforts (64). Recent discoveries of AID and APOBEC proteins and their potential for generating cancer and autoimmune disorders by indiscriminate cytidine deamination in the genome has focused interest on them (65–69). APOBEC3G, as described later, is of great interest in the pathophysiology of HIV infection as it inhibits the virus by cytidine deamination and is itself targeted by the HIV vif protein for degradation via the host cell proteasome pathway (19, 70–76).

The cytidine deamination reaction

The cytidine deamination reaction plays an integral role in the immune system. Cytidine deamination refers to the biochemical conversion of a (deoxy)cytidine (dC/C) into a (deoxy)uridine (dU/U) by enzymatic means. This conversion primarily occurs in the context of single stranded polynucleotides (DNA/RNA), but may also act on other substrates such as cytosine nucleoside analogs (60–63). The deamination reaction may be represented as a six step process that requires, in addition to the enzyme and substrate, one water molecule and releases one molecule of ammonia (deamination) per reaction (77, 78). The reaction starts with nucleophilic attack by the zinc hydroxide oxygen at C4 ring position of the bound substrate dC/C, followed by protonation of the N3 ring nitrogen by OE1 of the catalytic Glu residue, resulting in formation of the first tetrahedral intermediate. Proton transfer from the zinc activated hydroxyl group to the leaving NH_2^- group via OE2 of the catalytic Glu results in formation of the second tetrahedral intermediate. Following this step NH_3 escapes, leaving behind the enzyme-product complex that, after releasing the product dU/U, accepts a new water molecule that is deprotonated by the Glu carboxylate group and binds the Zn atom. A new substrate dC/C may now enter the active site (and the reaction cycle repeats) (Figure 1.2).

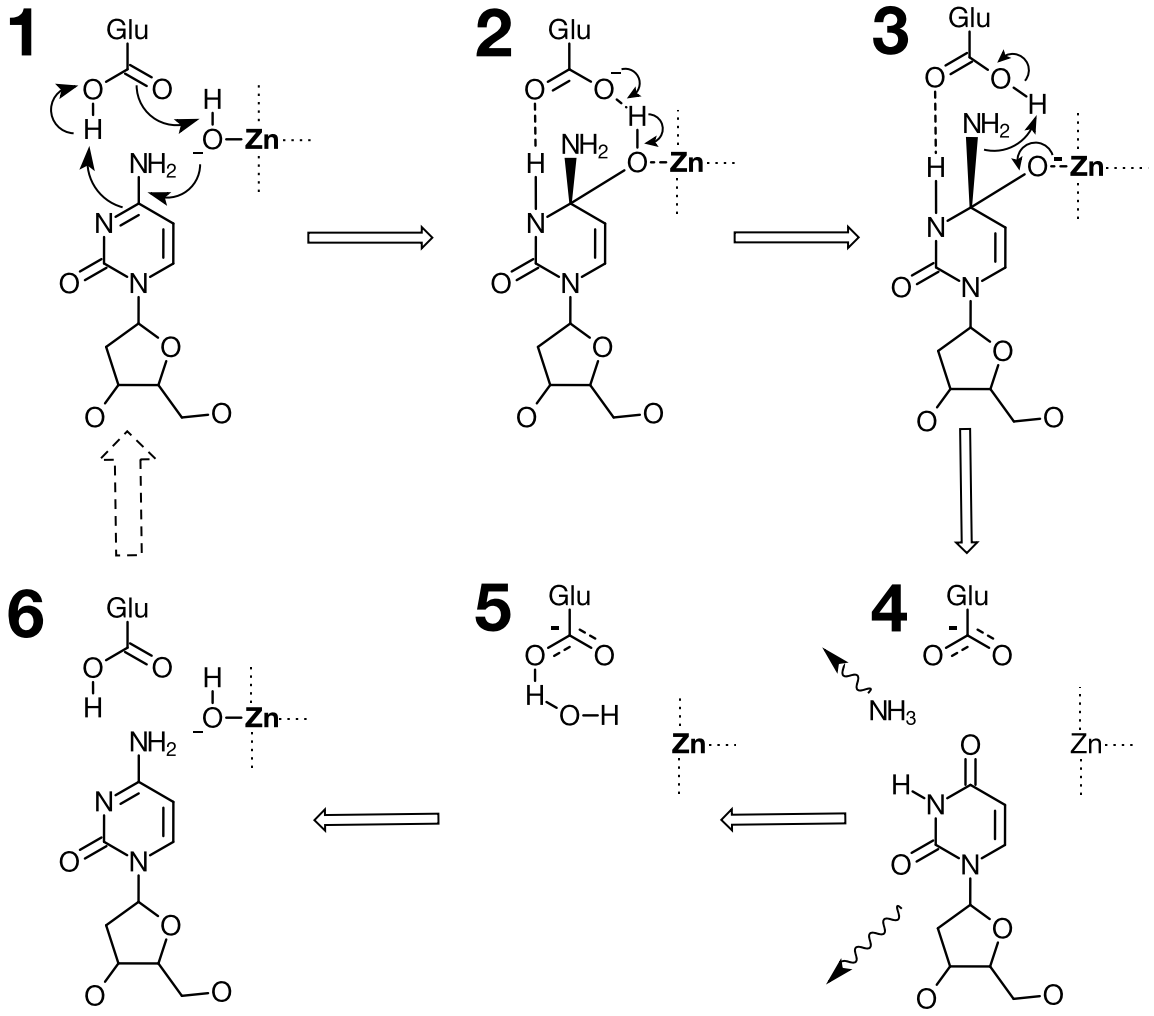

1
Glu
O
O
H
H
O–Zn
NH2
N
O
N
O
O
O
2
Glu
O
O
H
H
NH2
O--Zn
N
O
N
O
O
3
Glu
O
O
H
NH2
H
O--Zn
N
O
N
O
O
4
Glu
O
O
NH3
Zn
H
N
O
N
O
O
5
Glu
O
O
H
O–H
Zn
6
Glu
O
O
H
H
O–Zn
NH2
N
O
N
O
O

Figure 1.2

Figure 1.2 : Cytidine deamination in six steps: **(1)** Enzyme-Substrate complex (deoxycytidine). Glu is the catalytic glutamate residue (E259 in A3G), Zn is the active site zinc atom (coordinated by two Cys and one His in A3G). **(2)** First tetrahedral intermediate. **(3)** Second tetrahedral intermediate. **(4)** Enzyme-Product complex (deoxyuridine). **(5)** New water molecule approaches the active site. **(6)** New substrate molecule enters the active site. Based on : ref. (77, 78).

hCDA and the AID/APOBEC family

hCDA — The human cytidine deaminase gene is located on chromosome 1 and an evolutionarily conserved member of the pyrimidine salvage pathway. In addition to acting on cytidine, hCDA acts on nucleoside-analogue anti-viral and anti-cancer agents thus an important drug target for inhibitory adjuvants (60). Metabolism of anti-cancer agents by hCDA is also being explored as a means to alleviate the toxic effects of these agents on myeloid stem cells by using hCDA as a transgene via gene therapy approaches (79).

AID — The Activation Induced Deaminase expressed by the Activation Induced Cytidine Deaminase (AICDA) gene locus on chromosome 12 in humans, has been extensively investigated for its role in antibody diversification by somatic hypermutation (SHM) and class switch recombination (CSR) (54, 80). Recent investigations into AID function reveal a role in DNA demethylation driven epigenetic reprogramming by deamination of methylcytosines (81, 82).

APOBEC1 — The APOBEC1 protein, encoded by a gene locus on chromosome 12, was the first *polynucleotide* cytidine deaminase to be identified (47, 83). APOBEC1 edits the apolipoprotein B mRNA transcript by deaminating the cytidine at position 6666, converting a CAA codon into a stop codon, leading to production of truncated form of the apoB lipoprotein called apoB48 that is required for lipid transport in chylomicrons (84). Non-human APOBEC1 proteins may be involved in DNA cytidine deamination and defense against hepadnavirus (mHBV) infections (85). Additional evidence points to DNA demethylation activity in vertebrates (82, 86)

APOBEC2 — Expressed specifically in cardiac and skeletal muscles (49), APOBEC2 is encoded by a gene locus on chromosome 6 and although no DNA or RNA editing activity has been ascribed, APOBEC2 has been demonstrated to be essential for muscle development and maintenance in mice models (87) and play a role in vertebrate DNA demethylation

(86).

APOBEC3 — The APOBEC3 sub-family of proteins are encoded by a gene locus on chromosome 22 and discussed in the next section.

APOBEC4 — Bioinformatics methods have predicted the existence of an APOBEC4 gene locus on chromosome 1, however no functional significance has been ascribed to it as yet (50).

APOBEC5 — The existence of a new member of the APOBEC family, APOBEC5, was recently predicted by bioinformatics approaches in reptilian genomes (88) and demonstrated by experimental means in the bacterium *Xanthomonas oryzae* (89, 90). The presence and significance of this enzyme in mammals in general and humans in particular is as yet unknown.

The APOBEC3 sub-family

The APOBEC3 genes exist only in mammals (absent in mice), preferentially deaminate cytidines in DNA (91) and appear to have arisen as a result of tandem gene duplication events (92, 93). Chromosome 22 (locus q13.1) carries all seven human APOBEC3 genes enclosed between the genes CBX6 and CBX7 that are conserved in all vertebrates. The defining characteristic of all APOBEC3 genes is the presence of one or two "Z-domain" sequences carrying residues necessary for coordinating zinc and essential to deamination activity and/or substrate binding. The Z-domains may be classified by their signature sequences and represented schematically as in Figure 1.3 (94, 95).

Human APOBEC3A, 3C and 3H contain one Z-domain each, Z1, Z2 and Z3 respectively, whereas APOBEC3B (Z2-Z1), 3D/E (Z2-Z2), 3F(Z2-Z2) and 3G (Z2-Z1) contain two Z-domains. The Z-domain signatures do not signify the presence or absence of catalytic capability, or domain preponderance, for example, the N-terminal **Z2** domain in APOBEC3G is deami-

nation incompetent whereas the Z2 domain located C-terminally in APOBEC3F is capable of cytidine deamination (96). Similarly, APOBEC3B has a N-terminal Z2 and a C-terminal Z1 domain and both are catalytically active (97).

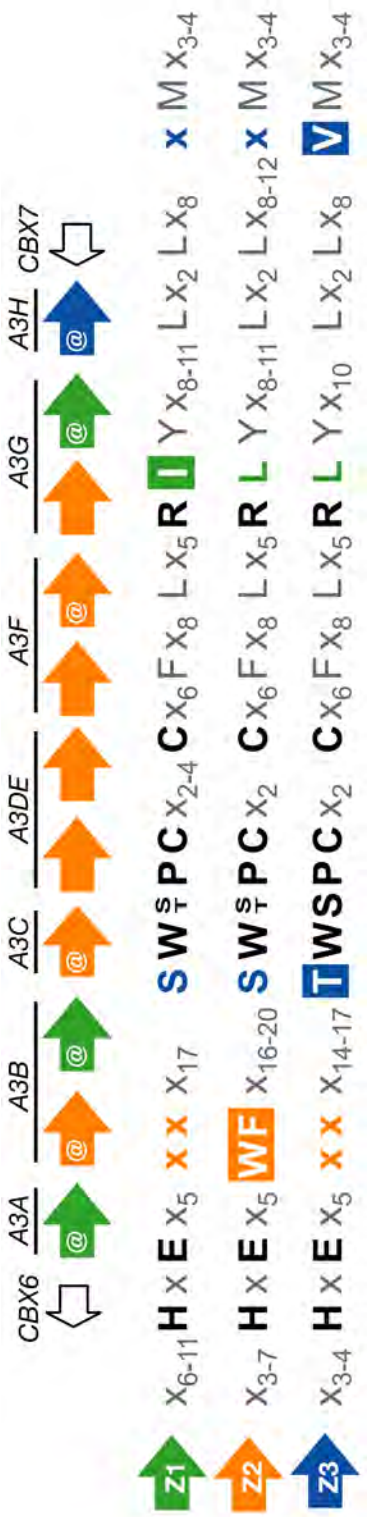

CBX6
A3A
A3B
A3C
A3DE
A3F
A3G
A3H
CBX7
Z1
Z2
Z3
X6-11 H X E X5 x x X17 S W S/T P C X2-4 C X6 F X8 L X5 R I Y X8-11 L X2 L X8 x M X3-4
X3-7 H X E X5 WF X16-20 S W S/T P C X2 C X6 F X8 L X5 R L Y X8-11 L X2 L X8-12 x M X3-4
X3-4 H X E X5 x x X14-17 T W S P C X2 C X6 F X8 L X5 R L Y X10 L X2 L X8 V M X3-4

Figure 1.3

Figure 1.3 : A schematic representation of the three Z-domain signature sequences in APOBEC3 proteins (Z1: green, Z: orange and Z3: blue). CBX6 and CBX7 genes flank the APOBEC3 locus on chromosome 22. The catalytically active domains are indicated by the @ sign and residues that uniquely identify each domain are enclosed in highlighting boxes. Based on : ref. (94, 95).

1.2.2 APOBEC3G

Human APOBEC3G (A3G), also called CEM15, is a cytidine deaminase discovered to exert an anti-retroviral effect against HIV in 2002 (57). HIV virions lacking the viral protein, vif, were demonstrated to be capable of infecting and replicating in "permissive" cell lines that lacked CEM15 (CEM-SS, Sup-T1 and 293T). Not only did the "non-permissive" cell lines (CEM, H9 and HUT78) not allow successful infection or replication of the Δvif HIV virions but expressing CEM15 in the "permissive" cells was sufficient for making them "non-permissive". Further investigation revealed that HIV vif (virion infectivity factor) binds to A3G and targets it to the proteasomal pathway for degradation (72, 98–103), by hijacking the Cul5-E3 ring ligase complex (76). Vif directed proteasomal degradation of A3G greatly reduces the number of A3G molecules packaged into assembling virions thus abrogating A3G's anti-retroviral effect during subsequent host cell infection (104).

A3G : mode of action

Expression of A3G in response to HIV infection was recently demonstrated to result from stimulation of TLR3 (Toll Like Receptor 3), a pattern recognition receptor (PRR) of the innate immune system that is expressed on dendritic cells and senses double stranded RNAs generated during retroviral replication. In addition to stimulating A3G expression, the downstream NF-κB and IRF3 driven Type-I Interferon (IFN-α/β) signaling also increased Tetherin and anti-HIV miRNA expression, further accentuating the viral restriction effect (105).

A3G is known to exist intracellularly as high molecular mass (HMM) and low molecular mass (LMM) ribonucleoprotein complexes. The HMM form is associated with RNPs, Staufen granules, Alu and hY RNAs and although enzymatically inactive, is shown to restrict Alu retrotransposons (106). RNAse action can convert the HMM form to enzymati-

cally active LMM that is responsible for anti-HIV activity (107). The LMM form may arise from disintegration of HMM complexes or by translation of fresh A3G transcripts in the cytoplasm and may then become associated with cell membrane lipid-rafts. This raft-associated LMM form was recently shown to be the primary method of packaging A3G into newly assembling HIV virions (108).

HIV encapsidated A3G is released into the host cells infected by the virion and by virtue of being closely associated with the viral genomic RNA is believed to have favorable access to the newly reverse transcribed DNA (109–113). The exact mechanism(s) for the anti-HIV effect of A3G are under dispute in the present literature (114–125) with the primary contention being the requirement of cytidine deaminase activity for exerting the anti-HIV effect. However, the general consensus remains that A3G is capable of exerting an anti-HIV effect and vif attempts to thwart A3G by targeted degradation.

A3G enzymatic activity has been demonstrated to be "processive", meaning, the enzyme does not release the substrate until multiple catalytic reactions have occurred, however it may change substrate polynucleotide strands by "intersegmental" transfer (126). A3G prefers a "signature" sequence **5'-CC<u>C</u>** for binding and deaminating within substrate polynucleotides in a 3' 5' direction, preferentially deaminating the underlined C. (127–130)

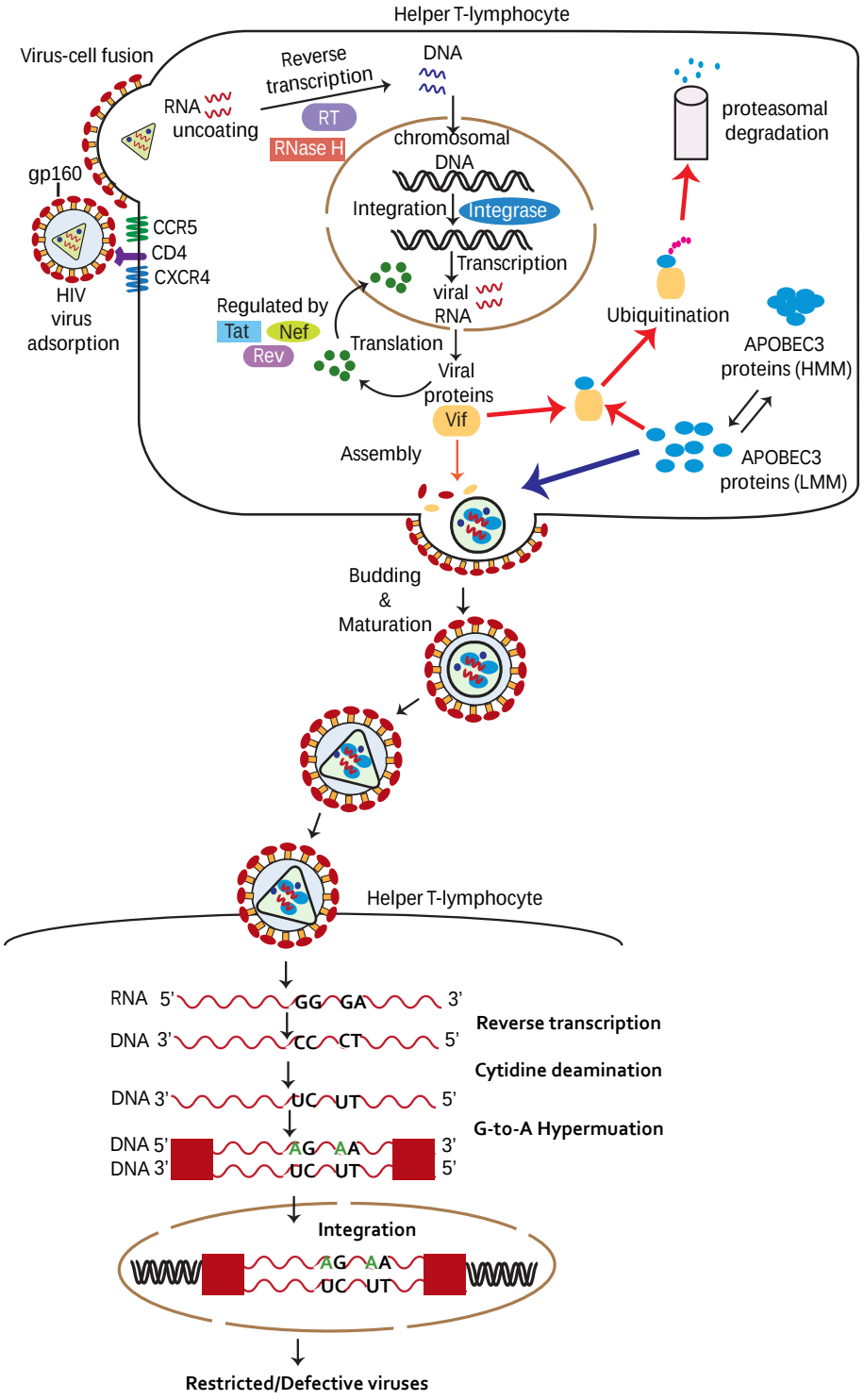

Helper T-lymphocyte
Virus-cell fusion
Reverse transcription
DNA
RNA uncoating
RT
RNase H
chromosomal DNA
Integration
Integrase
Transcription
viral RNA
proteasomal degradation
gp160
CCR5
CD4
CXCR4
HIV virus adsorption
Regulated by
Tat
Nef
Rev
Translation
Viral proteins
Vif
Ubiquitination
APOBEC3 proteins (HMM)
APOBEC3 proteins (LMM)
Assembly
Budding & Maturation
Helper T-lymphocyte
RNA 5'
GG GA
3'
DNA 3'
CC CT
5'
Reverse transcription
Cytidine deamination
DNA 3'
UC UT
5'
G-to-A Hypermuation
DNA 5'
DNA 3'
AG AA
UC UT
3'
5'
Integration
AG AA
UC UT
Restricted/Defective viruses

Figure 1.4

Figure 1.4 : APOBEC3G (A3G) activity in context of the HIV life-cycle. A3G (filled blue ellipses) is present in the cytoplasm as either a High-Molecular-Mass (HMM), primarily dormant form, or as Low-Molecular-Mass (LMM) form that is capable of restricting HIV. The viral *vif* protein (yellow), expressed in the host cell after viral genomic integration, targets A3G to the cellular proteasomal degradation pathway and attempts to thwart its packaging into the newly assembling virions at the cell memebrane. However, if A3G gets packaged into the assembling virions, it is capable of restrictiong viral replication by cytidine deamination when the virions infect the next host cell. Based on : ref. (131).

A3G and Vif

The virion infectivity factor (vif) is a 192 residue (23kD) non-enzymatic HIV protein that is predicted to be intrinsically disordered (132, 133) and leads to A3G degradation using the host-cell proteasomal machinery (76). Intrinsically disordered proteins (or, natively unstructured proteins) are characterized by a lack of well-defined tertiary structure amongst other features (134). Vif is predicated to posses a highly disordered C-terminus that, like most intrinsically disordered proteins, assumes a more ordered state upon oligomerization or in a complex with binding partners (135, 136). The intrinsic disorder creates unique challenges for structural analysis of vif and as of yet no full-length solution/NMR or crystal structures have been solved (137). However, experimental evidence accumulating over the previous decade has identified various features of vif that are involved in proteasome mediated degradation of A3G.

The Cullin5-Ring-Ligase (CRL5) complex is a E3 ubiquitin ligase, composed of the scaffold protein Cullin5, RING-finger protein Rbx2/ROC2, the adapter proteins Elongin B and C and a "substrate recognition protein" that targets the actual "substrate protein" to this complex (Figure 1.5). The "substrate protein" is ubiquitylated by an E2 ubiquitin-conjugation enzyme that also binds CRL5 (138). CRL5 recognizes a "SOCS-box" (Suppressor Of Cytokine Signaling) motif in the "substrate recognition protein", also present in HIV vif, that allows it to bind the CRL5 complex and present A3G for ubiquitylation (137, 139). Recent evidence suggests cellular protein CBFβ (core binding factor, beta subunit) may be involved in binding of vif to CRL5 (CSHL, Retroviruses 2011 meeting). Other studies suggest vif may inhibit A3G action without proteasomal degradation by non-competitively binding to substrate bound A3G and interfering with its catalytic activity (140, 141).

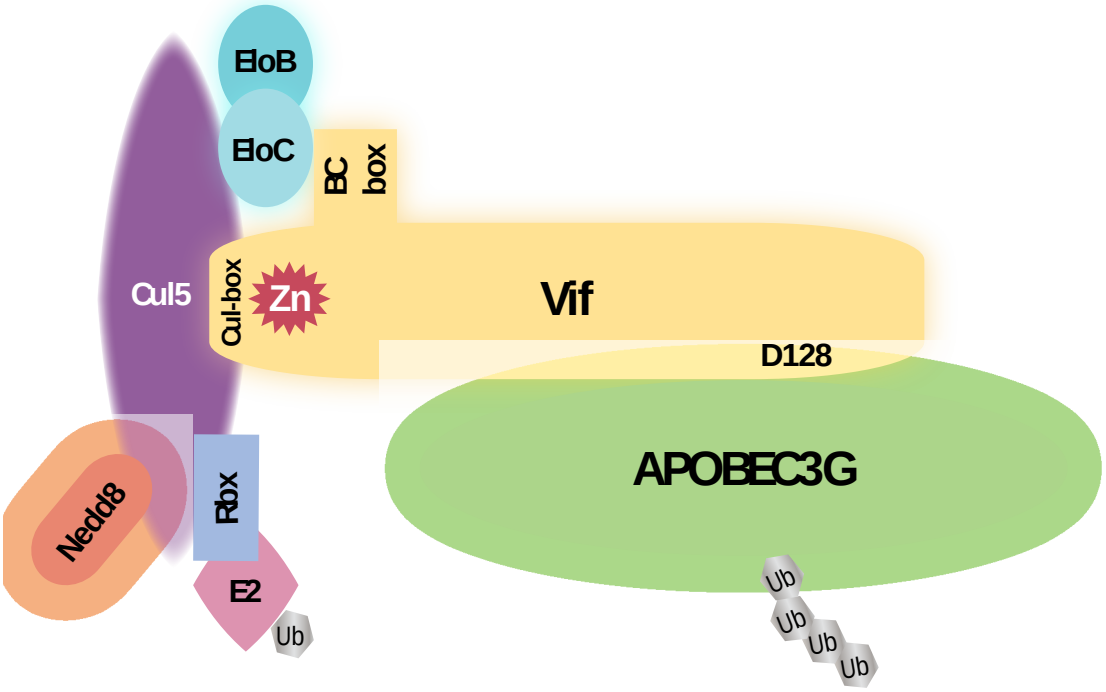

EloB
EloC
BC box
Cul5
Cul-box
Zn
Vif
D128
APOBEC3G
Nedd8
Rbx
E2
Ub
Ub
Ub
Ub
Ub

Figure 1.5

Figure 1.5 : A schematic representation of the Vif-Cullin5-RING-Ligase complex. Vif interacts with Elongin B/C as well as Cul5 via the C-terminal domain, whereas the N-terminal domain binds APOBEC3G and drives it to close proximity with the E2 ubiquitin ligase for poly-ubiquitylation. Based on : ref. (142).

A3G structures and implications for HIV

Currently, three NMR solution structures (PDB: 2JYW, 2KBO, 2KEM) and two X-ray crystallographic structures (PDB: 3E1U/3IQS, 3IR2) of the A3G C-terminal domain (CTD) are available from the Protein Data Bank (PDB/RCSB) (143–147) (Figure 1.6). The CTD is the catalytically active domain in A3G and the NTD, although containing a zinc-binding site (Z-domain), is catalytically inactive and is required for effectively binding the substrate polynucleotide sequence (148). The A3G NTD is predicted to be intrinsically disordered (149), as compared to the CTD, and may be a contributing factor to the challenges faced during soluble overexpression and purification of full-length A3G required for NMR structure solution or X-ray crystallography efforts.

The earliest A3G structure to be deposited in the PDB was the NMR solution structure (2JYW : 143) of the A3G-CTD, residues 198-384, with five solubility enhancing mutations, L234K, C243A, F310K, C321A and C356A that were elucidated earlier as not having any deleterious effect on enzymatic or biological activity (150). This solution structure established and corroborated features predicted from previous experiments, such as the arrangement of active residues that coordinate a zinc atom and secondary structure content of helices, sheets and loops. The X-ray crystallographic structure of the wild-type A3G-CTD residues 197-380 (3E1U : 146), with the last four residues truncated, was published the same year and confirmed the active site zinc coordination geometry as well as most secondary structure features observed in the NMR solution structure. Two structural features, a putative substrate binding "groove" and a large loop (*AC loop 3* per the authors), were in dispute between the two structures. Another NMR solution structure (2KBO : 144) of a slightly longer construct (residues 193-384) and subsequent modeling and experimental evidence (70) attempted to reconcile the differences in substrate ssDNA strand binding orientation, suggested by the previous structures, by introducing a

third putative substrate orientation. Similarly, the question of the *AC loop 3* shape could not be resolved from the myriad of orientations observed in solution. However, the authors demonstrated a 3' 5' order of cytidine deamination by A3G using time-chase ^{1}H NMR spectroscopy, corroborating previous observations in the literature (127–130). The question of (*AC loop 3*) structure was settled by the next NMR solution structure (2KEM : 145) of the A3G-CTD spanning residues 191-384 and incorporating the five solubility enhancing mutations enumerated earlier. The authors used NOE measurements to arrive at the distances between carbonyl oxygens and amide nitrogens of the polypeptide backbone in solution and confirm the arrangement of residues within the β2-β2′ "bulge" (author terminology). In addition, the authors also proposed a full-length A3G model with the A3G-NTD homology modeled from the existing CTD structures and some of the inter-domain linker region residues (191-196) now resolved in this structure. Our A3G-CTD structure in the PDB (3IR2 : 147) is the second X-ray crystallographic structure of A3G-CTD to be solved and it highlighted various errors in the previous crystal structure (3E1U), prompting a revision of the structure coordinates from the authors (3IQS : under same reference as 3E1U). A novel, inter-molecular, zinc binding site was also observed in this (3IR2) structure; further details are discussed in the next chapter of this dissertation.

The absence of a high-resolution full-length A3G structure or co-structure with substrate or inhibitors leaves open numerous questions regarding oligomerization, binding to vif (occurs at the NTD), substrate orientation at the active site and structural features pertinent to anti-HIV activity. Various models have been put forward in an attempt to answer some of the questions enumerated above and tested using different experimental methods. The veracity of such models may well be confirmed once reliable structures including the A3G NTD or co-structures with vif or polynucleotide substrates become available. However, the structural data available to date have helped answer some very impor-

tant questions such as A3G active-site architecture and putative A3G-CTD oligomerization interfaces, with direct implications for anti-HIV therapeutic approaches.

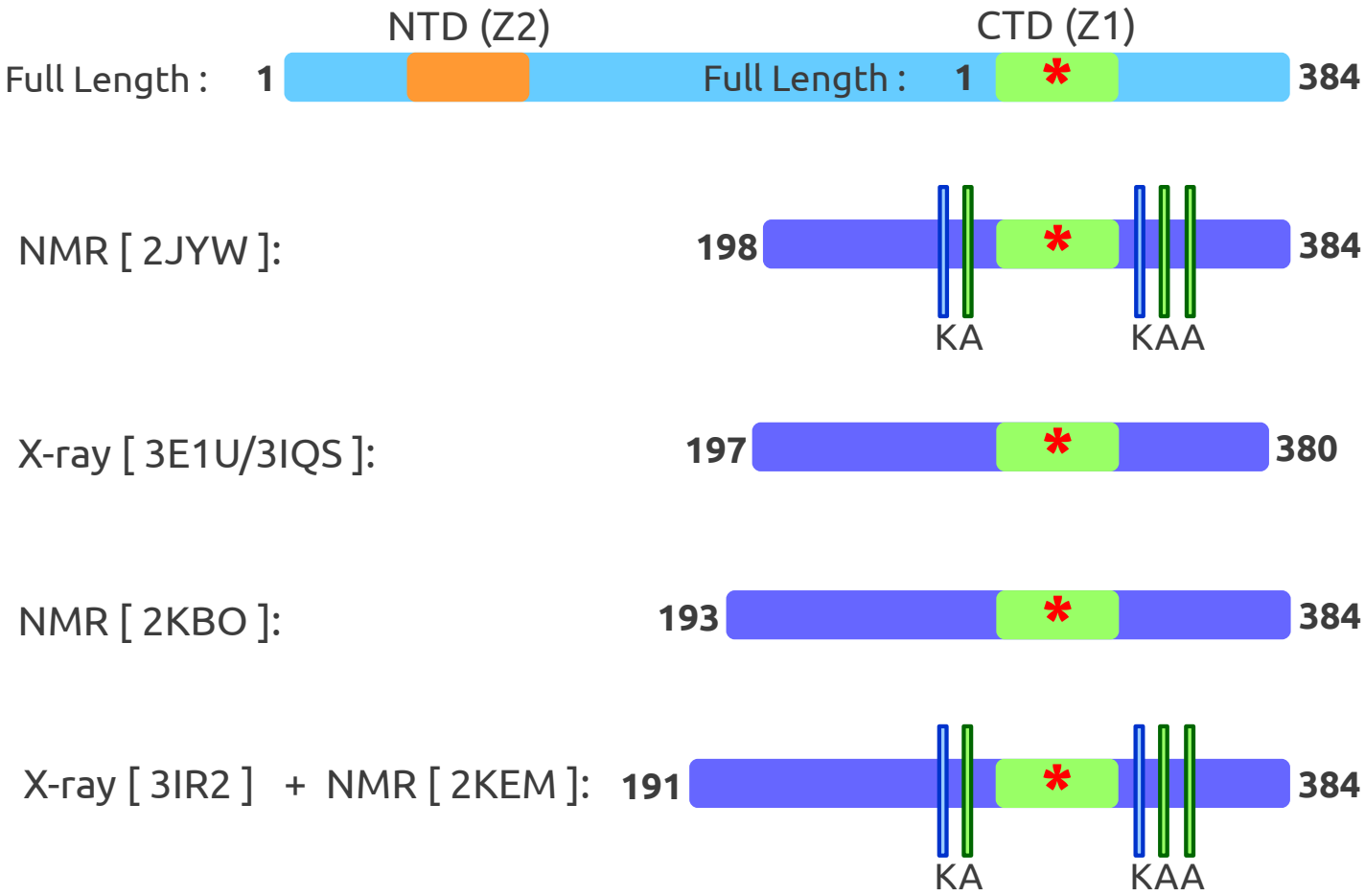

NTD (Z2)
CTD (Z1)
Full Length : 1
Full Length : 1
384
NMR [2JYW]:
198
384
KA
KAA
X-ray [3E1U/3IQS]:
197
380
NMR [2KBO]:
193
384
X-ray [3IR2] + NMR [2KEM]:
191
384
KA
KAA

Figure 1.6

Figure 1.6 : A schematic representation of the various A3G C-terminal domain constructs that have been used for structure elucidation. The full-length, wild-type, A3G is drawn at the top, followed by the CTD constructs. The PDB IDs of structures for each construct are listed within square brackets. Solubility enhancing mutations, L234K, C243A, F310K, C321A and C356A are indicated by vertical bars and letters K/A. Based on data from refs. (143 – 2JYW), (144 – 2KBO), (145 – 2KEM), (146 – 3E1U/3IQS), (147 – 3IR2).

1.3 Scope of the thesis

HIV/AIDS is a disease with extensive global impact and the search for curative treatment has proved to be a challenge despite intensive research, in large part due to viral mutations that lead to drug resistance and eventual failure of anti-retroviral therapy. Exclusively targeting HIV proteins for anti-retroviral drug discovery will, in all likelihood, pose the same problem. The complementary approach is to investigate viral-interacting host-cell proteins and disrupt protein-protein interaction(s) that are critical to viral life-cycle. A representative example is the interaction between HIV Vif and human APOBEC3G proteins, wherein, Vif binds APOBEC3G and targets it for proteasomal degradation. Why? Because APOBEC3G, a cytidine deaminase, is capable of hypermutating the viral genome to the detriment of HIV. In attempting to get rid of A3G, Vif has to hijack the host cell machinery and bind not only A3G, but also various other proteins that form the Cul5-RING-E3 ubiquitin ligase complex. Presently, the absence of any high-resolution structures of Vif preclude it being considered as a feasible target for anti-retroviral drug discovery approaches. Given the history of rampant and rapid emergence of drug resistance, Vif appears to be an even less attractive drug target, whereas, host proteins, being immutable, are far better candidates for the future of anti-HIV therapeutic research.

It is pertinent to mention here that two seemingly opposite schools of thought exist *vis-a-vis* the approach to A3G as an anti-HIV target:

— the 'activation' school of thought argues for approaches that will increase A3G activity in host cells to restrict HIV infection

— the 'inhibition' school suggests that inhibiting A3G activity may be of benefit in preventing sub-lethal mutagenesis that may potentially inflate the library of mutations in the HIV genome that allow the virus to escape therapeutic agents – thus causing drug resistance. Proponents of this school also point out the dangers to host genome from A3G

activation and unchecked cytidine deamination activity that may lead to erratic mutations and eventually carcinomas.

Both schools of thought have valid points favoring their respective approaches and the ideal approach may well lie in the middle, with neither unchecked activation, nor total inhibition but a well orchestrated 'modulation' of A3G action.

Chapter II : Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces : The work presented in this chapter attempts to explore and understand the structural details of the A3G catalytic domain, important for the anti-HIV effect exerted by full-length A3G in the host cell, by solving a high-resolution crystal structure. Various structural features observed in the crystal structure are discussed and results from experiments designed to interrogate the relevance of these structural feature to A3G function and anti-HIV activity are presented.

Chapter III : First-in-class small molecule inhibitors of the single-strand DNA cytidine deaminase APOBEC3G : This chapter describes the results of efforts to discover novel small-molecule ligands that may bind A3G and interfere with its cytidine deamination activity. A High Throughput Screen (HTS) was developed to assay the activity of compounds from a small, commercially available library and the best candidate compounds were used for co-crystallization with the A3G catalytic domain described earlier. Serendipitously, one co-crystal structure could be solved with a compound that bound covalently to a surface cysteine residue. Further investigation revealed the critical cysteine to be the one located near the active site and not the one on the surface. The cysteine required by the compound to exert an inhibitory effect had been mutated to an alanine residue in the construct used for co-crystallization during earlier attempts at improving protein solubility. Results from experiments designed to validate the above are described and discussed in this chapter. Numerous high-resolution crystal structures were solved over the course of

attempts at co-crystallization with A3G inhibitors, including an "apo" crystal structure, with the inhibition critical cysteine restored, that is presently the highest resolution crystal structure of the human A3G catalytic domain. In the absence of a co-crystal structure with the relevant cysteine residue, *in-silico* modeling was used to propose a putative mechanism of action arising from steric hinderance at the active site.

Chapter IV : Comparative analysis of Z-domain signature proteins in the human APOBEC3 subfamily : The availability of high-quality, high-resolution structural data for a protein enables modeling of related homologs with reasonable confidence. This chapter describes an attempt at *in-silico* homology modeling of some APOBEC3 subfamily proteins based on the high-resolution crystal structures of A3G catalytic domain. Similarities and difference across the models are discussed in light of data available in the current literature.

Preface to : Chapter II

This chapter of the thesis has been published in *Structure*

Shivender M.D. Shandilya, Madhavi N.L. Nalam, Ellen A. Nalivaika, Phillip J. Gross, Johnathan C. Valesano, Keisuke Shindo, Ming Li, Mary Munson, William E. Royer, Elena Harjes, Takahide Kono, Hiroshi Matsuo, Reuben S. Harris, Mohan Somasundaran, and Celia A. Schiffer (2010). **Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces.** *Structure* 18, 28-38. PMID: 20152150.

Author contributions : This work was a collaborative effort. **S.M.D.S** expressed, purified and crystallized the APOBEC3G catalytic domain and solved the crystal structure under the guidance of M.N.N., E.A.N., W.E.R., M.M. and C.A.S. Deaminase activity assays were conducted by P.J.C., J.C.V, K.S., M.L. and R.S.H. who also coordinated the Mass Spectrometry experiment. NMR experiments and subsequent analysis were carried out by E.H., T.K and H.M. Initial manuscript was written by S.M.D.S., M.S., R.S.H and C.A.S. **All authors** contributed to data analyses, figure constructions, and manuscript revisions.

Chapter II

Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces

Chapter II

Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces

2.1 Abstract

APOBEC3G is a DNA cytidine deaminase that has antiviral activity against HIV-1 and other pathogenic viruses. In this study the crystal structure of the catalytically active C-terminal domain was determined to 2.25Å. This structure corroborates features previously observed in nuclear magnetic resonance (NMR) studies, a bulge in the second beta strand and a lengthening of the second alpha helix. Oligomerization is postulated to be critical for the function of APOBEC3G. In this structure, four extensive intermolecular interfaces are observed, suggesting potential models for APOBEC3G oligomerization. The structural and functional significance of these interfaces was probed by solution NMR and disruptive variants were designed and tested for DNA deaminase and anti-HIV activities. The variant designed to disrupt the most extensive interface lost both activities. NMR solution data provides evidence that another interface, which coordinates a novel zinc site, also exists. Thus, the observed crystallographic interfaces of APOBEC3G may be important for both oligomerization and function.

2.2 Introduction

APOBEC3G (A3G), a DNA cytidine deaminase, belongs to the larger APOBEC family of proteins. Members include activation induced deaminase (AID), APOBEC1, APOBEC2, APOBEC4, and, in addition to A3G, six APOBEC3s (A3s), (48, 51, 92, 93); reviews by: (116, 120, 142, 151–153). These proteins have diverse biological functions that include editing mRNA

(APOBEC1), diversifying antibody gene DNA (AID), and restricting the mobilization of retroviruses and retrotransposons (A3s).

All the A3s (A, B, C, DE, F, G, and H) are single-strand DNA cytidine deaminases known to inhibit multiple retroelement substrates. The deaminases possess either one or two conserved zinc-coordinating (Z) motifs, with the consensus amino acid signature, $Hx_1Ex_{24-28}PCx_{2-4}C$ (48, 67, 94, 95). Zinc coordination is mediated by a histidine and two cysteines. The hydroxide that subsequently converts cytidine to uridine (C-to-U deamination) is generated when the active site glutamate removes hydrogen from water.

HIV-1, the retrovirus that causes AIDS, is the most prominent pathogen restricted by A3G (57, 117, 154–156). A3G suppresses HIV-1 infectivity by entering viral particles and deaminating the viral cDNA cytidines to uridines during reverse transcription. The uridines introduce adenosines in the complementary genomic strand which are often detected as the hallmark G-to-A hypermutations. These mutations produce stop codons and amino acid changes that can ultimately inactivate the virus. However, HIV-1 efficiently counteracts restriction by A3G, with the viral Vif protein binding and targeting the cytidine deaminase for proteasomal degradation; reviews by: (120, 142, 151, 153).

The life-or-death interaction between human A3G and HIV-1 Vif has made A3G the prototype member of the APOBEC family for biochemical and structural investigations. A3G has two consecutive Z-motifs giving the sequence of the enzyme an internal pseudo-symmetry. In the past two years, significant advances have been made in elucidating the structure of the C-terminal catalytic domain of A3G (A3G-CTD) but the conformations of several important functional regions differ between three NMR structures and a single X-ray crystal structure (143–146). The present study aims to resolve some of the controversy by determining a new, higher-resolution crystal structure of A3G-CTD. This new structure differs from the recently reported crystal structure (PDB ID: 3E1U with R-factor

25.2%, R-free 26.7%, Resolution 2.3Å) and strongly supports the structural integrity of the NMR structures (PDB IDs: 2JYW, 2KBO, 2KEM). Most importantly, this new crystal structure reveals four extensive interfaces, of which one or more may be important for A3G oligomerization and biological activity.

2.3 Results

The crystal structure of the human APOBEC3G catalytic domain, A3G191-384-2K3A – where 2K3A is L234K, C243A, F310K, C321A and C356A (145) – was determined at 2.25Å resolution. The A3G191-384-2K3A crystal structure was solved by molecular replacement, using a highly truncated model based on (PDB ID: 3E1U) (146). The crystal structure contains two molecules of A3G191-384-2K3A in the asymmetric unit. The final refinement statistics are: Rfactor (Rfree) 16.57% (20.82%) (Table 2.1).

The crystal structure of A3G191-384-2K3A has a core α-β-α fold consistent with other cytidine deaminases (143, 146, 157). As seen with the other A3G-CTD structures (145, 146), this structure has a five-stranded β-sheet surrounded on both sides by six α-helices (Figure 2.1). Secondary structural elements are numbered after the A3G191-384-2K3A NMR structure (PDB ID: 2KEM) (145) so that consistent comparisons can be made between the structures. The second β-strand is discontinuous, as also observed in the wild-type A3G-CTD (144) and A3G-CTD-2K3A (143, 145) NMR structures. As previously observed, the catalytic zinc is coordinated directly by H257, C288, and C291, and indirectly by the catalytic residue E259 via a water molecule. Thus, in terms of the overall fold, this structure is similar to the previously published structures but many key specific differences exist.

