Structural and Mechanistic Insight into APOBEC3G DNA Binding and Deamination

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

William Casey Solomon

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Hiroshi Matsuo

July 2015

© William Casey Solomon 2015

Acknowledgements

I would like to first and foremost thank my advisor, Hiroshi Matsuo, for his continued support and guidance throughout my graduate career. Without which, I know that I would not have been able to successfully complete this arduous four year journey. I would also like to thank Kylie Walters. You gave me the chance to join your lab as an undergraduate and you never treated me like one. When it came time for me to apply to graduate school and the admissions committee didn't want to give me an interview based on my lack of academic success, you stood up for me. I am eternally grateful for the opportunity that you gave me.

I would like to sincerely thank everyone who has been with me in both the darkness and the light. There have been many times during the last four years that I have questioned my abilities and my dedication to this pursuit, but at the end of the day I couldn't allow myself not to see this through to completion. I will attempt to name those who have been the most instrumental to my success here. It will by no means be an exhaustive list as I have been lucky to be constantly surrounded by individuals who saw more in me than I saw in myself.

Firstly, I'd like to thank Leah Randles who helped me begin this journey and was the most patient and supportive teacher that I've had the privilege to work with. I came to you at one of the most difficult points in your early graduate career, yet you made the time to teach me how to be a scientist. You have been a sounding board for my frustration and a shining example of perseverance through hardship. I will forever be thankful for that and your continuing friendship.

I would like to thank my best friends, Mike Hall, Nick Paulson, and Josh Lund. I can't begin to count the innumerable laughs and experiences that I've had with you all. Without those moments of levity to remind me that there is in fact a life outside of the lab, I know that I wouldn't have made it. You three are my brothers in the truest sense of the word. Thank you so much for accepting me at my worst and motivating me to be my best, I love you all.

To my parents, there is no way that I would be here without your sacrifices, your enduring love and support. You have taught me the importance of giving everything that I have and have demonstrated incredible resolution in the face of difficulty. Regardless of my goals, you have always been there to pick me up when I fail and cheer me on to success. I am proud to be your son and I love you.

To my partner, Natalie Wasiluk, you have seen me through the most difficult time in this whole process and you have done so with great compassion, patience, and love. I am forever in your debt for putting up with my weaknesses. Your kindness and caring are an inspiration and motivate me on a daily basis to be a better person. I would also like to thank Bailey Rejsa for giving me perspective. Regardless of what I achieve professionally, I know that the most important thing is to do right by you and be the best example of a human being that I possibly can be. I love you both.

Thank you to everyone that have been there for me over the past 29 years. I will never forget what you have done for me and I will do my best to be deserving of your support and love.

Dedication

I would like to dedicate this thesis to my family, both biological and chosen.

Thesis Abstract

Humans express the APOBEC3 family of proteins to defend against endogenous and exogenous DNA pathogens. APOBEC3 proteins display significant activity towards HIV-1 through incorporation into budding viral particle and interacting with HIV's RNA genome. In order to stably integrate into the human genome, HIV reverse transcribes the single stranded RNA genome into a double stranded DNA genome through a single stranded DNA intermediate. Once bound to the transient single stranded viral DNA intermediate, APOBEC3 proteins deaminate cytidines to uridines. Transcription over the resulting mutations results in G to A transversions in the coding sequence of the virus and non-functional gene products. APOBEC3 proteins are highly active on the single stranded DNA intermediate but lack catalytic activity on cytidines present in the viral RNA.

APOBEC3G is the most potent of these innate viral mutagens and selectively targets tri-cytidine hotspot motifs in viral genome intermediates. When I began my thesis research, the effects of A3G mutagenesis had been extensively studied but the structural interactions with single stranded DNA and the mechanism of RNA exclusion were unknown. My thesis research helped identify the amino acid responsible for pH dependent effects on substrate binding. The identification of this residue allowed for the development of a pH insensitive variant of A3G that provided indirect evidence for a reduced pH in the viral capsid during the initiation of reverse transcription. These results, along with the kinetic characterization of A3Gctd and determination of the substrate factors crucial for deamination, are described in Chapter 2.

In chapter 3, I continued to explore the effects of this pH dependent increase in binding affinity. I utilized NMR spectroscopy to identify the structural interactions of catalytic substrate binding. By generating a structurally stable but catalytically inactive mutant, I was able to identify differences between substrate interactions. This mutant also allowed me to explore structural interactions involved in substrate recognition and RNA exclusion. This lead to the identification of a novel substrate and a ribose sugar pucker dependent mechanism for target discrimination.

The role of APOBEC3 proteins in retroviral restriction as well as their connection to several types of cancer makes them a prime target for therapeutic interventions. My thesis research in trying to understand the structural and mechanistic components of the APOBEC3/DNA interaction provides information that may be useful in the development of treatments targeting this family of proteins.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

APOBEC3G and its interactions with oligonucleotides

1.1. Synopsis

During viral infections the body has two distinct layers of defense. The innate immune system is constantly present and responds immediately to viral infections while the adaptive immune system takes longer to respond to infection but is specifically targeted at the invading virus. Human APOBEC3G (A3G) is involved in the innate immune and as such is present in several types of cells at all times. A3G works by tagging along with the virus as it moves from cell to cell and analogous to a 'Trojan Horse' springs forth from the virus mutating the viral DNA which prevents the virus from being able to effectively infect new cells. Much of the mutational activity stems from A3G's ability to bind viral DNA in the proper catalytic orientation to allow A3G to mutate the DNA. The aim of this review is to compile the most current knowledge in the field as to how A3G accomplishes these goals and what physical characteristics are relevant to successful mutation.

1.2. Background

Human APOBEC3G (A3G) is a member of a family of cytidine deaminases including AID, APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3DE, APOBEC3F, APOBEC3G, APOBEC3H, and APOBEC4 which are involved in a range of cellular processes including antibody diversification, inhibition of transposable DNA elements and innate immunity. APOBEC3 proteins can be either single-domain (A3A, A3C, and A3H) or double-domain (A3B, A3DE, A3F, and A3G) proteins with one or two zinc coordinating domains respectively. A3G specifically demonstrates restrictive activity against HIV-1 in the absence of the viral accessory

protein Vif. A3G has been shown to be capable of retroviral restriction through both deaminase dependent and independent mechanisms of action [1]–[6]. However, A3G's HIV-1 restrictive capacity is dominated by the deamination dependent mechanism [5]. A3G exhibits a context specific deamination target first seen in characteristic GG->GA hypermutation patterns on the HIV-1 plus strand in infected patients. Further insight into deamination preferences in cell lines infected with Vif-deficient HIV-1 indicated a 5' CCC 3' target site with the underlined C indicating the preferred deamination site [1]–[3]. Extensive deamination of the viral genome introduces premature stop codons as well as generating missense mutations genome-wide while sub-lethal levels of mutation are thought to contribute to viral escape of host defenses. Recent studies have focused on the physical interactions of A3G with ssDNA in an attempt to understand the mechanism of A3G binding and target specificity.

1.3. APOBEC3 family

The human APOBEC3G protein belongs to a family of seven polynucleotide cytidine deaminases (A3A-A3H) [7], [8]. The APOBEC3 family share common catalytic mechanism, context dependent deamination preferences, and structure. APOBEC3 proteins deaminate cytidines in single stranded DNA polynucleotides to non-native uridine residues. The APOBEC family of proteins is present in all placental mammal and functions to restrict both endogenous transposable DNA elements and exogenous DNA viruses. The mammalian APOBEC3 locus in particular displays significant copy number variation and hallmarks of positive selection as a result of long term competition with viral pathogens [9]–[15]. The human APOBEC3 locus is organized linearly on

chromosome 22 [8]. The characteristic feature of APOBEC3 proteins are their zinc-coordinating domains (Z-domains), containing the conserved H-X-E-X_{25-31}-P-C-X_{2-4}-C zinc-binding motif (**Fig 1.1**). The APOBEC3 family is comprised of three evolutionarily distinct Z-domains (Z1, Z2, Z3), varying in amino acid composition within the zinc-binding motifs. Z1 and Z2 domains contain SW-S/T-C-X_{2-4}-C motifs while Z3 domains contain a TW-S-C-X_2-C motif. Z1 and Z2 domains are distinguished by H-X-E-X_6-V/I and H-X-W-X_5-W-F motifs, respectively [11].

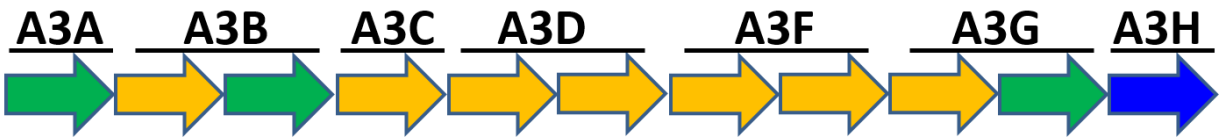

Fig. 1.1. APOBEC3 locus organization. APOBEC3 proteins are organized linearly on chromosome 22. Individual APOBEC3 protein compositions are shown. Each distinct Z-domain is indicated with a colored arrow representing evolutionary lineage, Z1 (green), Z2 (yellow), and Z3 (blue). (Adapted from LaRue et al. *Journal of Virology* 83(2) 2009, 494-497)

Cytidine deamination is highly context dependent with minimal target sites consisting of a dinucleotide 5'-XC motif, where the 5'-X is either another cytidine (A3G) or thymidine (A3A, A3B, A3C, A3DE, A3F, A3H) [16]. Cytidine deamination takes place through a conserved mechanism of nucleophilic zinc mediated hydrolysis of the amine group at the 4-position of the pyrimidine ring resulting in the replacement of the amine group with a carbonyl moiety. Outside of the minimal 5'-XC motif required for catalysis, APOBEC3 proteins require an extended polynucleotide containing a minimum of five nucleotides with three nucleotides preceding the target cytidine and one nucleotide following [17]. The composition of these flanking nucleotides result in different catalytic

rates (discussed later) but the presence of these flanking nucleotides is required for catalysis.

1.4. Structure

APOBEC3 proteins contain either one (A3A, A3C, and A3H) or two (A3B, A3DE, A3F, A3G) conserved zinc-coordinating domains. In the case of the double domain APOBEC3s the domains are likely connected by a flexible linker. The structures of several isolated APOBEC3 domains have been solved by X-ray crystallography or NMR (**Table 1.1**).