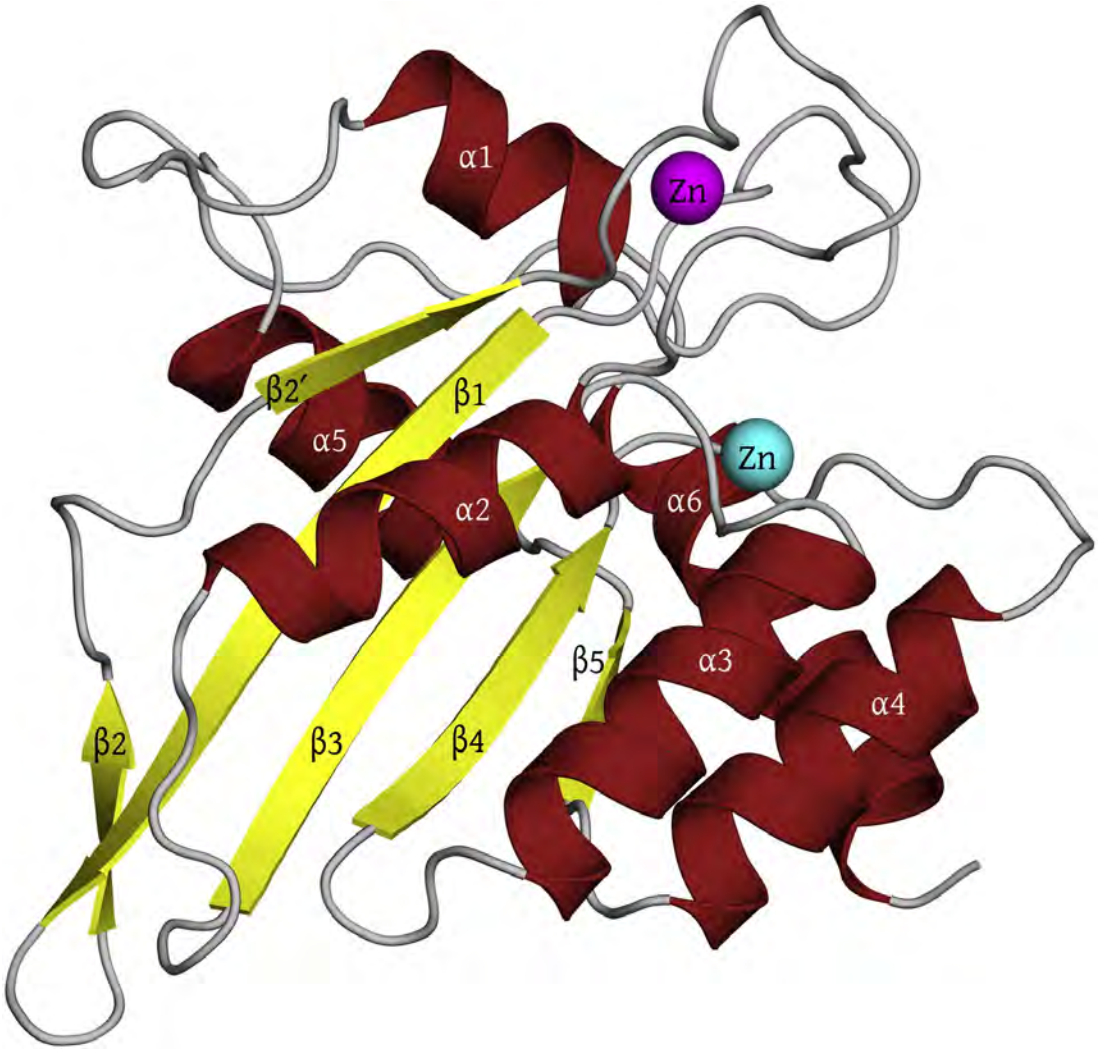


Figure 2.1

Figure 2.1 : Crystal structure of A3G191-384-2K3A with the α-helices highlighted in red and β-sheets highlighted in yellow, labeled as in (145). The catalytic zinc ion is shown in cyan and the intermolecular zinc ion is depicted in magenta in this and subsequent figures.

2.3.1 Crystal structure comparison

Comparison of A3G191-384-2K3A with the 3E1U crystal structure reveals numerous differences (highlighted in green and blue in Figure 2.2(A)). Some of these differences are in loop regions but other variations include large regions of secondary structure, notably the β1-β2 region (M227-Q237), the β2′-α2 region (H248-G255), and the α2-β3 region (P267-D274). In the A3G191-384-2K3A crystal structure, the second β-strand is discontinuous, with residues L235-R239 forming a prominent "bulge". This has been observed in both mutant and wild-type A3G-CTD NMR structures (143–145).

Crystallography and NMR spectroscopy are complementary techniques with well-established methods for verifying the structural integrity of protein structures. PDB-REDO recently organized some of the verification of crystal structures in the Protein Data Bank (158–160). Unfortunately, the only other A3G-CTD crystal structure in the database was flagged as having discrepancies in the PDB-REDO database (PDB ID: 3E1U with R-factor 25.2%, R-free 26.7%, Resolution 2.3Å). Many of the regions in the 3E1U structure (Holden et al., 2008) that differ from the NMR structures, are regions where the 3E1U structure's experimental data are the weakest (α-carbon B-factors shown in Figure 2.2(C)). The regions differing between these structures include β1-β2, β2′-α2 and α2-β3.

	A3G-191-384-2K3A
PDB-ID	3IR2
Resolution	2.25Å
Temperature	Cryogenic (-80°C)
Space group	P212121
Cell dimensions	
a	68.338Å
b	72.532Å
c	97.433Å
Molecules in AU	2
Completeness	96.9% (94.3%) (Last shell 2.33 - 2.25)
Total reflections	153207
Unique reflections	22967 21796 (working)
I/σ	9.6
Ave. redundancy	6.7
Rmerge	7.5%
RMSD in	
Bonds	0.0108Å
Angles	1.113°
Rfactor %	16.57
Rfree %	20.82
Water molecules per AU	379
Zinc atoms per AU	4
Ramachandran statistics	(such residues per molecule)
“additionally allowed” regions	12
“generously allowed” regions	0
“disallowed” regions	1

Table 2.1

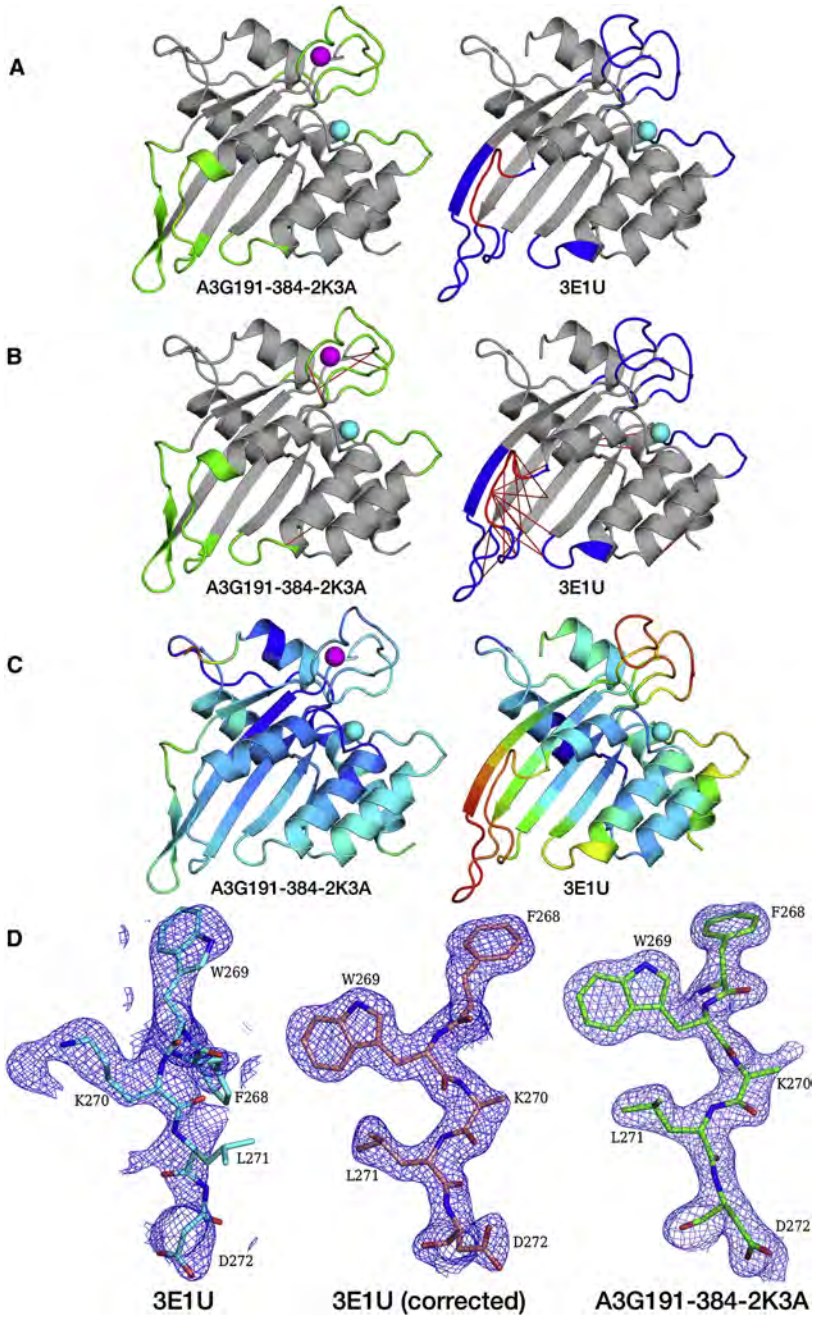

A
A3G191-384-2K3A
3E1U
B
A3G191-384-2K3A
3E1U
C
A3G191-384-2K3A
3E1U
D
W269
K270
F268
L271
D272
W269
F268
K270
L271
D272
W269
F268
K270
L271
D272
3E1U
3E1U (corrected)
A3G191-384-2K3A

Figure 2.2

Figure 2.2 : (A) Major regions of structural differences between A3G191-384-2K3A and 3E1U highlighted in blue and green. Residues F268-D272, which were mis-traced in the 3E1U structure are highlighted in red. (B) NOE violations of more than 6Å (red lines connecting α-carbon atoms) calculated based on the A3G191-384-2K3A structure's NMR data (PDB ID: 2KEM) (145). (C) Differences in α-carbon B-factors. Color spectrum blue to red depicts lowest to highest values. The overall α-carbon, B-factor range for the A3G191-384-2K3A crystal structure is 13.0-52.18Å^2, compared to 15.34–62.17Å^2 for the 3E1U crystal structure. (D) $2F_o$-F_c electron density maps, contoured at sigma level 1.0, show the placement of residues F268-D272 in three different structures: (i) the original and incorrect 3E1U, (ii) the rebuilt and corrected 3E1U, and (iii) the correct A3G191-384-2K3A crystal structure.

Evidence for ambiguity in the 3E1U structure is apparent in four ways: (1) 36 NOE restraints are violated in the 3E1U structure, by more than 6Å, with respect to NMR structural data; including in the very well-ordered α-helical and β-sheet regions that are well-established by NMR (Figure 2.2(B) and Experimental Procedures). (2) In addition to being flagged in PDB-REDO, analyses of the structure and experimental structure-factors of the 3E1U structure, downloaded from the Protein Data Bank (www.rcsb.org), revealed seven residues that were in unfavorable Ramachandran space (3 in disallowed and 4 in generously allowed regions). Most significantly, residues F268-D272 are completely mistraced in the electron density, burying K270 away from the surface (Figure 2.2(D)). These residues, located near the end of α2-helix, may be essential for the stability of the A3G-CTD active site because mutating W269 or L271 abrogates deaminase activity (143). When we re-fit this region in the 3E1U data, the final turn of α2-helix fit better within the electron density (Figure 2.2(D)). Thus, the structure in this region converged towards the conformation observed in the NMR structures not the one originally modeled in the 3E1U structure. The electron density in two other regions, β1-β2 and β2′-α2, remained ambiguous, making refitting unattainable. (3) In contrast to these regions in the 3E1U structure, our crystal structure of A3G-191-384-2K3A (R-factor 16.57%, R-free 20.82%, Resolution 2.25Å) is well-ordered. Our structure of A3G191-384-2K3A has only 9 NOE restraint violations to the NMR data; the α2-helix is well ordered; and the structure has lower α-carbon B-factors (Table 1) (Figure 2.2). (4) Mass spectrometry analysis validates that the β1-β2 structure is conserved in full-length A3G. In contrast to the 3E1U structure, all NMR structures and the A3G191-384-2K3A crystal structure show that residues M227 and W232 are adjacent to each other in the two neighboring beta-strands (β1-β2). This was verified by introducing two cysteine mutations: M227C in β1 and W232C in β2, into both full-length A3G-2K3A (Figure 2.11) and A3G (data not shown). Mass spectrometry determined that these two residues, M227C and W232C, form a disulfide bond and no peptides containing free M227C

or W232C are detected in either construct. In the 3E1U structure, M227 and W232 are not adjacent, rather R226 and L235 (part of the putative "continuous β2" strand) face each other. Applying similar mass spectrometry analysis to full-length A3G, with mutations R226C and L235C, did not result in a detectable peptide containing a disulfide bond (data not shown).

Authors of the 3E1U structure write: "Therefore, an intact full-length β2 strand and the five-stranded β-sheet core is probably the feature of wild-type APOBEC3G-CD2 and all other APOBEC proteins." (146). However, the A3G-191-384-2K3A crystal structure and the additional experimental data described above do not support this statement. These data suggest that the "bulge" observed between the β2-β2′ strands in NMR experiments and the A3G-191-384-2K3A crystal structure, is not merely an experimental artifact but an intrinsic feature of A3G-CTD structure.

2.3.2 Analysis of A3G191-384-2K3A crystal packing interfaces

The two molecules in the asymmetric unit in the crystal structure of A3G191-384-2K3A pack in such a way as to produce four major interfaces, all of which are the result of non- crystallographic symmetry, with surface areas of: 901Å^2, 604Å^2, 427Å^2 and 246Å^2, respectively (figures below). Multiple sequence alignment of the A3G-CTD to ten other homologs: A3G-NTD, A3A, A3B-CTD/NTD, A3C, A3DE-NTD/CTD, A3F-NTD/CTD and A3H reveals that many of the residues contributing more than 30Å^2 to these interfaces are unique to A3G-CTD (Figure 2.3).

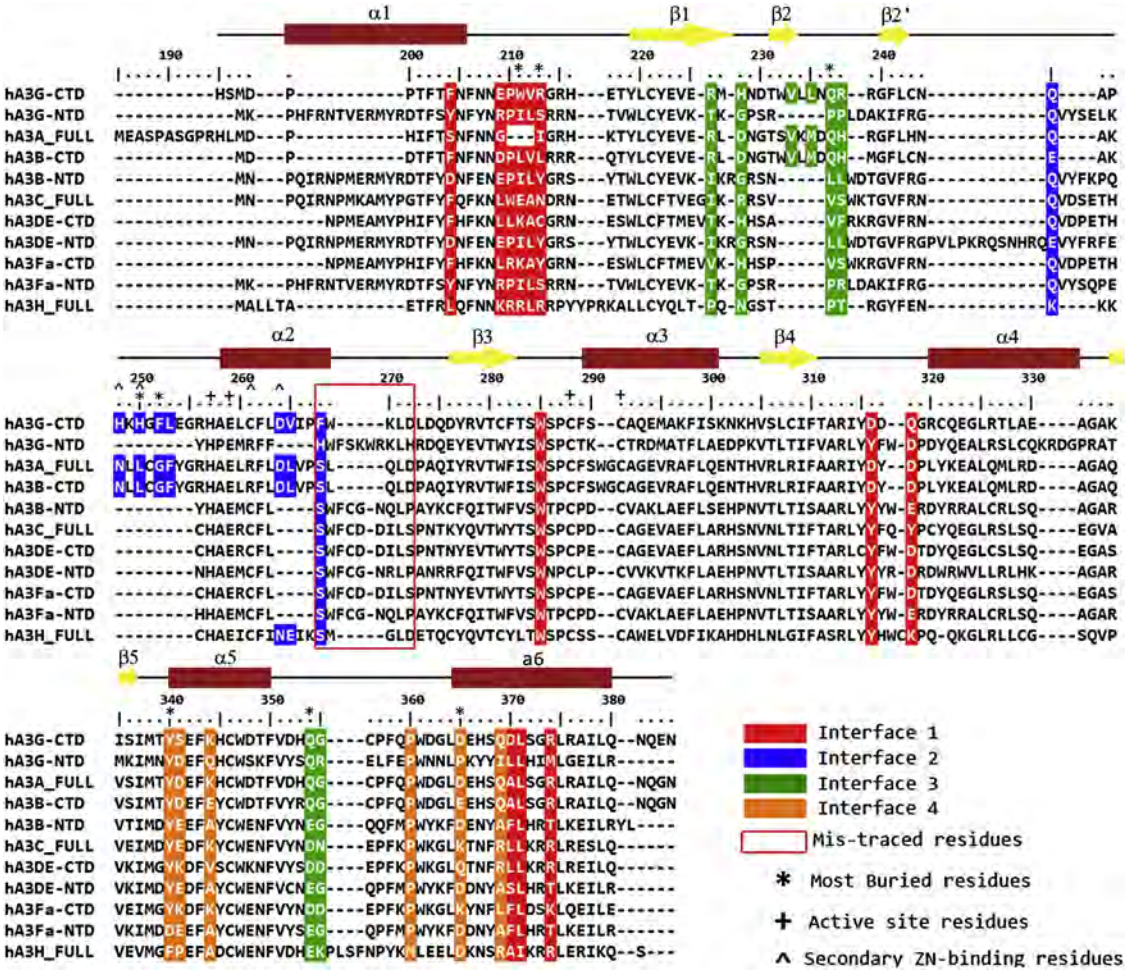

α1
β1
β2
β2'
α2
β3
α3
β4
α4
β5
α5
a6
hA3G-CTD
hA3G-NTD
hA3A_FULL
hA3B-CTD
hA3B-NTD
hA3C_FULL
hA3DE-CTD
hA3DE-NTD
hA3Fa-CTD
hA3Fa-NTD
hA3H_FULL
Interface 1
Interface 2
Interface 3
Interface 4
Mis-traced residues
* Most Buried residues
+ Active site residues
^ Secondary ZN-binding residues

Figure 2.3

Figure 2.3 : Multiple sequence alignment of all human APOBEC3 protein sequences (Refseq) by their respective NTDs and CTDs aligned to the A3G-CTD sequence. Residues contributing more than 30Å^2 to the buried surface area on an inter-molecular interface are highlighted. Secondary structure elements of the A3G191-384-2K3A structure are shown schematically above the sequence as brown rectangles for α-helices, yellow arrows for β-strands, and black lines for loop regions. The coloring scheme does not represent sequence conservation but only highlights residues involved in forming the four major interfaces.

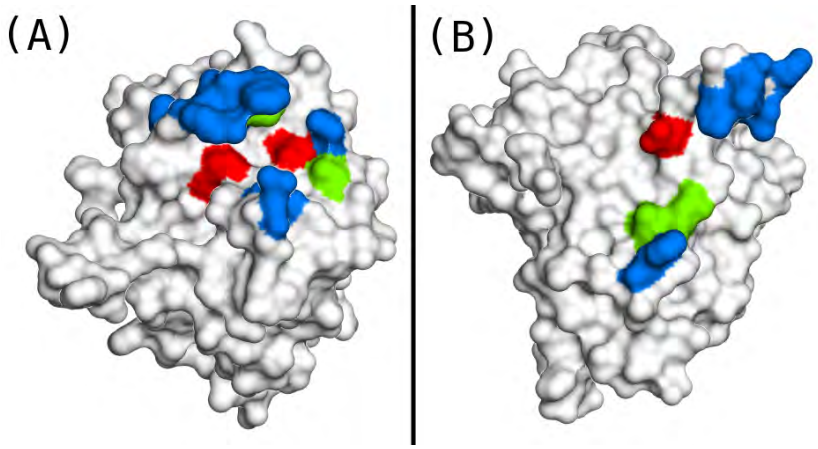


Figure 2.4

Figure 2.4 : (A). Interface 1 (Figure 2.5) and **(B).** Interface 2 (Figure 2.7) residues colored by sequence conservation based on the multiple sequence alignment in (Figure 2.3). Coloring scheme: Blue – Not conserved. Green – Partially conserved. Red – Fully conserved.

The largest interaction interface (901Å^2) displays excellent shape complementarity as observed from the surface representation of interfacing molecules (Figure 2.5(A)). Extensive surface contacts are observed, primarily between identical residues at the α1-loop-β1 from both molecules in the asymmetric unit. Twelve residues in each molecule contribute at least 30Å^2 to the interface, however, residues W211, R213 and Q318 together contribute over 450Å^2 of the interfacial area (Table S1) (Figure 2.5(B)). None of these three residues are conserved across the other APOBEC3's (Figure 2.3 and Figure 2.4). Three direct hydrogen bonds, nine water-mediated hydrogen bonds and one ionic interaction, occur across the entire interface (Table S2). All of these contacts verify the intimacy of this extensive interface. As the residues forming this extensive interface are not conserved among the other APOBEC3's or the NTD of A3G (Figure 2.3 and Figure 2.4), this interface may be unique to A3G-CTD.

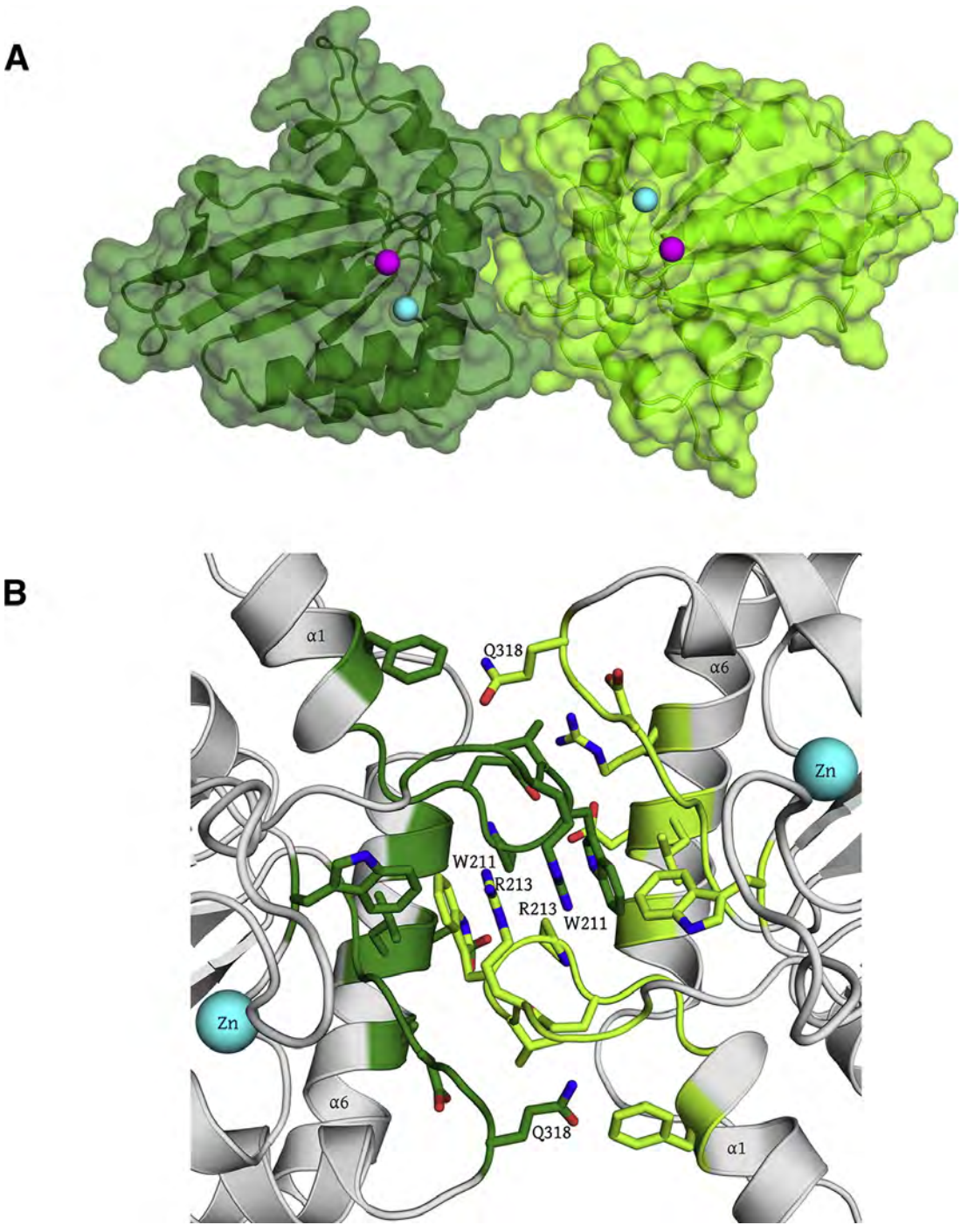


Figure 2.5

Figure 2.5 : Interface 1 between two A3G-CTD molecules in the asymmetric unit (A) Surface representation and (B) details of the largest interaction interface. Molecule A is shown in dark green, molecule B is shown in light green.

To assess the functional significance of this interface, a variant of A3G was made with the three amino acid substitutions W211A, R213A and R374E designed to profoundly disrupt the observed packing. This variant shows near undetectable levels (background) of DNA deaminase activity in vitro and abrogated anti-viral activity in the Vif-deficient HIV-1 reporter virus assay (Figure 2.6(A), Figure 2.6(B)). In contrast, the active site A3G-E259Q variant is catalytically dead and unable to inhibit Vif-deficient HIV-1, in agreement with prior studies (161, 162) (Figure 2.6(B)). All of the variants studied have near normal cellular expression levels and incorporate into viral particles (Figure 2.6(C)). Thus the residues that are integral to the first interface in the A3G191-384-2K3A crystal structure are essential to both A3G deaminase and anti-viral activity.

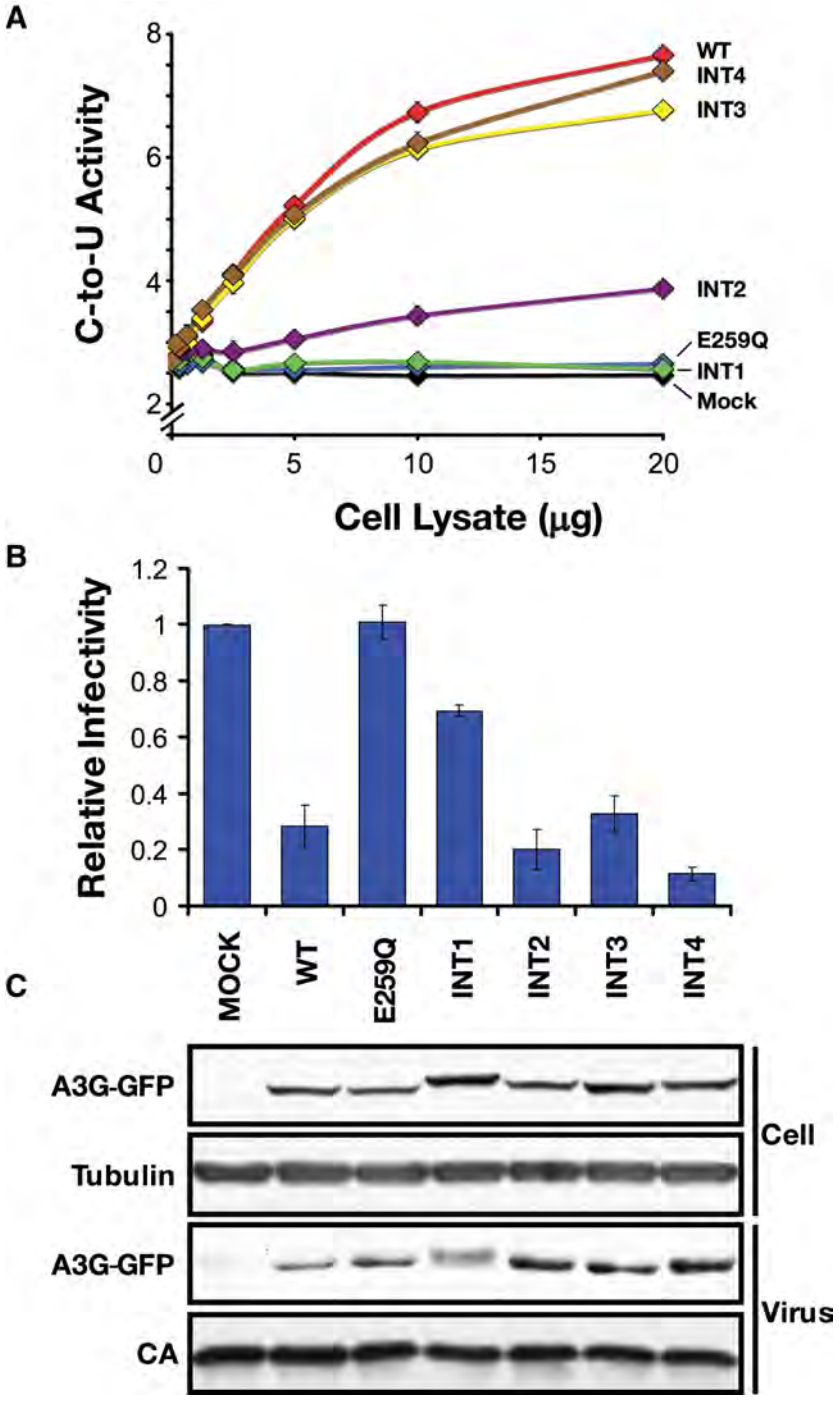

A
C-to-U Activity
8
6
4
2
0
5
10
15
20
Cell Lysate (µg)
WT
INT4
INT3
INT2
E259Q
INT1
Mock
B
Relative Infectivity
1.2
1
0.8
0.6
0.4
0.2
0
MOCK
WT
E259Q
INT1
INT2
INT3
INT4
C
A3G-GFP
Tubulin
A3G-GFP
CA
Cell
Virus

Figure 2.6

Figure 2.6 : DNA deaminase activity and HIV-1 restriction data. (A) A graph showing the results of a DNA oligonucleotide-based deamination assay with the indicated A3G-GFP constructs (MOCK: GFP only; WT: A3G-GFP; E259Q: A3G-E259Q-GFP; INT1: interface 1 A3G-GFP mutant (W211A, R213A, R374E); INT2: interface 2 A3G-GFP mutant (H248A, H250A, C261A); INT3: interface 3 A3G-GFP mutant (H228A, V233R, L235R); INT4: interface 4 A3G-GFP mutant (Y340A, S341A, K344A, D365A)). The y-axis reports the relative fluorescence level (C-to-U activity), and the x-axis shows the amount of A3G-expressing cell lysate used in each reaction. In most instances the SD of 3 replicas was smaller than the symbol. (B) A histogram showing the relative infectivity of Vif-deficient HIV-GFP produced in the presence the indicated A3G-GFP expression constructs or a GFP vector control. This histogram presents the average of 2 independent experiments, each with 3 replicas. The error bars show the relative difference between the 2 data sets. (C). Immunoblots of cellular lysates and virus particles probed for A3G-GFP, CA and Tubulin using anti-GFP, anti-p24 and anti-tubulin antibodies, respectively.

The second largest interface (604Å^2) involves the β2′-loop-α2 residues 247–254 in each of the two molecules of the asymmetric unit (Figure 2.7(A), Figure 2.7(B)). At this interface, eight residues of each molecule bury extensive surface area, with the largest surface area being buried by residues F252 and F268 (Table S1). This loop also coordinates an intermolecular zinc- binding site. H248 and H250 of one molecule in the asymmetric unit and the second molecule's C261 and D264 (through a water-mediated hydrogen bond) coordinate an intermolecular zinc ion (Table S2). The combination of these four residues occurring simultaneously is unique to A3G-CTD amongst the APOBEC3's (Figure 2.3). An additional four hydrogen/water mediated bonds are formed within this interface. This zinc-coordinating interface may provide insights into how a metal mediated switch could specifically modulate A3G oligomerization.

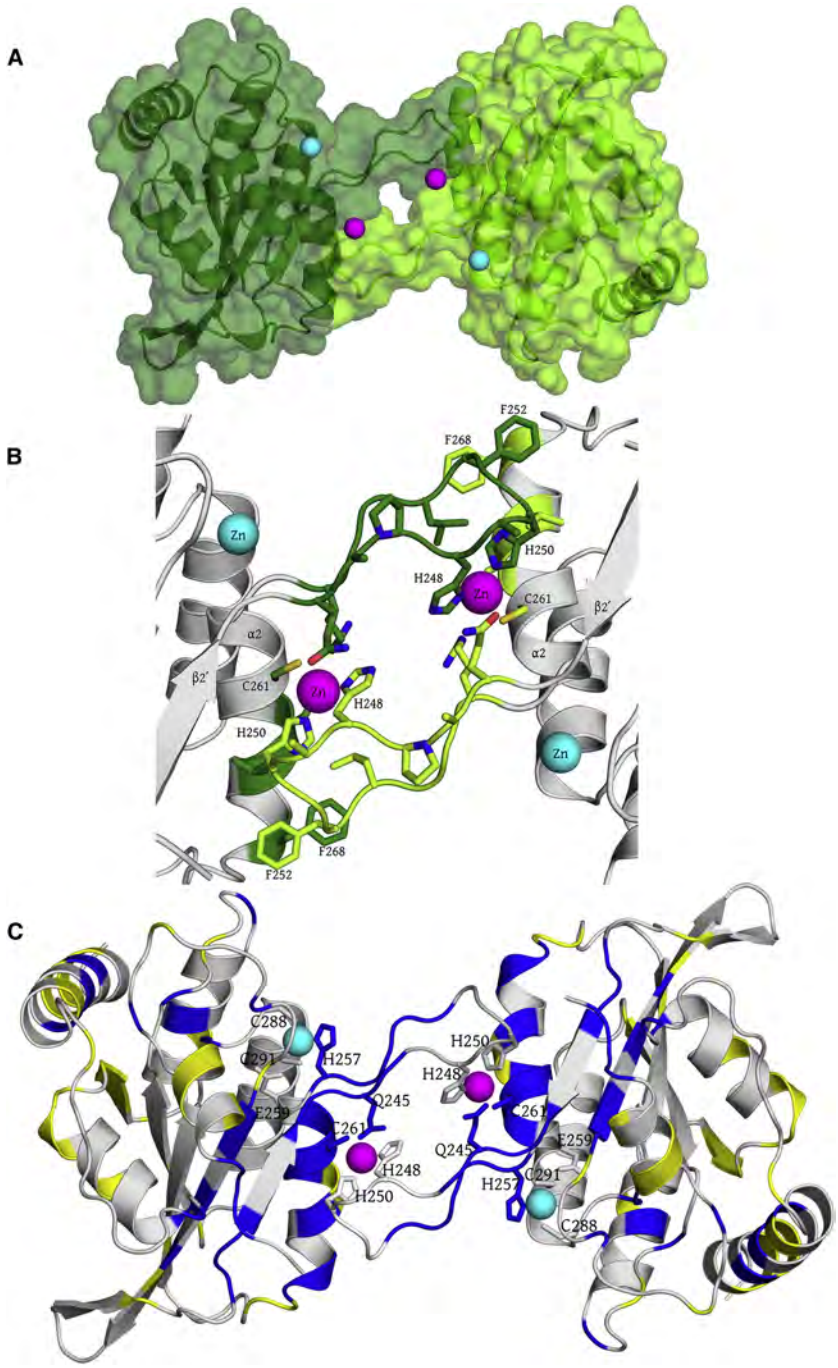

A
B
F252
F268
H250
H248
Zn
C261
β2′
α2
C261
H248
H250
F268
F252
C
C288
Q291
H257
E259
Q245
C261
H248
H250
H250
H248
C261
E259
Q245
H257
C291
C288

Figure 2.7

Figure 2.7 : Interface 2 between two A3G-CTD molecules in the asymmetric unit (A) Surface representation and (B) detailed interactions. Same coloring scheme as in Figure 2.5. (C) Interface 2 overview with coloring based on data from HSQC spectra. Blue residues: signal lost (I192, L193, S196, M197, D198, N208, R215, Y219, L220, Y222, V224, R238, R239, G240, A243, N244, Q245, A246, L253, E254, G255, R256, H257, L260, C261, F262, L263, V265, F268, W269, W285, S286, I314, G319, L325, F350, D352, S368, L371 and R374); Yellow residues: chemical shift indicating fast exchange (T201, T203, F204, W211, T218, E225, K234, D264, K270, L271, D316, T327, A333, I337, T339, S341, S372, G373, and R376).

NMR experiments support Zn^{2+} mediated oligomerization at this interface. Through titration of Zn^{2+} into isotope labeled A3G191-384-2K3A, loss of signal intensities and chemical shift perturbations were observed at 1mM Zn^{2+} (labeled blue, Figure 2.12). The residues most affected are located in the β2′-loop-α2 and α2 regions. In fact, these residues in the A3G191-384-2K3A crystal structure (blue, Figure 2.7(C)) map predominantly to the zinc-coordinating second interface, described above, indicating that this interface may contribute to zinc mediated oligomerization in solution.

To test the ability of zinc to modulate oligomerization a further series of NMR experiments was performed. A reference HSQC spectrum was taken at 50μM of Zn^{2+} concentration (Figure 2.8(A)) then, Zn^{2+} concentration was increased to 1.25mM causing the disappearance of amide proton NMR signals (Figure 2.8(B)). Next, 0.4 mM ethylenediaminetetraacetic acid (EDTA) was added to chelate the free Zn^{2+}. This caused the signals within the NMR spectra to reappear (Figure 2.8(C)). The similarity of chemical shifts, signal intensities and lineshapes of HSQC signals in (Figure 2.8(A) and (Figure 2.8(C))) combine to suggest that zinc mediates the equilibrium of A3G-CTD between oligomeric (Figure 2.8(C)) and monomeric states.

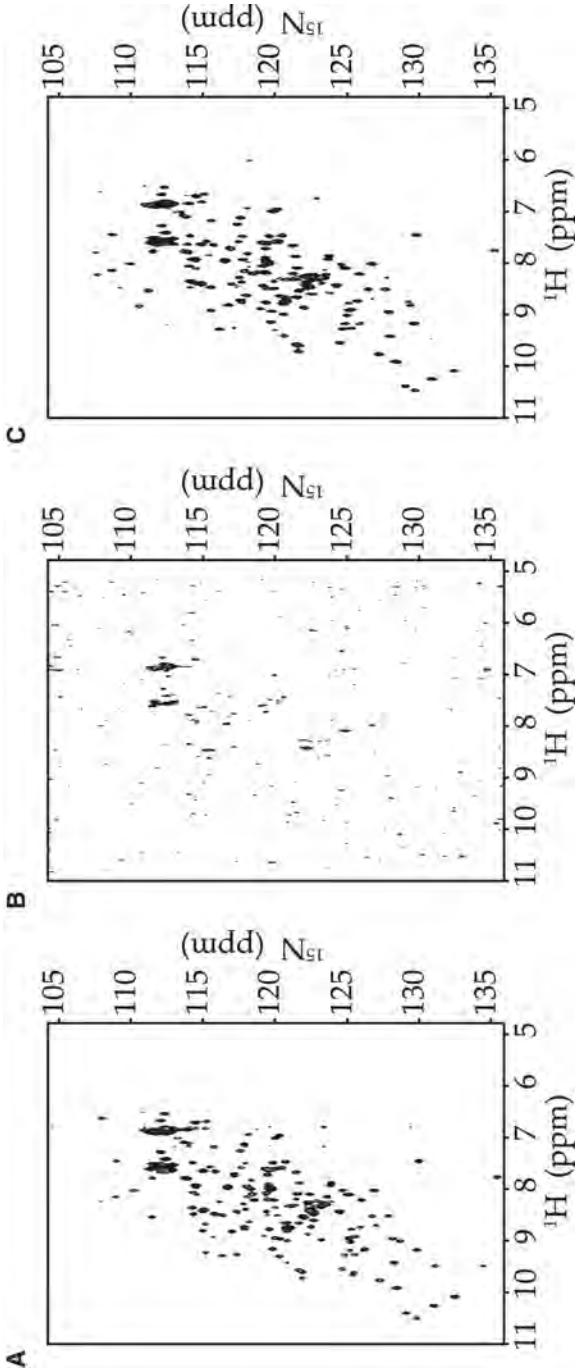


Figure 2.8

Figure 2.8 : Zn^{2+} dependent A3G191-384-2K3A aggregation is reversible. HSQC spectra of A3G191-384-2K3A with (A) 50µM Zn^{2+}, (B)1.25mM Zn^{2+}, and (C)1.25mM Zn^{2+} / 0.4mM EDTA. Protein concentration was 300µM for all spectra.

A variant designed to disrupt this zinc binding site in the second interface, H248A, H250A and C261A, shows normal expression levels and anti-viral activity (Figure 2.6(B), Figure 2.6(C)). These observations demonstrate that zinc-mediated oligomerization may not be essential for A3G's HIV-1 restriction activity. This variant is partly defective for DNA deaminase activity (Figure 2.6(A)). This alteration in DNA deaminase activity, may be attributable to C261A alone (Chen et al., 2008), and was not significant enough to compromise A3G anti-viral activity. Thus the residues that are integral to the second interface in the A3G191-384-2K3A crystal structure appear to impact A3G deaminase activity.

The third interface (427Å^2) involves residues at N-terminal (β1-β2 strands) and C-terminal ends of A3G191-384-2K3A (Figure 2.9(A), Figure 2.9(B)) (Table S1). Nine residues contribute extensively to the interface with Q354 burying the largest surface area at this interface followed by Q237, H228 and G355. In six of the eleven aligned APOBEC3 sequences (Q or E354)/G355 are conserved (Figure 2.3) but not H228 or Q237. Four direct and three water mediated hydrogen bonds are formed at this interface. Specifically, Q237 and R238 at the β2-β2′ "bulge" of one molecule (molecule A), form intermolecular hydrogen bonds to N-terminal residues R194 (side chain disordered) and S196 with the other molecule (molecule B). In addition, Q354 in the α5–α6 loop (molecule B) forms a hydrogen bond with R226 in β1 (molecule A), this interaction is potentially conserved in APOBEC3A and APOBEC3B-CTD. This interface would correspond closest to the β2-β2 interface observed in APOBEC2, but is clearly not formed and would entail extensive rearrangement of the α5 and α6 helices and their connecting loop. Of the four major interfaces, the packing of the third interface is least complementary and mutations at this interface (H228A, V233R and L235R) do not disrupt A3G's DNA deaminase or antiviral activities (Figure 2.6).

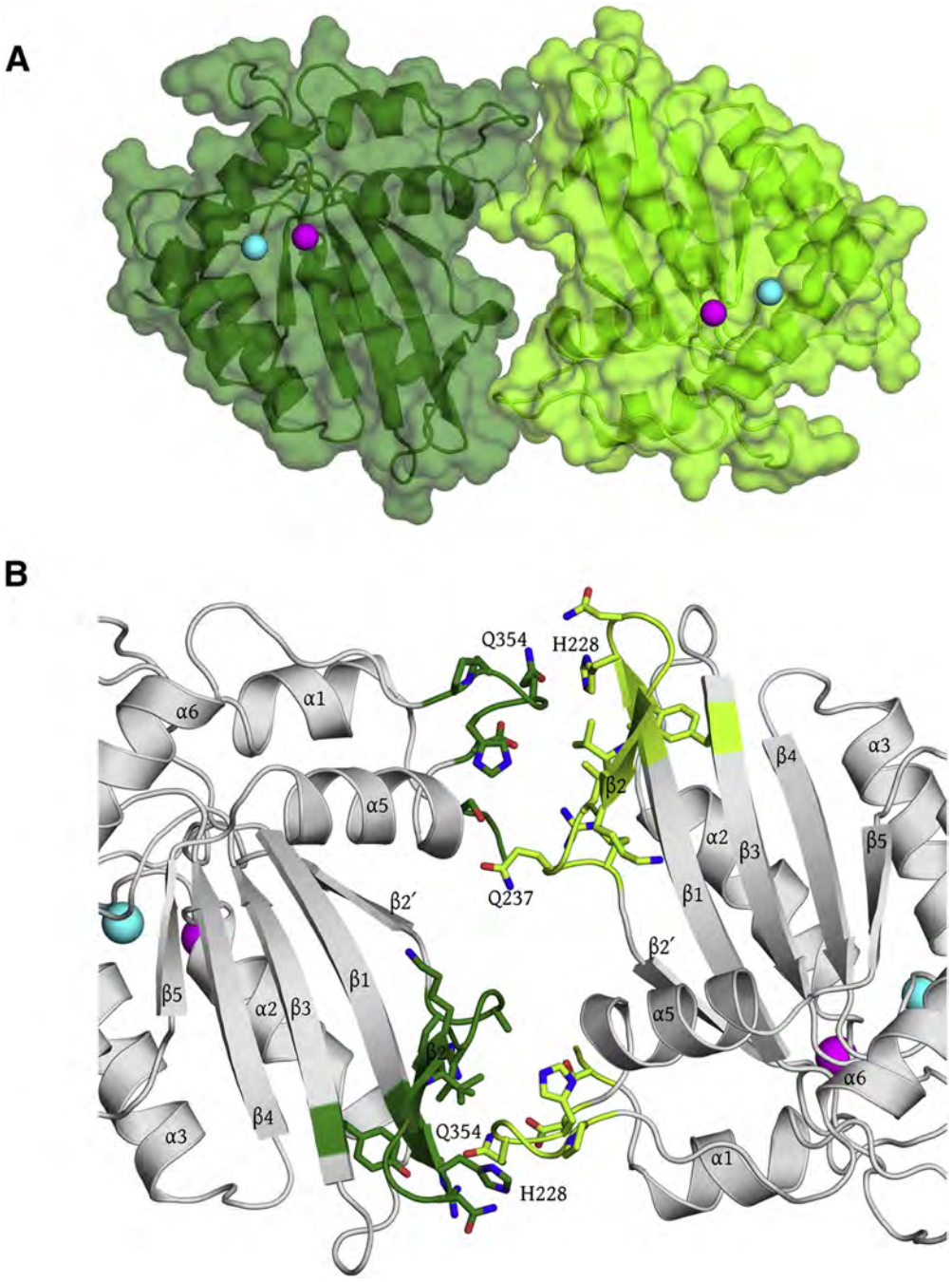

A
B
Q354
H228
α6
α1
α5
β2
α2
β3
β4
α3
β5
β2′
Q237
β1
β2′
β5
α2
β3
β1
α5
α6
β4
Q354
α3
α1
H228

Figure 2.9

Figure 2.9 : Interface 3 between two A3G-CTD molecules in the asymmetric unit (A) Surface representation and (B) detailed interactions. Same coloring scheme as in Figure 2.5.