Table 1.1. Structures of APOBEC family members

Protein	Method	Domain	Reference
APOBEC2	X-ray	Single Domain	2NYT [18]
APOBEC2	NMR	Single Domain	2RPZ [19]
APOBEC3A	NMR	Single Domain	2M65 [20]
APOBEC3A	X-ray	Single Domain	4XXO [21]
APOBEC3C	X-ray	Single Domain	3VOW [22]
APOBEC3F	X-ray	C-Terminal	4IOU [23]
APOBEC3G	NMR	C-Terminal	2KBO [24]
APOBEC3G	NMR	C-Terminal	2KEM [25]
APOBEC3G	X-ray	C-Terminal	3V4K [26]
APOBEC3G	NMR	N-Terminal	2MZZ [27]

Based upon the currently available structural information, APOBEC3 proteins adopt a conserved core fold consisting of six alpha-helices and five beta-strands (**Fig. 1.2**). The five beta-strands form an extended beta-sheet that comprises the hydrophobic core of the protein with the alpha-helices arranged around the core sheet. Unique to APOBEC3G is the presence of the non-continuous β2 sheet. The α2 and α3 helices contain the conserved H-X-E-X_{25-31}-P-C-X_{2-4}-C motif required for zinc coordination

(numbering based on N to C terminal locations of secondary structure elements). Loops of varying length and composition connect the secondary structure elements and have been implicated in substrate binding and target selection. These loops comprise the majority of the predicted ssDNA binding interface. The non-catalytic N-terminal domain of APOBEC3G structure displays the same core fold of the previously solved APOBEC catalytic domains. However, the α2 helix is shortened causing a significant difference in orientation resulting in a repositioning of the pseudo-catalytic glutamate residue away from the active site and loss of catalytic activity [27].

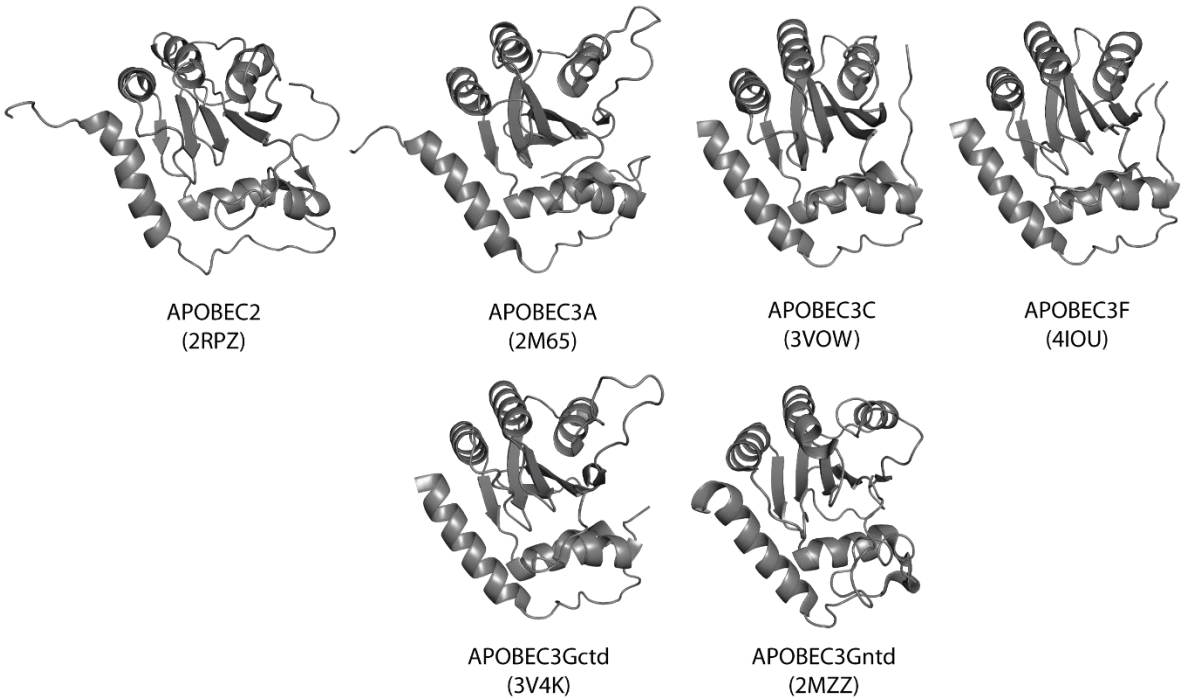

Fig. 1.2. Structural comparison of APOBEC family. Representative structures of APOBEC domains. Secondary structural elements ordered from N to C-terminus α1-β1-β2-(β2')*-α2-β3-α3-β4-α4-β5-α5-α6. *β2' only present in A3Gctd.

1.5. APOBEC3s and innate immunity to HIV-1

The most well characterized activity of APOBEC3 proteins is their ability to restrict HIV-1 viral infection in the absence of the viral Vif (viral infectivity factor) protein ([1]–[6]). During Vif-deficient HIV-1 infection, restrictive APOBEC3 proteins (A3DE, A3F, A3G, and A3H) are incorporated into budding viral particles (**Fig. 1.3**). Following viral maturation, virion binding to target cells, fusion with the target cell membrane, and injection of the viral capsid into the cytoplasm, HIV utilizes the viral reverse transcriptase to initiate the process of reverse transcription from the RNA genome which results in a negative sense ssDNA genome intermediate. The restrictive APOBEC3 proteins incorporated into the virion bind and deaminate target cytidines in this ssDNA intermediate. Subsequent transcription of the negative sense intermediate incorporates these mutations into the coding sequence of virus resulting in missense mutations in viral proteins as well as premature stop codons. This resulting hypermutation phenotype can render the virus incapable of further rounds of replication thus restricting virus effectively dead. Only APOBEC3 proteins from the initial producer cell, and not the APOBEC3 repertoire of the target cell, have access to the viral ssDNA.

In a Vif-proficient virus, the presence of the Vif protein abrogates APOBEC3 function. Vif recruits the endogenous proteins CBFβ, Elongin B, Elongin C, RBX1, and Cul5 into an E3 ligase complex that binds restrictive APOBEC3 proteins resulting in their polyubiquitination and subsequent degradation by the proteasome. This effectively limits the pool of resident restrictive APOBEC3 proteins and prevents their incorporation into the budding viral particles and allows for efficient viral replication.

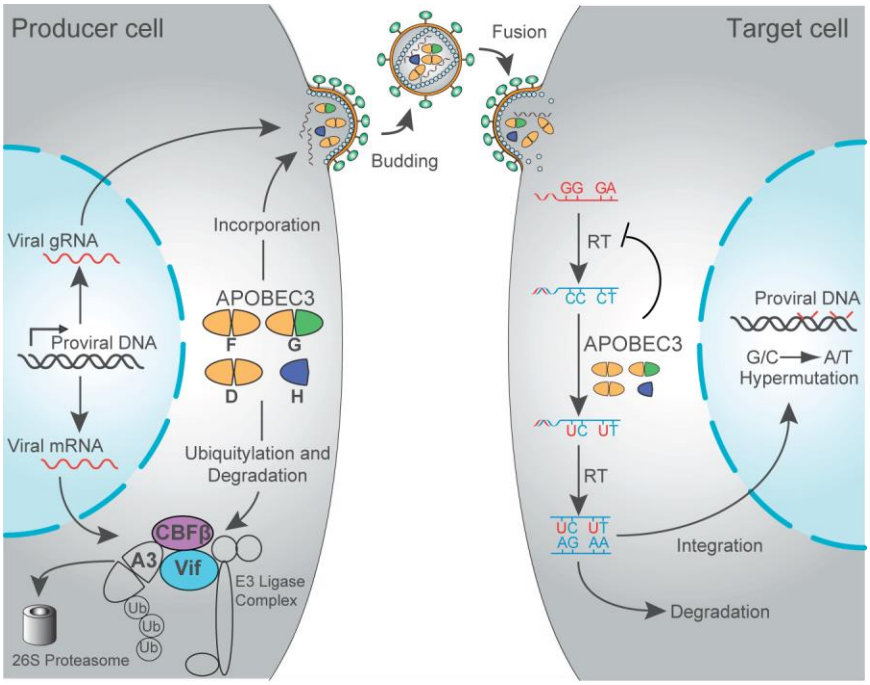

Fig. 1.3. Model for restrictive APOBEC3 function during the HIV-1 viral life cycle. (Adapted from Harris and Dudley, *Virology* 479-480 2015, 131–145)

1.6. APOBEC3G ssDNA binding interface

The catalytic activity of APOBEC3G comes exclusively from the C-terminal (CD-2) domain with the N-terminal (CD-1) domain assisting in DNA binding. The specificity of this binding is directly related to the ability of A3G to restrict viral replication through mutation. A3G exhibits a strong directionality in deamination preference stemming from a preferred orientation of the 5' CCC 3' target motif (underlined C indicates target deamination cytidine). Two independent orientations have been postulated for DNA binding to A3G [3], [28]. Understanding which of these two binding interfaces is accurate gives insight into A3G's mechanism and target cytosine positioning. A3G NMR and X-ray structures along with molecular modelling of the A3G-ssDNA complex display two potential DNA binding interfaces [25], [28], [29]. Unlike other DNA binding proteins, which tend to have positively charged surfaces that

make contact with the phosphate backbone of DNA, A3G binding interfaces direct the DNA into a negatively charged groove with several positively charged residues along the brim of the channel. One potential binding surface runs parallel to the parallel five beta sheet core fold falling between the alpha 1 and alpha 2 helices making contacts with the alpha0-beta1 loop (AC1 loop). Residues located in the AC1 loop and the evolutionarily conserved beta3-alpha2 loop appear to be critical for effectively positioning the substrate recognition sequence for deamination as evidenced by alanine scanning mutations at R213, R215, W285, R313, or R320 that either severely attenuate or eliminate deamination activity (**Fig. 1.4,** left) [29]. Several of these residues lie in a hydrogen bonding network which stabilize the contacts between AC1 and the β3-α2 loop positioning a required W285 residue so that it points into the binding groove allowing for potential aromatic ring stacking with DNA and RNA bases [30].