The final of the four major interfaces buries a total of 246Å^2 and involves substantial burial of six residues from each molecule and the formation of two hydrogen bonds and a salt bridge (Figure 2.10(A), Figure 2.2(B)) (Tables S1, S2). This interface is primarily made up of the tops of the α5 and α6 helices. Interestingly, residues in α5 and α6 are also affected by the addition of Zn^{2+}. These residues showed small chemical shift changes in HSQC spectra, suggesting that this region is involved in intermolecular interactions under a fast exchange (labeled yellow, Figure 2.7(C), Figure 2.12). At this interface, residues Y340 and D365 contribute most to the buried interface area. These two residues are well conserved among the various APOBEC3's (Figure 2.3) Y340 is conserved in all but one and (D/E/Q)365 is conserved in eight of the eleven APOBEC3 sequences. Both of these residues and the less conserved S341 are involved in hydrogen bonding; a salt-bridge is observed between the moderately conserved K344 and D365. Although extensive sequence conservation is present, this interface is not as large as the others and a variant of A3G with mutations Y340A, S341A, K344A and D365A has normal DNA deaminase and HIV-1 restriction activities (Figure 2.6).

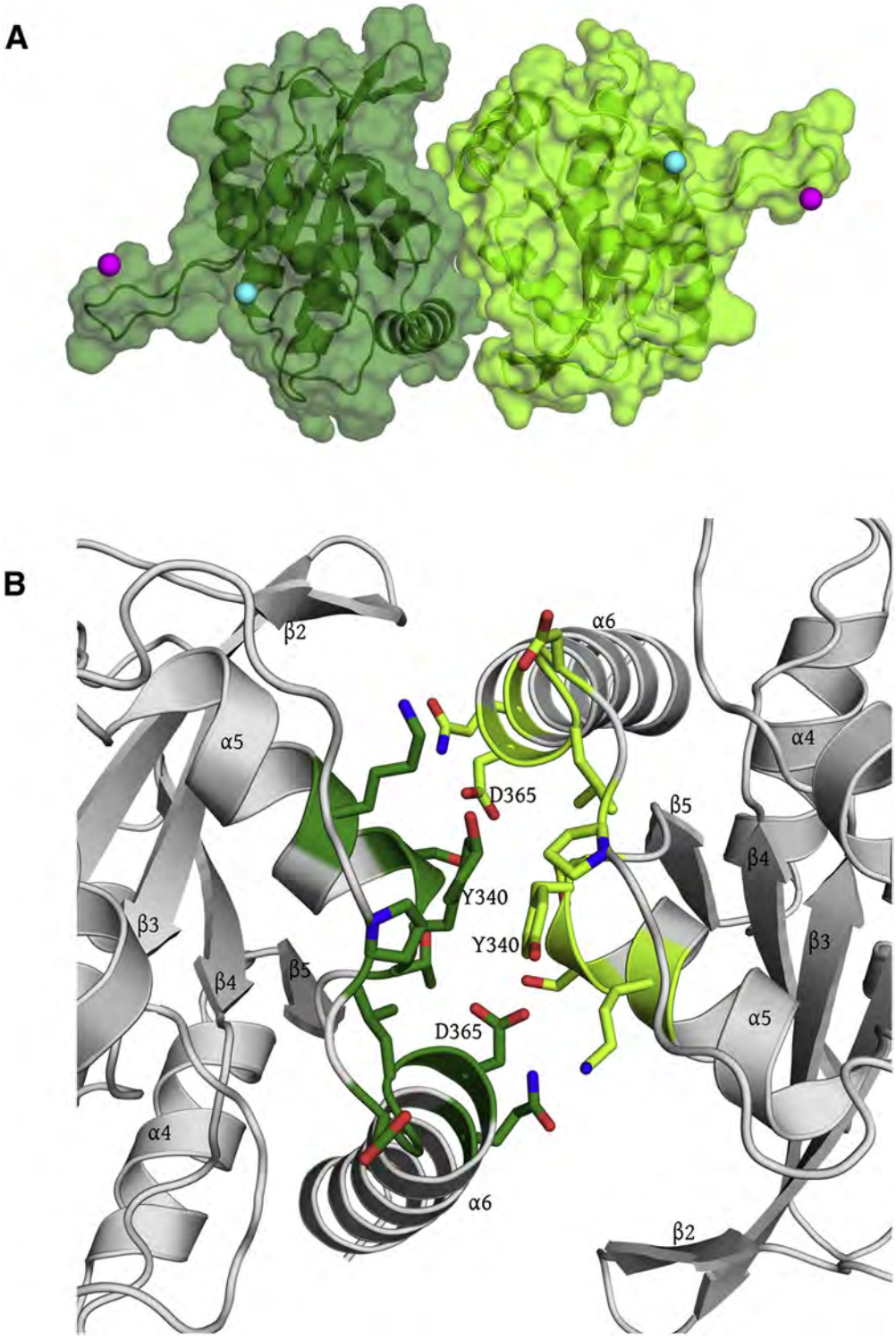


Figure 2.10

Figure 2.10 : Interface 4 between two A3G-CTD molecules in the asymmetric unit (A) Surface representation and (B) detailed interactions. Same coloring scheme as in Figure 2.5.

2.4 Discussion

In this study, the crystal structure of the catalytically active C-terminal domain of APOBEC3G was determined to 2.25Å. This structure differs significantly from the previously published crystal structure of this domain (PDB ID: 3E1U) (146). Here we have also demonstrated that the 3E1U structure's deposited experimental data do not unambiguously describe the structure deposited in the PDB. In contrast, our well-resolved crystal structure of A3G191-384-2K3A validates the published mutant and wild-type NMR structures (143–145), both in respect to the β2-β2′ "bulge" and a longer α2-helix with the residues F268-D272 properly placed. Thus, this structure represents the first clearly resolved crystal structure of the catalytic domain of APOBEC3G.

In addition mass spectrometry data support the observed β1-β2 conformation and presence of a bulge between the β2-β2′ strands of the A3G catalytic domain. We verified the register of the β1-β2 secondary structure elements by probing the proximity of key residues in this region through carefully engineered disulfide linkages. These results thereby confirm the structural integrity of the β1-β2 conformation observed in all NMR structures and the A3G191-384-2K3A crystal structure described here (143–145).

The unique interfaces observed between the asymmetric units of this crystal structure also lend insights into the potential role of oligomerization in APOBEC3G function. Full-length APOBEC3G sequence displays pseudo-symmetry (Figure 2.3), yet none of the largest interfaces observed in the crystal structure are conserved between the N- and C-terminal halves of the protein sequence nor are do they position the termini in a manner that could represent the full length enzyme. This is also consistent with the recent model proposed for the full-length A3G holoenzyme (145).

Beyond the full-length A3G enzyme, oligomerization – often dimerization or tetramerization – is a trend among the known deaminases. Escherichia coli cytidine deaminase

and Staphylococcus aureus TadA form dimers; Bacillus subtilis cytidine deaminase, yeast CDD1, and human cytidine deaminase form tetramers; and human ADAR1 and ADAR2 are homodimers (60, 77, 163–168). Multiple studies have shown that A3G can exist in states ranging from monomer to megadalton complexes (48, 116, 152, 169, 170). The importance of A3G oligomerization to HIV-1 suppression is controversial; some studies conclude that oligomerization is dispensable (123) while others claim that oligomerization is essential (171). The present study adds new insights into the potential roles of these interfaces for anti-viral and deaminase activities.

The surface(s) of A3G that forms these oligomers still remain unclear. SAXS, co-immunoprecipitation and FRET studies have implicated the C-terminal half of A3G in oligomerization (169, 170). This work potentially contrasts with more recent co-immuno-precipitation studies demonstrating that N-terminal residues are required for RNA-mediated dimerization (171, 172). Possibly both sets of studies have some merit and two distinct surfaces of A3G contribute to oligomerization.

Based on our analyses and data, we hypothesize that there are a variety of potential surfaces for oligomerization. The interactions observed at the two most extensive interfaces, the α1-loop-β1 (Figure 2.5) and the β2′-loop-α2 (Figure 2.7) which coordinates a zinc ion, represent unique A3G homo-oligomerization interfaces, while the tops of the α5 and α6 helices (Figure 2.10) could modulate both homo- and hetero-oligomerization of the different APOBEC3 family members.

A combination of solution NMR and the impact of variants designed to disrupt the interfaces support the possibility that three of these interfaces may have functional significance. Mutations introduced to disrupt the first interface reduce both DNA deaminase activity and anti-viral activity of A3G (Figure 2.6). However whether this is change in activity is due to a change in oligomerization or direct disruption in nucleic acid bind-

ing affinity remains to be determined. While mutations to disrupt the other interfaces do not alter the anti-viral activity of A3G, these interfaces still may impact oligomerization. The solution NMR data demonstrates a clear relationship between zinc-mediated EDTA-reversible oligomerization and specific chemical shifts of residues most strongly in the second interface that binds the intermolecular zinc but secondarily to the fourth interface (Figure 2.8 and Figure 2.12). The functional significance of the zinc mediated oligomerization remains to be elucidated.

Recent evidence indicates that binding of A3G to RNA may interfere with binding and deamination of ssDNA in a concentration dependent manner (173). High-Molecular-Mass (HMM) A3G complexes, the dormant form of A3G, are comprised of RNA and other ribonucleoproteins and the intermolecular interfaces described above may be required for binding of A3G to RNA and/or packaging into HMM complexes. Furthermore, the disruption of HMM complexes into Low-Molecular-Mass (LMM) active form of A3G may be achieved by small molecule ligands that weaken A3G–RNA binding by targeting the RNA binding regions in A3G or the A3G interfaces required for the formation of HMM complexes (174, 175).

The interfaces observed in this new A3G-CTD crystal structure may therefore have important functional implications in the role of oligomerization for the activity of A3G and will guide future studies.

2.5 Experimental Procedures

2.5.1 Protein expression and purification

BL21(DE3)-CodonPlus(RIL) (Stratagene, La Jolla, CA) bacterial cells, transformed with the pGEX6P1-A3G191-384-2K3A vector (145, 150), were grown to an O.D.600 of 0.6 [TB media, 100 µg/ml Ampicillin and 34 µg/ml Chloramphenicol]. They were induced overnight [0.5

mM IPTG, 18°C] to express GST-A3G191-384-2K3A. The cells were then harvested by centrifugation [5000 rpm, 15 min, 4°C] and re-suspended in lysis buffer [50 mM Na2HPO4/NaH2PO4 (pH 7.4) and protease inhibitor (Roche, Basel, Switzerland)]. Resuspension was followed by lysis with a cell disruptor. The lysate was centrifuged [43,000×g, 20 min, 4°C] and the supernatant incubated overnight with Glutathione-Sepharose (GE Healthcare, Piscataway, NJ) beads. The beads were washed with lysis buffer and incubated with PreScission Protease (GE Healthcare, Piscataway, NJ) in cleavage buffer [50 mM Na2HPO4/NaH2PO4 (pH 7.4), 0.005% Tween20, and 1 mM dithiothreitol]. A3G191-384-2K3A in the supernatant was further purified by FPLC (Pharmacia/GE Healthcare, Uppsala, Sweden) using a Superdex-75 size-exclusion column.

2.5.2 Protein crystallization

Purified A3G191-384-2K3A protein was exchanged into crystallization buffer [50 mM HEPES pH 7.5, 1 mM tris(2-carboxyethyl)phosphine, 50 µM zinc acetate, and 150 mM NaCl] with a PD-10 column (GE Healthcare, Piscataway, NJ). The protein was concentrated to 10 mg/ml using Amicon centrifugal filters (Millipore, Billerica, MA). Crystals were grown by the hanging-drop/vapor-diffusion method on greased VDX plates (Hampton Research, Aliso Viejo, CA) at 4°C with the reservoir solution [100 mM HEPES (pH 7.5), 10% Polyethylene glycol 4000, and 100 mM magnesium chloride].

2.5.3 Data collection and processing

Diffraction data for the A3G191-384-2K3A protein crystals were collected by synchrotron radiation at Argonne National Laboratory (Advanced Photon Source, Chicago, IL) BioCARS 14-BMC under cryogenic conditions. Diffraction images were indexed and scaled using the program HKL2000 (176); the structure was solved in $P2_12_12_1$ space group at 2.25Å.

2.5.4 Structure solution and refinement

The molecular replacement solution, calculated by PHASER (177) with the final translation function Z-score of 41.1, contained two molecules in the asymmetric unit. Phases were subsequently improved by building solvent molecules using ARP/wARP (178). Refinement was carried out with REFMAC5 (179) in the CCP4 suite (180), with cycles of rigid-body and restrained refinement; Non-Crystallographic Symmetry (NCS) restraints were applied in the first two cycles and TLS parameters (181) were used in subsequent refinement cycles. A free R-value with 5% of the data was used to limit the possibility of over-refinement. Electron density was viewed by the program COOT (182) and was used for model optimization.

2.5.5 Structure comparison and analysis

Graphical visualization and analysis was achieved using PyMOL (183). The interaction interfaces formed between molecules in the A3G191-384-2K3A crystal were analyzed by the Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (184). The NOE violations were calculated using accept.inp protocol of CNS (185). The A3G-CTD crystal structure (PDB ID: 3E1U) was used as the input coordinate file and the NOE files used to calculate the A3G191-384-2K3A NMR structure (PDB ID: 2KEM), were set as NOE distance restraints. The algorithm PRANK, as implemented in the software PRANKSTER, performed the multiple sequence alignments, with the +F option and default values (186). All human APOBEC3 sequences were downloaded from the NCBI Refseq database (APOBEC3A: NP_663745.1; APOBEC3B: NP_004891.3; APOBEC3C: NP_055323.2; APOBEC3DE: NP_689639.2; APOBEC3F: NP_660341.2; APOBEC3G: NP_068594.1; APOBEC3H: NP_861438.1) (187).

2.5.6 DNA deaminase activity assays

The A3G-GFP and A3G-E259Q-GFP expression constructs have been described (161, 188). All derivatives were constructed by combinatorial or sequential site-directed mutagenesis (Stratagene), with primer sequences available on request. The indicated A3G-GFP or interface mutant expression constructs were transfected into HEK-293 cells using Fugene (Roche). Cell lysates were prepared and DNA deaminase activities were measured as described (189). The sequence of the single-strand DNA substrate was: 5′-(6-FAM)-AAA-TTC-TAA-TAG-ATA-ATG-TGA-(TAMRA).

2.5.7 HIV-1 infectivity studies and immunoblots

These were performed as described (Harjes2009). HIV-GFP reporter viruses were produced by Fugene-mediated transfection (Roche) of 293T cells with a five plasmid cocktail (190). The HIV-GFP proviral plasmid CS-CG, the Gag-Pol expression plasmid, the Rev expression plasmid, and the VSV-G envelope expression plasmid constituted 0.9 µg of the cocktail, and the vector control or the A3G expression plasmid constituted the remaining 0.1 µg. Viruses containing supernatants were harvested 48 h post transfection by filtering cell-free supernatants (0.22-µm PVDF; Millipore) and pelleting by centrifugation (20,000 g, 2 hrs). The resulting viral pellet was resuspended directly in SDS gel loading buffer, fractionated by SDS-PAGE, transferred to a PVDF membrane (Millipore), and probed with an anti-GFP antibody JL-8 (Invitrogen) to detect the A3G-GFP constructs. An anti-p24 monoclonal antibody (191) provided by M. Malim through the National Institutes of Health AIDS Research and Reference Reagent Program was used as loading control. Both monoclonal antibodies were detected using a horseradish-peroxidase-conjugated goat anti-mouse IgG serum (Bio-Rad), followed by chemiluminescent imaging (Roche). After harvesting the viral supernatants, A3G levels in virus-producing cells were monitored by extracting soluble proteins with RIPA buffer (1 h, 4 °C, gentle rotation), removing particulates by cen-

trifugation (10 min, 20,000 g), and immunoblotting, as described above. An anti-tubulin monoclonal antibody (Covance) was used as cellular lysate loading control.

2.5.8 NMR Zn^{2+} titration experiments

^{15}N-labeled A3G191-384-2K3A protein was expressed and purified (in 1mM DTT, 200mM NaCl, 0.005% Tween 20 (Fisher scientific) and 50mM tris(hydroxymethyl)aminomethane buffer, pH 8.0) as described in (145). $ZnCl_2$ was titrated into 300 µM ^{15}N-labeled A3G191-384-2K3A at concentrations of 50µM, 200 µM, 400 µM, 600 µM, 800 µM, 1mM and 1.25mM. A heteronuclear single quantum coherence (HSQC) spectrum was recorded at each Zn^{2+} concentration, which enabled specific amino acid chemical shift perturbations to be detected. Bruker 700MHz equipped with a cryoprobe was used to collect all NMR spectra.

2.5.9 Mass spectrometry of cysteine mutants

Cysteine substitution variants were constructed using QuikChange protocol (Stratagene) and verified by DNA sequencing. A3G191-384-2K3A, A3G191-384-2K3A-M227C-W232C, A3G1-384-M227C-W232C and A3G1-384-R226C-L235C were produced using the same procedure as described in the protein expression and purification. Proteins were isolated from sodium dodecyl sulfate (SDS) polyacrylamide gel after Coomassie Brilliant Blue (CBB) staining. Isolated protein bands were digested using trypsin and subjected to mass spectrometry (MS) at the Taplin Biological Mass Spectrometry Facility (Harvard Medical School). Both high resolution full MS scans and high resolution MS/MS scans were performed using LTQ-FT and LTQ-Orbitrap (Thermo Electron) to verify the peptide.

2.6 Supplementary Data

2.6.1 Mass Spectrometry data

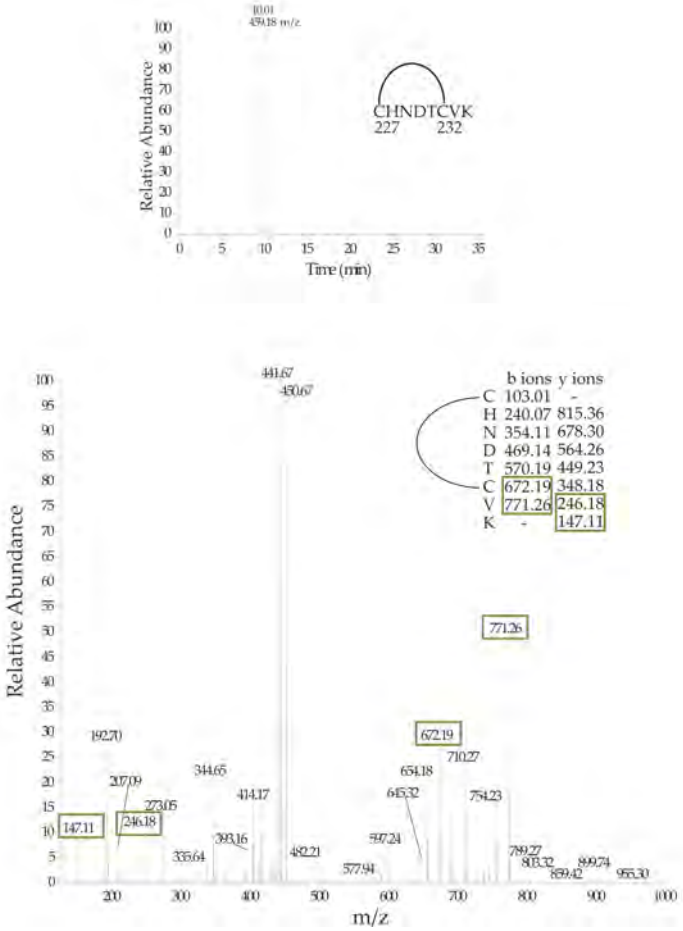


Figure 2.11

Figure 2.11 : (A). An ion-selective chromatogram for the predicted mass of the disulfide peptide in A3G1-384-2K3A-M227C-W232C. A peptide containing the predicted M227C-W232C disulfide bond was detected. The sequence of the peptide (C227HNDTCVK234) is shown in the inset, similar data was seen for the wild-type construct (data not shown). (B). A high resolution MS/MS fragmentation experiment to verify the peptide containing a disulfide bond. The only predicted ions that were detected are the ions outside the disulfide bridge (green ovals). This is expected, as the presence of a disulfide bond hinders peptide fragmentation. The peptide sequence is shown in the figure inset.

2.6.2 HSQC data

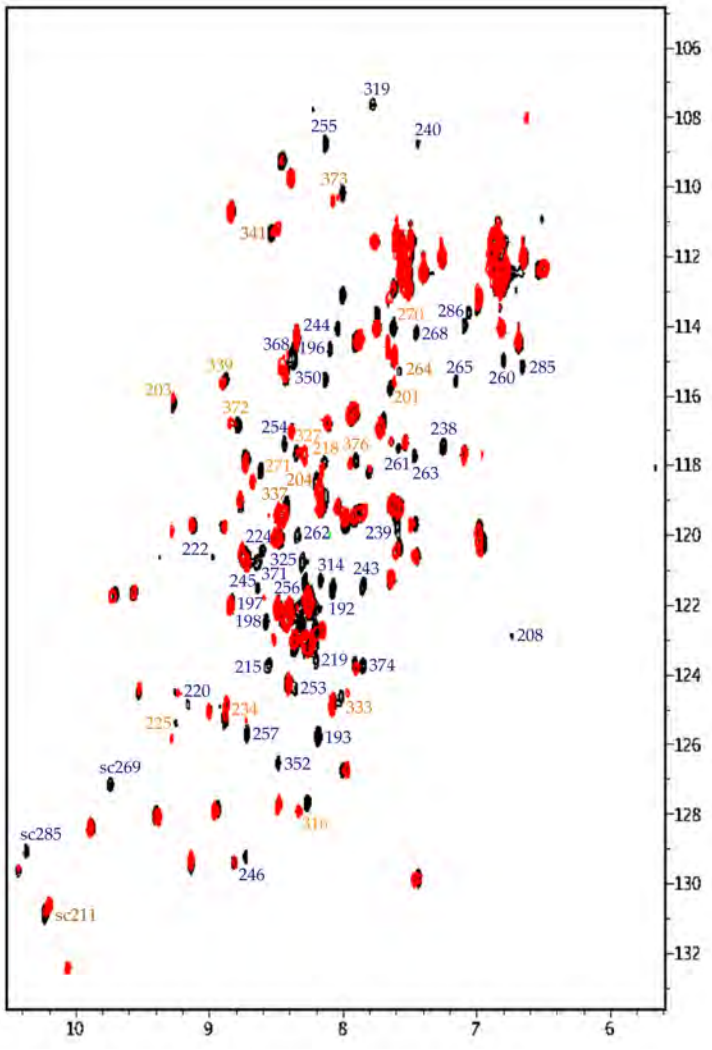


Figure 2.12

Figure 2.12 : Superimposition of the HSQC spectra of 15N-labeled A3G191-384-2K3A with 50µM Zn^{2+} (black) and 1mM Zn^{2+} (red). Residues numbers of signals that disappeared from spectrum at 1mM Zn^{2+} concentration are labeled in blue. Residues numbers of signals whose chemical shifts were moved at 1mM Zn^{2+} concentration are labeled in yellow.

Preface to : Chapter III

This chapter of the thesis is a manuscript *under preparation*

***Ming Li , *Shivender M.D. Shandilya ,** Michael A. Carpenter, Anurag Rathore, William L. Brown, Kuan-Ming Chen, Hiroshi Matsuo, Angela L. Perkins, Daniel A. Harki, Jonathan Solberg, Derek J. Hook, Jeff Johnson, Nevan J. Krogan, Mohan Somasundaran, Akbar Ali, Celia A. Schiffer, and Reuben S. Harris

***** These authors contributed equally to this work.

Author contributions : This work was a collaborative effort. R.S.H. and M.L. orchestrated all aspects of these studies and drafted the manuscript, **S.M.D.S.**, A.A., M.S. and C.A.S. contributed crystallographic and modeling data, J.J. and N.K. contributed MS data, K.C. and H.M. contributed NMR data, and J.S. and D.H. assisted with assay optimization and high-throughput screening. A.L.P. and D.A.H. performed the small molecule synthesis and conducted the glutathione reactivity experiments. A.R. and W.L.B. assisted with protein preparation and biochemical experiments. **All authors** contributed to data analyses, figure constructions, and manuscript revisions.

Chapter III

First-in-class small molecule inhibitors of the single-strand DNA cytidine deaminase APOBEC3G

Chapter III

First-in-class small molecule inhibitors of the single-strand DNA cytidine deaminase APOBEC3G

3.1 Abstract

The human APOBEC3G protein, a single-stranded DNA cytosine deaminase, functions as part of the innate immune response to inhibit virus and transposon replication. APOBEC3G is the prototype for the larger mammalian polynucleotide deaminase family, but no specific inhibitors exist to modulate this important enzyme. High-throughput screening identified 34 compounds that inhibit APOBEC3G catalytic activity. 19/34 small molecules contained catechol moieties, which are known to be sulfhydryl reactive following oxidation to the orthoquinone state. Located proximal to the active site, C321 was identified as the binding site for the inhibitors by a combination of structural analysis and mutational screening. Covalent binding of a representative inhibitor to C321 was verified by mass spectrometry. Bulkier substitutions C321-to-L, F, Y, or W were less active, mimicking chemical inhibition. Thus, a structural model is presented in which the mechanism of inhibition appears competitive, by sterically blocking substrate DNA cytosines from entering the APOBEC3G active site.

3.2 Introduction

APOBEC3G (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G; A3G) is a single-strand (ss)DNA cytosine to uracil (C-to-U) deaminase, which belongs to a larger family of polynucleotide DNA and RNA editing enzymes with a variety adaptive and innate immune functions (120, 131, 192). A3G has become the prototype for understanding the retrovirus and retrotransposon restriction activity of several family members, in large

part because it potently inhibits Vif-deficient HIV-1 replication. Current working models posit that A3G packages into assembling viral particle through a RNA-Gag interaction, travels with the particle until a new target cell becomes infected, and then interferes with viral cDNA synthesis by deamination-independent mechanisms4 (likely by binding viral genomic RNA and impeding reverse transcriptase progression) and deamination-dependent mechanism (115, 121, 161). The hallmark of A3G restriction is minus strand ssDNA C-to-U deamination events that are encoded as plus (genomic) strand G-to-A hypermutations. The predominant means by which HIV-1 can replicate in A3G-expressing cells is by expressing its accessory protein Vif, a natural antagonist of A3G that recruits an E3-ubiquitin ligation complex and promotes A3G degradation (76). The host-pathogen conflict between cellular APOBEC3s and lentiviral Vif is not specific to HIV-1, as strong evidence supports its existence in every other mammal that is infected with a HIV-related virus (i.e., many types of non-human primates, carnivores such as cats, and agricultural animals such as sheep and cattle (193–196)).

A3G is also the prototype for polynucleotide cytosine deaminase structure-function studies, because it is the only catalytically active family member with high resolution NMR and X-ray crystallography structures available to date (143–147,(and this study)). Multiple structures and a strong similarity to other family members have produced a range of scientifically useful models. However, the multiple interactions with A3G and both RNA and DNA polynucleotides during the biological activity of this enzyme make it imperative that these macromolecular associations become understood in as much detail as possible. Moreover, major fundamental questions remain unanswered, such as why are single strand DNA Cs and not RNA Cs deaminated and what is the oligomeric state of the anti-viral form of A3G? Here, we explore a new approach toward understanding A3G and its mechanism of DNA deamination by conducting high-throughput screens (HTS) to

identify small molecules that inhibit its C-to-U deaminase activity. We have identified a class of structurally similar compounds containing catechol moieties that react with an A3G catalytic domain cysteine and most likely inhibit DNA deamination by a competitive steric inhibition mechanism.

3.3 Results

3.3.1 Specific A3G inhibitors identified by HTS and sub-screening against the related DNA deaminase APOBEC3A

To screen for small molecule A3G inhibitors, we first optimized and miniaturized a fluorescence-based DNA deamination assay (189, 197, 198) (Figure 3.1). Full-length human A3G was purified from HEK293T cells as a myc-His6 epitope-tagged protein (Figure 3.21). To initiate the assay, recombinant enzyme is incubated with a single-stranded DNA (ssDNA) oligonucleotide containing a single target cytosine, a 6-FAM fluorophore at the 5' end, and TAMRA quenching molecule at the 3' terminus. Deamination of the target cytosine to uracil (C-to-U) is followed by uracil excision by E. coli uracil DNA glycosylase and subsequent phosphodiester backbone cleavage by hydroxide, which releases the 6-FAM fluorophore from the TAMRA quench. Deaminase activity is quantified directly with a fluorescence plate reader. Using DMSO as a negative control and the non-specific inhibitor aurintricarboxylic acid (ATA) found in preliminary screens as a positive control, the average Z-score in 384 well plates was 0.85 indicating that the assay is both robust and reproducible (FluorDeaminaseAssayWorks and Figure 3.20).

The 1280 compound library of pharmacologically active compounds (LOPAC, Sigma) was used to screen for chemical inhibitors of A3G. A total of 34 small molecules inhibited ssDNA deaminase activity by at least 40% in two independent experiments (raw 384 well plate data in Figure 3.2). Dose response curves were generated to re-confirm in-

hibitors and determine IC50 values, which mostly ranged from low- to mid-micromolar levels (representative data in Figure 3.3; complete dose response set in Figure 3.23). Compound structures are shown in Figure 3.4, and a summary of the 34 compounds relevant to this study can be found in Figure 3.5.

By analogy to kinase inhibitor studies, which often use related enzymes to gauge compound specificities (199–202), we performed secondary dose response screens using APOBEC3A (A3A), which is 65% identical (73% similar) at the amino acid level to the C-terminal catalytic domain of A3G (Figure 3.21). Surprisingly, the majority (24/34) of compounds failed to inhibit A3A, indicating a strong specificity for A3G (Figure 3.3 (B), (C) and Figure 3.5). As an additional specificity test, all candidate active compounds were tested against the second essential enzyme in our HTS reactions, uracil DNA glycosylase (UDG), by substituting uracil for the normal target cytosine in the ssDNA substrate (Figure 3.1, Figure 3.3 and Figure 3.5). Only ATA (MN1) blocked UDG activity, concordant with prior studies showing that this compound is a non-specific nucleic acid enzyme inhibitor (203–205) (Figure 3.3 (A) and Figure 3.5).

A total of 24 A3G-specific inhibitors were identified in LOPAC screens. One obvious chemical trend emerged with 19/24 of these small molecules containing a common catechol (ortho-dihydroxyphenyl) structure (Figure 3.5). Approximately 1% of all known drugs possess a catechol pharmacophore, including at least 17 FDA-approved drugs (206). Two prominent members of this family, apomorphine (MN2) and carbidopa (MN9), were identified as A3G inhibitors. Based on this conserved structural feature and the established clinical relevance of catechol pharmacophores, we focused additional studies on this class of small molecule inhibitors with emphasis on methyl-3,4-dephostatin (MN30), which re-confirmed through commercial sources and de novo synthesis (Figure 3.3 (C), Figure 3.4, Figure 3.12, and Figure 3.19).

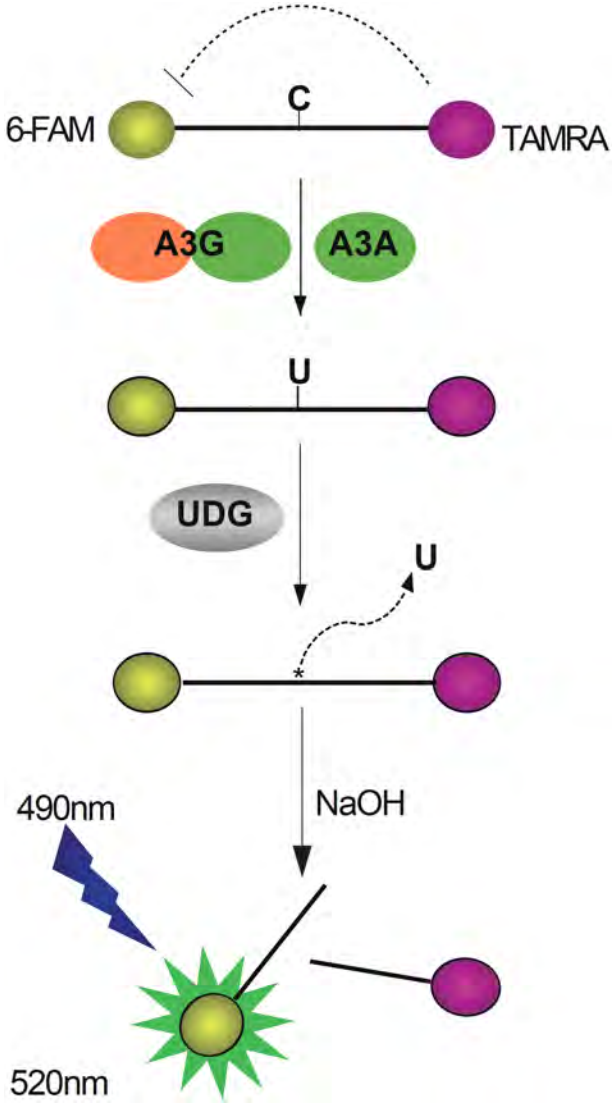


Figure 3.1

Figure 3.1 : Schematic of the fluorescence-based single-strand DNA deamination assay. APOBEC3G or APOBEC3A deaminates the cytosine in the substrate ssDNA oligo to a uracil, which is excised by uracil DNA glycosylase (UDG). The remaining abasic site is cleaved by basic hydrolysis so that the 5' labeled 6-FAM fluorophore is released from 3' labeled TAMRA quencher and emits fluorescence at 520nm under excitation at 490 nm.

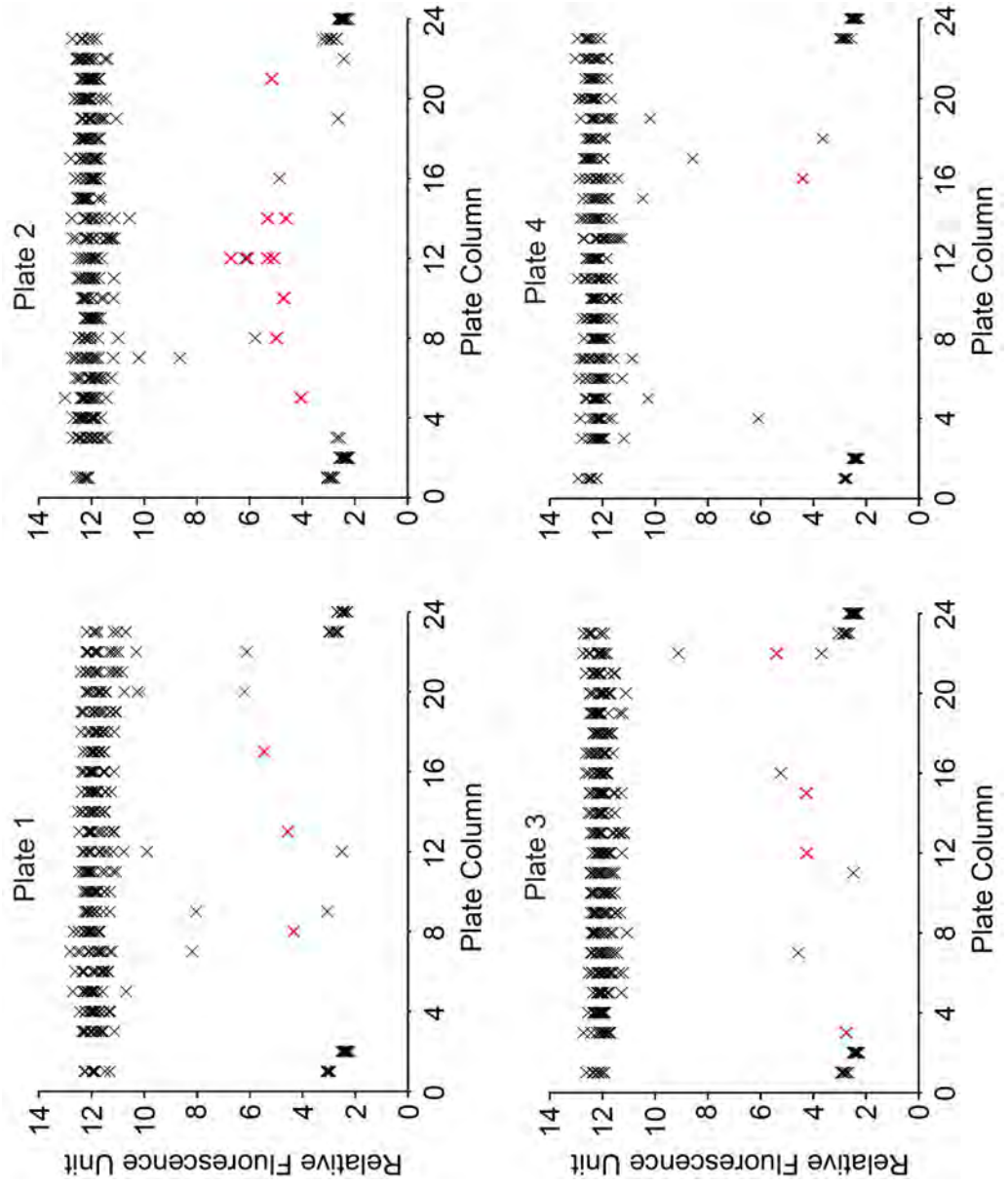


Plate 1
Plate 2
Plate 3
Plate 4
Relative Fluorescence Unit
Plate Column
0
2
4
6
8
10
12
14
16
20
24

Figure 3.2

Figure 3.2 : Representative High-Throughput-Screening (HTS) data. The screening with 1,280 LOPAC compounds was run in four 384 well plates. 40.5 nM of A3G, 6.7×10^{-4} Unit of UDG, 333.3 nM of A3G substrate oligo (5'-6-FAM-AAA-CCC-AAA-GAG-AGA-ATG-TGA-TAMRA-3') and 10 μM of each compound were transferred into the reaction plates. The first and last two columns in each LOPAC 384 well plate were filled with negative controls (DMSO), positive controls (10 μM of ATA) or non-enzyme controls. Each X represents a single data point and the red-shaded X's signify confirmed inhibitors.

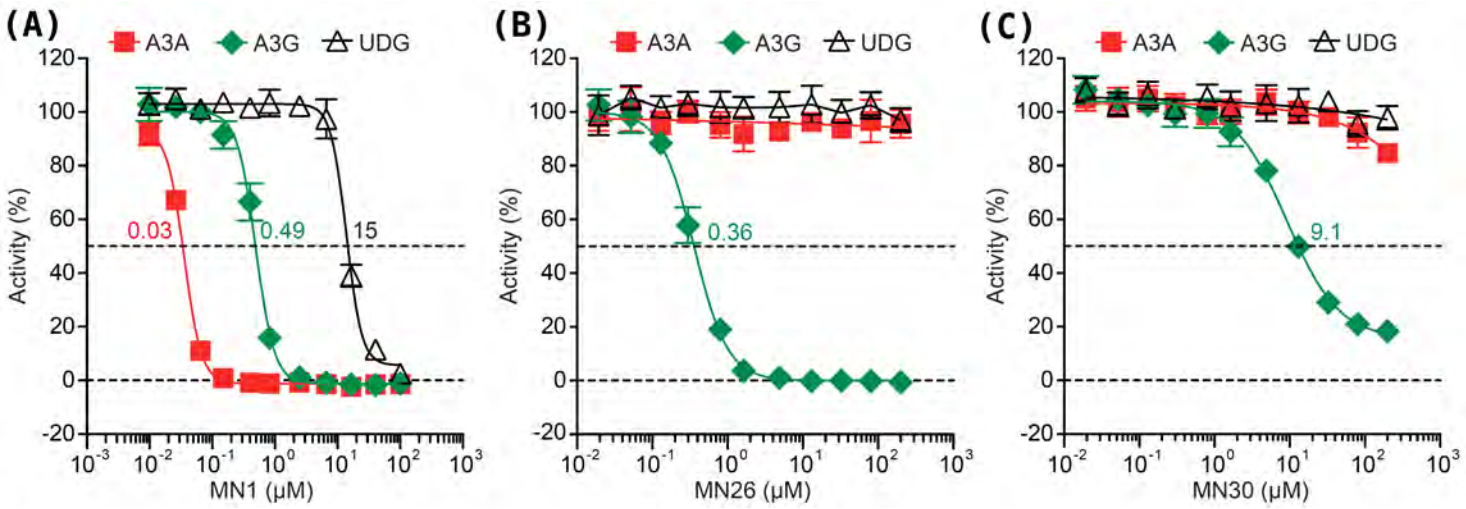

(A)
A3A
A3G
UDG
Activity (%)
0.03
0.49
15
MN1 (μM)
(B)
A3A
A3G
UDG
Activity (%)
0.36
MN26 (μM)
(C)
A3A
A3G
UDG
Activity (%)
9.1
MN30 (μM)

Figure 3.3

Figure 3.3 : Dose response assay of three compounds MN1 (A), MN26 (B) and MN30 (C) against A3G, A3A and UDG were run with dual-substrate oligo (5'-6-(FAM)-AAA-TAT-TCC-CTA-ATA-GAT-AAT-GTG-A-(TAMRA)-3') or UDG substrate oligo (5'-6-FAM-AAA-CCU-AAA-GAG-AGA-ATG-TGA-TAMRA-3'). IC_{50} values are listed next to the corresponding curves.

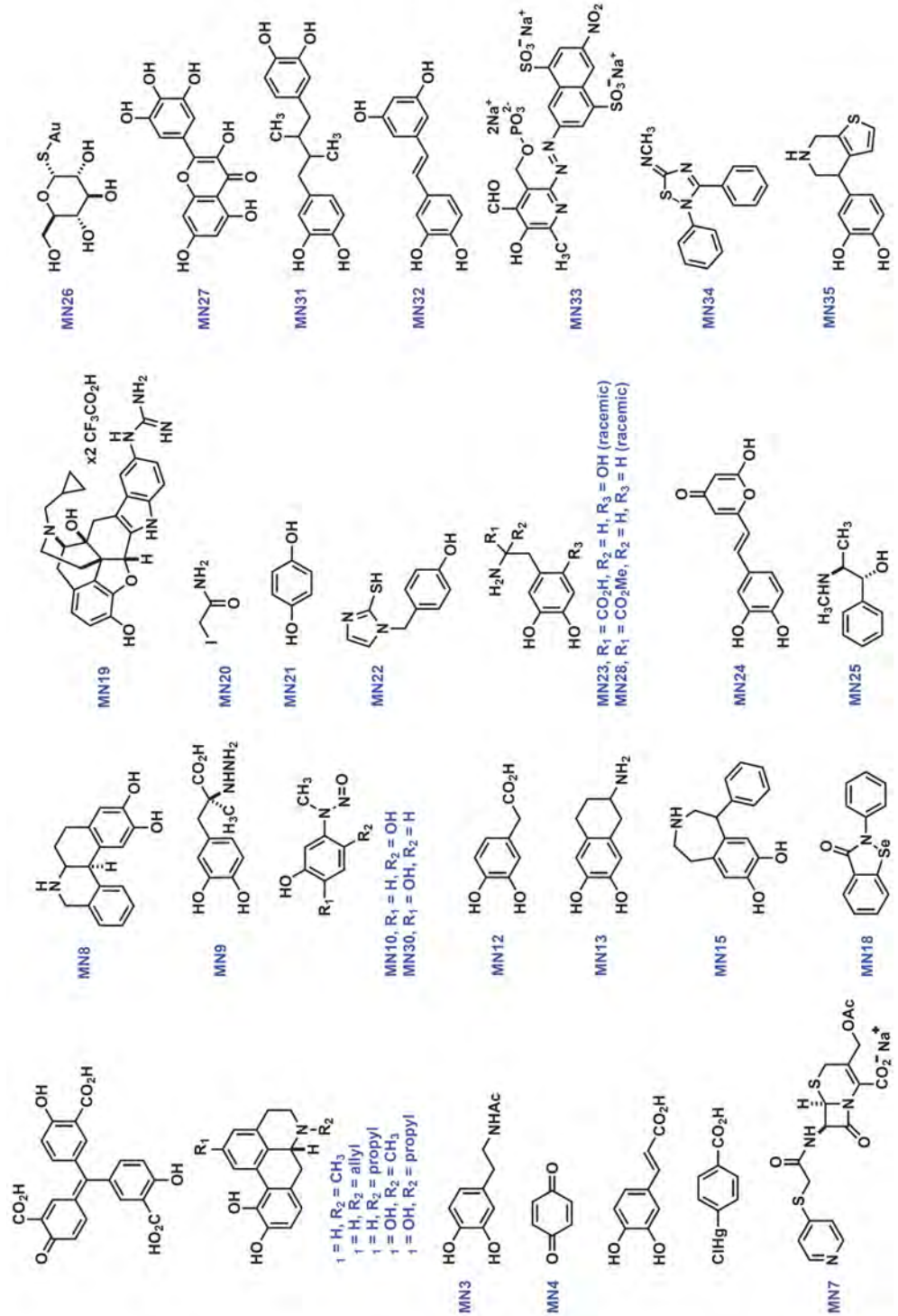

MN26
MN27
MN31
MN32
MN33
MN34
MN35
MN19
MN20
MN21
MN22
MN23, R1 = CO2H, R2 = H, R3 = OH (racemic)
MN28, R1 = CO2Me, R2 = H, R3 = H (racemic)
MN24
MN25
MN8
MN9
MN10, R1 = H, R2 = OH
MN30, R1 = OH, R2 = H
MN12
MN13
MN15
MN18
MN3
MN4
MN7
R1 = H, R2 = CH3
R1 = H, R2 = allyl
R1 = H, R2 = propyl
R1 = OH, R2 = CH3
R1 = OH, R2 = propyl

Figure 3.4

Figure 3.4 : Representative APOBEC3G inhibitors identified using the HTS assay.

Entry	Common Name	Linear Formula	MW	IC50 Values (μM)		
				A3A	A3G	UDG
MN1	Aurintricarboxylic acid	C22H14O9	422.34	0.03	0.49	14.81
MN2	R-(−)-Apomorphine hydrochloride hemihydrate	$C_{17}H_{17}NO_2 \cdot HCl \cdot 1/2H_{2O}$	312.79	>200	2.6	>200
MN3	N-Acetyldopamine monohydrate	$C_{10}H_{13}NO_3 \cdot H_2O$	213.23	>200	0.5	ND*
MN4	p-Benzoquinone	C6H4(=O)2	108.09	>200	0.2	
MN5	Caffeic acid	$HO)_2C_6H_3CH=CHCO_2H$	180.16	>200	85.4	>200
MN6	4-Chloromercuribenzoic acid	ClHgC6H4CO2H	357.16	0.26	0.13	ND
MN7	Cephapirin sodium salt	$C_{17}H_{16}N_3O_6S_2Na$	445.45	12.0	7.5	ND
MN8	Dihydrexidine hydrochloride	$C_{17}H_{16}NO_2 \cdot HCl$	302.78	>200	0.59	>200
MN9	S-(−)-Carbidopa	$C_{10}H_{14}N_2O_4$	226.23	>200	5.26	ND
MN10	Dephostatin	$C_7H_8N_2O_3$	168.15	149.2	0.43	>200
MN11	R(−)-N-Allylnorapomorphine hydrobromide	C19H19NO2 · HBr	374.27	>200	2.9	ND
MN12	3,4-Dihydroxyphenylacetic acid	(HO)2C6H3CH2CO2H	168.15	>200	19.0	>200
MN13	(±)-2-Amino-6,7-dihydroxy-1,2,3,4-tetrahydronaphthalene hydrobromide	C10H13NO2 · HBr	260.13	>200	0.7	>200
MN14	R(−)-Propylnorapomorphine hydrochloride	C19H21NO2 · HCl	331.84	>200	6.4	ND
MN15	(±)-SKF-38393 hydrochloride, 6-PHENYL-4-AZOBICYCLO[5.4.0]UNDECA-7,9,11-TRIENE-9,10-DIOL	C16H17NO2 · HCl	291.77	>200	27.0	>200
MN16	R(−)-2,10,11-Trihydroxyaporphine hydrobromide	C17H17NO3 · HBr	364.23	>200	1.7	>200
MN17	R(−)-2,10,11-Trihydroxy-N-propyl-noraporphine hydrobromide hydrate	C19H21NO3 · HBr · xH2O	392.29 (anhydrous basis)	>200	13.3	ND
MN18	Ebselen	C13H9NOSe	274.18	30.6	2.8	>200
MN19	5′-Guanidinonaltrindole di(trifluoroacetate) salt hydrate	C27H29N5O3 · 2C2HF3O2 · xH2O	699.60 (anhydrous basis)	0.42	6.4	>200
MN20	Iodoacetamide	ICH2CONH2	184.96	>200	3.0	ND
MN21	Hydroquinone	C_6H_4-1,4-$(OH)_2$	110.11	>200	2.6	ND
MN22	1-(4-Hydroxybenzyl)imidazole-2-thio	$C_{10}H_{10}N_2OS$	206.26	>200	3.5	>200
MN23	6-Hydroxy-DL-DOPA	$C_9H_{11}NO_5$	213.19	0.30	4.0	>200
MN24	Hispidin	$C_{13}H_{10}O_5$	246.22	>200	2.0	>200
MN25	(1*R*,2*S*)-(−)-Ephedrine	$C_6H_5CH[CH(NHCH_3)CH_3]OH$	165.23	~54	1.3	ND
MN26	Aurothioglucose hydrate	$C_6H_{11}AuO_5S \cdot xH_2O$	392.18 (anhydrous basis)	>200	0.4	>200
MN27	Myricetin	$C_{15}H_{10}O_8$	318.24	0.60	3.4	>200
MN28	(±)−3-(3,4-Dihydroxyphenyl)-2-methyl-DL-alanine	$C_{10}H_{13}NO_4 \cdot 1.5H_2O$	238.24	>200	13.3	>200
MN30	Methyl-3,4-dephostatin	$C_7H_8N_2O_3$	168.15	>200	9.1	>200
MN31	Nordihydroguaiaretic acid	$[-CH(CH_3)CH_2C_6H_3$-1,2-$(OH)_2]_2$	302.36	>200	8.8	ND
MN32	Piceatannol	$C_{14}H_{12}O_4$	244.24	200.6	1.8	>200
MN33	PPNDS tetrasodium	C18H11N4O14PS2 · Na4	694.36	3.3	5.6	
MN34	SCH-202676 hydrobromide	$C_{15}H_{16}BrN_3S$	350.28	307	~28.8	>200
MN35	SKF-89626 hydrobromide	$C_{13}H_{13}NO_2S \cdot HBr$	328.22	>200	1.3	>200

*ND, not determined

Figure 3.5

Figure 3.5 : APOBEC3G inhibitor data.

3.3.2 DNA binding assays suggest two mechanisms of A3G inhibition

To begin to functionally group the A3G inhibitors, we performed a series of ssDNA binding experiments using electrophoretic mobility shift assays (EMSAs). Prior studies have shown that the A3G holoenzyme binds ssDNA with an affinities ranging from 50μM to 70μM and that this activity mostly emanates from the N-terminal pseudocatalytic domain, rather than the C-terminal catalytic domain (126, 127, 207). The ssDNA binding activity of A3G is evident by several higher mass ssDNA-protein complexes visible in native gels (207–209). Interestingly, a 80-fold molar excess of inhibitor (50 μM) enabled a rough partition of all molecules into one of two general classes, those that block A3G from binding ssDNA and those that do not (Figure 3.6). Compounds representative of each class were re-confirmed in dose response EMSAs (Figure 3.7). All but two of the repurchased catechol analogues showed no inhibition of A3G's ssDNA binding activity suggesting that they function by binding within the C-terminal catalytic domain (e.g. MN30). The only exceptions, MN10 and MN35, might also bind within the N-terminal domain. However, these and the other molecules that may function by inhibiting ssDNA binding were not considered further because we do not yet have the capacity to prepare sufficiently pure quantities of A3G N-terminal nucleic acid binding domain or holoenzyme for structural studies.

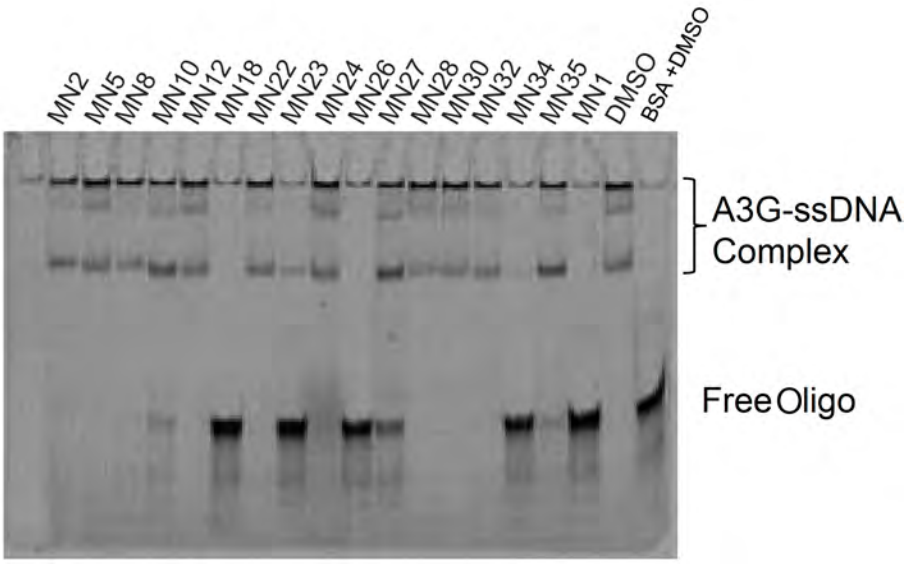

MN2
MN5
MN8
MN10
MN12
MN18
MN22
MN23
MN24
MN26
MN27
MN28
MN30
MN32
MN34
MN35
MN1
DMSO
BSA +DMSO
A3G-ssDNA
Complex
Free Oligo

Figure 3.6

Figure 3.6 : A3G-ssDNA complex formation is differentially affected by the indicated small molecules (0.64 µM A3G, 0.08 µM ssDNA, 50 µM compound). As controls, DMSO did not interfere with complex formation, and BSA did not cause complex formation.

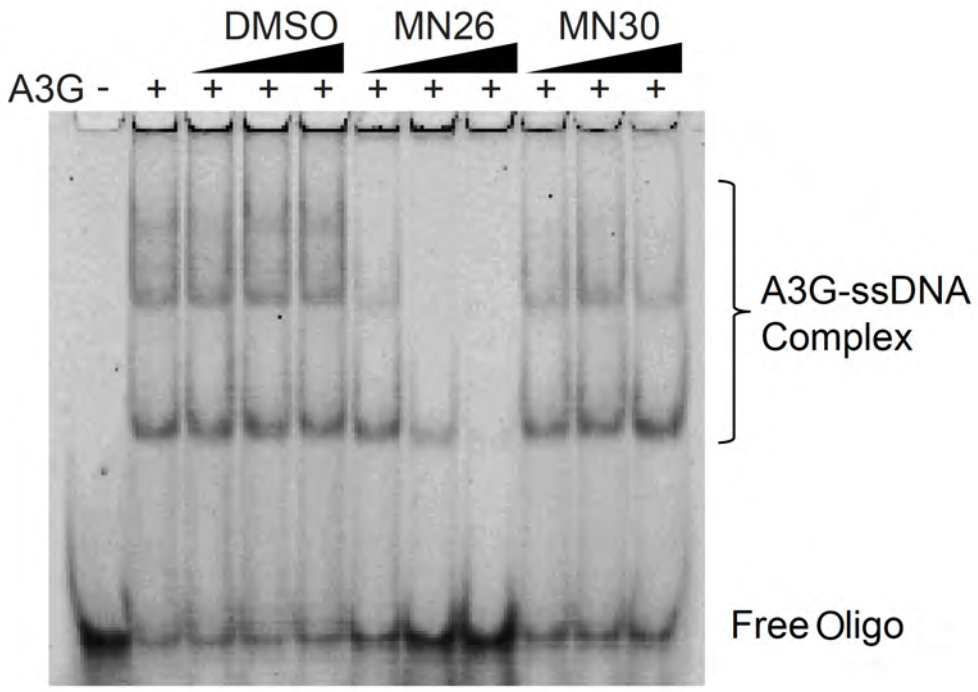


Figure 3.7

Figure 3.7 : Representative EMSAs with A3G, ssDNA, and 2, 5, and 10 μM of MN26 or MN30. Most catechol containing compounds such as MN30 do not influence A3G-ssDNA complex formation.

3.3.3 Crystal structure indicates covalent binding of MN30 to APOBEC3G

We next leveraged the fact that the A3G C-terminal catalytic domain is well investigated by structural studies using NMR spectroscopy and X-ray crystallography (143–147). Several of the A3G inhibitors indentified above were used for co-crystallization experiments with A3G-191-384-2K3A (L234K, C243A, F310K, C321A and C356A) (145, 147). A3G-191-384-2K3A crystals were soaked with inhibitors and diffraction data obtained to 2.5Å. Positive electron density was observed in the Fo-Fc map adjacent to C308 for MN30-soaked crystals (Figure 3.8, Figure 3.9 and Figure 3.12). The refinement statistics and electron density maps were improved by modeling the MN30 chemical structure onto the C308 region (Figure 3.9). Surprisingly, the 1.7Å estimated distance between the sulfhydryl group of C308 and the 5-position carbon of MN30 suggested a covalent linkage. Additional experiments with MN30, MN30 analogues, and glutathione as a sulfhydryl substrate (one of the most abundant in cells) demonstrated that MN30 autooxidizes under aerobic (O_2) conditions to the orthoquinone and then most likely forms a covalent bond with cysteine through a Michael addition reaction (Figure 3.10 and S4). These results demonstrate that MN30 can indeed form a covalent linkage with the sulfhydryl group of cysteine, and specifically with at least one cysteine within the A3G catalytic domain. However, we were surprised to find in parallel biochemical experiments that an A3G-C308A mutant still showed wildtype susceptibility to MN30 inhibition (Fig. 4d), strongly suggesting that our crystal-soaking experiments identified a secondary binding site.

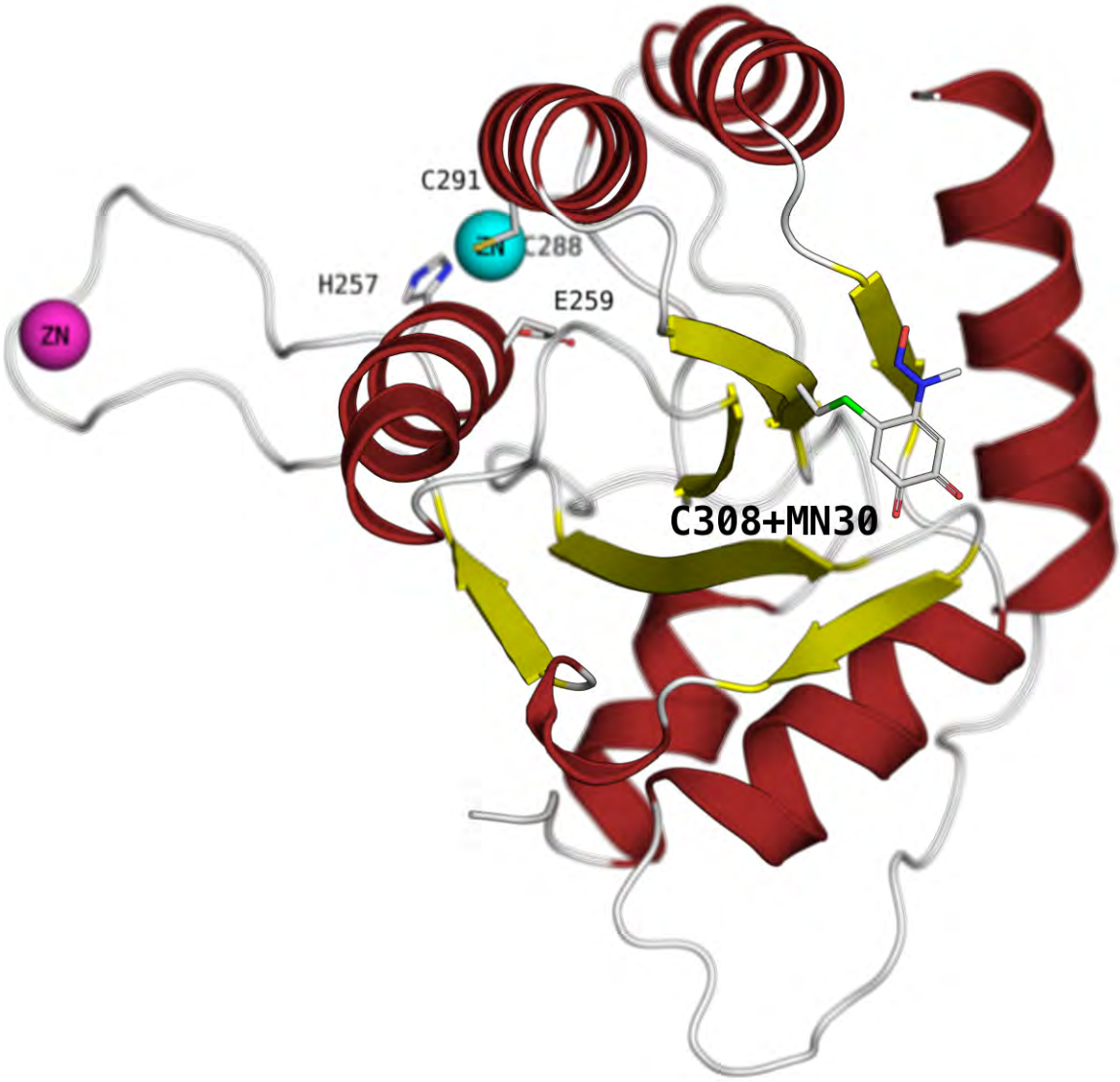


Figure 3.8

Figure 3.8 : Crystal structure of A3G-191-384-2K3A displaying the MN30 binding region. The C308+MN30 adduct is displayed as a ball-and-stick representation with the C308 thiol sulfur colored green. α-helices are colored red, β-sheets are colored yellow and the active site Zn^{2+} and intermolecular Zn^{2+} atoms are represented as cyan and magenta spheres respectively. Active site Zn^{2+} coordinating residues and the catalytic E259 are shown in a ball-and-stick representation.

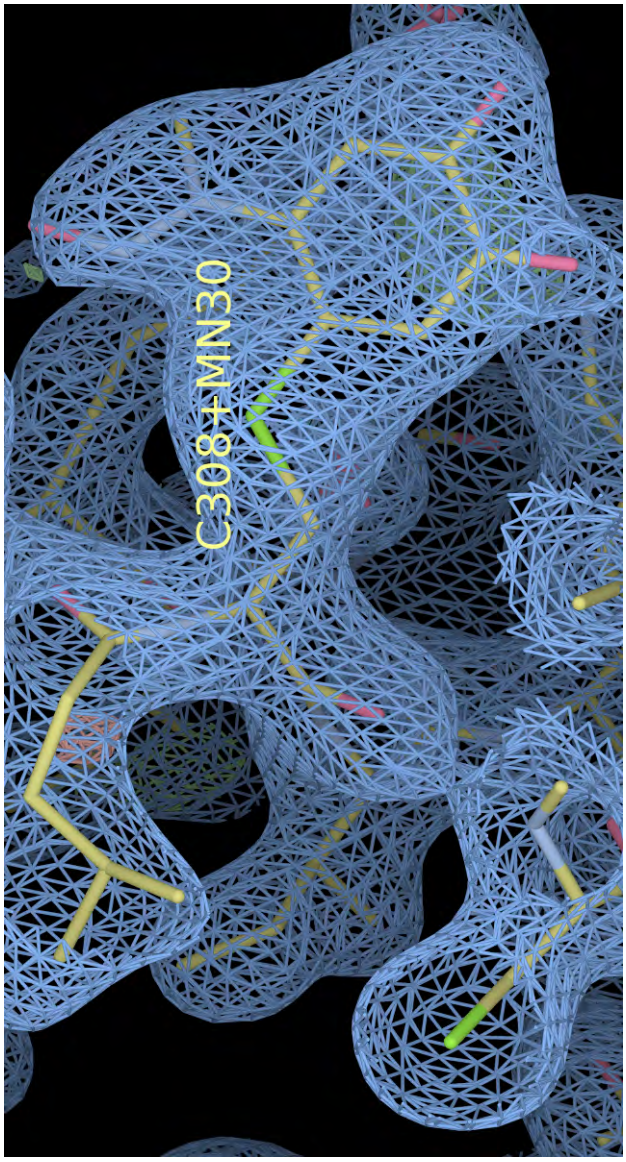


Figure 3.9

Figure 3.9 : The 2Fo-Fc electron density map (blue mesh, contoured at sigma level 1.0), at the C308+MN30 adduct region in the A3G-191-384-2K3A crystal structure. Atoms and bonds are represented as sticks, with the C308 thiol sulfur colored green.

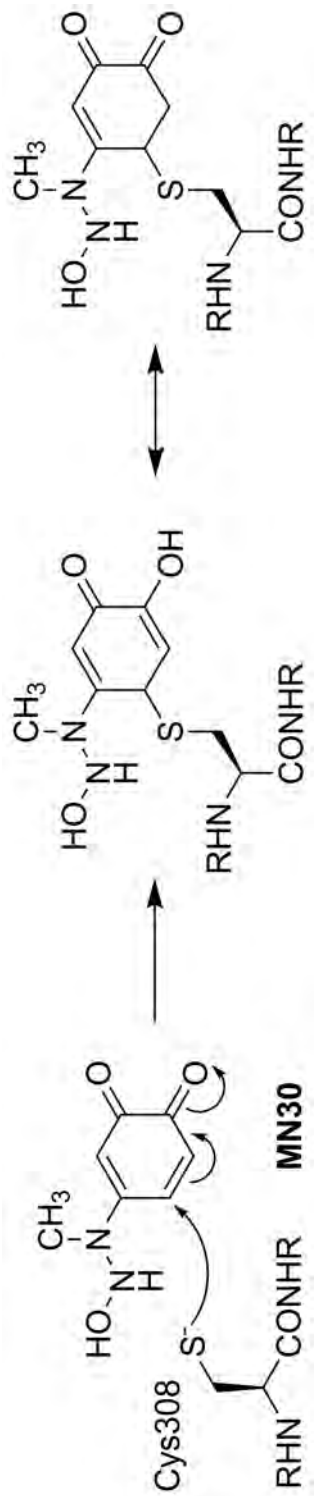


Figure 3.10

Figure 3.10 : Chemical schematic showing the Michael's addition reaction leading to formation of MN30 adduct with a Cysteine residue.

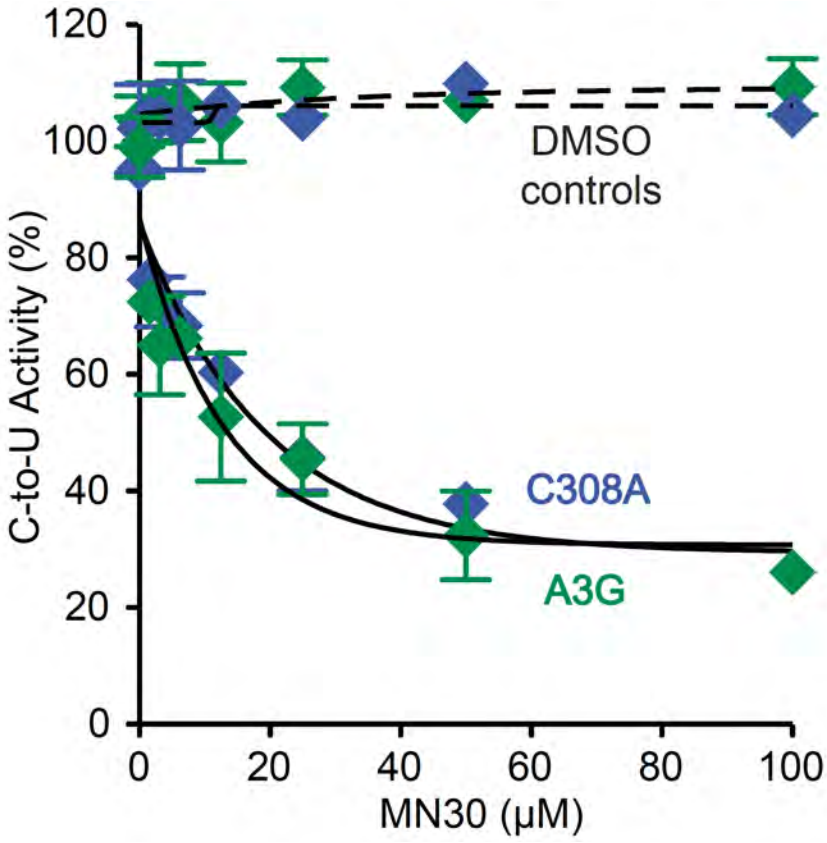

120
100
80
60
40
20
0
C-to-U Activity (%)
DMSO
controls
C308A
A3G
0
20
40
60
80
100
MN30 (μM)

Figure 3.11

Figure 3.11 : A3G C308A mutation did not affect the deaminase activity. The full length A3G-MycHis and A3G C308A-MycHis proteins showed similar activity and were both inhibited by MN30.

	A3G191-384-2K3A	A3G191-380-2K2A
PDB ID	----	----
Resolution	2.04 Å	1.38 Å
Temperature	Cryogenic (-80° C)	Cryogenic (-80° C)
Space group	$P12_11$	$P2_12_12_1$
Cell dimensions :		
a	53.08 Å	68.19 Å
b	67.76 Å	72.17 Å
c	64.47 Å	96.81 Å
α	90.0°	90.0°
β	107.38°	90.0°
γ	90.0°	90.0°
Molecules in AU	2	2
Completeness	96.5% (94.3%) (Last shell 2.04 – 2.11)	93.7% (92.6%) (Last shell 1.38 – 1.41)
Total reflections	26969	280403
Unique reflections	28034 25587 (working)	99113 86726 (working)
I/σ	13.6	12.3
Ave. redundancy	5.0	3.0
Rmerge	5.6%	4.8%
RMSD in:		3.3%
Bonds	0.0096 Å	0.0095 Å
Angles	1.337°	1.409°
Rfactor %	23.11	18.26
Rfree %	28.37	21.47

Figure 3.12

Figure 3.12 : Crystallographic statistics for the A3G-191-384-2K3A (MN30 + Cys308) and A3G191-380-2K2A (apo) crystal structures.

3.3.4 Alanine mutagenesis revealed C321 as the A3G inhibitor-binding site

The A3G catalytic domain has four surface cysteines C243, C308, C321 and C356, in addition to partially exposed C288 and C291 that coordinate the active site zinc. The A3G catalytic domain variant used above for our initial studies, A3G-191-384-2K3A, has three cysteine to alanine substitutions, C243A, C321A and C356A. In the context of the full-length A3G protein, these three substitutions have no discernable effect on localization, deamination, oligomerization, or Vif-deficient HIV-1 restriction activities (145, 147). This full-length construct was purified and tested in parallel with wildtype enzyme against MN30. Interestingly, this mutant enzyme was fully resistant to MN30 and all other catechol-based inhibitors obtained from commercial sources (MN2, 5, 8, 10, 12, 13, 15, 16, 24, 28, 30, 32 and 35; MN3, 9, 11, 14, 17 and 31 were not tested) (Figure 3.13). Thus, one of the three cysteines mutated in the 2K3A construct was most likely the biological target of this class of A3G inhibitors.

To identify the inhibitor binding site, a series of three single cysteine to alanine mutants was constructed, purified to >80% homogeneity from HEK293 cells as myc-His tagged proteins, and tested against representative inhibitors in the fluorescence-based ssDNA C-to-U activity assay. Dose response data clearly revealed a single cysteine, C321, as the site targeted by MN30 and related catechol-type inhibitors, as this single amino acid substitution was alone sufficient to render A3G fully resistant to chemical inhibition (e.g., MN30 data in Figure 3.14). Although the C321A mutant showed nearly the same level of catalytic activity as wildtype A3G, variants with bulkier amino acid substitutions in place of C321 (to L, F, Y, and W) showed significantly diminished ssDNA deaminase activity, mimicking the effects of MN30 and demonstrating that this amino acid position can impact catalytic activity (Figure 3.24).

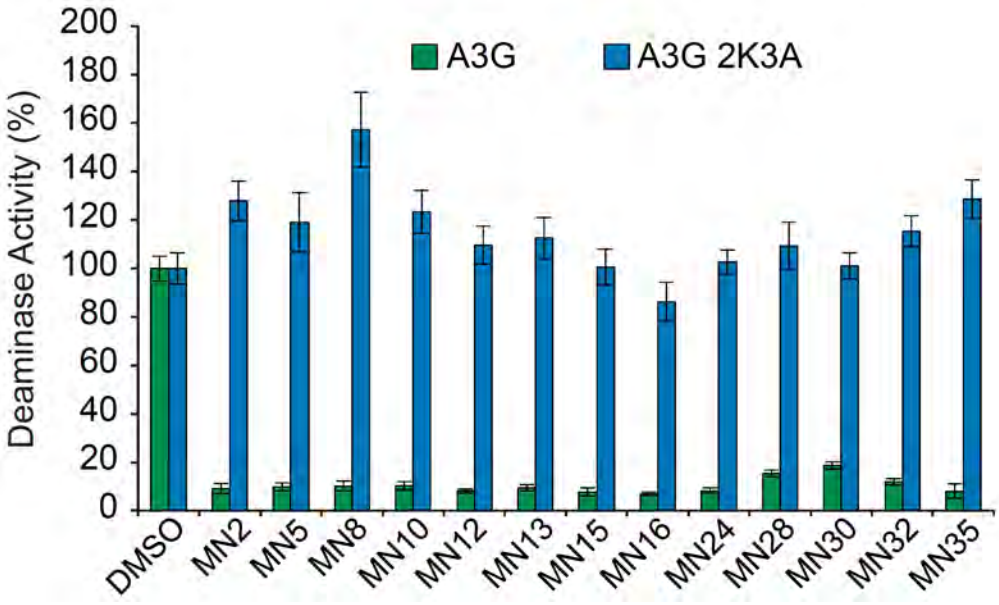

A3G
A3G 2K3A
Deaminase Activity (%)
200
180
160
140
120
100
80
60
40
20
0
DMSO
MN2
MN5
MN8
MN10
MN12
MN13
MN15
MN16
MN24
MN28
MN30
MN32
MN35

Figure 3.13

Figure 3.13 : Full-length A3G-2K3A completely resists inhibition by all tested catechol series inhibitors. The mean and SD of triplicate deaminase assays with 50 µM compound, 0.0675 µM A3G, 0.33 µM ssDNA, and excess UDG are shown relative to the DMSO only controls.

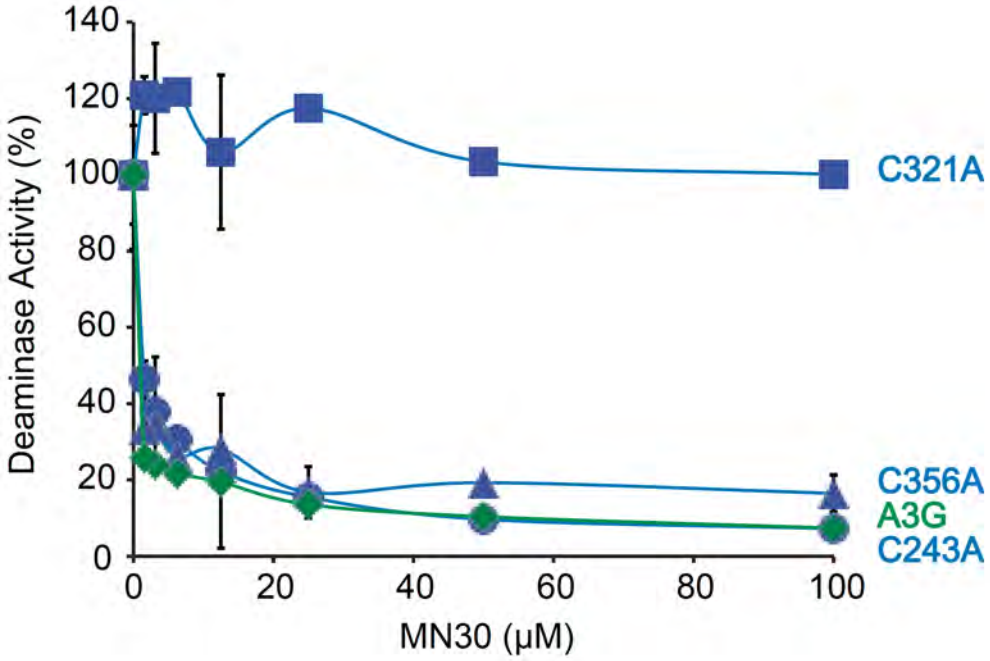

140
120
100
80
60
40
20
0
Deaminase Activity (%)
0
20
40
60
80
100
MN30 (μM)
C321A
C356A
A3G
C243A

Figure 3.14

Figure 3.14 : A single amino acid substitution C321A defines the binding site in A3G for catechol series inhibitors. Wildtype A3G and the indicated single alanine derivatives were purified as described in the methods and Figure 3.21 and assayed in parallel with varying concentrations of MN30. Data from triplicate assays were normalized to the DMSO only controls, and the relative mean activities are presented with SD. Several other catechol series molecules were also effective against A3G-C321A (data not shown).

3.3.5 Additional proof that MN30 forms a covalent bond with A3G Cys321

To directly ask whether MN30 forms a covalent bond with A3G, we incubated it with full length A3G, purified the protein by gel fractionation, performed in-gel digestion with trypsin, and analyzed the resulting peptides by mass spectrometry. The C321 containing tryptic peptide C*QEGLRTLAEAGAK showed a 167.11 Da addition demonstrating covalent modification by MN30 (Figure 3.15). Both 2+ and 3+ charge states of the peptide were detected and fragmented; the MS/MS spectra for both charge states resulted in highest-scoring identification that corresponded to the C*QGELRTLAEAGAK peptide. The dominant b-ion series observed in the MS/MS spectrum of the 3+ peptide suggests that MN30 is protonated in the gas phase, and the 167.11 Da mass reflects this. Similar LC-MS/MS results were obtained with MN30 and the A3G191-380-2K2A protein (see below and data not shown).

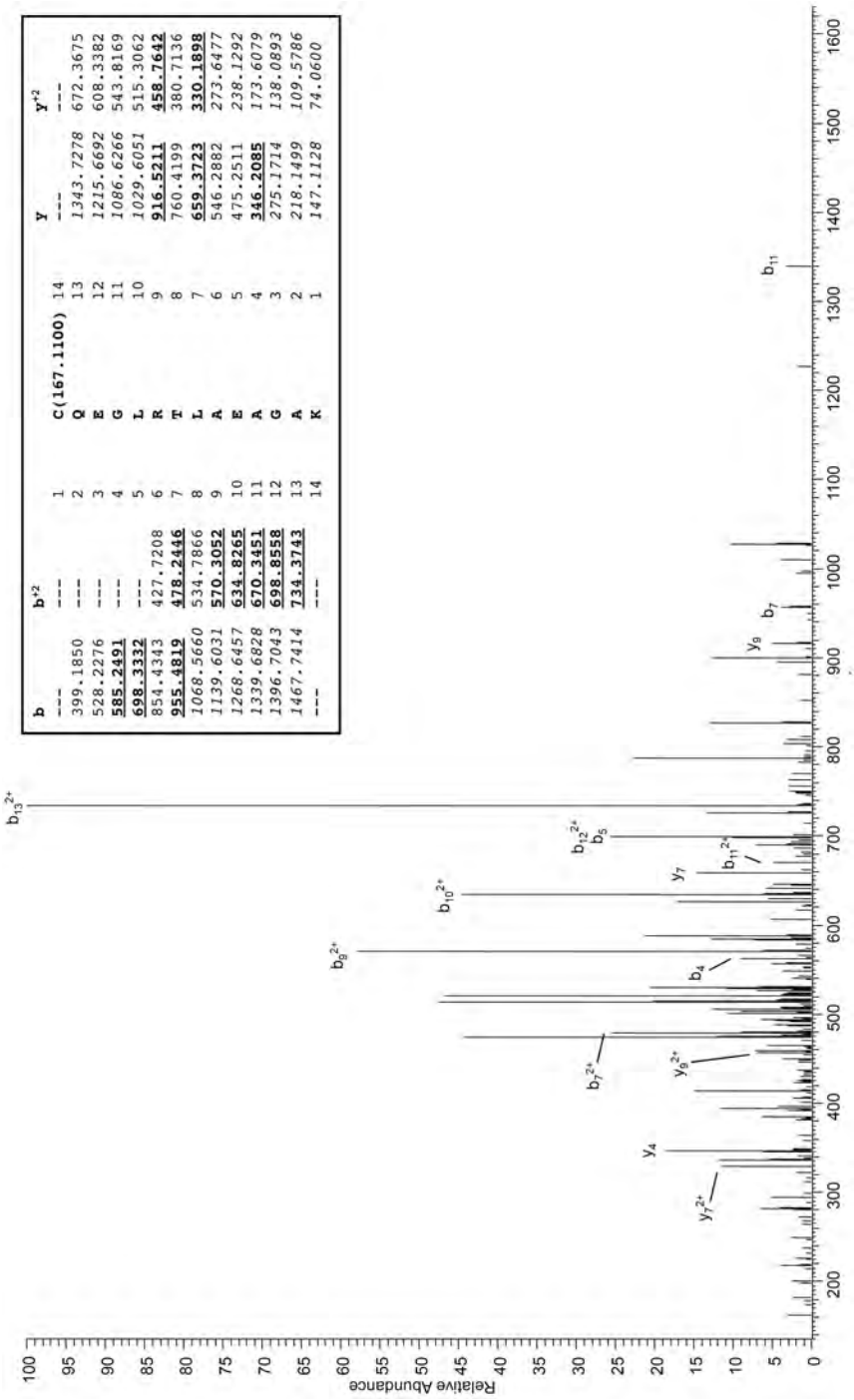

b
b+2
C(167.1100)
y
y+2
399.1850 --- 2 Q 13 1343.7278 672.3675
528.2276 --- 3 E 12 1215.6692 608.3382
585.2491 --- 4 G 11 1086.6266 543.8169
698.3332 --- 5 L 10 1029.6051 515.3062
854.4343 427.7208 6 R 9 916.5211 458.7642
955.4819 478.2446 7 T 8 760.4199 380.7136
1068.5660 534.7866 8 L 7 659.3723 330.1898
1139.6031 570.3052 9 A 6 546.2882 273.6477
1268.6457 634.8265 10 E 5 475.2511 238.1292
1339.6828 670.3451 11 A 4 346.2085 173.6079
1396.7043 698.8558 12 G 3 275.1714 138.0893
1467.7414 734.3743 13 A 2 218.1499 109.5786
--- --- 14 K 1 147.1128 74.0600
Relative Abundance
m/z

Figure 3.15

Figure 3.15 : The MS/MS spectrum of the peptide C(167.11)QEGLRTLAEAGAK is depicted with main b and y sequence ions annotated. The inset provides a table of predicted b and y ion m/z ratios with detected ions indicated (bold, underlined). Masses in italics in the table are not predicted to be present based on residues carrying charge in the peptide and/or the mass range of the instrument.

3.3.6 A structural model for competitive active site inhibition

To further understand the mechanism of MN30 action, a new A3G C-terminal domain construct was generated with C321 restored (A3G-191-380-2K2A). The "apo" crystal structure of A3G-191-380-2K2A was solved to a resolution of 1.38Å (Figure 3.16 and Figure 3.12). This is the highest resolution crystal structure solved to date for the A3G catalytic domain enabling full resolution of most amino acid side chains. As described previously, the enzyme is globular with a hydrophobic β-sheet core surrounded by six α-helices. Numerous attempts were also made to co-crystallize or soak MN30 into existing A3G-191-380-2K2A crystals, but these were unsuccessful. We suspect that upon binding C321, MN30 causes conformational changes that may no longer permit our current constructs to crystallize or it may be interfering with crystal packing.

Nevertheless, we took advantage of this new high-resolution structure to generate a lowest energy model of MN30 covalently bound to C321 (Figure 3.17 and Figure 3.18). This model suggests a mechanism of competitive inhibition through steric hindrance. The ring structure of MN30 is predicted to be oriented towards the catalytic residues and cause Y315 to flip such that it contacts W285, an essential residue lining the active site (143). The net result is a model in which MN30 causes partial active site occupation by Y315, which in turn is likely to prevent ssDNA cytosine from being bound and/or positioning appropriately for catalysis.

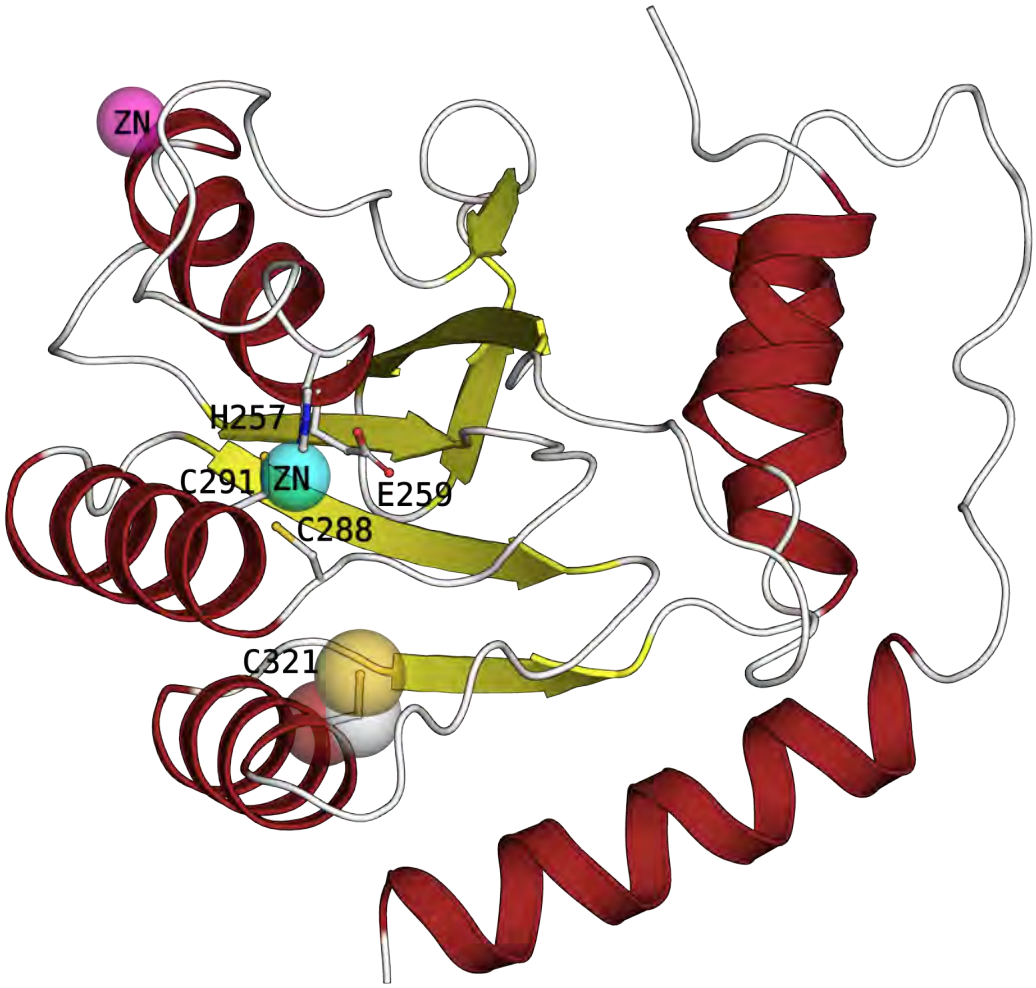


Figure 3.16

Figure 3.16 : Crystal structure of A3G-191-380-2K2A showing the location and orientation of C321, displayed in ball-and-stick and transparent CPK representations. α-helices are colored red, β-sheets are colored yellow and the active site Zn^{2+} and intermolecular Zn^{2+} atoms are represented as cyan and magenta spheres respectively. Active site Zn^{2+} coordinating residues and the catalytic E259 are shown in a ball-and-stick representation.