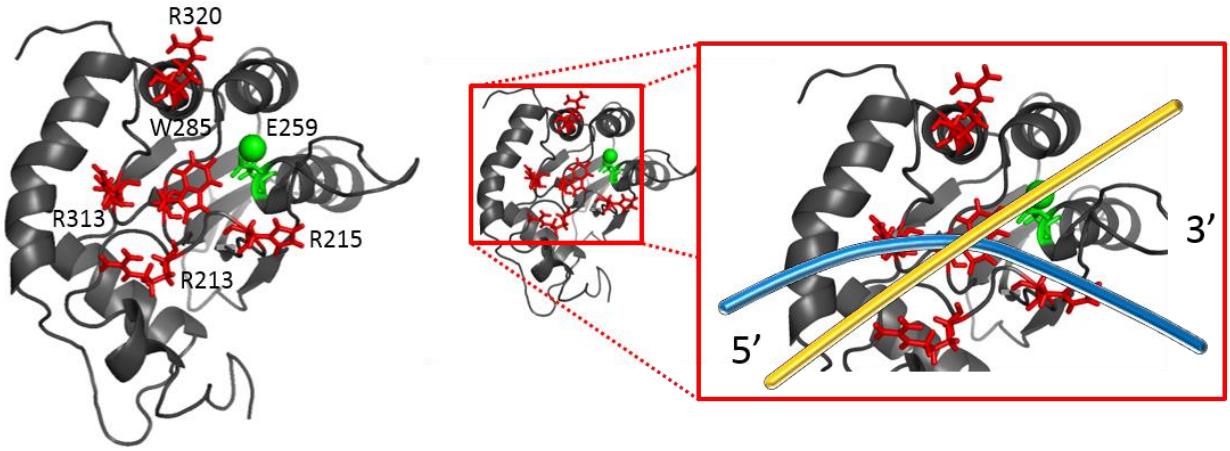

Fig. 1.4. A3Gctd-2K3A deamination mutants and potential binding orientations. Residues that result in significantly reduce mutation frequencies are indicated on the structure of A3Gctd-2K3A in red (left) [29]. Predicted binding orientation of ssDNA on A3Gctd (right), yellow strand indicates 'brim' binding orientation, blue strand indicates 'kink' binding orientation.

Computer simulation of the A3G-DNA binding interface indicates that the 5' end of ssDNA is directed away from the loop region and the 3' end extends into the groove making contact with the catalytic E259 residue [29]. This orientation requires the flipping out of the target cytidine residue in order for catalytically necessary contact to be made as has been seen previously in the TadA adenosine deaminase protein [29], [31]. Simulations also indicate extensive contact between H257 and the cytidine residue 5' of the target C. Chemical shift perturbations of this residue have been seen in both the wild-type and 2K3A mutant proteins upon addition of ssDNA. Another potential interaction interface extends from the same 5' contacts with residues R313 and R320 which then exhibits a 45 degree kink at the catalytic residue, possibly correlated with the base flip out predicted for catalysis [30], with the 3' tail again extending past the groove between alpha1 and alpha2 towards AC loops 1 & 3 (**Fig. 1.4,** right) [28]. All the current data available from NMR chemical shift analysis is unable to distinguish between these two conformations as both conformations have potential contacts at a majority of the shifted residues depending on base specific orientation in the recognition sequence. Co-crystallization or a more targeted NMR approach is required to further elucidate the A3G-DNA interface.

1.7. Target Site Recognition

Efficiently locating specific target sequences in long stretches of nucleotides is crucial for A3G to effectively restrict viral replication. Once the target sequence is located it must also be deaminated and then released so that the enzyme may acquire

another target. *In vitro* analysis using purified A3G and short oligonucleotides containing the target triple cytidine site also revealed the impact of the 5' and 3' flanking nucleotides on deaminase efficiency. Adenosine is the preferred residue in both the 5' and 3' position with a more pronounced effect when located at the 3' position as these residues extend into the core of the protein and will make considerably more contacts than the 5' end. This effect does not exhibit a general purine specificity as guanidine substitutions at these positions also create a two to three-fold reduction in the specific activity of the protein [32]. Further examination of the active site through nucleoside analogs highlighted specific aspects of the 5' and 3' nucleotides surrounding the target cytidine that are important for target recognition. Paramount among these interactions are the 3 and 4 positions in the purine ring of the cytidine at the -1 position. Substrate analogs that maintain the highest levels of activity contain hydrogen bond acceptors at ring position 3 and acceptors at position 4. These electrostatic interactions, potentially mediated through tyrosine 315, likely aid in the positioning of the target C for efficient catalysis [33]. Interestingly, a wholesale swap of the beta4-alpha4 loop of A3G (I314-Q321) with the homologous loop in A3F (L306-Q315) changes target sequence specificity from CCC (A3G) to TTC (A3F) indicating that these residues are solely responsible for mediating important 5' DNA contacts [34]. However, without high resolution structural information on the A3G-DNA complex there is no way to conclusively determine the protein residues responsible for these interactions.