Figure 3.17 : The 2Fo-Fc electron density map (blue mesh, contoured at sigma level 1.0), around C321 and the active site region, in A3G-191-380-2K2A crystal structure. Atoms and bonds are represented as sticks.

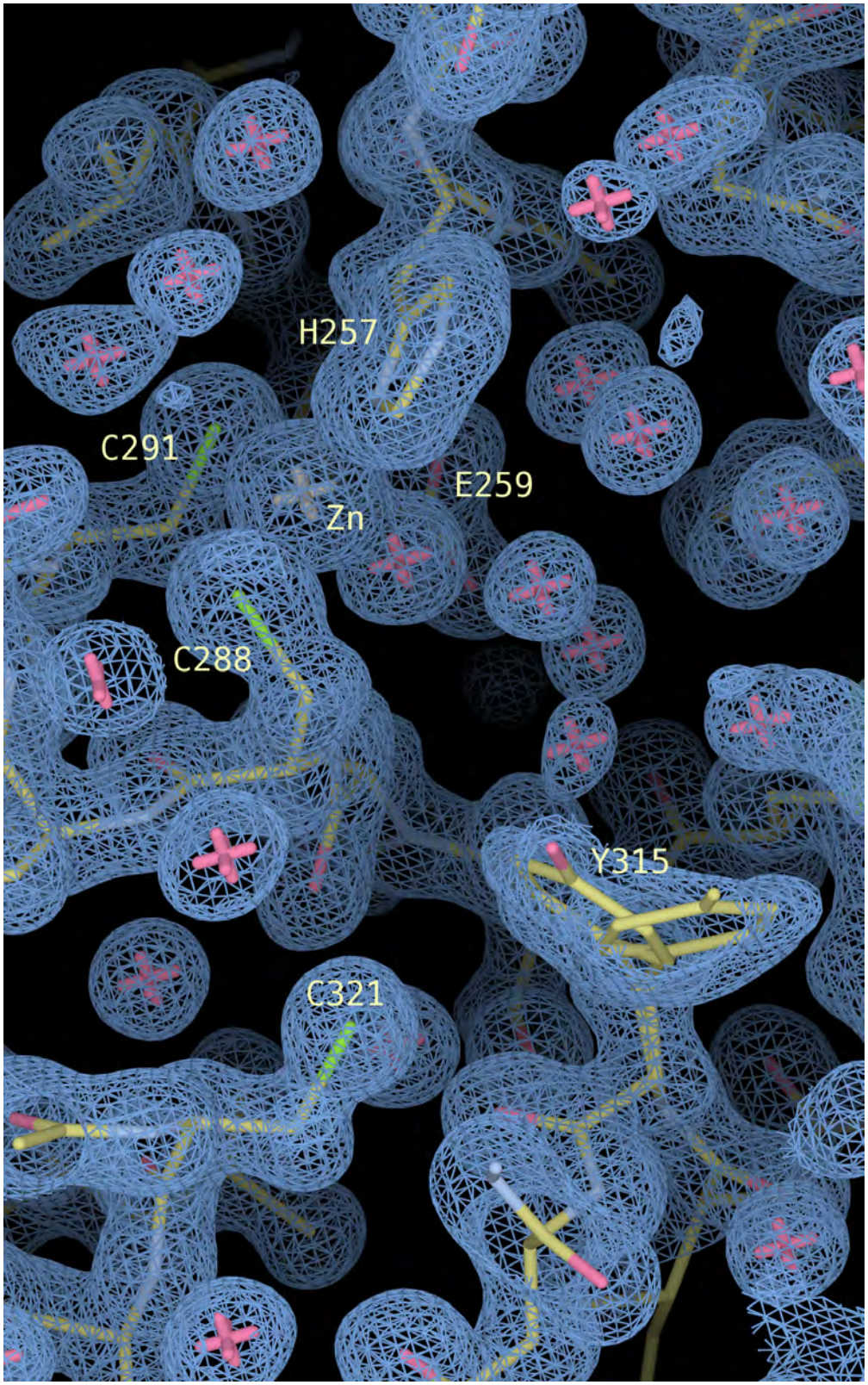


Figure 3.17

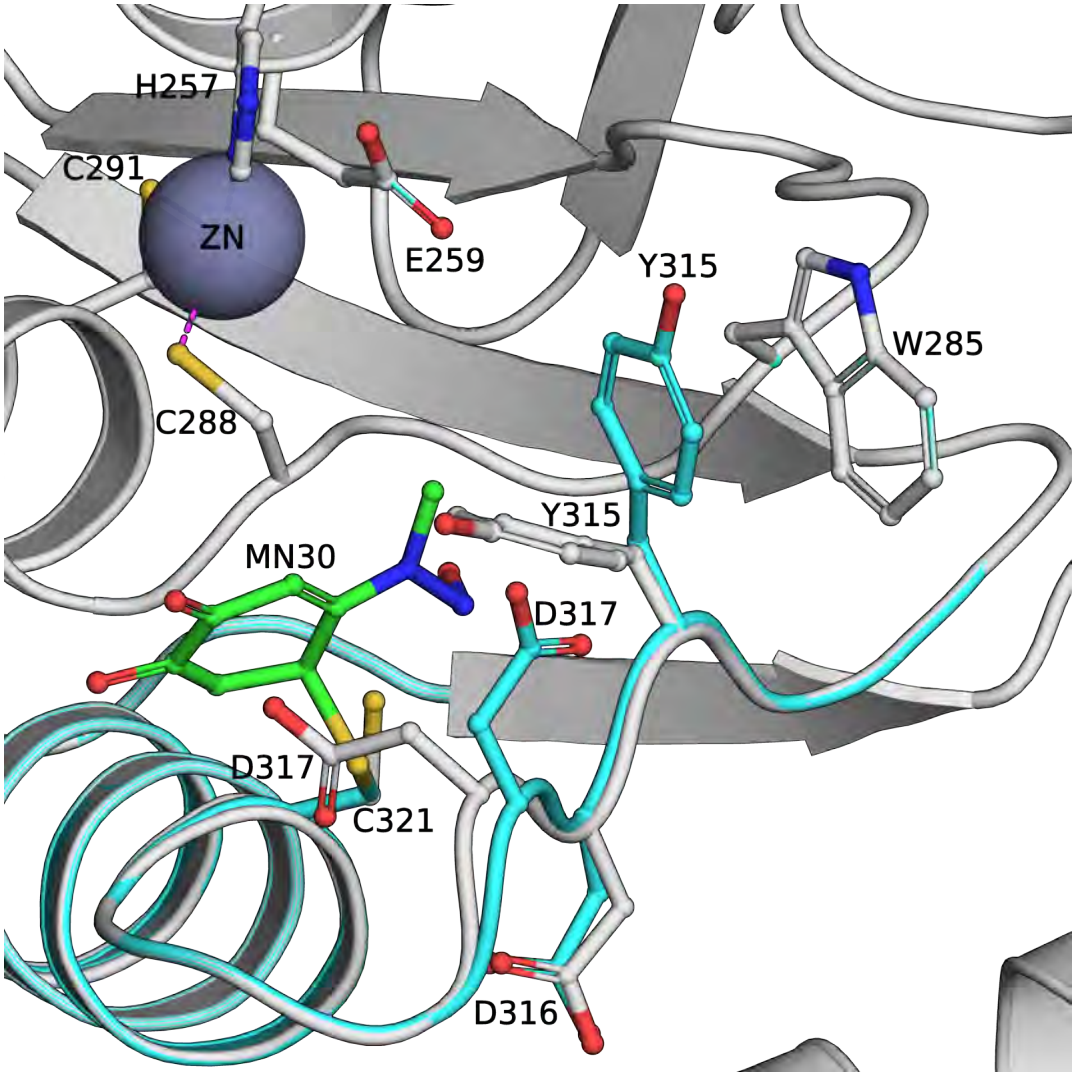

H257
C291
ZN
E259
Y315
W285
C288
Y315
MN30
D317
D317
C321
D316

Figure 3.18

Figure 3.18 : The MN30+C321 adduct model, based on the A3G-191-380-2K2A crystal structure. The final model is displayed as cyan colored ribbons and ball-and-stick, with MN30 colored green, superimposed onto the A3G-191-380-2K2A crystal structure (white ribbons). Identical residues are displayed in both the model and the crystal structure and labeled accordingly.

3.4 Discussion

APOBEC3G is the prototypical enzyme for understanding the molecular biology, biochemistry, and the structural biology of DNA deaminase-mediated mechanisms of innate immunity to a variety of transposons and viruses including the retrovirus HIV-1. Here, we report the results of HTS and follow-up studies that culminate in identifying the first 36 small molecule inhibitors of A3G deaminase activity. 22 of these compounds appear to be A3G-specific inhibitors, because they do not block the activity of the related DNA deaminase A3A or the unrelated E. coli uracil DNA glycosylase. Remarkably, 18/22 A3G-specific compounds contain a catechol (ortho-dihydroxyphenyl) structure, which is a chemical feature of approximately 1% of all known drugs (206). A combination of genetic and physical techniques, with emphasis on methyl-3,4-dephostatin (MN30), combined to demonstrate that this class of small molecules blocks A3G activity by covalently linking to the C321 sulfhydryl group and thereby sterically interfering with active site amino acid residues (particularly the essential residues Y315 and W285 (143)).

The fact that MN30 and most likely all of the other catechol-type compounds described here can react with two cysteines in A3G (C308 and C321) and one in glutathione suggests that these small molecules are not completely specific. However, the fact that they do not inhibit the related deaminase A3A (5 cysteines) or the second enzyme in our deaminase assays uracil DNA glycosylase (1 cysteine) suggests that these compounds may at least be semi-specific. It is likely that the cysteine reactivity information can be used together with the new 1.38Å A3G catalytic domain crystal structure also described here to perform virtual docking and molecular dynamic simulations of related catechol containing molecules to identify a subset with improved specificity for the A3G active site.

Interestingly, methyl-3,4-dephostatin (MN30) is a rationally-designed analogue of dephostatin (MN10). Dephostatin is a natural product isolated from Streptomyces that in-

terferes with protein tyrosine phosphatase activity (210–212), but it proved too unstable for most biological applications (213). In an effort to identify dephostatin analogues with enhanced potency, methyl-3,4-dephostatin (MN30) was synthesized and found to possess similar protein phosphatase inhibitory activity as compared to its parent natural product MN10 (214). The molecular mechanisms of MN30-binding to protein phosphatases have been well studied and this compound has demonstrated efficacy in a murine model of diabetes following oral dosing (215). Our future studies will be directed toward expanding DNA deaminase inhibitor HTS, directed medicinal chemistry, and using MN30 and derivative compounds as molecular probes to further dissect the innate immune and antiviral functions of A3G.

3.5 Experimental Procedures

Proteins:

A3G-myc-His and A3A-myc-His were purified from stably transfected or transiently transfected HEK293T cells using the C-terminal hexahistine tags, as described (197, 198).

C-to-U assay:

The fluorescence-based assay has been reported and refined (189, 197, 198). Recombinant human A3A-myc-His and A3G-myc-His proteins were diluted with 50 mM Tris.Cl, pH 7.4, 150 mM NaCl, 10% Glycerol and 0.5% Triton X-100 to intent concentration. 20 µl of the solution were incubated with 10 pmol substrate DNA oligo (Biosearch Technologies, Inc.) and 0.02 unit of UDG (NEB) diluted in 20 µl of 50 mM Tris.Cl, pH7.4, 10 mM EDTA for 2 h at 37 °C in Nunc 384-well black plates. 3 µl of 4N NaOH was added, followed by mixing and incubating at 37°C for another 30 min. 3 µl 4 N HCl and 27 µl 2M Tris.Cl (pH7.9) was then added for neutralization and the relative deaminase activity was quantified by reading fluorescence with excitation at 490 nm and emission at 520 nm on Synergy

Mx Monochromator-Based Multi-Mode (BioTek Instruments, Inc.) or LJL Analyst AD (LJL BioSystems, Inc.) microplate readers.

HTS including specifics on instrumentation:

(i) Compound transfer: 40 nl of library compound (10 mM in DMSO), aurintricarboxylic acid (ATA, 10 mM in DMSO; positive control), or DMSO (negative control) were tranfsered into Nunc 384 well plate (#262260) using an ECHO 550 Liquid Handler (Labcyte Inc.).

(ii) Protein addition: Add 20 µl of recombinant A3A-Myc-His A3G-Myc-His assay solution to each well of the 384 plate using a Perkin Elmer Flexdrop Reagent Dispenser; mix by shaking on a DPC micromix shaker for 1min (form 20, amplitude 7), and incubate for 15 min at 37°C.

(iii) Substrate addition: Add 20 µl single-strand DNA oligo solution with E.coli UDG to each well of 384 well plate using a Perkin Elmer Flexdrop Reagent Dispenser; mix by shaking on a DPC micromix shaker for 1 min (form 20, amplitude 7), and incubate for 2 hours at 37°C.

(iv) Product cleavage: Add 3 µl of cleavage solution (4N NaOH) using a Perkin Elmer Flexdrop Reagent Dispenser, mix by shaking on a DPC micromix shaker for 1 min (form 20, amplitude 7), and incubate for 30 minutes at 37°C.

(v) Reaction neutralization: Add 30 µl of neutralization solution (1.8M Tris-Cl (pH 7.9), 0.4N HCl) using a Perkin Elmer Flexdrop Reagent Dispenser; mix by shaking on a DPC micromix shaker for 1 min (form 20, amplitude 7) at room temperature.

(vi) Read-out: Quantify fluorescence with excitation at 490 nm and emission at 520 nm using a LJL Analyst AD (LJL BioSystems, Inc.).

Electrophoretic mobility shift assays:

A3G-mycHis protein was incubated together with fluorescein labeled single strand oligo at 37°C for 30 min. The reaction products were subjected to electrophoresis in TBE poly-

acrylamide gel of indicated concentration. The oligo shift were imaged at 463 nm.

Crystallization and High-resolution crystal structures:

A3G-191-384-2K3A and A3G-191-380-2K2A were expressed with GST-tags and purified and crystallized as described earlier (147). MN30 dissolved in DMSO at 1 mM stock concentration was added to crystallization drops with A3G-191-384-2K3A crystals and incubated for 12 hours followed by flash freezing in liquid nitrogen and data collection under cryogenic conditions. Data for A3G-191-384-2K3A crystals soaked with MN30 was collected on a R-AXIS IV detector system (Rigaku Inc.) at the Univ. Mass. Med. Sch. and for the A3G-191-380-2K2A "apo" crystals at GM/CA-CAT beamline at the Advanced Photon Source (APS, ANL, Chicago, IL). Data processing and reduction for MN30 soaked A3G-191-384-2K3A crystals was done with HKL2000 (176, 216) and that for A3G-191-380-2K2A "apo" crystals with XDS (217) as provided via the *Xia2* package (218). In both cases, further refinement was carried out using Refmac5 (179, 219) as provided in the CCP4 software suite (180) and the model building done using Coot (182). PyMOL (Schrodinger LLC) was used for viewing final structures and generating images.

Modeling the C321+MN30 adduct:

The MN30+Cys321 adduct was modeled into the A3G-191-380-2K2A crystal structure using the Prime "Covalent Docking" module (220–222) as provided in the software Maestro (223), following standard protein and ligand preparation modules and protocols as provided by the vendor (Schrodinger LLC). The ligand interaction diagram was also generated using Maestro.

Mass Spectrometry:

Purified A3G-mycHis reacted with MN30 was separated by SDS-PAGE and stained with Coomassie. A3G bands were excised from the gel and digested with trypsin (224, 225). Extracted peptides were resuspended in 20 µl of 0.1% formic acid and analyzed by LC-

MS/MS on an LTQ Orbitrap XL mass spectrometer (Thermo Scientific) equipped with a nanoACQUITY autosampler and chromatography system (Waters). 5 µl of each sample was injected onto a nanoACQUITY Symmetry C18 trap (5 um particle size, 180 um x 20 mm) in buffer A (0.1% formic acid) at a flow rate of 4 µl/min and then separated over a nanoACQUITY BEH C18 analytical column (1.7 µm particle size, 100 µm x 100 mm) over two hrs with a gradient from 2% to 25% buffer B (99.9% ACN, 0.1% formic acid) at a flow rate of 0.4 µl/min. The mass spectrometer continuously collected data in a data-dependent manner, collecting a survey scan in the Orbitrap mass analyzer at 40,000 resolution with an automatic gain control (AGC) target of 1 x 10^6 followed by collision-induced dissociated (CID) MS/MS scans of the 10 most abundant ions in the survey scan in the ion trap with an AGC target of 5,000, a signal threshold of 1,000, a 2.0 Da isolation width, and 30 ms activation time at 35% normalized collision energy. Charge state screening was employed to reject unassigned or 1+ charge states. Dynamic exclusion was enabled to ignore masses for 30 s that had been previously selected for fragmentation. The collected mass spectra were searched against the SwissProt database (downloaded 1/11/11) using the Protein Prospector software suite (226). To identify potentially MN30-modified peptides, a secondary search was performed against on the sequence for A3G allowing mass modifications between 160 to 170 Da to any cysteine.

APOBEC3A Homology Model:

The A3G-191-380-2K2A crystal structure was used as the template for modeling the APOBEC3A sequence (RefSeq GI:300869638) using the comparative homology modeling module PRIME (220–222), as available in the software Maestro (223). The resulting model was subjected to Polak-Ribière conjugate gradients (PRCG) energy minimization, also available in Maestro, until the gradient converged. The final model satisfied Ramachandran plot constraints and no steric clashes were present.

3.6 Supplementary Data

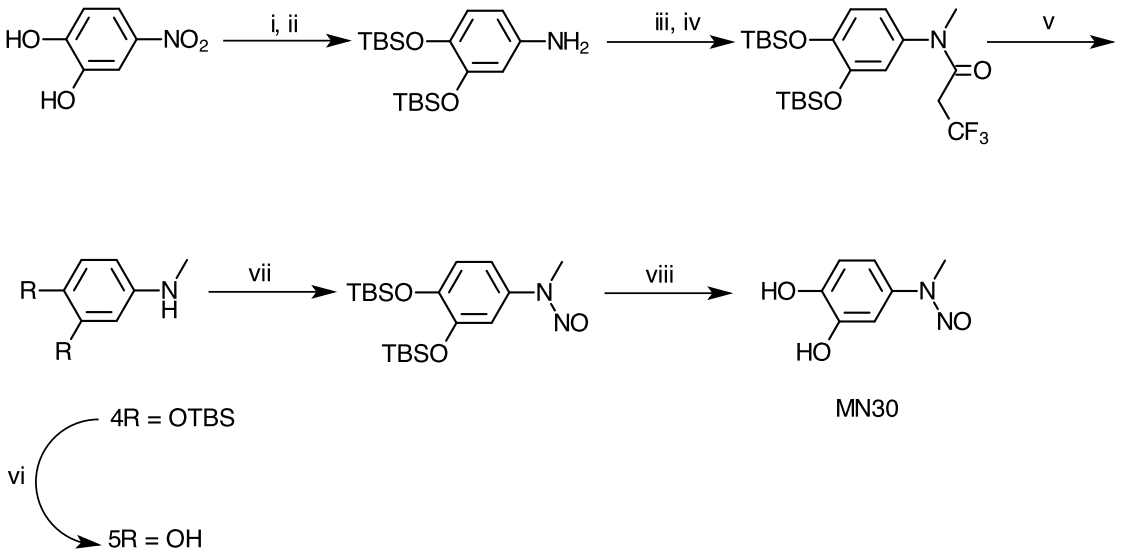


Figure 3.19

Figure 3.19 : Synthesis of MN30 and des-nitroso 5. Reagents and conditions: i. TBSCl, imidazole, DMAP, DMF, 89%; ii. H_2 (1 atm), Pd/C, EtOH, 91%; iii. $(CF_3CO)_2O$, pyridine, CH_2Cl_2, 96%; iv. NaH, DMF; MeI 76%; v. K_2CO_3, MeOH, H_2O, 66%; vi. TBAF, THF, AcOH 63%; vii. $NaNO_2$, HCl, THF, 76%; viii. NaH-HF buffer, 72%.

Category	Parameter	Description
Assay	Type of assay	Biochemical
	Target	A3G ssDNA cytosine deaminase
	Primary measurement	Fluorescence intensity
	Key reagents	A3G-mycHis protein, Fluorophore labeled ssDNA oligo
	Assay protocol	See methods section
	Additional comments	
Library	Library size	1,280
	Library composition	Pharmacologically active compounds
	Source	Sigma Aldrich
	Additional comments	
Screen	Format	Nunc 384-well plate (catalog #262260)
	Concentration(s) tested	10 µM
	Plate controls	Purified A3G-mycHis + ssDNA oligo Purified A3G-mycHis + ssDNA oligo + 10µM MN1 Assay buffer + ssDNA oligo
	Reagent/ compound dispensing system	Reagent dispensing: Thermo MultiDrop Compound dispensing: LabCyte ECHO 550
	Detection instrument and software	LJL Analyst
	Assay validation/QC	n = 576: Purified A3G-mycHis + ssDNA oligo: %CV: 2.83 n = 576: Assay buffer + ssDNA oligo: %CV: 3.92
	Correction factors	None
	Normalization	None
	Additional comments	
Post-HTS analysis	Hit criteria	40%
	Hit rate	2.8%
	Additional assay(s)	8-point dose response assay (n = 2 at each concentration) with purified A3G-mycHis and purified A3A-mycHis; Subsequent 12-poind dose response assay (n = 3 at each concentration) of repurchased compounds with purified A3G-mycHis and purified A3A-mycHis
	Confirmation of hit purity and structure	
	Additional comments	

Figure 3.20

Figure 3.20 : HTS data in tabular format.

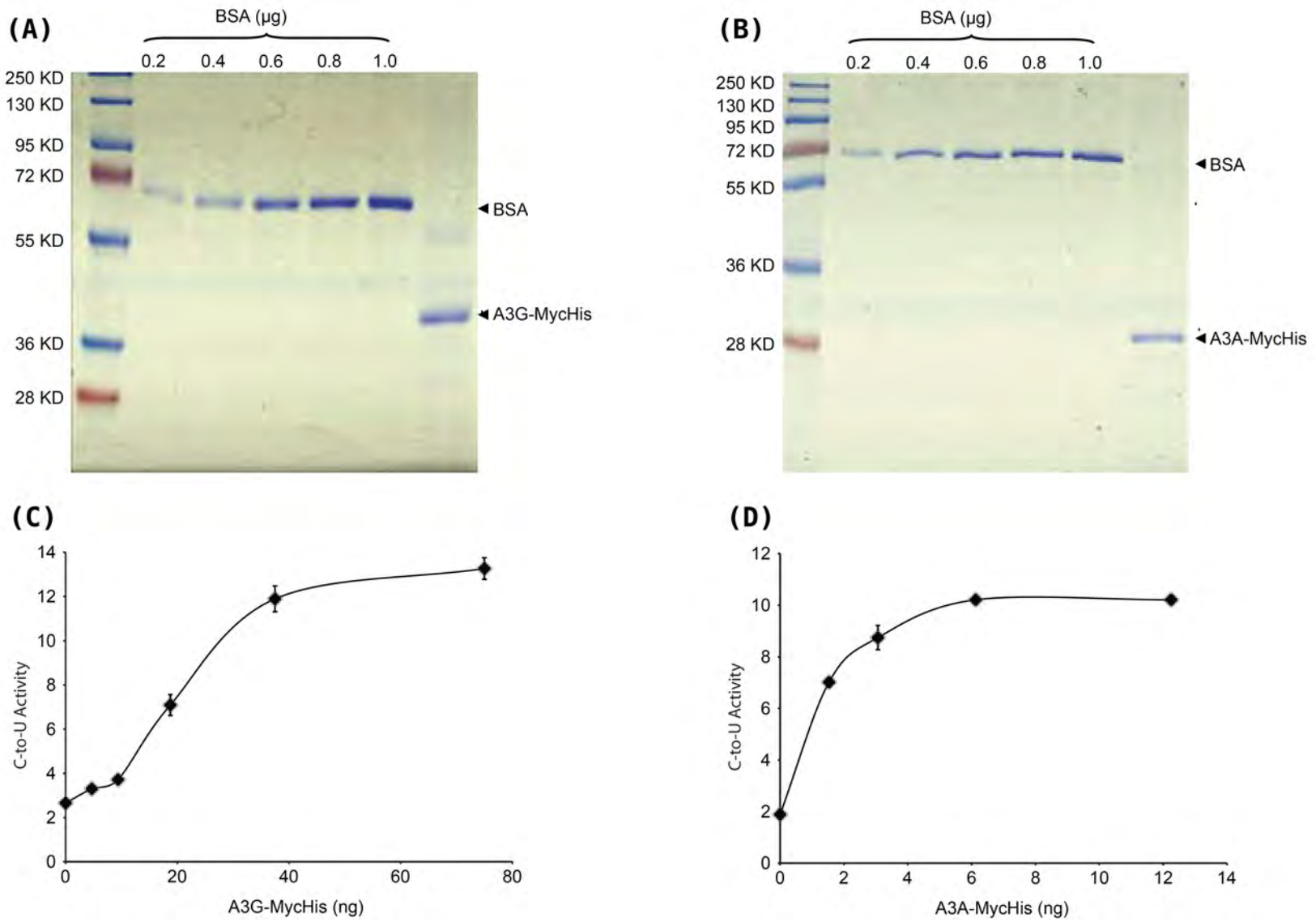

(A)
BSA (µg)
0.2
0.4
0.6
0.8
1.0
250 KD
130 KD
95 KD
72 KD
55 KD
36 KD
28 KD
BSA
A3G-MycHis
(B)
BSA (µg)
0.2
0.4
0.6
0.8
1.0
250 KD
130 KD
95 KD
72 KD
55 KD
36 KD
28 KD
BSA
A3A-MycHis
(C)
C-to-U Activity
A3G-MycHis (ng)
(D)
C-to-U Activity
A3A-MycHis (ng)

Figure 3.21

Figure 3.21 : **(A),(B)** Image of Coomassie Blue-stained A3G- and A3A-mycHis proteins purified from HEK-293T cells. BSA standard is shown for comparison. **(C),(D)** Activities of the indicated proteins in the fluorescence-based DNA deaminase assay. The A3G assay used 5'-6-FAM-AAA-CCC-AAA-GAG-AGA-ATG-TGA-(TAMRA)-3' and the A3A assay 5'-6-FAM-AAA-TTC-AAA-GAG-AGA-ATG-TGA-(TAMRA)-3'.

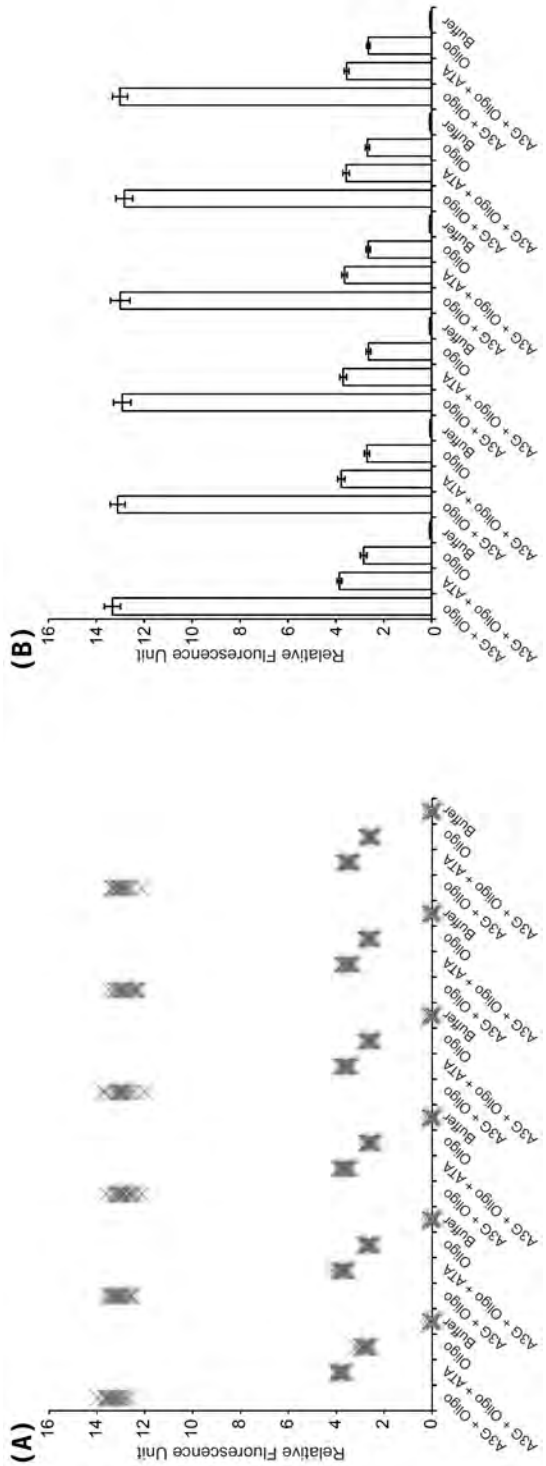

(A)
Relative Fluorescence Unit
16
14
12
10
8
6
4
2
0
A3G + Oligo
Oligo + ATA
Oligo
Buffer
A3G + Oligo
A3G + Oligo + ATA
(B)
Relative Fluorescence Unit
16
14
12
10
8
6
4
2
0
A3G + Oligo
Oligo + ATA
Oligo
Buffer
A3G + Oligo
A3G + Oligo + ATA

Figure 3.22

Figure 3.22 : **(A)** The relative fluorescence readings for an entire 384 well plate with 16 identical samples in each column replicated six times through the plate. Columns 1, 5, 9, 13, 17, 21 report uninhibited reactions: 0.0405 µM A3G-mycHis, 0.0333 µM ssDNA substrate (5'-6-FAM-AAA-CCC-AAA-GAG-AGA-ATG-TGA-TAMRA-3'), and excess 6.7×10^{-4} units UDG. Columns 2, 6, 10, 14, 18, 22 report ATA-inhibited reactions, otherwise identical to the positive reactions. Columns 3, 7, 11, 15, 19, 23 report reactions with no A3G, but otherwise identical to positive reactions. Columns 4, 8, 12, 16, 10, 24 report reactions with no ssDNA substrate, but otherwise identical to positive reactions. The raw fluorescence reading of a 384 well plate, 16 samples in each column. **(B)** The mean value ±S.D. of each column of raw data shown in **(A)**. The average Z-score in 384 well plates was 0.85.

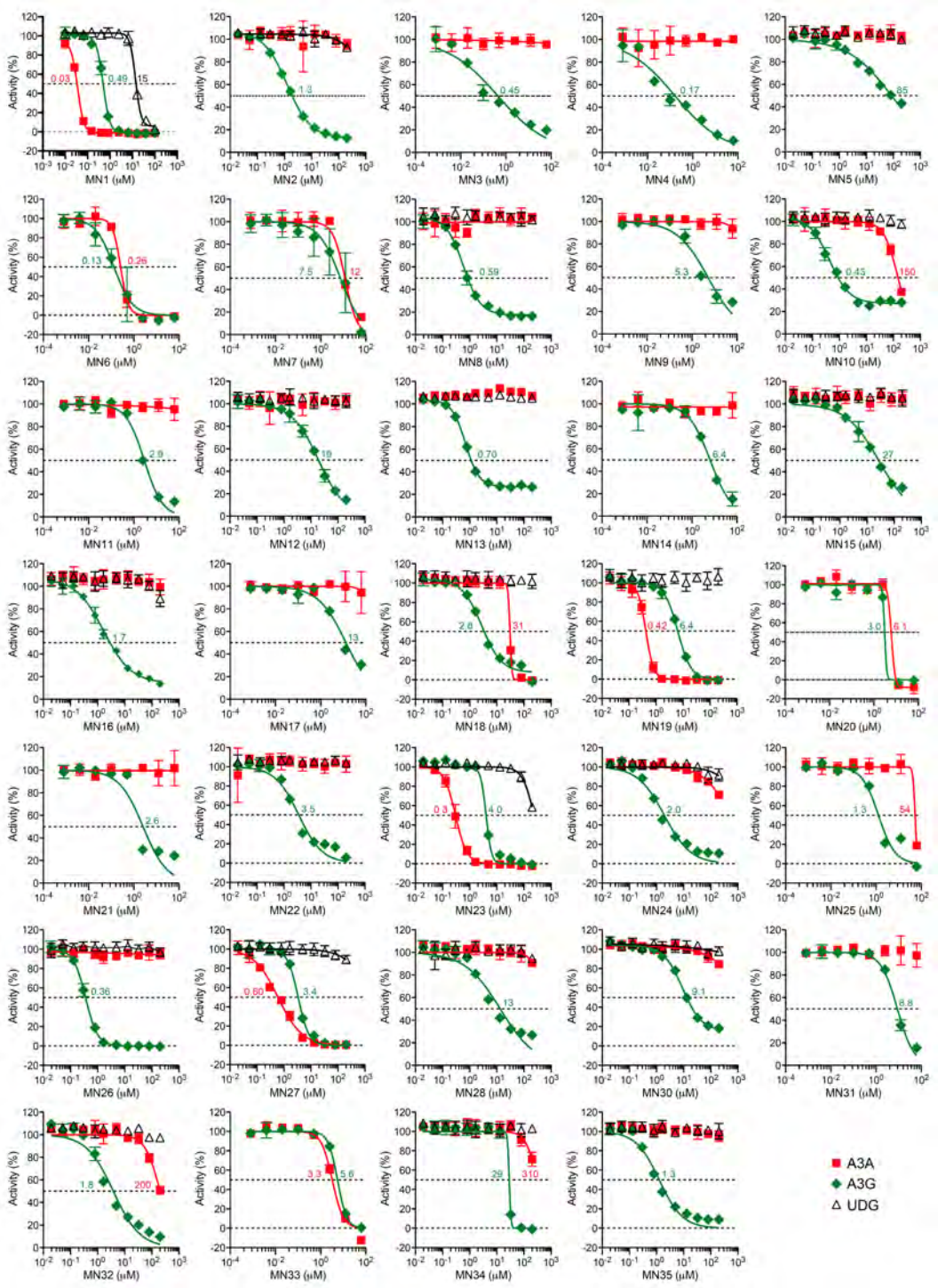

Activity (%)
MN1 (μM)
MN2 (μM)
MN3 (μM)
MN4 (μM)
MN5 (μM)
MN6 (μM)
MN7 (μM)
MN8 (μM)
MN9 (μM)
MN10 (μM)
MN11 (μM)
MN12 (μM)
MN13 (μM)
MN14 (μM)
MN15 (μM)
MN16 (μM)
MN17 (μM)
MN18 (μM)
MN19 (μM)
MN20 (μM)
MN21 (μM)
MN22 (μM)
MN23 (μM)
MN24 (μM)
MN25 (μM)
MN26 (μM)
MN27 (μM)
MN28 (μM)
MN30 (μM)
MN31 (μM)
MN32 (μM)
MN33 (μM)
MN34 (μM)
MN35 (μM)
A3A
A3G
UDG

Figure 3.23

Figure 3.23 : Dose response data for 34 compounds. All 34 compounds were tested with 0.0405 µM A3G-mycHis, 0.0038 µM A3A-mycHis, or 6.7×10^{-4} units/µl UDG in dose response assays using 0.033 µM ssDNA substrate (dual substrate oligo 5'-6-(FAM)-AAA-TAT-TCC-CTA-ATA-GAT-AAT-GTG-A-(TAMRA)-3' for A3G and A3A; U oligo 5'-6-FAM-AAA-CCU-AAA-GAG-AGA-ATG-TGA-(TAMRA)-3' for UDG). All dose responses were done in triplicate, and the mean ±S.D. is shown for each concentration. 8-point (not shown) and 12-point dose response curves were generated for all repurchased compounds; only 8-point dose response curves with A3G and A3A (not UDG) were generated with non-repurchased small molecules. Relevant IC_{50} values are indicated here and all IC_{50} values are listed in Table 1.

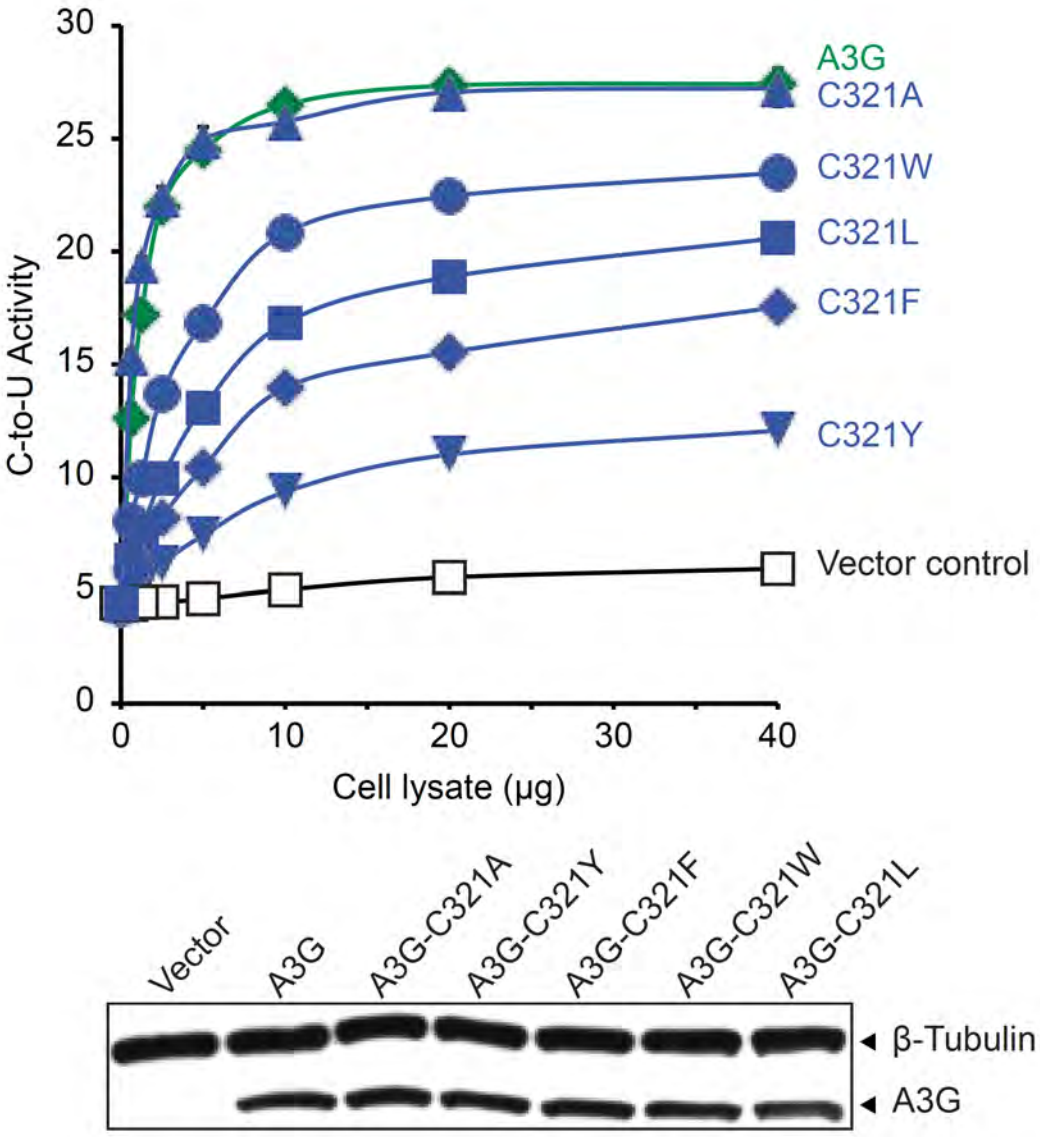

30
25
20
15
10
5
0
C-to-U Activity
0
10
20
30
40
Cell lysate (µg)
A3G
C321A
C321W
C321L
C321F
C321Y
Vector control
Vector
A3G
A3G-C321A
A3G-C321Y
A3G-C321F
A3G-C321W
A3G-C321L
β-Tubulin
A3G

Figure 3.24

Figure 3.24 : Wildtype A3G-mycHis or the indicated C321 mutant derivatives were expressed in HEK293T cells. 48 hrs post-transfection, soluble cell extracts were prepared, quantified, and assayed in triplicate for DNA C-to-U activity using a fluorescently labeled dual A3A/A3G ssDNA substrate oligo as described (189, 197, 198). Anti-c-myc immunoblotting showed that each protein was expressed at similar levels, and anti-β-tubulin was used as a loading control.

Chapter IV

Comparative analysis of Z-domain signature proteins in the human APOBEC3 subfamily

Chapter IV

Comparative analysis of Z-domain signature proteins in the human APOBEC3 subfamily

4.1 Introduction

The APOBEC3 subfamily of proteins, a subset of the greater APOBEC family of cytidine deaminases, are essential for controlling endogenous retrotransposons (e.g. Long Interspersed Nuclear Element 1: LINE-1) (227, 228) and for anti-retroviral immune response (e.g. HIV-1) (57, 229). This subfamily of APOBEC proteins, found in all primates and most mammals, has been under positive evolutionary selection over the past 33 million years (92). Human chromosome 22 (locus q13.1) has been identified to carry the entire APOBEC3 gene family, consisting of seven members, abbreviated as A3A, A3B, A3C, A3D/E, A3F, A3G and A3H (48, 230). The intracellular localization and functions of APOBEC3 family members are distinct, for example, A3A has been observed cell-wide with a nuclear preference, A3C is cell-wide, A3F and A3G are predominantly cytoplasmic and A3H is again cell-wide but with a propensity for nucleoli (142, 227). The protein sequences of APOBEC3 proteins reveal a shared underlying architecture characterized by zinc-binding signature sequences or Z-domains (95).

The three Z-domain signature sequences are:

Z1: $[x_{6-11}HxEx_5xxx_{17}SWS/TPCx_{2-4}Cx_6Fx_8Lx_5RIYx_{8-11}Lx_2Lx_8xMx_{3-4}]$

Z2: $[x_{3-7}HxEx_5WFx_{16-20}SWS/TPCx_2Cx_6Fx_8Lx_5RLYx_{8-11}Lx_2Lx_{8-12}xMx_{3-4}]$

Z3: $[x_{3-4}HxEx_5xxx_{14-17}TWSPCx_2Cx_6Fx_8Lx_5RLYx_{10}Lx_2Lx_8VMx_{3-4}]$

Four of the seven human APOBEC3 proteins contain two Z-domains, A3B (Z2/Z1), A3D/E (Z2/Z2), A3F (Z2/Z2) and A3G (Z2/Z1), that are believed to have arisen as a result of gene duplication events (94). The remaining three carry a single Z-domain: A3A (Z1), A3C (Z2)

and A3H (Z3) (Figure 1.2.1). The Z-domain signatures are merely sequence patterns and do not signify the presence or absence of catalytic ability. For instance A3A, A3C and A3H each contain a single catalytically active Z-domain, Z1, Z2 and Z3 respectively. A3F and A3G, both contain two Z-domains with only the C-terminal domain being catalytically active and the catalytically-inactive domain being more relevant to substrate binding. A3B also contains two domains and both are catalytically active.

The availability of high-resolution structures of various proteins has enabled modeling of related proteins by homology modeling, an approach that has been validated and employed successfully over the last three decades (231, 232). Recent literature describes homology modeling attempts based on APOBEC2 (157) and A3G-CTD structures, for example the homology models of A3G-NTD (70, 144, 145, 233), APOBEC3A (234) and APOBEC3F-CTD (70, 235). Homology models of multiple related proteins are useful for comparing shared structural features or to highlight differences and distinguishing characteristics amongst family members. Furthermore, these models may be used for generating new hypotheses that may be satisfied by prior experimental data or be verified by further experiments designed to test one or more aspects of such hypotheses.

4.2 Results

4.2.1 Sequence Analysis

The high-resolution crystal structures of human APOBEC3G catalytic/C-terminal domain, discussed in the previous chapters (147), have the potential to serve as reliable templates for modeling homologous APOBEC3 family members, as evident from the multiple-sequence alignment of six APOBEC3 domains to the crystallized (mutated) A3G-CTD and wild-type A3G-CTD sequences (Figure 4.1). The sequences were selected to represent all three Z-domains, including a catalytically inactive **Z2** domain (A3G-NTD). The sequence align-

ment highlights the features shared across the listed APOBEC3 family members, such as, the active-site catalytic glutamate (Glu259 in A3G-CTD) and the zinc coordinating residues (His257, Cys288 and Cys291 in A3G-CTD). In addition to the corresponding Z-domain signature residues, highlighted in Figure 4.1 by colored dots and corresponding to Figure 1.3, several other residues are also absolutely conserved across the sequences.

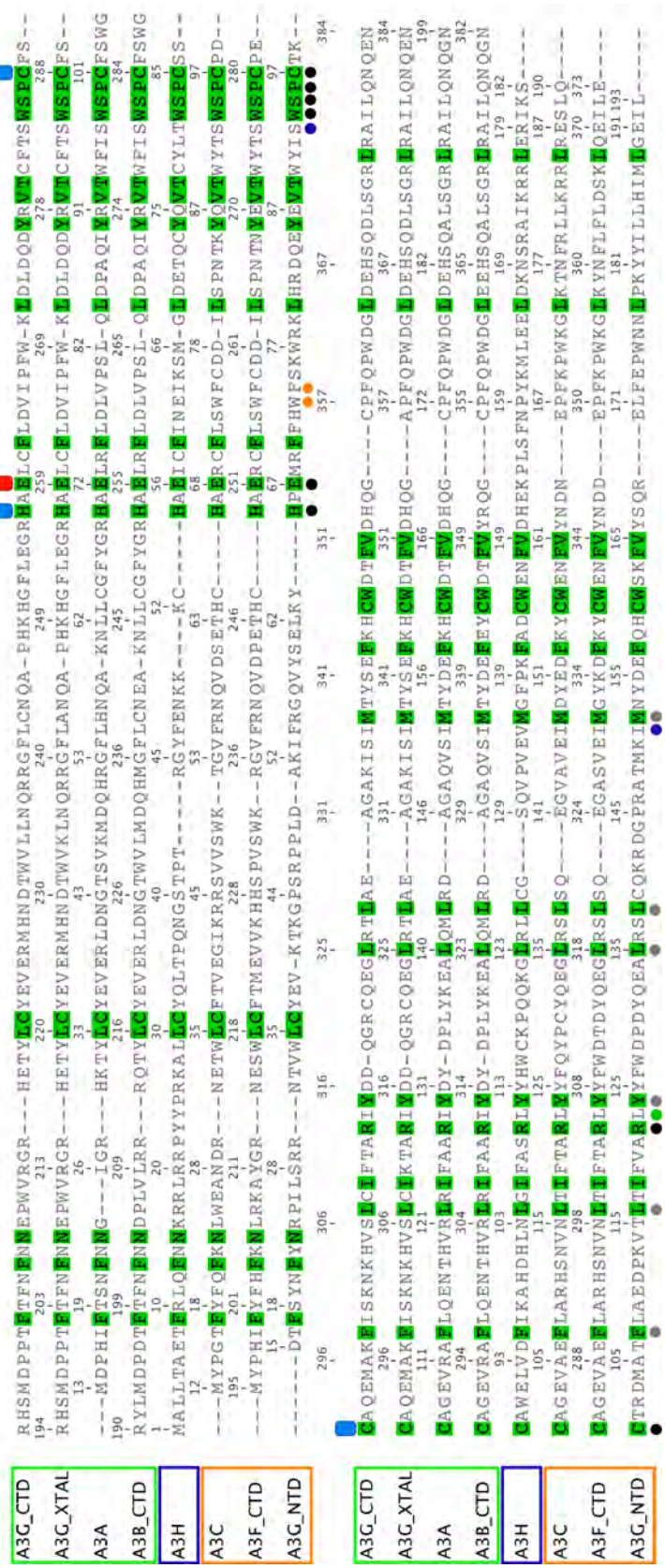


Figure 4.1

Figure 4.1 : Sequence alignment of the homology modeled human APOBEC3 sequences to the template A3G-CTD crystal structure sequence. Identical residues are highlighted in green, zinc coordinating residues are indicated by blue rectangles and the catalytic glutamate residue is indicated by the red rectangle. The colored circles represent residues listed in each Z-domain signature, with the colors matching Figure 1.3 as per references (94, 95).

4.2.2 Phylogenetic Analysis

Not surprisingly, phylogenetic analysis based on the multiple-sequence alignment (Figure 4.2), reveals a partitioning or clustering amongst the sequences that corresponds to the three Z-domain signatures, likely arising as a consequence of their evolutionary history (94). Based on branch lengths, the **Z1** domain proteins, A3A, A3B-CTD and A3G-CTD appear more closely related to each other and to the **Z3** domain protein A3H, whereas, the **Z2** domain proteins A3C, A3F-CTD and A3G-NTD form a distinct group, as has also been observed in previous analysis (94). This phylogenetic analysis suggests a framework for expected divergence in homology models built based on the A3G-CTD crystal structure template (sequence: A3G-XTAL), for example, the A3G-NTD model may diverge far more from the template than the A3A model.

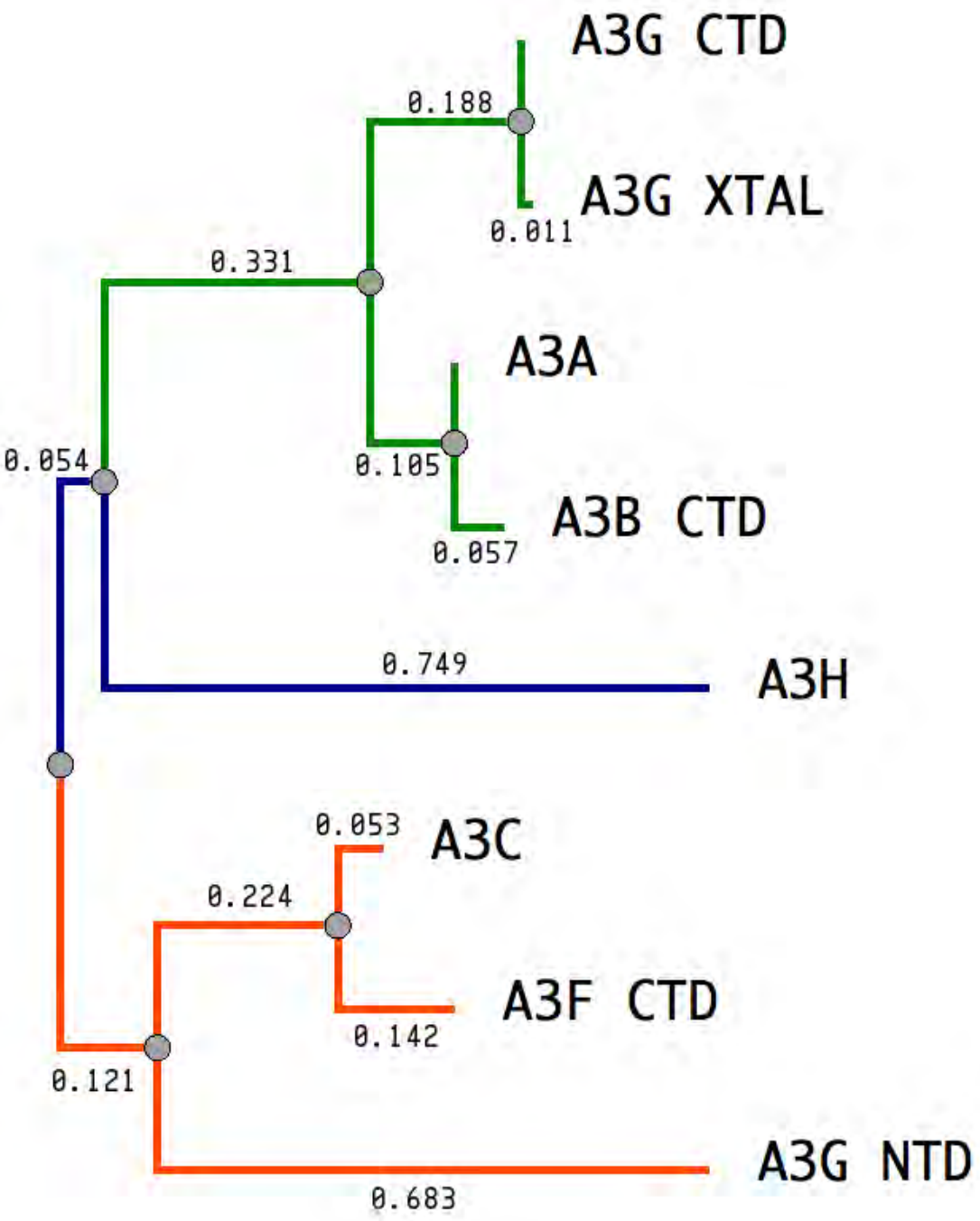

A3G CTD
0.188
A3G XTAL
0.011
0.331
A3A
0.054
0.105
A3B CTD
0.057
0.749
A3H
0.053
A3C
0.224
A3F CTD
0.142
0.121
A3G NTD
0.683

Figure 4.2

Figure 4.2 : Phylogenetic tree based on sequence alignment from Figure 4.1. The branches are colored as per the Z-domain coloring in Figure 1.3 (Z1: green, Z2: orange, Z3: blue).

4.2.3 Features of APOBEC3 homology models

Common structural core

All APOBEC3 homology models share secondary structural features with the template structure but differ primarily in the length and orientation of loop regions, with A3G-NTD model diverging most from the template. The best-fit average Cα RMSD values for the six models with respect to the template A3G-CTD crystal structure are 0.921Å – A3A, 1.028Å – A3B-CTD, 2.645Å – A3H, 1.578Å – A3C, 1.741Å – A3F-CTD and 5.42Å – A3G-NTD. Analyzing these homology models using computational statistical methods (*core.find* method in ref. 236), reveals a common minimal "core" of structural similarity across all models (Figure 4.3). Structural alignment of all APOBEC3 homology models and the template structure shows very good agreement around the "core" region (Figure 4.4). As may be expected, a majority of residues making the "core" are completely conserved across the template and target APOBEC3 protein sequences (Figure 4.1).

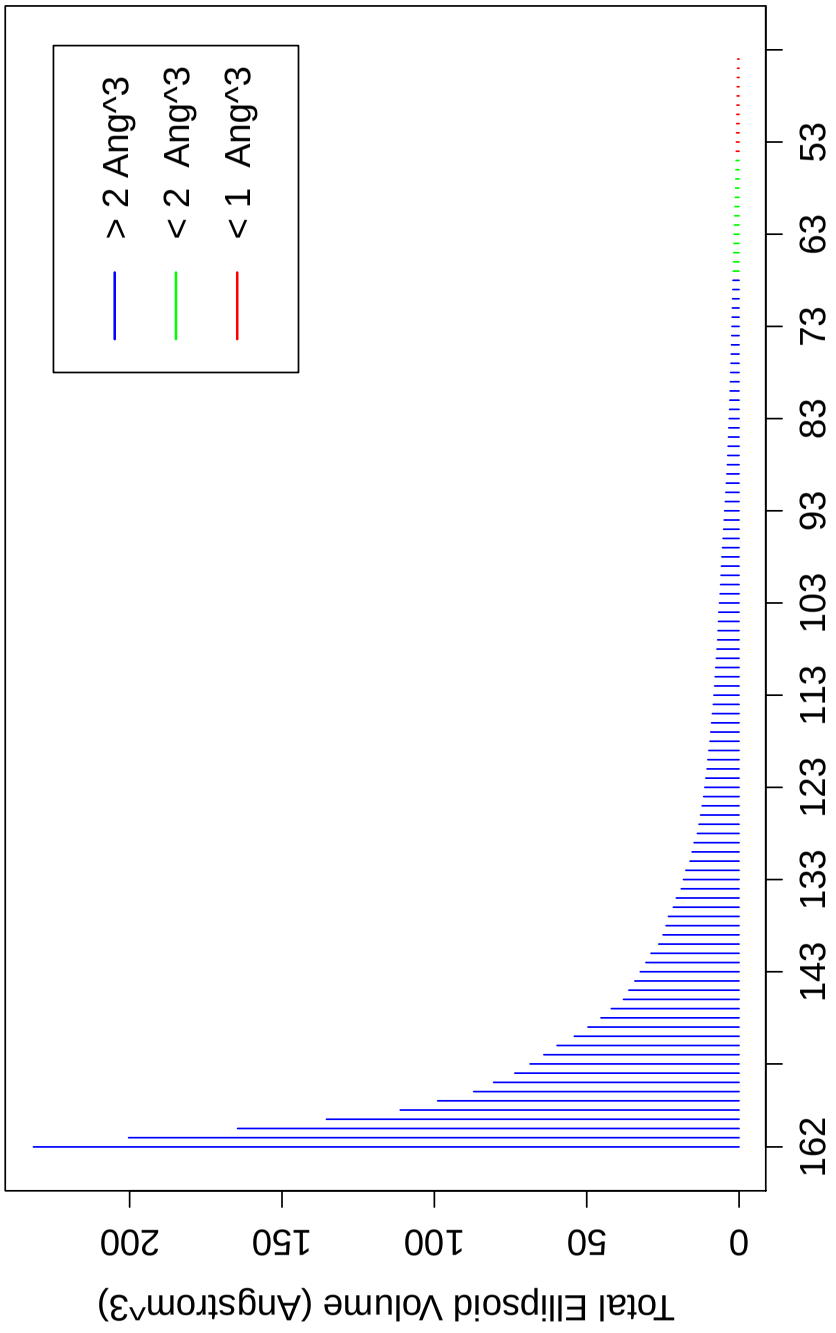
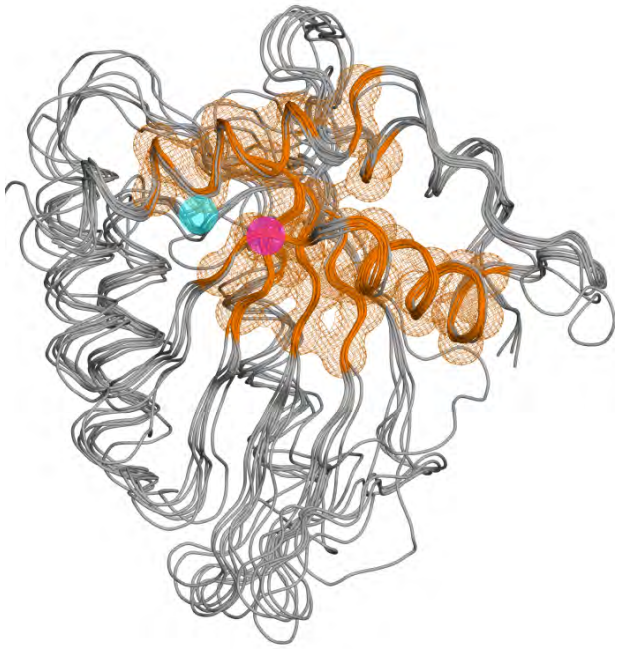
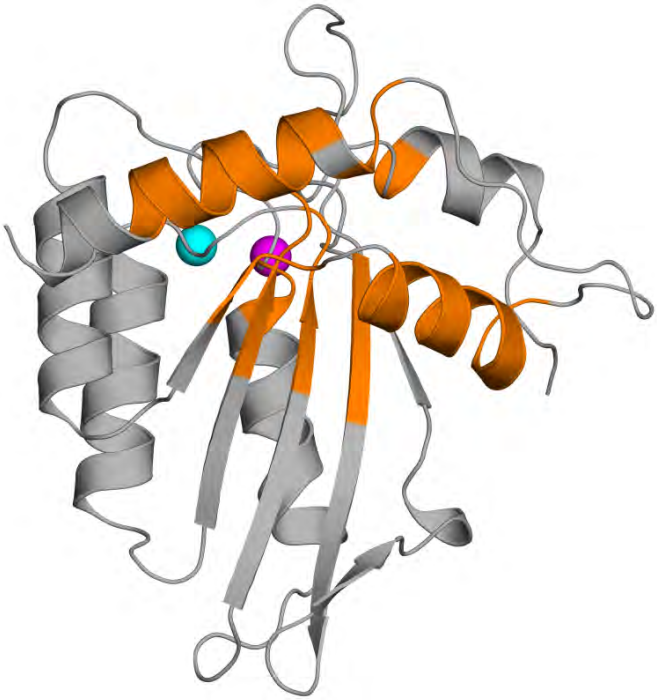

> 2 Ang^3
< 2 Ang^3
< 1 Ang^3
Total Ellipsoid Volume (Angstrom^3)
0
50
100
150
200
Core Size (Number of Residues)
162
143
133
123
113
103
93
83
73
63
53

Figure 4.3

Figure 4.3 : A histogram of the "core" ellipsoid volumes against the number of residues considered as a part of the core in all APOBEC3 homology models. Based on data generated using Bio3D (236). The Y-axis represents the total ellipsoid volume as defined by (237) and the X-axis represents the number of residues contributing to the ellipsoid volume. A "core" comprises a set of residues, across multiple structures, that have minimal structural deviation with respect to each other and thus a minimal total volume as calculated after three-dimensional superimposition.



(a) All APOBEC3 homology models



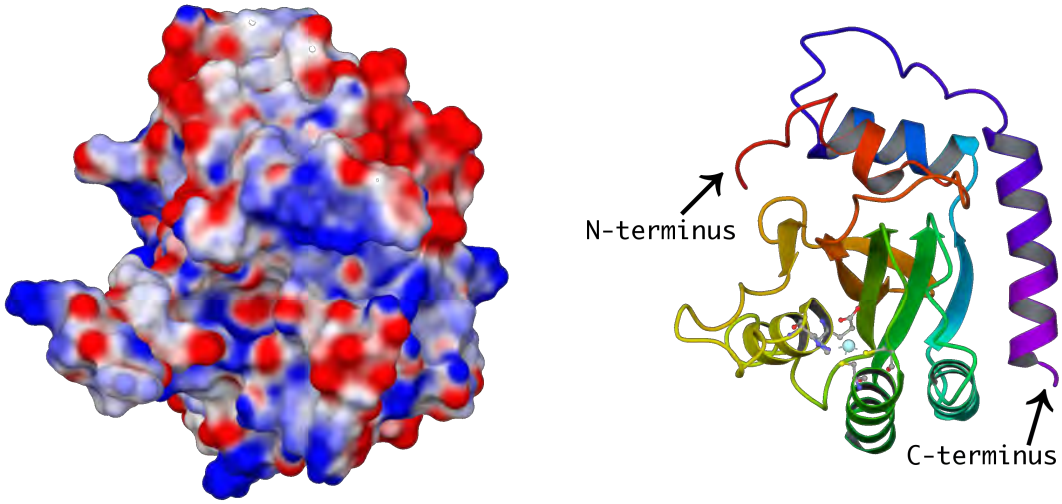
(b) Core region in A3G-CTD

Figure 4.4

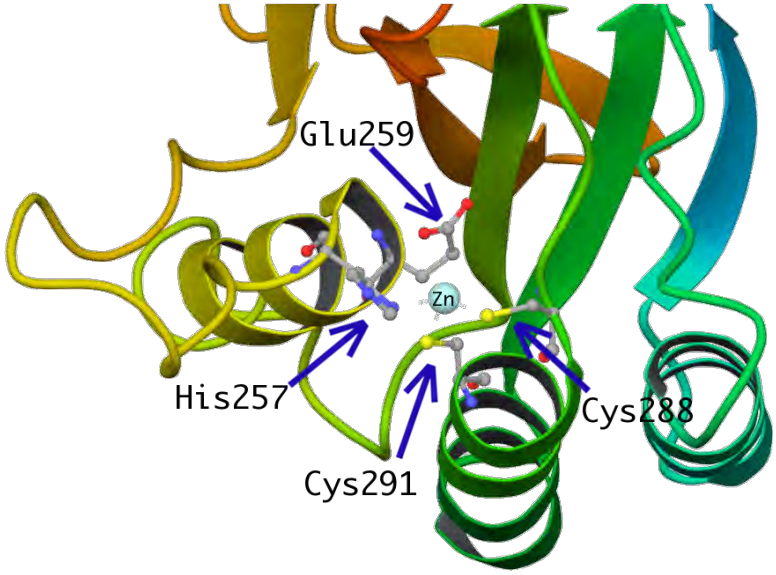
Figure 4.4 : The "core" region is highlighted in all APOBEC3 homology models (orange mesh) and the template APOBEC3G-CTD crystal structure (orange cartoon). The atoms in the "core" region correspond to atoms from the following A3G-CTD residue numbers: 205, 206, 219–222, 257, 258, 281–284, 309–314, 336–338, 340–351, 363, 364 and 366–375. Active-site Zn^{2+}: Cyan spheres. Secondary-zinc-binding-site Zn^{2+}: Magenta spheres.

Surface electrostatic potential

The most striking difference across the homology models lies in the surface charge distribution as suggested by electrostatic potential calculations (Figure 4.5). The orientation depicted in Figure 4.5 and Figure 4.6 is a direct "top" view looking at the enzyme active-site region and shows great heterogeneity in electrostatic potential across the six models and the template. APOBEC3H, a single-domain **Z3** domain protein, shows very large positively charged regions around the active site compared to all other models (Figure 4.6f). In contrast, APOBEC3C, another single-domain protein, but with a **Z2** signature, is not as strongly charged around the active site, with the majority of positive charge localized around the C-terminal helical region (Figure 4.6c). Similarly, APOBEC3A (Figure 4.6a), with a single **Z1** domain, appears more positively charged near the active site region adjoining the N-terminus as compared to the other two **Z1** domains, APOBEC3B-CTD (Figure 4.6b) and the A3G-CTD template (Figure 4.5a). The same is true for the APOBEC3G-NTD (Figure 4.6e) that appears more positively charged compared to A3G-CTD.



(a) A3G-CTD overview

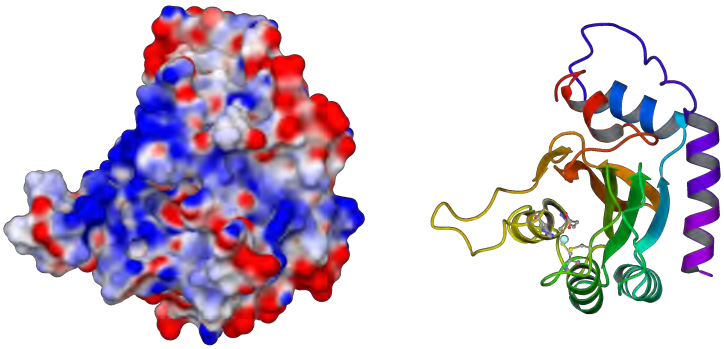


(b) Active site region

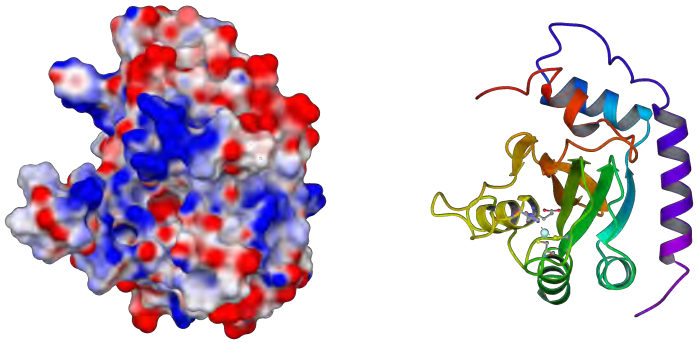
Figure 4.5

Figure 4.5 : Electrostatic potential surface representation of the template APOBEC3G-CTD crystal structure. Red represents negative and blue represents positive potential with the numerical range between -0.125 — +0.125 kcal/M

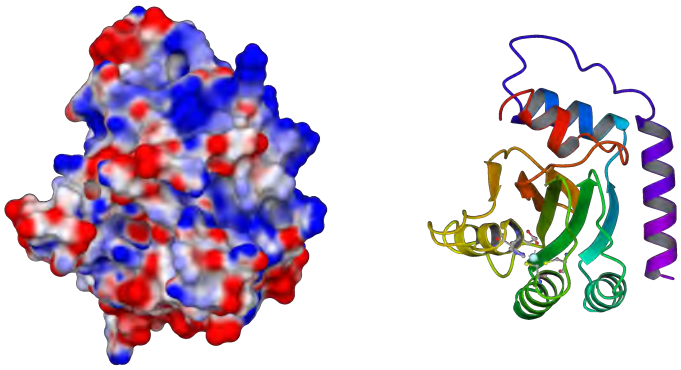
.



(a) APOBEC3A

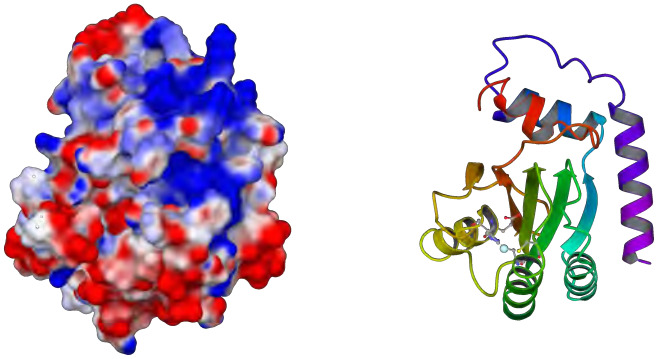


(b) APOBEC3B-CTD

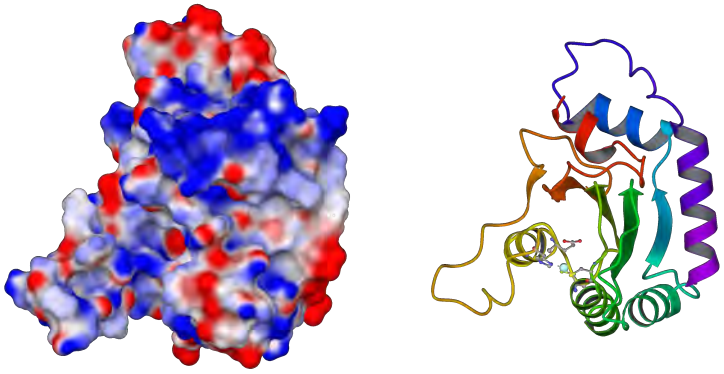


(c) APOBEC3C

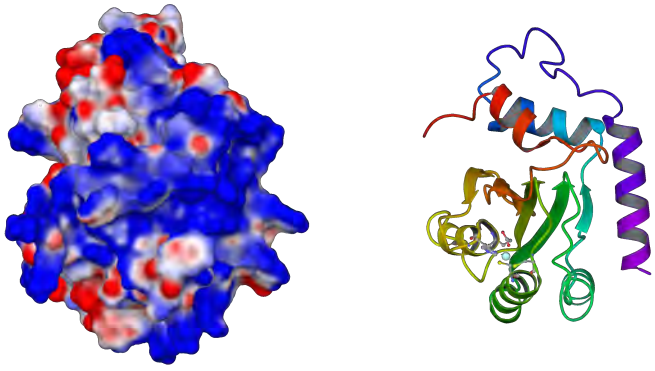
Figure 4.6



(d) APOBEC3F-CTD



(e) APOBEC3G-NTD



(f) APOBEC3H

Figure 4.6

Figure 4.6 : The human APOBEC3 family homology models – **Top view** – looking down into the active site region.

4.3 Discussion

The APOBEC3 sequences selected for this study represent all three Z-domain signatures and are catalytically active except for one **Z2** domain member, APOBEC3G-NTD. Multiple sequence alignment displays the consensus in regions defining each of the Z-domain signatures, whereas, the differences suggest regions that may play important roles in substrate specificity and binding avidity across these proteins. The phylogenetic analysis identifies a spontaneous clustering of these sequences based on their Z-domain signatures, with A3H (**Z3**) on a separate branch from the A3G-CTD, A3A and A3B-CTD (**Z1**) and the A3G-NTD, A3F-CTD and A3C (**Z2**) branches. The phylogenetic distance between the modeled proteins and the template sequence correlated with the RMS deviations between the template structure and final homology models. The fact that A3G-CTD and A3G-NTD are on distinct branches suggests "unequal crossing-over" gene-duplication events as discussed and illustrated in (94).

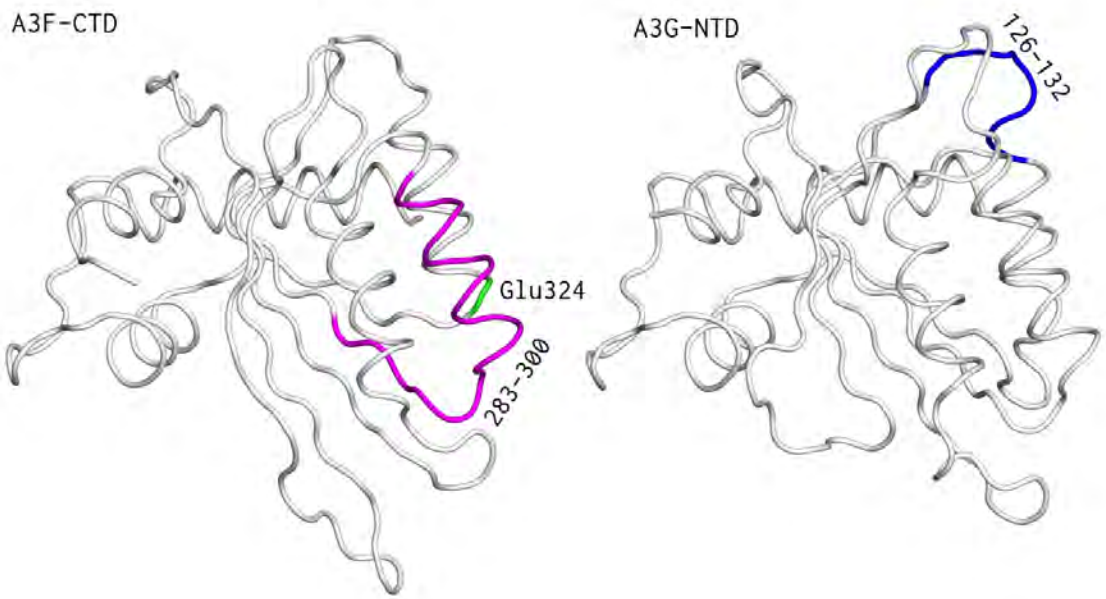
Structural dissimilarity across the homology models was contrasted by a structurally similar "core" region that loosely corresponds to the conserved residues in the multiple sequence alignment. This "core" may signify residues retained during evolution to maintain structural integrity and provide the fundamental scaffold required for proper folding, structural stability, oligomerization, substrate binding and enzymatic function.

The electrostatic potential surfaces highlight the structural variation across the APOBEC3 homology models that may contribute to substrate binding affinities, oligomerization propensity and interactions with human and viral (HIV-1 vif) proteins or nucleic acids. For instance, APOBEC3H may bind more readily to the negatively charged DNA/RNA phosphodiester backbone due to the highly positive electrostatic potential on its surface, suggesting a possible explanation for binding and catalysis by a single-domain APOBEC3 protein. Based on similar observations, the argument may be extended to other two single-

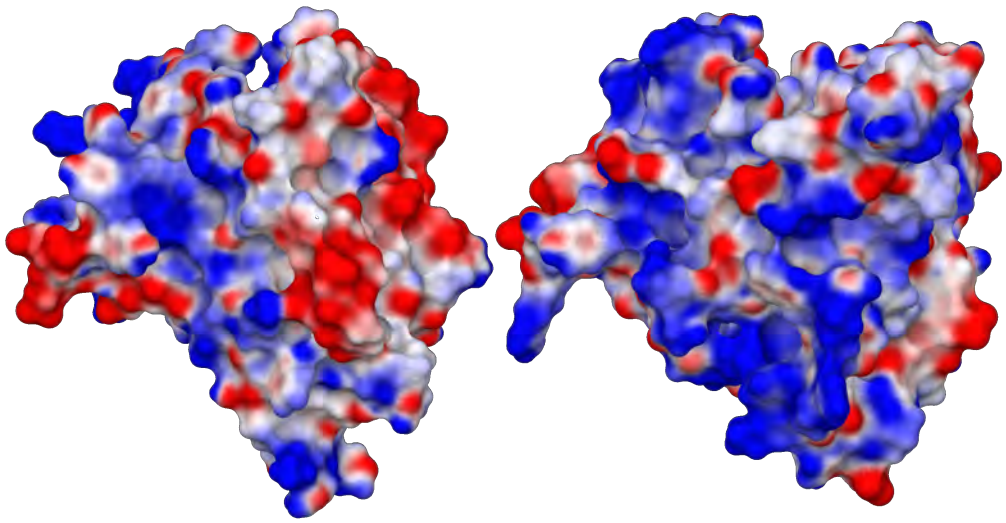
domain APOBEC3 proteins, A3A and A3C, by proposing that extent of positive electrostatic potential on the protein surface determines binding affinity to substrates and other proteins.

HIV-1 vif is an important determinant of viral infectivity as it binds A3G and targets it for proteasomal degradation. Residues 126–132 in A3G-NTD and 283–300 and Glu324 in A3F-CTD have been implicated as specificity determinants in binding to vif (Figure 4.7) (131, 233, 235). Comparing these regions in the A3G-NTD and A3F-CTD to A3G-CTD, the role of electrostatic potential in directing this interaction appears to be important. Furthermore, HIV-1 vif, the nemesis of human A3G, is known to contain large swaths of positively charged surface residues that are required for binding negatively charged residues on A3G-NTD surface and directing it towards proteasomal degradation (131, 238). Comparing the A3G-NTD model's surface electrostatic potential to the highly positively charged A3H model (Figure 4.8) immediately suggests that unfavorable charge interactions, even electrostatic repulsion, may be a major reason for A3H to escape binding and neutralization by HIV-1 vif (239).

The homology models described and analyzed in this study may serve as a foundation for building experimentally testable hypotheses to further our understanding of the human APOBEC3 family of proteins in terms of structural features, binding interactions and biological activity.

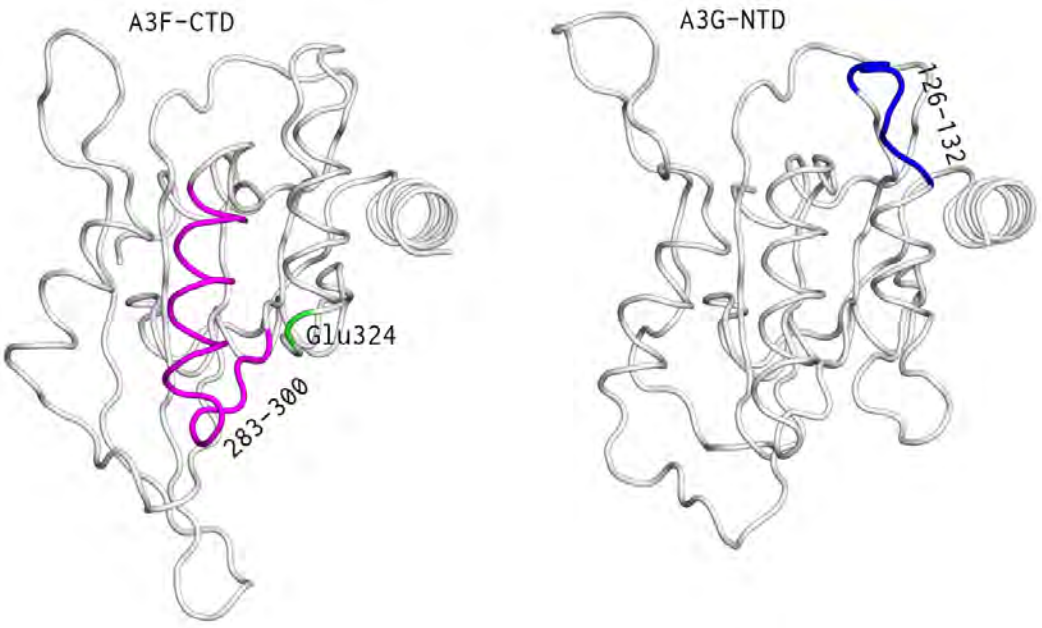


(a) A3F-CTD vs A3G-NTD

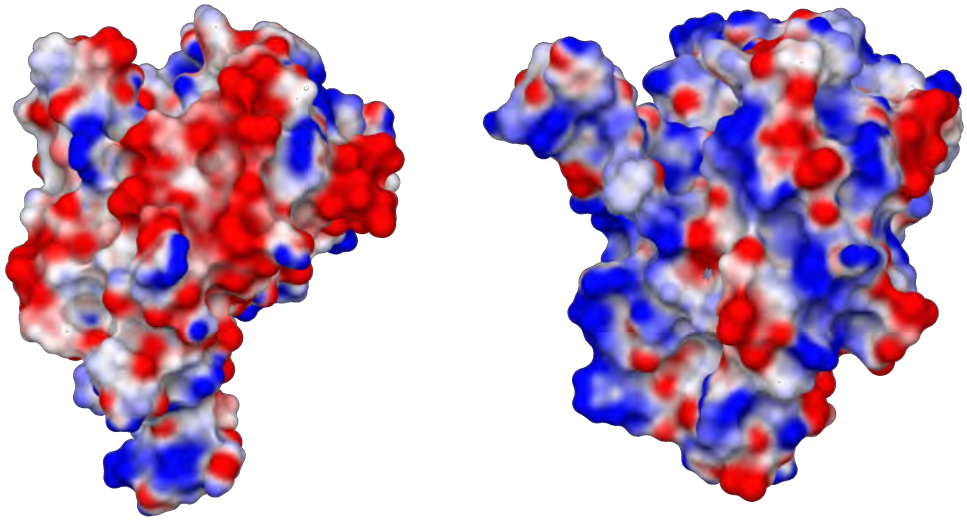


(b) A3F-CTD vs A3G-NTD Electrostatic Surface

Figure 4.7



(c) A3F-CTD vs A3G-NTD



(d) A3F-CTD vs A3G-NTD Electrostatic Surface

Figure 4.7

Figure 4.7 : HIV-1 vif binding residues highlighted in A3F-CTD (magenta, green (Glu324)) and A3G-NTD (blue) homology models. Based on data from 131, 233, 235.

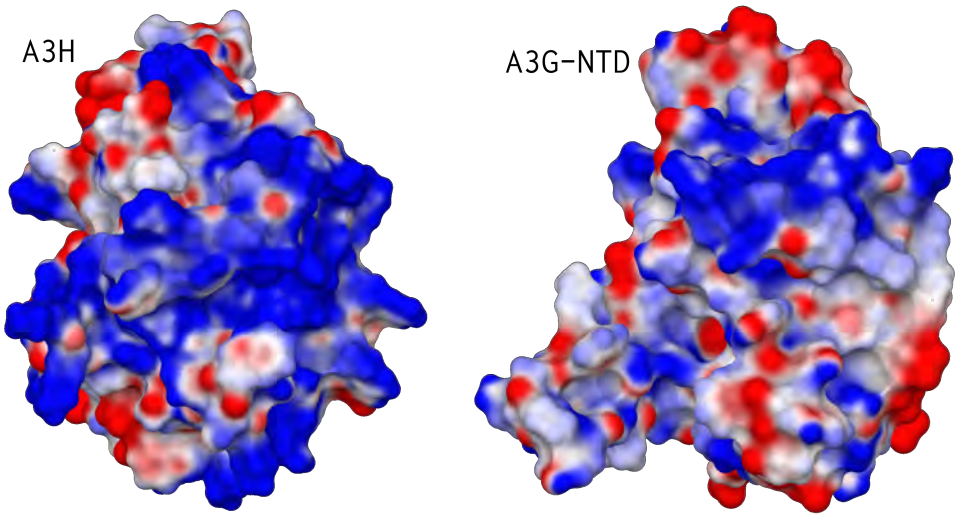


Figure 4.8

Figure 4.8 : A side-by-side comparison of the of the A3H and A3G-NTD models displaying the contrast in surface electrostatic potential that may be critical to binding and degradation by HIV-1 vif (239).

4.4 Methods

4.4.1 Sequence and Phylogenetic analysis

Sequences for the human APOBEC3 proteins used for this study were downloaded from NCBI RefSeq database (187). Dual-domain sequences were iteratively pairwise-aligned after dividing the sequences at putative domain boundaries using MUSCLE (240, 241), until an acceptable alignment between the two domains was obtained. The phylogenetic tree was constructed from the multiple sequence alignment based on tools available at the website http://www.phylogeny.fr (242–246) and visualized using the UGENE software (Unipro UGENE: an open-source bioinformatics toolkit; http://ugene.unipro.ru).

4.4.2 Homology Modeling and Electrostatic potential surfaces

All homology models were generated using the high resolution APOBEC3G-CTD crystal structure, based on the A3G-191-380-2K2A construct, discussed in the previous chapters. The homology modeling software module, PRIME (220–222), as available in the Maestro software suite (223), was used for generating all the models that were later subjected to Polak-Ribière Conjugate Gradient (PRCG) energy minimization until gradient convergence using the MacroModel software module (247), also within Maestro. Molecular surfaces were calculated in Maestro and colored by electrostatic potential calculated from partial charges assigned to atoms used for the calculating the surface.

Chapter V

Discussion and Perspectives

Chapter V
Discussion and Perspectives

HIV/AIDS, a disease with global impact and importance, has proved to be a challenging therapeutic target decades after discovery and extensive scientific inquiry (1, 4). The greatest hurdle in discovering a "cure" arises from the virus' ability to mutate, leading to changes in molecular recognition that allow viral proteins to function normally while evading small molecule inhbitors, thus causing drug resistance — the predominant cause of failed anti-HIV therapy (5, 248, 249).

Targeting viral proteins alone for therapy is an almost Sisyphean task — resistance to novel drugs often emerges during initial clinical trials, greatly diminishing the impact of drug discovery efforts on the disease landscape (19). An alternative to investigating viral proteins exclusively for drug discovery is to target host proteins (host factors) that interact with HIV during the viral life-cycle (13, 28). Anti-HIV host factors have the distinct advantage of side-stepping drug resistance due to mutational alteration and may prove to be more fruitful anti-HIV drug discovery targets.

The innate immune response may be triggered by sensing of pathogen-associated molecular patterns (PAMP) such as nucleic acids by Pattern Recognition Receptors (PRR) for example Toll-Like Receptors (TLR). TLR3 and TLR7/8 recognize dsRNA and ssRNA respectively triggering downstream signaling that leads to type 1 interferon (IFN-α/β) production and A3G up-regulation (105). Rogue nucleic acid fragments, be they in the form of endogenous retrotransposons (LTRs/LINE-1) or external pathogenic retroviruses (HIV-1), pose a significant enough threat to normal functioning of eukaryotic cells that numerous innate immune response pathways and factors that neutralize such threats have been conserved throughout evolution.

Role of APOBEC3 proteins APOBEC3 (A3) proteins are an important part of the host innate immune response, specifically directed against genomic retroelements and retroviruses. Members of the human A3 subfamily of proteins have proved to be important anti-HIV host factors that hypermutate the reverse transcribed HIV ssDNA, by deaminating Cytidines to Uridines, rendering it crippled or ineffectual (57, 117). The importance of this class of host factors may be gauged from the fact that the HIV genome encodes a protein, Vif (virion infectivity factor), that targets A3G and A3F to a Cul5-E3 ubiquitin ligase complex for subsequent proteasomal degradation (76, 250). Extensive and detailed structural data on the A3G-Vif interaction may provide the understanding required for its disruption by small molecule ligands, and allowing A3G to impede the viral life-cycle (13, 131, 251).

Vif is predicted to be intrinsically disordered and thus challenging to express and purify in quantities required for high-resolution structure determination. However, partial structural data available in the literature (132, 137) may be used in conjunction with homology models of Vif-binding A3 protein domains, such as A3G-NTD and A3F-CTD, to generate testable hypotheses regarding interaction determinants. Furthermore, the recent discovery of CBFβ (Core Binding Factor - beta subunit, a transcription factor) as a novel member of the Vif-Cul5-E3 ubiquitin ligase complex is an encouraging development, that may herald the discovery of new structural details relevant to Vif-A3G binding, A3G-polyubiquitylation and subsequent proteasomal degradation (252).

Other human APOBEC3 proteins, notably A3C (253), A3H (239), and some A3A isoforms (254) have been demonstrated to exert an anti-HIV-1 effect. The impact of these and other APOBEC3 subfamily members on HIV-1 warrants extensive investigation (255). The structural features of A3C have been investigated by biochemical methods and computational models used to generate testable hypotheses on substrate binding (256, 257). Similar stud-

ies on A3A propose the presence of specific structural features that may enable binding to polynucleotide substrates (234).