1.8. Single Stranded Polynucleotide Binding

A3G exhibits stringent requirements in the locating and positioning of the DNA substrate for catalysis, the opposite is true of general single stranded polynucleotide binding. A3G is capable of binding to both ssDNA and ssRNA with a higher affinity for ssRNAs. While these ssRNAs are responsible for A3G incorporation into budding virions [35]–[39] they inhibit A3G catalytic activity [38]. This activity is restored both *in vitro* and *in vivo* through RNAse digestion of the inhibitory RNAs. Full length A3G is capable of non-specifically binding ssDNA longer than 9nt regardless of base content or the presence of the target motif. The kD across a range of 10-69nt long ssDNA is also constant at 50nM for full length A3G in contrast with deamination efficiency which shows a drastic decrease from full deamination with 69nt ssDNA to ~10% at 10nt ssDNA. This indicates that DNA binding is not the sole factor in determining deamination efficiency [32].

1.9. A3G Exhibits Locational/Contextual Differential Deamination Rates

A3G deamination rates on DNA substrates containing target motifs with varying distances from the 3' end show a strong location dependence with a 3' -> 5' deamination polarity. This coincides with the model of random ssDNA binding followed by local scanning with a preferred deamination orientation. A3G specific activity on ssDNAs with target motifs less than 30nt from the 3' end show a marked reduction in deamination efficiency compared to identical target motifs greater than 30nt away [40]. The 30nt dead zone is likely a region required for binding of the N-terminal non-catalytic domain in

order to facilitate the catalytically active orientation. In concert with the 3' -> 5' locational dependent deamination rates A3G shows a distinct 3' -> 5' pattern in the order of deamination within a target 5'-CCC-3' motif [3], [32]. The second and third cytidines in the target sequence are deaminated at markedly different rates *in vitro*. The 3' cytidine is fully deaminated in solution with 10-fold excess target DNA in less than 30 minutes while the middle cytidine is deaminated 20 times slower with the 5' C completely retained. The functional implication of this deamination order is the creation of premature stop codons when the preferred target CCCA sequence is in frame. Deamination of the 3' C in this context changes an in frame tryptophan codon (TGG) on the plus strand to a stop codon (TAG). Further deamination of the middle C maintains this in frame stop codon (TAA) [28]. Generation of these stop codons, in concert with non-sense mediated decay initiated by extensive deamination, in part explain why A3G is so effective in restricting HIV-1 infection. This combination of activity gradients along with the target sequence work together in the context of viral replication to efficiently deaminate the short lived ssDNA intermediate following RNAse digestion of viral RNA.

1.10. Oligomerization

APOBEC3G has been shown to exist in a range of oligomeric forms ranging from monomers to high molecular mass aggregates with the oligomeric composition being implicated in viral restriction, DNA binding, and catalytic efficiency [32], [40]–[42]. Until recently the predicted structural organization of A3G oligomers was based on the crystal structure of the single domain APOBEC2 (A2) protein. A2 crystallized as a

homotetramer with the asymmetric unit being a dimer of dimers in an elongated conformation atypical of other characterized human cytidine deaminases which formed square shaped tetramers [43]. However, the A2 crystal structure utilized a truncated version of A2 which lacked the N-terminal 40 residues. A recent solution NMR of the full length A2 protein calls the validity of using the A2 crystal structure as a model for the double domain deaminases of the APOBEC3 family into question. The full length solution structure containing the N-terminal extension positions these residues over the β2-β2 dimer contacts seen in the crystal structure. In addition to the physical obstruction of the predicted binding interface additional biochemical data shows A2 to be a monomer in solution [44]. This coincides with an extended structure of the A3G C-terminal domain in which the alpha1 helix is fully formed to stabilize the core fold of the protein with part of the alpha0 helix present and extending into the interface predicted from the A2 structure [25]. In combination with the discontinuous nature of the beta2 strand of A3G these data make the idea of the A2 β2-β2 "rib cage" model of the full length protein unlikely. An extended full length structure has been postulated from the NMR data and appears to be corroborated in SAXS data showing an extended envelope [25], [45]. In the absence of a high resolution structure of a full length double domain APOBEC3 protein the exact nature of the monomer and oligomers remains unclear.

Regardless of the exact physical nature of the monomeric and oligomeric forms of APOBEC3G there is still extensive debate over the physiologically active and relevant form of the enzyme. Full-length wild-type A3G purifies as a combination of 18% monomers, 80% unresolved mixture of dimers to tetramers, and 2% high molecular weight aggregates. Deaminase activity has been observed directly in this complex

mixture and in mutant constructs specifically designed to be monomeric in solution. The monomer mutant of A3G retains both the specific activity and 30nt dead zone characteristics of the wild-type indicating that dimerization is not critical to A3G activity [41]. There is some contention that higher order oligomers are required for efficient deamination and viral restriction [46]. However, even the C-terminal domain alone demonstrates considerable deaminase activity allowing it to be used in isolation to determine kinetic parameters [28]. The most widely accepted active form is a dimer as there is the largest amount of direct and indirect data supporting this conformation. Atomic force microscopy measurements on ssDNA substrates in solution show a predominantly dimeric population attached to the target sites. A3G monomers were seen to readily bind ssDNA but rapidly dimerized, dimers were also seen to release a single monomer leaving the other monomer still bound to the DNA. While this technique is unable to distinguish which form was active in catalysis it indicates that the dimer is readily and even preferentially formed on substrate DNA [42]. AFM has also shown an interesting correlation between available binding area and multimeric state. On substrates shorter than the previously described, monomers have been seen binding in roughly equal proportions to dimers whereas dimers dominate on longer ssDNA segments indicating that monomer contact is sufficient to acquire the substrate but dimer is the preferred multimeric state [47].