An emerging functional role for APOBEC proteins in controlling DNA methylation suggests that the scope of this class of enzymes extends beyond the immune system to gene expression control and genome regulation (55, 258, 259). Furthermore, since cytidine deaminases are known to act as polynucleotide mutators (68, 166), their role in carcinogenesis merits further investigation.

The importance of Structural data High-resolution structures of various proteins, determined over the past few decades, have proved critical in enhancing our understanding of many aspects related to their function, such as higher-order organization (oligomerization, complex formation), enzymatic action (substrate binding, enzyme chemistry, inhibitor binding) and regulation (post-translational modifications). Most importantly, such structures have laid the foundation for an entirely new approach to drug discovery, being used as targets for *in-silico* docking and virtual screening to discover novel ligands. The structures serve not only as the staring point for structure-based drug design, an approach that has led to numerous clinically useful drugs and promising drug candidates (260–262), but when co-crystallized with ligands, enable verification of the ligand binding modes, thus helping narrow down the search for novel chemical entities and optimizing chemical library design efforts. Data from multiple high-resolution co-crystal structures of a protein with various ligands may be useful in identifying structural features and chemical moieties that lead to drug resistance in rapidly evolving targets, such as employed in the case of HIV-1 protease (248, 249, 263).

The high-resolution crystal structures of A3G catalytic C-terminal domain (A3G-CTD) have established a framework for answering various questions relating to A3G function (147,Chapter III). Details regarding secondary structure and tertiary fold highlight the

structural differences between A3G-CTD and human APOBEC2, the only other member of the human APOBEC protein family with a high-resolution structure (157). The A3G-CTD structures confirm the active site architecture is similar to known cytidine deaminases, with a catalytically critical Zn^{2+} atom coordinated, in A3G-CTD, by one histidine and two cysteine residues positioned near a catalytically essential Glutamate residue, suggesting a shared deamination reaction mechanism (77).

The intermolecular interfaces identified from the A3G-CTD crystal structures, especially interfaces 1 and 2, may be of functional importance. Interface 1, the largest in terms of buried surface area, is located near the active site region and a zone of positively charged residues that coordinate a Cl^- ion in the crystal structures (Chapter II). The negatively charged Cl^- may be incorporated as a result of the crystallization screen containing $MgCl_2$. However, this also suggests that a zone of positive charge near the active site region is a fundamental property of the A3G-CTD that may be functionally relevant to bind the negatively charged phosphodiester backbone of nucleic acid substrates. Further exploration regarding the formation of this interface in solution, for example, by studying the oligomerization propensity of interface disrupting mutants, that do not directly affect structural stability, using biophysical and biochemical methods, may help establish the functional importance of this interface.

Interface 2, the second largest observed in the A3G-CTD crystal structures, incorporates an intermolecular Zn^{2+} coordination site that appears specific to A3G-CTD. The Zn^{2+} coordination at this site is different than at the active site, with two histidine residues contributed from the longest loop of one A3G-CTD molecule, and one cysteine residue, on the outer surface of a helix, from the reciprocating second A3G-CTD molecule. The Zn^{2+} coordinating loop is of variable length in other human A3 proteins, as seen from the multiple sequence alignment and homology models, and may be incapable of such

intermolecular Zn^{2+} coordination, thus suggesting a unique role this region may play in defining the characteristics of each protein.

The full-length structure of A3G will help answer critical questions regarding oligomerization, substrate binding and activity. For example, the propensity to oligomerize preferentially via the NTD or the CTD may impact substrate binding depending on the charge profile presented to the substrate polynucleotides. Full-length A3G has been demonstrated to bind single-stranded DNA primarily as oligomers, using Atomic Force Microscopy (AFM) (198). Delineating the interfacial interactions from high-resolution structures of full-length A3G may help explain the driving force for such observations and suggest putative models for substrate binding and activity.

Drug discovery efforts In conjunction with, or even in the absence of, structure-based drug design, high-throughput screening assays may help identify chemical entities that bind the target protein. Co-crystal structures, of the target protein and the "hits" so discovered, are critical to delineate the binding mode and to focus on specific chemical classes for further investigation. The co-crystal structure of A3G-CTD, with a "hit" (MN30) from a high-throughput assay, was solved after soaking the "apo" A3G-CTD crystals with concentrated solutions of MN30. The "hit" compound, MN30, binds covalently to the only cysteine residue (C308) on the protein surface that was 'free', i.e. not buried and not involved in Zn^{2+} coordination. Further investigation revealed that the binding of MN30 to another cysteine residue (C321) near the active site region underlies this inhibition, most likely arising from steric-hinderance to substrate – ssDNA – binding. Thus structural data from co-crystallization experiments, though providing no direct evidence of inhibitory activity, served as a foundation for further investigation and modeling a highly probable MN30 binding mode. Furthermore, these experiments suggest future application of this class of compounds as molecular probes to investigate A3G and other related A3 proteins

may be a reasonable approach.

Binding of a small molecule to a protein, such as MN30 in the case of A3G-CTD, may interfere with not just function but also other characteristics of the protein. The covalent modification of sulfhydryl groups on the 'free' cysteine residues in A3G-CTD by catechol moieties, as present in MN30, may lead to not just enzymatic inhibition but also prevent co-crystallization of A3G-CTD incubated with MN30 by binding to C261, required for the intermolecular Zn^{2+} coordinating site at interface 2, and thus hinder the formation of crystal contacts required for crystal nucleation and growth. The final result being a lack of co-crystals by pre-incubating A3G-CTD with MN30 and lower quality data from A3G-CTD crystals soaked with MN30 due to possible disruption of internal crystallographic order.

The structural mechanisms that prevent or disrupt intermolecular interactions, for example between an enzyme and its ligand, may also be exploited for enhancing protein activity. This is especially relevant in the case of A3G where the two forms of the protein exist in the cell, an active low molecular mass (LMM) form and a dormant high molecular mass (HMM) complex composed of A3G oligomers with RNA and other proteins (107, 108). Drug-like small molecules that disrupt the HMM complexes so as to release A3G for packaging into HIV particles are being explored further as a promising approach to exploit the anti-retroviral effect of A3G (174, 175). The A3G-CTD structures may be used as templates for a full-length A3G homology model to better explore the mechanism(s) of action of such compounds.

More importantly, the attempted neutralization of A3G by Vif may not be absolute and thus result in a background level of A3G driven hypermutations that allow a subset of mutated viral populations to survive and eventually give rise to drug resistant HIV variants (21, 124, 264). Coupled with the error-prone viral reverse transcriptase, sublethal

mutagenesis may contribute to the repertoire of mutations that assist HIV in escaping anti-retroviral drugs, leading to treatment failure. In the absence of a promising strategy to disable Vif mediated degradation of A3G and create a detriment to HIV (140, 265), approaches that inhibit sublethal mutagenesis arising from background A3G activity, may reduce the mutational advantage available to the virus. In this context, structure-based drug-design efforts to inhibit A3G activity may provide a useful, even if accessory, approach to removing this mutational advantage.

Future direction Structural and biophysical exploration of APOBEC3 proteins is critical to understanding the basis of their function and to exploit their potential, by inhibition or activation, for abrogating HIV infection. Structural data on APOBEC2 and A3G-CTD may be used for modeling the structures of related APOBEC proteins using *in-silico* methods and for structure-based drug design by high-throughput virtual screening. Substrate polynucleotides may be docked to these structures and homology models to gain insights into putative binding modes and further using that information to develop novel drug-discovery strategies. The same may also assist in designing protein constructs with mutations to improve crystallization or co-crystallization outcomes.

Appendices

Appendix A

A reservoir-free crystallization method

1.1 Preface

The following method, based on modification of the method described in "Rayment, I. (2002) Structure 10, 147-151." (266), helped improve the size and diffraction quality of A3G-CTD crystals, in conjunction with the A3G-191-380-2K2A construct, based on conditions where crystals were obtained previously. The principal premise of the publication (266) being that the nucleation and growth of protein crystals under vapor diffusion (VD) conditions occurs under conditions that may never be precisely defined. Deviations in micro-environmental conditions may affect the rate of VD, thus influencing transition from nucleation promoting to growth promoting conditions in the crystallization drops. The final concentration of reservoir solution components such as precipitant, buffer and salt, are not identical between the drop and the reservoir during nucleation and growth, thus crystals nucleate and grow in drop conditions that are imprecisely quantified. The advantage to defining these condition as precisely as possible would be an improved likelihood of consistent crystal nucleation and growth during subsequent trials and ideally a systematic method for improving crystal size and quality. The method described here is a modified high-throughput version of the approach described in the publication referred to above (266).

1.2 Method and Results

The high-throughput method was setup using a crystallization-tray setup robot (Phoenix – Art Robbins Instruments, LLC) in a 96-well tray format. The primary variable was the salt ($MgCl_2$) concentration, keeping the concentration of all other screen components constant (12% PEG4000, 0.1 M HEPES pH 7.5) as well as the protein concentration (10 mg/mL)

and volume (500 nL) identical in all drops. Thus, $MgCl_2$ concentration varied across the tray, in a non-linear fashion (Figure 1.2), whereas, the total drop volumes varied as per the volume of screen used on a per-row basis (Figure 1.1).

The results from the above trial identified one highly favorable condition providing a large single crystal (Figure 1.3(A)) and a small set of conditions that provided reasonable hits (Figure 1.3(B)). The calculated effective $MgCl_2$ concentration was similar, but not identical in the drops that provided the crystals, the minor variations in concentration were accentuated by the differences in total drop volume — perhaps a more important variable than $MgCl_2$ concentration alone. The majority of crystals could be harvested and diffracted to below 2.0Å resolution under synchrotron sourced X-ray radiation.

(A) MgCl2 concentration(M) in screen solutions used for drop setup:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
2	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
3	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
4	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
5	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
6	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
7	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525
8	0.250	0.275	0.300	0.325	0.350	0.375	0.400	0.425	0.450	0.475	0.500	0.525

(B) Screen volume(nL)used for each drop: Protein volume is fixed in all drops (500 nL). (No reservoir.)

	1	2	3	4	5	6	7	8	9	10	11	12
1	100	100	100	100	100	100	100	100	100	100	100	100
2	200	200	200	200	200	200	200	200	200	200	200	200
3	300	300	300	300	300	300	300	300	300	300	300	300
4	400	400	400	400	400	400	400	400	400	400	400	400
5	500	500	500	500	500	500	500	500	500	500	500	500
6	600	600	600	600	600	600	600	600	600	600	600	600
7	700	700	700	700	700	700	700	700	700	700	700	700
8	800	800	800	800	800	800	800	800	800	800	800	800

(C) Effective MgCl2 concentration(M) after drop setup:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0417	0.0458	0.0500	0.0542	0.0584	0.0626	0.0667	0.0709	0.0751	0.0793	0.0835	0.0877
2	0.0714	0.0786	0.0857	0.0929	0.1000	0.1072	0.1143	0.1215	0.1287	0.1358	0.1430	0.1501
3	0.0938	0.1031	0.1125	0.1219	0.1313	0.1407	0.1500	0.1594	0.1688	0.1782	0.1876	0.1970
4	0.1111	0.1222	0.1333	0.1445	0.1556	0.1667	0.1778	0.1889	0.2001	0.2112	0.2223	0.2334
5	0.1250	0.1375	0.1500	0.1625	0.1750	0.1875	0.2000	0.2125	0.2250	0.2376	0.2501	0.2626
6	0.1364	0.1500	0.1636	0.1773	0.1909	0.2046	0.2182	0.2318	0.2455	0.2591	0.2728	0.2864
7	0.1458	0.1604	0.1750	0.1896	0.2042	0.2188	0.2334	0.2479	0.2625	0.2771	0.2917	0.3063
8	0.1538	0.1692	0.1846	0.2000	0.2154	0.2308	0.2462	0.2616	0.2769	0.2923	0.3077	0.3231

Figure 1.1

Figure 1.1 : The screen is setup in a 96-well ARI INTELLI-PLATE 96-3 (Art Robbins Instruments, LLC) and consists of 12% PEG4000, 0.1M HEPES pH 7.5 and $MgCl_2$ varying in concentration: **(A)** $MgCl_2$ concentration is the same in each column, linearly increasing in steps of 0.025 M. **(B)** the screen volume is the same in each row, increasing linearly in steps of 100 nL. The protein (A3G-CTD at 10 mg/mL) is constant throughout the plate at 500 nL per drop. **(C)** Effective $MgCl_2$ concentration varies throughout the tray as tabulated. Thus, two entities that vary are : (i). $MgCl_2$ concentration varies throughout the plate, and (ii). the *total* drop volume varies across the rows. The best crystal hit obtained with this screen occurred in the condition colored in green (see image: Figure 1.3(A)), additional hits (but no single or as large crystals) occurred at conditions marked in yellow (see image: Figure 1.3(B)).

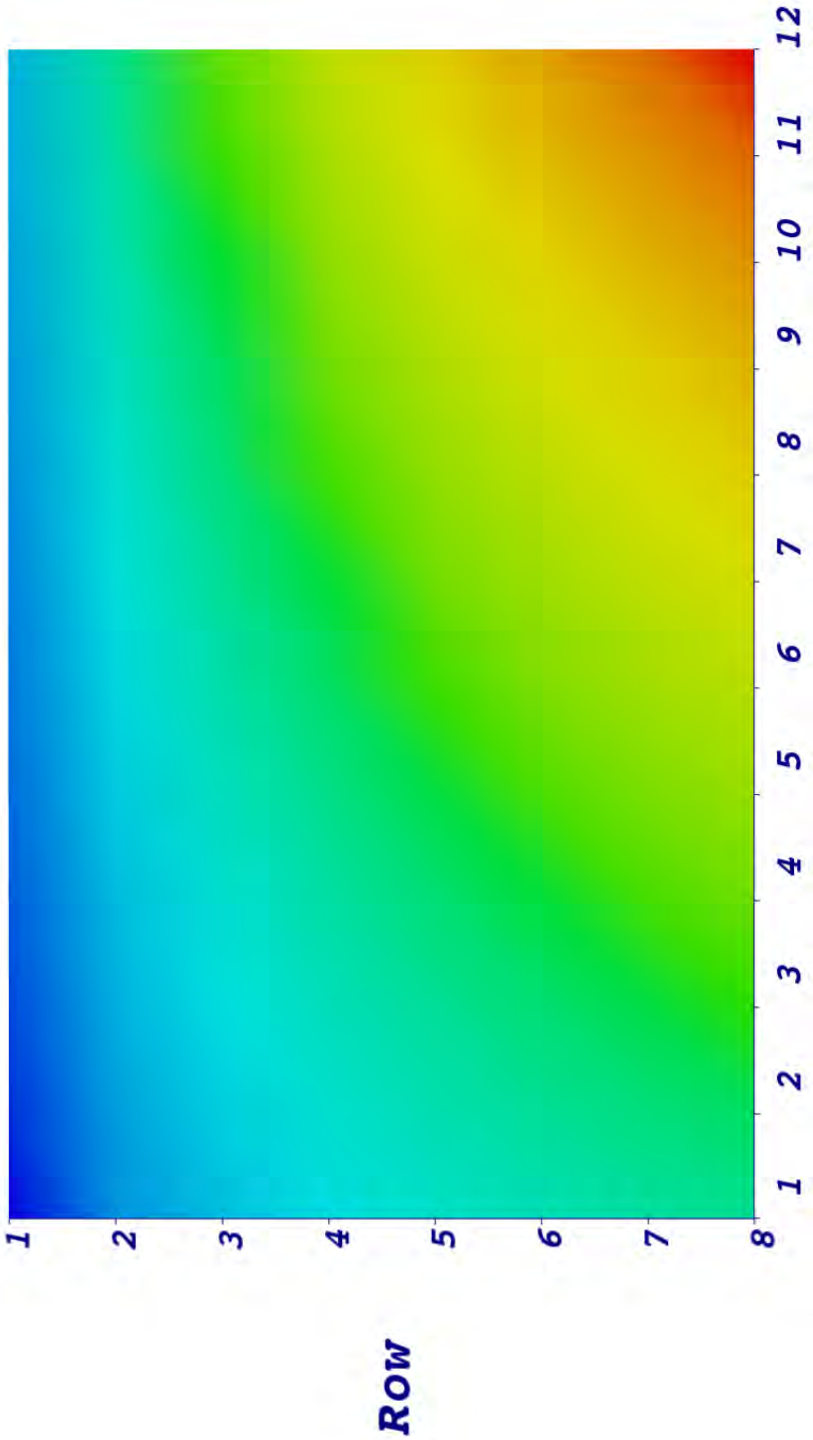


Figure 1.2

Figure 1.2 : The variation in effective $MgCl_2$ concentration across the 96-well tray varies in a *non-linear* manner as evidenced by this colored heatmap, with concentration increasing from Blue to Red. Based on data from Figure 1.1(C). Image created with Mayavi2 (267).

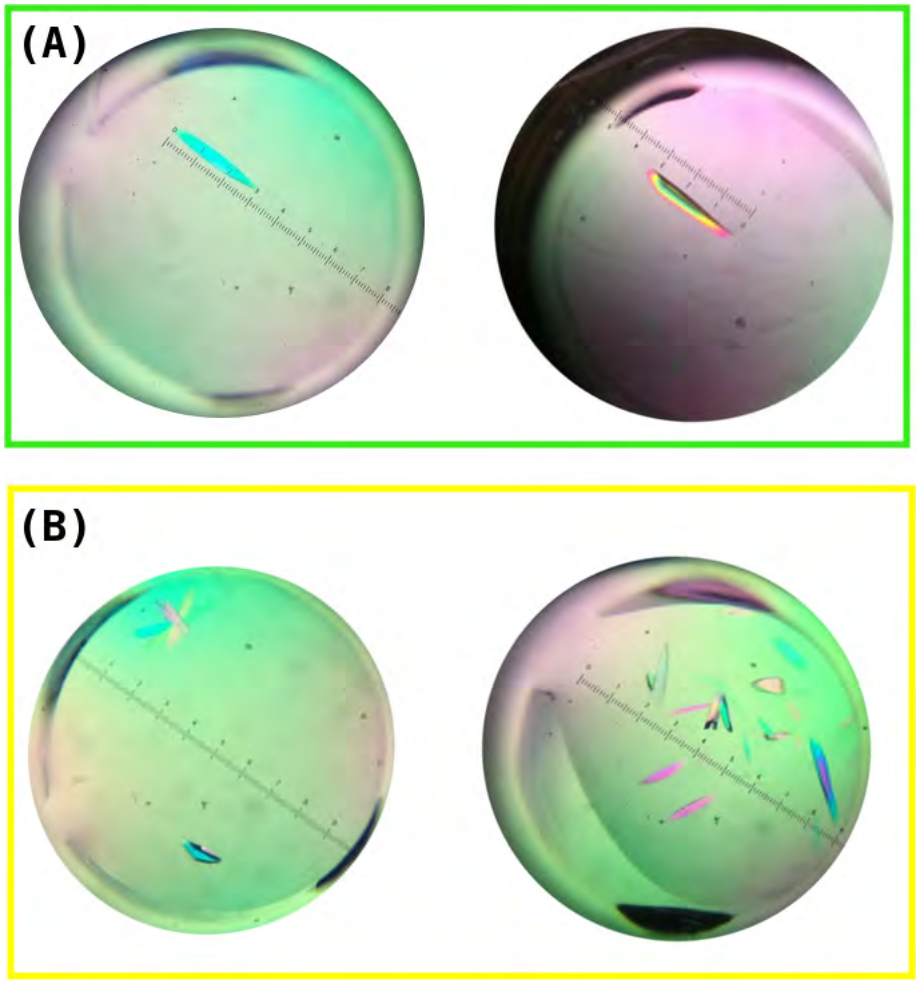


Figure 1.3

Figure 1.3 : **(A)** Large, single crystal obtained in the well colored green in Figure 1.1. The two images are of the same crystal, 48 hours apart. **(B)** Multiple, harvestable crystals obtained in conditions colored yellow in Figure 1.1.

Appendix B

Additional crystal structures of A3G-191-380-2K2A

2.1 Preface

A subset of various compounds identified as A3G inhibitors in the HTS assay (Chapter 3), were used for co-crystallization by pre-incubating A3G-191-380-2K2A with these compounds before setting up crystallization trials. However, no crystal hits were observed in these trials. Subsequently, A3G-191-380-2K2A crystals, obtained using methods defined in Appendix A, were soaked with solutions of these compounds in an attempt to obtain structures of protein-compound complexes. Multiple X-ray diffraction datasets were obtained for the soaked crystals. However, upon solving the structures, no evidence for the presence of any of these compounds could be observed in the electron density maps. The crystallographic data and refinement statistics for various such crystals are listed in Figure 2.1.

Number	DatasetName	Resolution (Å)	Spacegroup	a (Å)	b (Å)	c (Å)	alpha (°)	beta (°)	gamma (°)	Mol/AU	Completeness (Overall)	Completeness (Last shell)	Total Reflections	Unique Reflections	Working Reflections	I/sigma (Overall)	Average Redundancy	Rmerge (%) (Overall)	Rpim (%) (Overall)	RMSD (Å) Bonds	RMSD (°) Angles	Rfactor (%)	Rfree (%)
1	MN30_ON_3	1.47	P21	59.46	67.67	60.84	90.00	106.74	90.00	2	97.00	87.10	270272	76223	69401	11.9	3.5	6.5	4.1	0.0107	1.326	23.04	26.01
2	2K2A380_MN30_50k	1.48	P212121	68.11	72.20	96.78	90.00	90.00	90.00	2	93.63	95.21	316227	77879	71840	14.9	4.1	4.7	2.7	0.0094	1.359	20.01	22.78
3	MN30_ON_5	1.73	P21	59.86	67.66	60.80	90.00	106.55	90.00	2	81.20	87.70	95839	39087	32048	8.5	2.5	7.2	5.1	0.0120	1.433	24.91	28.33
4	MN30_ON_6	2.07	P21	59.42	67.75	60.65	90.00	106.58	90.00	2	98.40	99.30	87056	27728	25545	6.5	3.1	12.6	8.1	0.0097	1.299	25.54	30.12

Figure 2.1

Figure 2.1 : Table summarizing the crystallographic statistics for various "apo" A3G-191-380-2K2A crystals that diffracted and the strucutres solved. The "DatasetName" includes the compound identifier for the compound solution that was used for soaking each crystal. None of the solved structures had any electron density indicating the presence of the soaked compound (hence "apo" crystals).

Appendix C

In-silico model of the CaM + Kv channel complex

3.1 Preface

The following work was done with Dr. William Kobertz at U. Mass. Med. School. The Kobertz lab works on ion channels, their interaction partners and methods to modulate channel function. The crystal structures of voltage gated potassium channels (Kv) and calcium-calmodulin (CaM) are solved, However, no known co-crystal structures or other high-resolution structural data is available that may help delineate their interaction. Preliminary data from experiments that tested for interaction between Kv and CaM using polymeric hydrocarbon probes (268, 269) reveled that T34 and T110 of CaM may lie 42Å and 52Å respectively from the "bottonm" end of the channel pore. The CaM + Kv channel complex was modeled using these data and available crystal structure in order to better understand this interaction.

3.2 Method and Results

The "ring" of CaM subunits (1G4Y) was docked to the rKV channel structure (2A79) by visual inspection in PyMOL so as to prevent any clashes amongst the crystal structures. A pseudo-atom was placed at the intracellular end of the channel pore, at a distance of 20Å from the closest potassium ion in pore crystal structure, so as to define a region where a chemical probe – depending on its length – may block the pore. One CaM subunit was modeled in to satisfy distance constraints for the two Threonine residues, T34 and T110, to the pseudo-atom, within a tolerance of ±5Å. Three other CaM subunits were generated by symmetry operation commands within PyMOL and nudged into position to satisfy the same distance constraints.

The placement and orientation of CaM subunits in relation to the Kv channel could not be unambiguously determined based on two distance constraints. The CaM subunits could be oriented with the additional constraint that their N-termini face the intracellular end of the channel pore. Beyond these constraints, the orientation of CaM subunits was completely ambiguous in terms of the *rotational orientation* of the CaM subunit "ring" with respect to the Kv channel along the channel pore axis as in Figure 3.1 and Figure 3.3. Similarly, the rotational orientation of each CaM subunit along its internal axes was ambiguous and multiple conformations could be modeled that satisfied the two distance constraints (Figure 3.1, Figure 3.2, Figure 3.3 and Figure 3.4).

The final models are being used to generate testable hypotheses and define other distance constraints to delineate the CaM + Kv channel complex interactions.

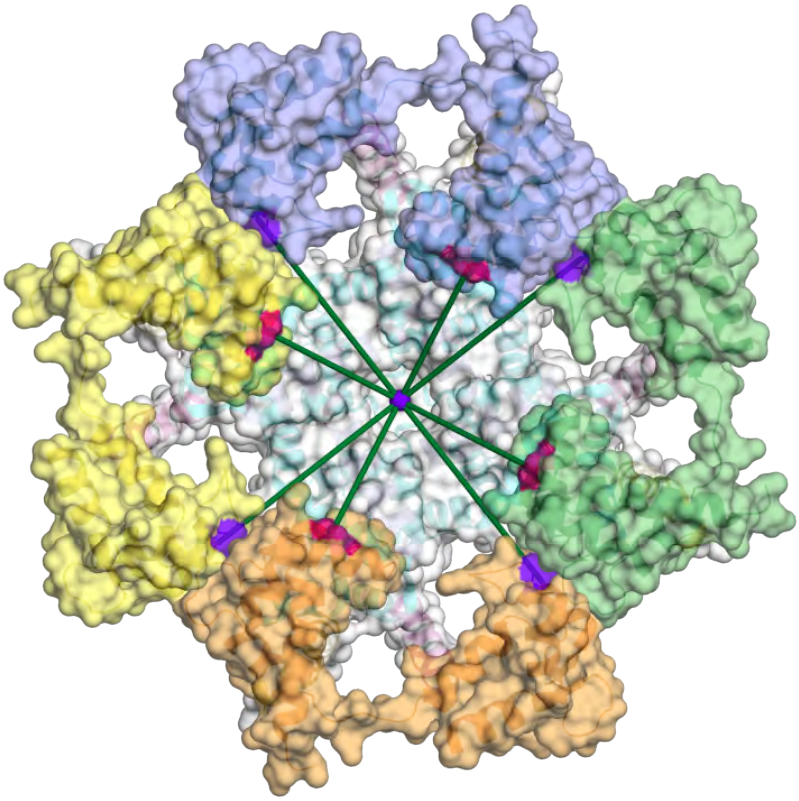


Figure 3.1

Figure 3.1 : CaM + rKv conformation #1 – looking up : Modeled CaM + rKv conformation based on distance data. View looking at the channel pore from intracellular side. CaM T34 – Magenta surface, CaM T110 – Purple surface. Pseudo-atom – Purple sphere at the center.

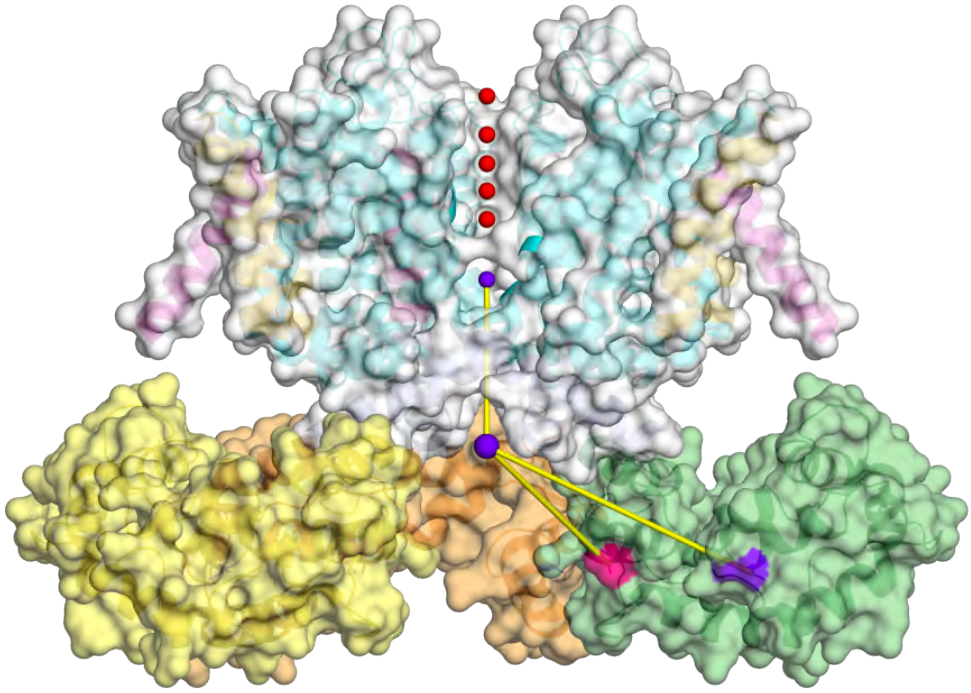


Figure 3.2

Figure 3.2 : Modeled CaM + rKv conformation based on distance data. View looking at a cross-section of the channel pore. Kv potassium ions – Red. CaM T34 – Magenta surface, CaM T110 – Purple surface. Pseudo-atom – Larger purple sphere.

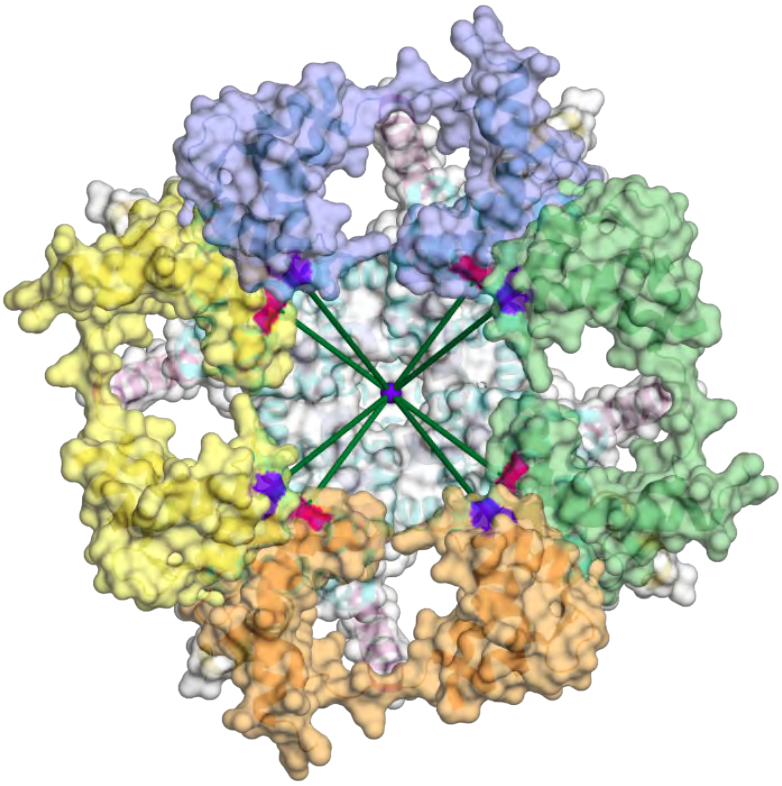


Figure 3.3

Figure 3.3 : Modeled CaM + rKv conformation based on distance data. View looking at the channel pore from intracellular side. CaM T34 – Magenta surface, CaM T110 – Purple surface. Pseudo-atom – Purple sphere at the center.

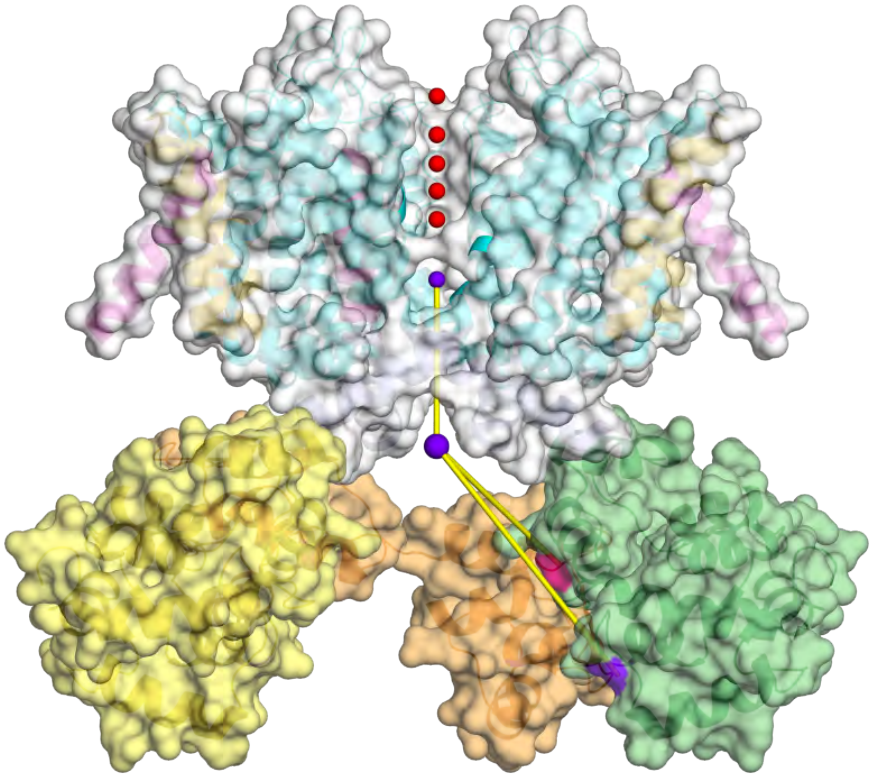


Figure 3.4

Figure 3.4 : Modeled CaM + rKv conformation based on distance data. View looking at a cross-section of the channel pore. Kv potassium ions – Red. CaM T34 – Magenta surface, CaM T110 – Purple surface. Pseudo-atom – Larger purple sphere.

Appendix D

Theoretical and computational analysis of HIV-1 Vif

This appendix is based in part on the paper published in
Proteins: Structure, Function, and Bioinformatics

Jared R. Auclair, Karin M. Green, **Shivender Shandilya**, James E. Evans, Mohan Somasundaran, Celia A. Schiffer. (2007) **Mass spectrometry analysis of HIV-1 Vif reveals an increase in ordered structure upon oligomerization in regions necessary for viral infectivity.** *Proteins: Structure, Function, and Bioinformatics* 69(2):270-84. PMID: 17598142.

My contribution to this publication was the analysis of HIV-1 Vif in terms of predicted intrinsic disorder: **Figure 4.1.**

4.1 Preface

HIV-1 Vif (virion infectivity factor) plays a critical role in targeting human APOBEC3G, an anti-HIV protein, for degradation via the cellular proteasomal pathway (76). Furthermore, HIV-1 Vif is predicted to be intrinsically disordered by numerous software algorithms (Figure 4.1 and Figure 4.2), a property that likely has important implications for structure, function and biochemical properties.

4.2 Method and Results

Prior to analysis of the cross-linked HIV-1 Vif samples by MS, predictors of natural disordered regions (PONDR), was used to determine regions of intrinsic disorder in HIV-1 Vif by submitting the HXB2 HIV-1 Vif sequence (270, 271). This program uses four different predictor routines to suggest regions of intrinsic disorder. Only one of the four predictors, VL-XT, predicted the N-terminus of HIV-1 Vif to be disordered, and two of the predictors, VL-XT and XL1-XT, predicted two short regions (possibly extended loops) between residues 50–63 and 87–100 to be disordered. However, all four of the predictors, unanimously predicted the C-terminus to be disordered (Figure 4.1). Therefore, this suggests the C-terminus of HIV-1 Vif is intrinsically disordered. In addition to predicting

regions of intrinsic disorder, HIV-1 Vif secondary structure was predicted using the Pole BioInformatique Lyonnais Network Protein Sequence Analysis (NPS) secondary structure consensus prediction program because intrinsically disordered proteins often do have defined secondary structure. The majority of HIV-1 Vif secondary structure is predicted to be random coil, however, there is predicted secondary structure: the relatively ordered N-terminus is predicted to consist of mostly β-sheets and one α-helix and the disordered C-terminus is predicted to consist of mostly α-helices.

The same HIV-1 Vif sequence was also submitted to FoldIndex (149) for disorder prediction and correlated broadly with PONDR VL-XT predictions (Figure 4.1 and Figure 4.3). The MS cross-linking data when projected onto the FoldIndex predictions correlated with the majority of cross-links in predicted disordered regions and *few* cross-links in predicted *ordered regions* (Figure 4.4).

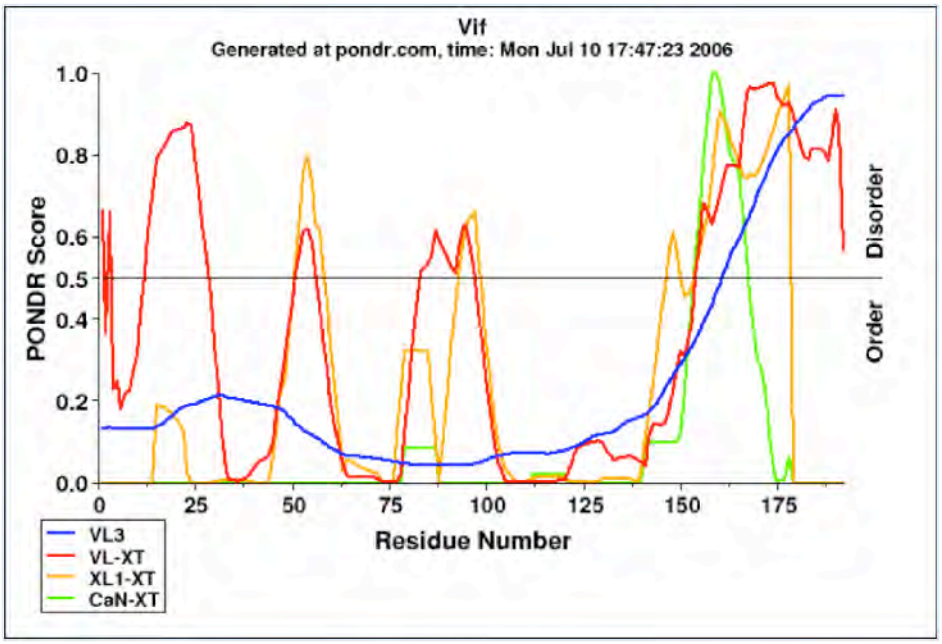

Vif
Generated at pondr.com, time: Mon Jul 10 17:47:23 2006
PONDR Score
1.0
0.8
0.6
0.5
0.4
0.2
0.0
0
25
50
75
100
125
150
175
Residue Number
Disorder
Order
VL3
VL-XT
XL1-XT
CaN-XT

Figure 4.1

Figure 4.1 : Predicted regions of intrinsic disorder for HIV-1 Vif using PONDR, Predictors of Natural Disordered Regions (270, 271). The four colors indicate four separate predictor algorithms: blue VL3, red VL-XT, yellow XL1-XT, and green CaN-XT.

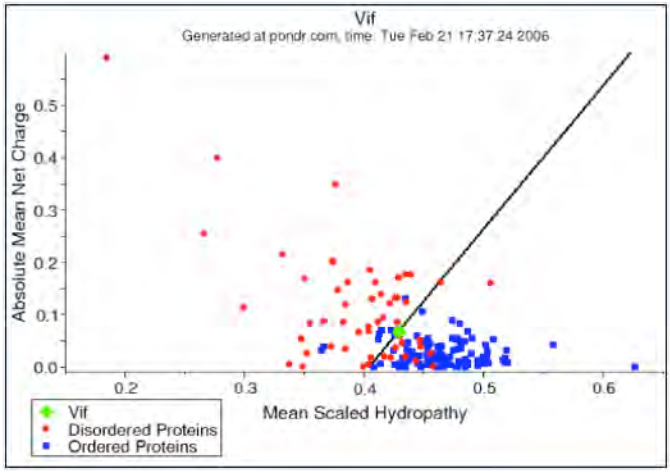

Vlf
Generated at pondr.com, time: Tue Feb 21 17:37:24 2006
Absolute Mean Net Charge
0.0
0.1
0.2
0.3
0.4
0.5
0.2
0.3
0.4
0.5
0.6
Mean Scaled Hydropathy
Vlf
Disordered Proteins
Ordered Proteins

Figure 4.2

Figure 4.2 : Location of HIV-1 Vif in terms of protein charge *vs* hydrophobicity using PONDR, Predictors of Natural Disordered Regions (270, 271).

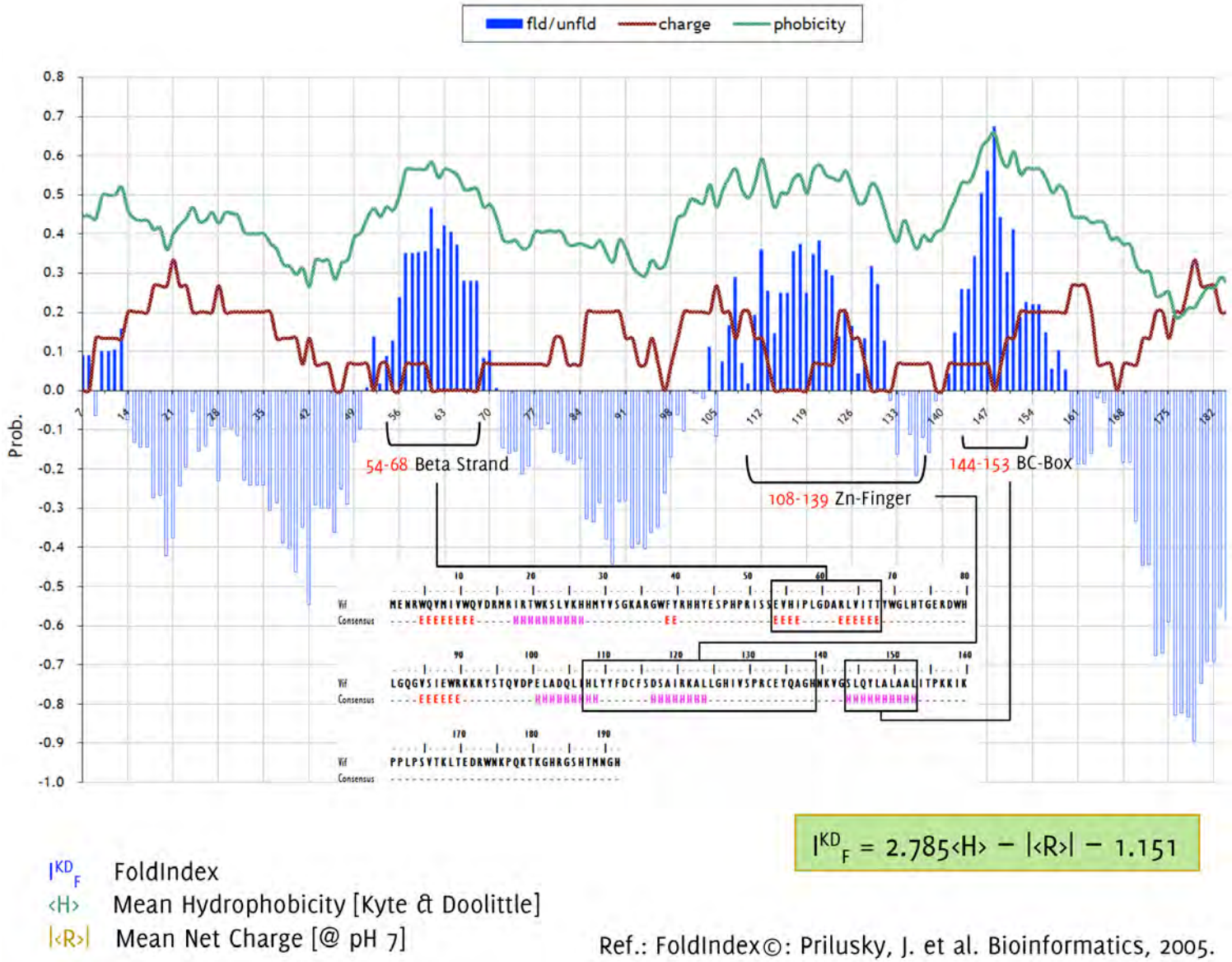

fld/unfld
charge
phobicity
Prob.
54-68 Beta Strand
108-139 Zn-Finger
144-153 BC-Box
$I^{KD}_F = 2.785\langle H\rangle - |\langle R\rangle| - 1.151$
I^{KD}_F FoldIndex
⟨H⟩ Mean Hydrophobicity [Kyte & Doolittle]
|⟨R⟩| Mean Net Charge [@ pH 7]
Ref.: FoldIndex©: Prilusky, J. et al. Bioinformatics, 2005.