Dimer formation has the possibility of occurring in a number or orientations, CD1-CD1, CD2-CD2, and CD1-CD2 (**Fig. 1.5**). CD2-CD2 dimers have been postulated on the basis of SAXS data[45], but such dimers are not seen in catalytically active CD2 only mutants used for NMR structural determination [28], [29]. A3G proteins that were

mutated in the N-terminal and C-terminal domains indicate that the N-terminal CD1-CD1 dimers are the predominant form. Mutations made in the homologous regions of both the CD1 and CD2 domain contacts predicted by the A2 crystal structure only yielded a significant increase in monomer formation with the CD1 mutants. Residual dimer formation was observed in the CD1 mutant, attributable to minimal retained CD1-CD1 affinity as well as CD1-CD2 and CD2-CD2 binding, but higher order oligomer formation was completely abolished [41].

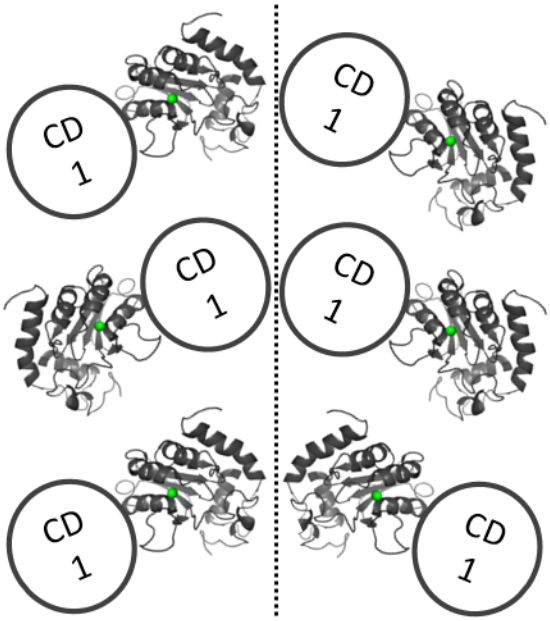

Fig. 1.5. Potential fl-A3G dimer interactions. Potential dimerization orientations for oligomerization of fl-A3G monomers into dimer, middle orientation (CD1-CD1 interface) constitutes likely dimer form. Dimerization interface indicated by dashed line.

1.11. Processivity

A3G incorporated into budding HIV-1 virions in the producer cell has been shown to be necessary for active A3G deamination and restriction of HIV-1 in target cells [36], [37], [39], [48]. The amount of A3G incorporated into these virions correlates with

the extent of HIV-1 proviral DNA deamination in a linear fashion [49]. Incorporation of a small number of A3G active units into an HIV-1 virion can exhibit extensive deamination capacity. A3G active units incorporated into the virion show characteristic deamination profiles and multiple deaminations are visible upon incorporation of a single A3G active unit [49], [50]. As the active state of A3G (monomer, dimer or higher order oligomer) is unknown the term active unit refers to the oligomeric state able to cause deamination. Since it is clear that minimal amounts of A3G incorporation can lead to high levels of proviral DNA mutations the question of the mechanism of A3G action on extended sequences becomes relevant.

A3G has demonstrated capacity to deaminate multiple target sequences in ssDNA substrates under single substrate binding conditions [32], [40], [41], [51], [52]. Sequential double deamination reactions are observed at a 6-7 fold excess over the predicted values of the rate of double deamination resulting from two independent A3G molecules. The range over which correlated single A3G molecule double deaminations are observed is from 3-100nt of intervening ssDNA [32]. This range is further confirmed by real-time AFM data showing single A3G units, in this case dimers, moving across ssDNA sequences of up to 69nt without dissociation [47]. In concert with the scanning motion exhibited by A3G productive binding results in extremely long dwell times on a single DNA target extending up to 10 minutes [51]. This extended dwell time coupled with scanning gives the opportunity for multiple coordinated deaminations.

There is much discussion concerning the mechanism of A3G processivity with the two competing mechanisms being sliding/micro jumping and intersegmental transfer. Intersegmental transfer makes the case for the simultaneous binding of either two discreet

ssDNAs or the folding of a single target ssDNA in two positions. This simultaneous binding could be facilitated through the non-catalytic CD-1 or through the release of ssDNA by one subunit from a dimer [52]. Although for this to be the mechanism of proccessive deamination in the case of a single DNA strand the interactions between A3G and ssDNA would have to bring the tail end of the target into a position where it could be bound by the free CD-1 or the CD-2 domain of the un-bound monomer. Single molecule FRET (smFRET) experiments show that there is no significant contraction of the DNA when bound to A3G [51]. These experiments were done under conditions where A3G is predominantly monomeric so there is still a possibility that the dimer is capable of facilitating this contraction through an unknown mechanism. The same smFRET experiments also indicate that A3G is capable of scanning ssDNA with and without a target motif. A3G exhibits the same degree of processivity on targets ranging from 3-100 intervening bases giving weight to the scanning hypothesis [32]. The contribution of micro jumping in solution complicates the possibility for a single mechanism being responsible for coordinated deaminations. A3G has been shown to be able to catalyze coordinated deaminations on substrates with dsDNA segments annealed between target sites indicating that the A3G active unit must dissociate from the DNA and reattach to the same strand bypassing the double stranded region [40]. This micro jumping could be accomplished by either complete dissociation or by intersegmental transfer. In reality it is very likely that A3G translocation between adjacent sites is dictated by a combination of these two mechanisms depending on distance and environmental factors.