Figure 4.3

Figure 4.3 : FoldIndex (149) predicted intrinsic disorder in HIV-1 Vif protein sequence correlated with various features of Vif known from literature. Predicted secondary structure: H – α-helices and E – β-strands.

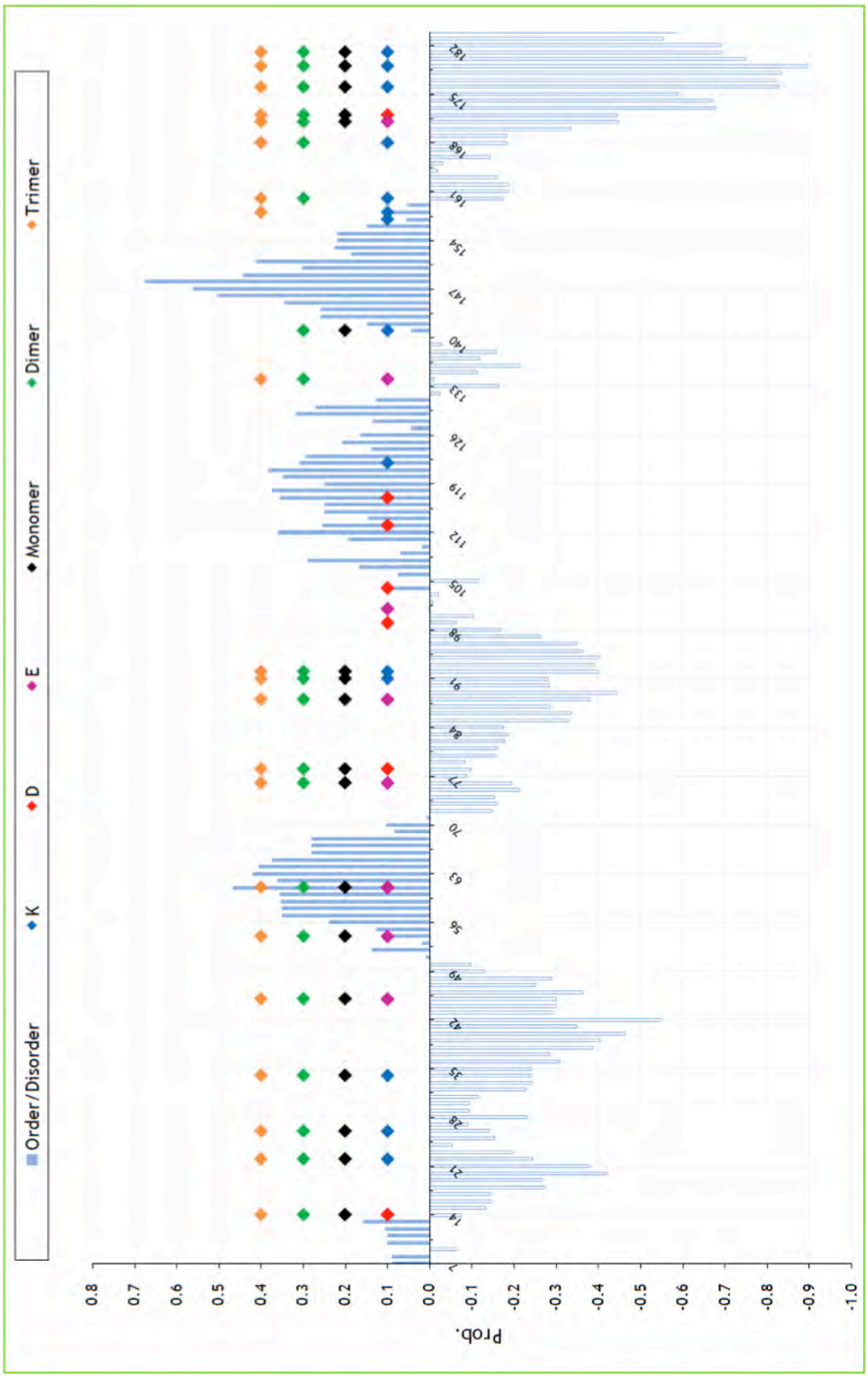

Order/Disorder
K
D
E
Monomer
Dimer
Trimer
Prob.
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
182
175
168
161
154
147
140
133
126
119
112
105
98
91
84
77
70
63
56
49
42
35
28
21
14
7

Figure 4.4

Figure 4.4 : Mass Spectrometry data on HIV-1 Vif coss-linking, from the publication listed at the beginning of this appendix (132), was mapped onto the FoldIndex (149) predicted intrinsic disorder in the protein sequence. All cross-linkable residues are indicated by blue (Lysine), red (Aspartate) or purple (Glutamate) colored diamonds. Cross-links deduced from MS data are indicated by black, green or orange colored diamonds for Vif monomers, dimers and trimers respectively.

References

[1] Basavapathruni, A. & Anderson, K. S. Reverse transcription of the HIV-1 pandemic. *The FASEB journal official publication of the Federation of American Societies for Experimental Biology* **21**, 3795–3808 (2007).

[2] Gallo, R. C. *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220**, 865–867 (1983).

[3] Barre-Sinoussi, F. *et al.* Isolation of A T-Lymphotropic Retrovirus from A Patient at Risk for Acquired Immune-Deficiency Syndrome (Aids). *Science* **220**, 868–871 (1983).

[4] Sidibe, M. & UNAIDS. UNAIDS Report on the global AIDS epidemic 2010 (2010).

[5] Anderson, J., Schiffer, C., Lee, S.-K. & Swanstrom, R. Viral protease inhibitors. *Handbook of experimental pharmacology* **1**, 85–110 (2009).

[6] Berger, A. *et al.* Interaction of Vpx and apolipoprotein B mRNA-editing catalytic polypeptide 3 family member A (APOBEC3A) correlates with efficient lentivirus infection of monocytes. *The Journal of biological chemistry* **285**, 12248–54 (2010).

[7] Blanco, J.-L., Varghese, V., Rhee, S.-Y., Gatell, J. M. & Shafer, R. W. HIV-1 integrase inhibitor resistance and its clinical implications. *The Journal of infectious diseases* **203**, 1204–14 (2011).

[8] Johnson, V. A. *et al.* Update of the drug resistance mutations in HIV-1: December 2010. *Topics in HIV medicine : a publication of the International AIDS Society, USA* **18**, 156–63 (2010).

[9] Tomescu, C., Abdulhaqq, S. & Montaner, L. J. Evidence for the innate immune response as a correlate of protection in human immunodeficiency virus (HIV)-1 highly exposed seronegative subjects (HESN). *Clinical and experimental immunology* **164**, 158–69 (2011).

[10] Aouizerat, B. E., Pearce, C. L. & Miaskowski, C. The search for host genetic factors of HIV/AIDS pathogenesis in the post-genome era: progress to date and new avenues for discovery. *Current HIV/AIDS reports* **8**, 38–44 (2011).

[11] Arhel, N. Revisiting HIV-1 uncoating. *Retrovirology* **7**, 96 (2010).

[12] Lever, A. M. L. & Jeang, K.-T. Insights into Cellular Factors That Regulate HIV-1 Replication in Human Cells. *Biochemistry* **50**, 920–31 (2011).

[13] Tavassoli, A. Targeting the protein-protein interactions of the HIV lifecycle. *Chemical Society reviews* **40**, 1337–46 (2011).

[14] Chiu, I. M. *et al.* Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature* **317**, 366–368 (1985).

[15] Gomez, C. & Hope, T. J. The ins and outs of HIV replication. *Cellular microbiology* **7**, 621–6 (2005).

[16] Gorry, P. R. & Ancuta, P. Coreceptors and HIV-1 pathogenesis. *Current HIV/AIDS reports* **8**, 45–53 (2011).

[17] Marchand, C., Johnson, A. A., Semenova, E. & Pommier, Y. Mechanisms and inhibition of HIV integration. *Drug discovery today Disease mechanisms* **3**, 253–260 (2006).

[18] Fulcher, A. J. & Jans, D. a. The HIV-1 Tat transactivator protein: a therapeutic target? *IUBMB life* **55**, 669–80 (2003).

[19] Flexner, C. HIV drug development: the next 25 years. *Nature Reviews Drug Discovery* **6**, 959–966 (2007).

[20] Klatt, E. *Pathology of AIDS* (Mercer University, School of Medicine, Savannah, 2011), version 22 edn.

[21] Sadler, H. a., Stenglein, M. D., Harris, R. S. & Mansky, L. M. APOBEC3G Contributes to HIV-1 Variation Through Sublethal Mutagenesis. *Journal of virology* **84**, 7396–7404 (2010).

[22] Erickson, J. W., Gulnick, S. V. & Markowitz, M. Protease inhibitors: resistance, cross-resistance, fitness and the choice of initial and salvage therapies. *AIDS* **13**, S189—-S204 (1999).

[23] Schinazi, R. F., Larder, B. A. & Mellors, J. W. Mutations in retroviral genes associated with drug resistance. *Internatl Antiviral News* **5**, 129–142 (1997).

[24] Schouten, J. T. HIV drug resistance and the other causes of treatment failure. *STEP Perspect* **9**, 5–8 (1997).

[25] Vella, S. Clinical implications of resistance to antiretroviral drugs. *AIDS Clinical Care* **9** (1997).

[26] Kolli, M., Lastere, S. & Schiffer, C. A. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* **347**, 405–409 (2006).

[27] Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M. & Schiffer, C. A. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of Virology* **78**, 12446–12454 (2004).

[28] Adamson, C. S. & Freed, E. O. Novel approaches to inhibiting HIV-1 replication. *Antiviral research* **85**, 119–41 (2010).

[29] Frey, G. *et al.* Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 13938–43 (2006).

[30] Hatziioannou, T., Perez-Caballero, D., Yang, A., Cowan, S. & Bieniasz, P. D. Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10774–9 (2004).

[31] Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848–53 (2004).

[32] Zhang, F., Hatziioannou, T., Perez-Caballero, D., Derse, D. & Bieniasz, P. D. Antiretroviral potential of human tripartite motif-5 and related proteins. *Virology* **353**, 396–409 (2006).

[33] Black, L. R. & Aiken, C. TRIM5alpha disrupts the structure of assembled HIV-1 capsid complexes in vitro. *Journal of virology* **84**, 6564–9 (2010).

[34] Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361–365 (2011).

[35] Emerman, M. How TRIM5alpha defends against retroviral invasions. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5249–50 (2006).

[36] Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425–30 (2008).

[37] Perez-Caballero, D. *et al.* Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **139**, 499–511 (2009).

[38] Tokarev, A., Skasko, M., Fitzpatrick, K. & Guatelli, J. Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS research and human retroviruses* **25**, 1197–210 (2009).

[39] Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–33 (2009).

[40] Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology* **11**, 228–34 (2009).

[41] Hariharan, M., Scaria, V., Pillai, B. & Brahmachari, S. K. Targets for human encoded microRNAs in HIV genes. *Biochemical and Biophysical Research Communications* **337**, 1214–1218 (2005).

[42] Houzet, L. & Jeang, K.-T. MicroRNAs and human retroviruses. *Biochimica et biophysica acta* **xxx**, 1–8 (2011).

[43] Kumar, A. The silent defense: micro-RNA directed defense against HIV-1 replication. *Retrovirology* **4**, 26 (2007).

[44] Sung, T.-L. & Rice, A. P. miR-198 inhibits HIV-1 gene expression and replication in monocytes and its mechanism of action appears to involve repression of cyclin T1. *PLoS pathogens* **5**, e1000263 (2009).

[45] Triboulet, R. *et al.* Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science (New York, N.Y.)* **315**, 1579–82 (2007).

[46] Davidson, N. O. *et al.* Proposed nomenclature for the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme: APOBEC-1. *RNA (New York, N.Y.)* **1**, 3 (1995).

[47] Navaratnam, N. *et al.* The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *The Journal of biological chemistry* **268**, 20709–12 (1993).

[48] Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–296 (2002).

[49] Liao, W. *et al.* APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochemical and Biophysical Research Communications* **260**, 398–404 (1999).

[50] Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I. & Koonin, E. V. D. A. S. E. P. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* **4**, 1281–1285 (2005).

[51] Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. & Neuberger, M. S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Molecular Biology and Evolution* **22**, 367–377 (2005).

[52] Hall, R. S. *et al.* Three-dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry* **50**, 5077–85 (2011).

[53] Manis, J. P., Tian, M. & Alt, F. W. Mechanism and control of class-switch recombination. *Trends in immunology* **23**, 31–9 (2002).

[54] Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–63 (2000).

[55] Fritz, E. L. & Papavasiliou, F. N. Cytidine deaminases: AIDing DNA demethylation? *Genes & development* **24**, 2107–14 (2010).

[56] Chen, H. *et al.* APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. *Current Biology* **16** (2006).

[57] Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).

[58] Gallo, A. & Locatelli, F. ADARs: allies or enemies? The importance of A-to-I RNA editing in human disease: from cancer to HIV-1. *Biological reviews of the Cambridge Philosophical Society* **ePub**, ePub (2011).

[59] Tsutsumi, S. *et al.* Wobble inosine tRNA modification is essential to cell cycle progression in G(1)/S and G(2)/M transitions in fission yeast. *The Journal of biological chemistry* **282**, 33459–65 (2007).

[60] Chung, S. J., Fromme, J. C. & Verdine, G. L. Structure of human cytidine deaminase bound to a potent inhibitor. *Journal of Medicinal Chemistry* **48**, 658–660 (2005).

[61] Drake, J. C., Hande, K. R., Fuller, R. W. & Chabner, B. A. Cytidine and deoxycytidylate deaminase inhibition by uridine analogs. *Biochemical Pharmacology* **29**, 807–811 (1980).

[62] Schröder, J. K. *et al.* Constitutive overexpression of the cytidine deaminase gene confers resistance to cytosine arabinoside in vitro. *Leukemia official journal of the Leukemia Society of America Leukemia Research Fund UK* **10**, 1919–1924 (1996).

[63] Steuart, C. D. & Burke, P. J. Cytidine deaminase and the development of resistance to arabinosyl cytosine. *Nature New biology* **233**, 109–110 (1971).

[64] Deng, L.-Y., Wang, J.-P., Gui, Z.-F. & Shen, L.-Z. Antitumor activity of mutant bacterial cytosine deaminase gene for colon cancer. *World journal of gastroenterology : WJG* **17**, 2958–64 (2011).

[65] Cascalho, M. Advantages and disadvantages of cytidine deamination. *Journal of immunology (Baltimore, Md. : 1950)* **172**, 6513–8 (2004).

[66] Chelico, L., Pham, P. & Goodman, M. F. Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **364**, 583–593 (2009).

[67] Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229 (2008).

[68] Petit, V., Vartanian, J.-P. & Wain-Hobson, S. Powerful mutators lurking in the genome. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**, 705–15 (2009).

[69] Zaheen, A. & Martin, A. Activation-induced cytidine deaminase and aberrant germinal center selection in the development of humoral autoimmunities. *The American journal of pathology* **178**, 462–71 (2011).

[70] Bransteitter, R., Prochnow, C. & Chen, X. S. The current structural and functional understanding of APOBEC deaminases. *Cellular and molecular life sciences : CMLS* **66**, 3137–47 (2009).

[71] Gandhi, S. K., Siliciano, J. D., Bailey, J. R., Siliciano, R. F. & Blankson, J. N. Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. *Journal of Virology* **82**, 3125–3130 (2008).

[72] Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Medicine* **9**, 1398–1403 (2003).

[73] Mariani, R. *et al.* Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21–31 (2003).

[74] Mehle, A. *et al.* Identification of an APOBEC3G Binding Site in HIV-1 Vif and Inhibitors of Vif-APOBEC3G Binding. *Journal of Virology* **in press**, 13235–13241 (2007).

[75] Stopak, K. & Greene, W. C. Protecting APOBEC3G: a potential new target for HIV drug discovery. *Current Opinion Investigational Drugs* **6**, 141–147 (2005).

[76] Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056–1060 (2003).

[77] Betts, L., Xiang, S., Short, S. A., Wolfenden, R. & Carter Jr., C. W. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. *Journal of Molecular Biology* **235**, 635–656 (1994).

[78] Davidson, N. O., Anant, S. & MacGinnitie, A. J. Apolipoprotein B messenger RNA editing: insights into the molecular regulation of post-transcriptional cytidine deamination. *Current opinion in lipidology* **6**, 70–4 (1995).

[79] Hébrard, C., Dumontet, C. & Jordheim, L. P. Development of gene therapy in association with clinically used cytotoxic deoxynucleoside analogues. *Cancer gene therapy* **16**, 541–50 (2009).

[80] Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N. a. & Honjo, T. Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Advances in immunology* **94**, 1–36 (2007).

[81] Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–7 (2010).

[82] Morgan, H. D., Dean, W., Coker, H. a., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of biological chemistry* **279**, 52353–60 (2004).

[83] Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science (New York, N.Y.)* **260**, 1816–9 (1993).

[84] Powell, L. M. *et al.* A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831–40 (1987).

[85] Renard, M., Henry, M., Guétard, D., Vartanian, J.-P. & Wain-Hobson, S. APOBEC1 and APOBEC3 cytidine deaminases as restriction factors for hepadnaviral genomes in non-humans in vivo. *Journal of molecular biology* **400**, 323–34 (2010).

[86] Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–12 (2008).

[87] Sato, Y. *et al.* Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *The Journal of biological chemistry* **285**, 7111–8 (2010).

[88] Severi, F., Chicca, A. & Conticello, S. G. Analysis of Reptilian APOBEC1 Suggests that

RNA Editing May Not Be Its Ancestral Function. *Molecular Biology and Evolution* **28**, 1125–1129 (2010).

[89] Lada, A. G., Iyer, L. M., Rogozin, I. B., Aravind, L. & Pavlov, I. I. [Vertebrate immunity: mutator proteins and their evolution]. *Genetika* **43**, 1311–27 (2007).

[90] Lada, A. G. *et al.* Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry. Biokhimiĭa* **76**, 131–46 (2011).

[91] Prochnow, C., Bransteitter, R. & Chen, X. S. APOBEC deaminases-mutases with defensive roles for immunity. *Science in China. Series C, Life sciences / Chinese Academy of Sciences* **52**, 893–902 (2009).

[92] Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient Adaptive Evolution of the Primate Antiviral DNA-Editing Enzyme APOBEC3G. *PLoS Biology* **2**, E275 (2004).

[93] Zhang, J. & Webb, D. M. Rapid evolution of primate antiviral enzyme APOBEC3G. *Human Molecular Genetics* **13**, 1785–1791 (2004).

[94] LaRue, R. S. *et al.* The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* **9**, 104 (2008).

[95] LaRue, R. S. *et al.* Guidelines for naming nonprimate APOBEC3 genes and proteins. *Journal of Virology* **83**, 494–497 (2009).

[96] Hache, G., Liddament, M. T. & Harris, R. S. The Retroviral Hypermutation Specificity of APOBEC3F and APOBEC3G is Governed by the C-terminal DNA Cytosine Deaminase Domain. *Journal of Biological Chemistry* **280**, 10920–10924 (2005).

[97] Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486–493 (2007).

[98] Conticello, S. G., Harris, R. S. & Neuberger, M. S. The Vif Protein of HIV Triggers Degradation of the Human Antiretroviral DNA Deaminase APOBEC3G. *Current Biology* **13**, 2009–2013 (2003).

[99] Dang, Y., Siew, L. M. & Zheng, Y. H. APOBEC3G is degraded by the proteosomal pathway in a Vif-dependent manner without being polyubiquitylated. *Journal of Biological Chemistry* **in press**, 13124–13131 (2008).

[100] Kao, S. *et al.* The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *Journal of Virology* **77**, 11398–11407 (2003).

[101] Mehle, A. *et al.* Vif Overcomes the Innate Antiviral Activity of APOBEC3G by Promoting Its Degradation in the Ubiquitin-Proteasome Pathway. *Journal of Biological Chemistry* **279**, 7792–7798 (2004).

[102] Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The Antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature Medicine* **9**, 1404–1407 (2003).

[103] Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Molecular Cell* **12**, 591–601 (2003).

[104] Liu, B., Yu, X. F. X.-F., Luo, K. & Yu, Y. Influence of primate lentiviral Vif and proteasome inhibitors on human immunodeficiency virus type 1 virion packaging of APOBEC3G. *Journal of Virology* **78**, 2072–2081 (2004).

[105] Zhou, Y. *et al.* A critical function of toll-like receptor-3 in the induction of anti-human immunodeficiency virus activities in macrophages. *Immunology* **131**, 40–9 (2010).

[106] Esnault, C. *et al.* APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430–433 (2005).

[107] Chiu, Y. L. *et al.* High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc. Natl. Acad. Sci USA* **103**, 15588–15593 (2006).

[108] Ma, J. *et al.* The cellular source for APOBEC3G's incorporation into HIV-1. *Retrovirology* **8**, 2 (2011).

[109] Burnett, A. & Spearman, P. APOBEC3G multimers are recruited to the plasma membrane for packaging into human immunodeficiency virus type 1 virus-like particles in an RNA-dependent process requiring the NC basic linker. *Journal of Virology* **81**, 5000–5013 (2007).

[110] Goila-Gaur, R., Khan, M. A., Miyagi, E. & Strebel, K. Differential sensitivity of "old" versus "new" APOBEC3G to human immunodeficiency virus type 1 vif. *Journal of Virology* **83**, 1156–1160 (2009).

[111] Huthoff, H. & Malim, M. H. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *Journal of Virology* **81**, 3807–3815 (2007).

[112] Khan, M. A. *et al.* Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *Journal of Virology* **75**, 7252–7265 (2001).

[113] Khan, M. a. *et al.* Encapsidation of APOBEC3G into HIV-1 virions involves lipid raft association and does not correlate with APOBEC3G oligomerization. *Retrovirology* **6**, 99 (2009).

[114] Bishop, K. N., Holmes, R. K. & Malim, M. H. Antiviral Potency of APOBEC Proteins Does Not Correlate with Cytidine Deamination. *Journal of Virology* **80**, 8450–8458 (2006).

[115] Browne, E. P., Allers, C. & Landau, N. R. Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology* **387**, 313–21 (2009).

[116] Chiu, Y. L. & Greene, W. C. Multifaceted antiviral actions of APOBEC3 cytidine deaminases. *Trends Immunol* **27**, 291–297 (2006).

[117] Harris, R. S. *et al.* DNA Deamination Mediates Innate Immunity to Retroviral Infection. *Cell* **113**, 803–809 (2003).

[118] Iwatani, Y. *et al.* Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* **in press** (2007).

[119] Li, X. Y., Guo, F., Zhang, L., Kleiman, L. & Cen, S. APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription. *Journal of Biological Chemistry* **in press** (2007).

[120] Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **364**, 675–687 (2009).

[121] Miyagi, E. *et al.* Enzymatically Active APOBEC3G is Required for Efficient Inhibition of HIV-1. *Journal of Virology* **in press**, 13346–13353 (2007).

[122] Newman, E. N. *et al.* Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Current Biology* **15**, 166–170 (2005).

[123] Opi, S. *et al.* Monomeric APOBEC3G Is Catalytically Active and Has Antiviral Activity. *Journal of Virology* **80**, 4673–4682 (2006).

[124] Russell, R. A., Moore, M. D., Hu, W. S. & Pathak, V. K. APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA. *Retrovirology* **6**, 16 (2009).

[125] Shindo, K. *et al.* The Enzymatic activity of CEM15/APOBEC3G in the regulation of the infectivity of HIV-1 Virion, But Not a Sole Determinant of Its Antiviral Activity. *Journal of Biological Chemistry* **278**, 44412–44416 (2003).

[126] Nowarski, R., Britan-Rosich, E., Shiloach, T. & Kotler, M. Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat Struct Mol Biol* **15**, 1059–1066 (2008).

[127] Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. APOBEC3G DNA deaminase acts processively 3' –> 5' on single-stranded DNA. *Nat Struct Mol Biol* **13**, 392–399 (2006).

[128] Chelico, L., Pham, P. & Goodman, M. F. Mechanisms of APOBEC3G-catalyzed processive deamination of deoxycytidine on single-stranded DNA. *Nat Struct Mol Biol* **16**, 454–456 (2009).

[129] Chelico, L., Prochnow, C., Erie, D. a., Chen, X. S. & Goodman, M. F. Structural

model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *The Journal of biological chemistry* **285**, 16195–205 (2010).

[130] Coker, H. A. & Peterson-Mahrt, S. K. The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. *DNA Repair (Amst)* **6**, 235–243 (2006).

[131] Smith, J. L., Bu, W., Burdick, R. C. & Pathak, V. K. Multiple ways of targeting APOBEC3-virion infectivity factor interactions for anti-HIV-1 drug development. *Trends in pharmacological sciences* **30**, 638–46 (2009).

[132] Auclair, J. R. *et al.* Mass spectrometry analysis of HIV-1 Vif reveals an increase in ordered structure upon oligomerization in regions necessary for viral infectivity. *Proteins* **69**, 270–284 (2007).

[133] Reingewertz, T. H., Shalev, D. E. & Friedler, A. Structural disorder in the HIV-1 Vif protein and interaction-dependent gain of structure. *Protein and peptide letters* **17**, 988–98 (2010).

[134] Uversky, V. N. Intrinsically disordered proteins from A to Z. *The international journal of biochemistry & cell biology* **43**, 1090–103 (2011).

[135] Marcsisin, S. R. *et al.* On the Solution Conformation and Dynamics of the HIV-1 Viral Infectivity Factor. *Journal of molecular biology* **410**, 1008–22 (2011).

[136] Reingewertz, T. H., Benyamini, H., Lebendiker, M., Shalev, D. E. & Friedler, A. The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. *Protein engineering design selection PEDS* **22**, 281–287 (2009).

[137] Stanley, B. J. *et al.* Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *Journal of Virology* **82**, 8656–8663 (2008).

[138] Sarikas, A., Hartmann, T. & Pan, Z.-Q. The cullin protein family. *Genome biology* **12**, 220 (2011).

[139] Yu, Y., Xiao, Z., Ehrlich, E. S. & Yu, X. F. X.-F. Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. *Genes & Development* **18**, 2867–2872 (2004).

[140] Britan-Rosich, E., Nowarski, R. & Kotler, M. Multifaceted Counter-APOBEC3G Mechanisms Employed by HIV-1 Vif. *Journal of molecular biology* **410**, 1065–76 (2011).

[141] Santa-Marta, M., da Silva, F. A., Fonseca, A. M. & Goncalves, J. HIV-1 Vif can directly inhibit apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G-mediated cytidine deamination by using a single amino acid interaction and without protein degradation. *The Journal of biological chemistry* **280**, 8765–75 (2005).

[142] Chiu, Y. L. & Greene, W. C. APOBEC3G: an intracellular centurion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **364**, 689–703 (2009).

[143] Chen, K.-M. M. *et al.* Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119 (2008).

[144] Furukawa, A. *et al.* Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO Journal* **28**, 440–451 (2009).

[145] Harjes, E. *et al.* An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *Journal of Molecular Biology* **389**, 819–832 (2009).

[146] Holden, L. G. *et al.* Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121–124 (2008).

[147] Shandilya, S. M. D. *et al.* Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **18**, 28–38 (2010).

[148] Feng, Y. & Chelico, L. Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the non-catalytic domain. *The Journal of biological chemistry* **286**, 11415–11426 (2011).

[149] Prilusky, J. *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).

[150] Chen, K. M. *et al.* Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. *FEBS Letters* **581**, 4761–4766 (2007).

[151] Goila-Gaur, R. & Strebel, K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* **5**, 51 (2008).

[152] Chiu, Y. L. & Greene, W. C. APOBEC3 cytidine deaminases: distinct antiviral actions along the retroviral life cycle. *Journal of Biological Chemistry* **281**, 8309–8312 (2006).

[153] Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annual Review of Immunology* **in press**, 317–353 (2007).

[154] Zhang, H. *et al.* The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94–98 (2003).

[155] Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. Hypermutation of HIV-1 DNA in the Absence of the Vif Protein. *Science* **300**, 1112 (2003).

[156] Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).

[157] Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**, 447–451 (2007).

[158] Joosten, R. P., Womack, T., Vriend, G. & Bricogne, G. Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallographica. Section D: Biological Crystallography* **65**, 176–185 (2009).

[159] Joosten, R. P. *et al.* PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *Journal of Applied Crystallography* **42**, 376–384 (2009).

[160] Sanderson, K. New protein structures replace the old. *Nature* **459**, 1038–1039 (2009).

[161] Schumacher, A. J., Hache, G., Macduff, D. A., Brown, W. L. & Harris, R. S. The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *Journal of Virology* **82**, 2652–2660 (2008).

[162] Hache, G., Shindo, K., Albin, J. S. & Harris, R. S. Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Current Biology* **18**, 819–824 (2008).

[163] Carlow, D. C., Carter, C. W. J., Mejlhede, N., Neuhard, J. & Wolfenden, R. Cytidine deaminases from B. subtilis and E. coli: compensating effects of changing zinc coordination and quaternary structure. *Biochemistry* **38**, 12258–65 (1999).

[164] Johansson, E., Mejlhede, N., Neuhard, J. & Larsen, S. Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 A resolution. *Biochemistry* **41**, 2563–2570 (2002).

[165] Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153–159 (2006).

[166] Navaratnam, N. & Sarwar, R. An overview of cytidine deaminases. *International Journal of Hematology* **83**, 195–200 (2006).

[167] Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *Journal of Biological Chemistry* **282**, 16054–16061 (2007).

[168] Xie, K. *et al.* The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc. Natl. Acad. Sci USA* **101**, 8114–8119 (2004).

[169] Bennett, R. P., Presnyak, V., Wedekind, J. E. & Smith, H. C. Nuclear Exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *Journal of Biological Chemistry* **283**, 7320–7327 (2008).

[170] Wedekind, J. E. *et al.* Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *Journal of Biological Chemistry* **281**, 38122–38126 (2006).

[171] Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F. & Malim, M. H. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog* **5**, e1000330 (2009).

[172] Friew, Y. N., Boyko, V., Hu, W. S. & Pathak, V. K. Intracellular interactions between

APOBEC3G, RNA, and HIV-1 Gag: APOBEC3G multimerization is dependent on its association with RNA. *Retrovirology* **6**, 56 (2009).

[173] McDougall, W. M. & Smith, H. C. Direct evidence that RNA inhibits APOBEC3G ss-DNA cytidine deaminase activity. *Biochemical and biophysical research communications* (2011).

[174] Smith, H. C. OyaGen Inc. (2011).

[175] Smith, H. C., Prohaska, K. & McDougall, W. M. Identification of small molecule activators of APOBEC3G as candidate compounds for therapeutic development and HIV/AIDS prevention. In *Retroviruses - CSHL Meeting*, chap. Abstracts, 247 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2011).

[176] Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307–326 (1997).

[177] McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica. Section D: Biological Crystallography* **63**, 32–41 (2007).

[178] Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171–1179 (2008).

[179] Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica. Section D: Biological Crystallography* **D53**, 240–255 (1997).

[180] Collaborative-Computational-Project, N. . The CCP4 suite: programs for protein crystallography. *Acta Crystallographica. Section D: Biological Crystallography* **50**, 760–763 (1994).

[181] Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica. Section D: Biological Crystallography* **62**, 439–450 (2006).

[182] Emsley, P. & Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D - Biological Crystallography* **60**, 2126–2132 (2004).

[183] DeLano, W. L. . *The PyMOL User's Manual* (DeLano Scientific LLC, San Carlos, CA, USA., 2002).

[184] Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774–797 (2007).

[185] Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica. Section D: Biological Crystallography* **D54**, 905—-921. (1998).

[186] Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci USA* **102**, 10557–10562 (2005).

[187] Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**, D61–D65 (2007).

[188] Stenglein, M. D., Matsuo, H. & Harris, R. S. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *Journal of Virology* **82**, 9591–9599 (2008).

[189] Thielen, B. K. *et al.* T Cells Contain an RNase-Insensitive Inhibitor of APOBEC3G Deaminase Activity. *PLOS Pathog* **21**, e135 (2007).

[190] Liddament, M. T., Brown, W. L., Schumacher, A. J. & Harris, R. S. APOBEC3F Properties and Hypermutation Preferences Indicate Activity against HIV-1 in Vivo. *Current Biology* **14**, 1385–1391 (2004).

[191] Simon, J. H. *et al.* The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. *Journal of Virology* **71**, 5259–5267 (1997).

[192] Albin, J. S. & Harris, R. S. Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert reviews in molecular medicine* **12**, e4 (2010).

[193] Bogerd, H. P., Tallmadge, R. L., Oaks, J. L., Carpenter, S. & Cullen, B. R. Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. *Journal of virology* **82**, 11889–901 (2008).

[194] Larue, R. S., Lengyel, J., Jónsson, S. R., Andrésdóttir, V. & Harris, R. S. Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *Journal of virology* **84**, 8193–201 (2010).

[195] Münk, C. *et al.* Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. *Genome biology* **9**, R48 (2008).

[196] Schmitt, K. *et al.* Comparison of the replication and persistence of simian-human immunodeficiency viruses expressing Vif proteins with mutation of the SLQYLA or HCCH domains in macaques. *Virology* **404**, 187–203 (2010).

[197] Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature structural & molecular biology* **17**, 222–9 (2010).

[198] Shlyakhtenko, L. S. *et al.* Atomic Force Microscopy Studies Provide Direct Evidence for Dimerization of the HIV Restriction Factor APOBEC3G. *Journal of Biological Chemistry* **286**, 3387 (2011).

[199] Caffrey, D. R., Lunney, E. a. & Moshinsky, D. J. Prediction of specificity-determining residues for small-molecule kinase inhibitors. *BMC bioinformatics* **9**, 491 (2008).

[200] Fedorov, O. *et al.* A systematic interaction map of validated kinase inhibitors with

Ser/Thr kinases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20523–8 (2007).

[201] Fischer, P. M. The design of drug candidate molecules as selective inhibitors of therapeutically relevant protein kinases. *Current medicinal chemistry* **11**, 1563–83 (2004).

[202] Kim, B.-H. *et al.* NSC114792, a novel small molecule identified through structure-based computational database screening, selectively inhibits JAK3. *Molecular cancer* **9**, 36 (2010).

[203] Benchokroun, Y., Couprie, J. & Larsen, A. K. Aurintricarboxylic acid, a putative inhibitor of apoptosis, is a potent inhibitor of DNA topoisomerase II in vitro and in Chinese hamster fibrosarcoma cells. *Biochemical pharmacology* **49**, 305–13 (1995).

[204] González, R. G., Haxo, R. S. & Schleich, T. Mechanism of action of polymeric aurintricarboxylic acid, a potent inhibitor of protein–nucleic acid interactions. *Biochemistry* **19**, 4299–303 (1980).

[205] Hashem, A. M. *et al.* Aurintricarboxylic acid is a potent inhibitor of influenza A and B virus neuraminidases. *PloS one* **4**, e8350 (2009).

[206] Bak, A., Magdziarz, T., Kurczyk, A. & Polanski, J. Mapping drug architecture by MoStBioDat: rapid screening of intramolecular hydrogen bonded motifs in catechols. *Drug Development Research* **72**, 209–218 (2011).

[207] Iwatani, Y., Takeuchi, H., Strebel, K. & Levin, J. G. Biochemical Activities of Highly Purified, Catalytically Active Human APOBEC3G: Correlation with Antiviral Effect. *Journal of Virology* **80**, 5992–6002 (2006).

[208] Demorest, Z. L., Li, M. & Harris, R. S. Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of aid and APOBEC3G. *The Journal of biological chemistry* **286**, 26568 –26575 (2011).

[209] McDougall, W. M., Okany, C. & Smith, H. C. Deaminase activity on ssDNA occurred in vitro when APOBEC3G forms homotetramers and higher-order complexes. *The Journal of biological chemistry* (2011).

[210] Imoto, M. *et al.* Dephostatin, a novel protein tyrosine phosphatase inhibitor produced by Streptomyces. I. Taxonomy, isolation, and characterization. *The Journal of antibiotics* **46**, 1342–6 (1993).

[211] Kakeya, H. *et al.* Dephostatin, a novel protein tyrosine phosphatase inhibitor produced by Streptomyces. II. Structure determination. *The Journal of antibiotics* **46**, 1716–9 (1993).

[212] Uesugi, Y. *et al.* Inhibition of ATRA-induced myeloid differentiation in acute promyelocytic leukemia by a new protein tyrosine phosphatase inhibitor, 3,4-dephostatin. *Journal of experimental & clinical cancer research : CR* **19**, 363–6 (2000).

[213] Watanabe, T. *et al.* Structure–Activity Relationship and Rational Design of 3,4-Dephostatin Derivatives as Protein Tyrosine Phosphatase Inhibitors. *Tetrahedron* **56**, 741–752 (2000).

[214] Watanabe, T., Takeuchi, T., Otsuka, M., Tanaka, S. & Umezawa, K. Synthesis and protein tyrosine phosphatase inhibitory activity of dephostatin analogs. *The Journal of antibiotics* **48**, 1460–6 (1995).

[215] Umezawa, K., Kawakami, M. & Watanabe, T. Molecular design and biological activities of protein-tyrosine phosphatase inhibitors. *Pharmacology & Therapeutics* **99**, 15–24 (2003).

[216] Otwinowski, Z. Oscillation data reduction program. In Sawyer, L., Isaacs, N. & Bailey, S. (eds.) *Proceedings of the CCP4 study weekend - Data collection and processing*, 56–62 (SERC Daresbury Laboratory, England, 1993).

[217] Kabsch, W. Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125–32 (2010).

[218] Winter, G. Xia2 : an Expert System for Macromolecular Crystallography Data Reduction. *Journal of Applied Crystallography* **43**, 186–190 (2009).

[219] Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods in Enzymology* **374**, 300–321 (2003).

[220] Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *Journal of molecular biology* **320**, 597–608 (2002).

[221] Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–67 (2004).

[222] PRIME. PRIME (2011).

[223] Maestro. Maestro (2011).

[224] Rosenfeld, J., Capdevielle, J., Guillemot, J. C. & Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Analytical biochemistry* **203**, 173–9 (1992).

[225] Hellman, U., Wernstedt, C., Góñez, J. & Heldin, C. H. Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical biochemistry* **224**, 451–5 (1995).

[226] Clauser, K. R., Baker, P. & Burlingame, A. L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry* **71**, 2871–2882 (1999).

[227] Lovsin, N. & Peterlin, B. M. APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. *Annals of the New York Academy of Sciences* **1178**, 268–75 (2009).

[228] Muckenfuss, H. *et al.* APOBEC3 proteins inhibit human LINE-1 retrotransposition. *Journal of Biological Chemistry* **281**, 22161–22172 (2006).

[229] Farrow, M. A. & Sheehy, A. M. Vif and Apobec3G in the innate immune response to HIV: a tale of two proteins. *Future Microbiol* **3**, 145–154 (2008).

[230] OhAinle, M., Kerns, J. A., Malik, H. S. & Emerman, M. Adaptive Evolution and Antiviral Activity of the Conserved Mammalian Cytidine Deaminase APOBEC3H. *Journal of Virology* **80**, 3853–3862 (2006).

[231] Acharya, K. R., Stuart, D. I., Phillips, D. C. & Scheraga, H. A. A critical evaluation of the predicted and X-ray structures of alpha-lactalbumin. *Journal of protein chemistry* **9**, 549–63 (1990).

[232] di Luccio, E. & Koehl, P. A quality metric for homology modeling: the H-factor. *BMC bioinformatics* **12**, 48 (2011).

[233] Russell, R. A., Smith, J., Barr, R., Bhattacharyya, D. & Pathak, V. K. Distinct domains within APOBEC3G and APOBEC3F interact with separate regions of human immunodeficiency virus type 1 Vif. *Journal of Virology* **83**, 1992–2003 (2009).

[234] Bulliard, Y. *et al.* Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *Journal of virology* **85**, 1765–76 (2011).

[235] Albin, J. S. *et al.* A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. *The Journal of biological chemistry* **285**, 40785–92 (2010).

[236] Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics (Oxford, England)* **22**, 2695–6 (2006).

[237] Gerstein, M. & Altman, R. B. Average core structures and variability measures for protein families: application to the immunoglobulins. *Journal of molecular biology* **251**, 161–75 (1995).

[238] Chen, G., He, Z., Wang, T., Xu, R. & Yu, X.-F. A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. *Journal of virology* **83**, 8674–82 (2009).

[239] Dang, Y. *et al.* Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *The Journal of biological chemistry* **283**, 11606–14 (2008).

[240] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

[241] Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

[242] Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology* **55**, 539–52 (2006).

[243] Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17**, 540–52 (2000).

[244] Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* **36**, W465–9 (2008).

[245] Dereeper, A., Audic, S., Claverie, J.-M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC evolutionary biology* **10**, 8 (2010).

[246] Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696–704 (2003).

[247] MacroModel. MacroModel (2011).

[248] Nalam, M. N. & Schiffer, C. A. New approaches to HIV protease inhibitor drug design II: testing the substrate envelope hypothesis to avoid drug resistance and discover robust inhibitors. *Curr Opin HIV AIDS* **3**, 642–646 (2008).

[249] Schiffer, C. A. *Combating drug resistance – identifying resilient molecular targets and robust drugs*, chap. 7, 127–132 (Royal Society of Chemistry, 2007).

[250] Cadima-Couto, I., Saraiva, N., Santos, A. C. C. & Goncalves, J. HIV-1 Vif Interaction with APOBEC3 Deaminases and its Characterization by a New Sensitive Assay. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology* **6**, 296–307 (2011).

[251] Zhan, P., Li, W., Chen, H. & Liu, X. Targeting protein-protein interactions: a promising avenue of anti-HIV drug discovery. *Current medicinal chemistry* **17**, 3393–409 (2010).

[252] Shindo, K. *et al.* Cellular CBF is a HIV Vif cofactor required for APOBEC3G degradation. In *Retroviruses - CSHL Meeting*, chap. Abstracts, 103 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2011).

[253] Bourara, K., Liegler, T. J. & Grant, R. M. Target Cell APOBEC3C Can Induce Limited G-to-A Mutation in HIV-1. *PLoS Pathog* **3**, e153 (2007).

[254] Thielen, B. K. *et al.* Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *The Journal of biological chemistry* **285**, 27753–66 (2010).

[255] Refsland, E. W. *et al.* Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic acids research* **38**, 4274–84 (2010).

[256] Langlois, M.-A., Beale, R. C. L., Conticello, S. G. & Neuberger, M. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G antiretroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Research* **33**, 1913–1923 (2005).

[257] Stauch, B. *et al.* Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12079–84 (2009).

[258] Guo, J. U., Su, Y., Zhong, C., Ming, G.-L. & Song, H. Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* 423–434 (2011).

[259] Wang, Y. J. *et al.* Expression and regulation of antiviral protein APOBEC3G in human neuronal cells. *Journal of Neuroimmunology* **206**, 14–21 (2009).

[260] Murray, C. W. & Blundell, T. L. Structural biology in fragment-based drug design. *Current opinion in structural biology* **20**, 497–507 (2010).

[261] Powers, R. Structure-Based Drug Discovery Edited by Harren Jhoti (Astex Therapeutics, Cambridge, U.K.) and Andrew R. Leach (GlaxoSmithKline, Stevenage, U.K.). Springer: Dordrecht. 2007. xii + 250 pp. 169.00. ISBN 978-1-4020-4406-9. *Journal of the American Chemical Society* **129**, 10042–10043 (2007).

[262] Zhang, S. Computer-aided drug discovery and development. *Methods in Molecular Biology* **716**, 23–38 (2011).

[263] Prabu-Jeyabalan, M. *et al.* Substrate envelope and drug resistance: crystal structure of RO1 in complex with wild-type human immunodeficiency virus type 1 protease. *Antimicrobial Agents and Chemotherapy* **50**, 1518–1521 (2006).

[264] Kim, E.-Y. *et al.* Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *Journal of virology* **84**, 10402–5 (2010).

[265] Nathans, R. *et al.* Small-molecule inhibition of HIV-1 Vif. *Nature biotechnology* **26**, 1187–92 (2008).

[266] Rayment, I. Small-Scale Batch Crystallization of Proteins RevisitedAn Underutilized Way to Grow Large Protein Crystals. *Structure* **10**, 147–151 (2002).

[267] Ramachandran, P. & Varoquaux, G. Mayavi2 User Guide (2008).

[268] Blaustein, R. O., Cole, P. A., Williams, C. & Miller, C. Tethered blockers as molecular 'tape measures' for a voltage-gated K+ channel. *Nature structural biology* **7**, 309–11 (2000).

[269] Ahern, C. a. & Kobertz, W. R. Chemical tools for K(+) channel biology. *Biochemistry* **48**, 517–26 (2009).

[270] Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).

[271] Vucetic, S. *et al.* DisProt: A Database of Protein Disorder. *Bioinformatics* **21**, 137–140 (2005).