1.12. Summary

Many structural factors contribute to A3G's demonstrated capacity to restrict Vif-deficient HIV-1 replication. High affinity for single stranded nucleotides leads to both the effective incorporation into nascent viral particles as well as efficient deamination of target ssDNAs. Active center loop 3 gives A3G target specificity leading to the introduction of premature stop codons. The N-terminal domain allows for multimerization leading to catalytically oriented active sites and increases ssDNA dwell time permitting multiple proccessive deaminations from a single A3G active unit. Current data in the field gives good insight into some mechanisms of A3G action, yet more information is still necessary for a complete picture. A high resolution structure of the N-terminal non-catalytic domain would give insight into potential oligomerization interfaces. While a complex structure of the A3G/DNA interface would go a long way in determining the mechanism of target site recognition.

Chapter 2

Impact of H216 on the DNA Binding and Catalytic Activities of the HIV Restriction Factor APOBEC3G

Stefan Harjes, **William C. Solomon***, Ming Li, Kuan-Ming Chen, Elena Harjes, Reuben S. Harris, Hiroshi Matsuo*

Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, USA
**These authors contributed equally to this work.*

Reprinted with permission from American Society for Microbiology

Copyright © American Society for Microbiology, Journal of Virology, Vol 87, 2013, Issue 12, pp 7008-14 10.1128/JVI.03173-12[1]

William C. Solomon contributed to Figure 2.4, generation of mutants, NMR experiments, and manuscript editing

[1] We thank Kylie Walters for helpful discussion. This work was supported by the National Institutes of Health (grants AI073167 to H.M., GM091743 to H.M. and R.S.H., and AI064046 to R.S.H.). Funding forNMRinstrumentation was provided by the Office of the Vice President for Research, Medical School, College of Biological Science, NSF (BIR-961477), and the Minnesota Medical Foundation. The University of Minnesota Supercomputing Institute provided computational resources.

2.1 Synopsis

Full length APOBEC3G (fl-A3G) binds single stranded DNA (ssDNA) oligonucleotides and in a context specific manner deaminates a cytosine base to a deoxyuracil. In this chapter we report the catalytic parameters of the isolated C-terminal domain of APOBEC3G-2K3A (A3Gctd-2K3A), the minimal substrate for A3Gctd-2K3A, the pH dependence of the catalytic rate, and the residue responsible for modulating the pH effects. fl-A3G, wild-type A3Gctd (wt-A3Gctd) and A3Gctd-2K3A all demonstrate a notable increase in the catalytic rate at low pH which allows for determination of Michaelis-Menten type kinetic parameters for the deamination of ssDNA under conditions suitable for NMR analysis. A3Gctd displays a reduced binding affinity for ssDNA while retaining a catalytic turnover rate similar to the holoenzyme. The C-terminal domain recognizes a minimal 5mer ssDNA substrate containing a centrally located 5'-CCC-3' hotspot motif that is flanked by a single nucleotide on both ends. A single amino acid H216 was shown to be responsible for the pH dependence both A3Gctd-2K3A and fl-A3G.

2.2 Introduction

Human APOBEC3G (A3G) is a member of a family of Zn2+-dependent polynucleotide cytosine deaminases. This family was named after APOBEC1 (apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1) and also includes the antibody gene diversification enzyme AID (activation-induced cytidine deaminase) (reviewed in references [53]–[57]). A3G can restrict HIV-1 replication by packaging into

assembling viral particles for delivery to target cells, where it deaminates cytosine to uracil in newly transcribed viral DNA. These cDNA uracils base pair with adenine during plus-strand synthesis and result in G-to-A hypermutation and, in turn, inactivation of the viral genome. A3G has two Zn2+ binding domains that span residues 1 to 196 and 197 to 384, but only the C-terminal domain is catalytically active [58]–[60]. The N-terminal domain interacts with HIV-1 Vif, RNA, and single-stranded DNA (ssDNA) (e.g., see references [59] and [61]–[63]). A3G predominantly deaminates the 3′ cytosine (underlined) in a 5′-CC<u>C</u> sequence, although the middle cytosine can also be deaminated in subsequent reactions following deamination of the 3′ cytosine [2], [3], [52], [64], [65]. In longer ssDNA substrates with multiple 5′-CCC sites, A3G deamination exhibits a 3′→5′ spatial preference in vitro [61], [41], [66]. In the present study, we use the catalytic domain of A3G (A3Gctd) to determine kinetic parameters. Our results provide kinetic constants for two independent deaminations within a 5′-CCC sequence, which explain A3G's catalytic site preference for the 3′ cytosine. We identify a strong pH dependence of the reaction speed, which implies that a histidine residue is involved in substrate binding. In addition, we identify the shortest-length ssDNA substrate for A3Gctd to be a pentanucleotide.

2.3 Results

2.3.1 Time-resolved NMR analysis of DNA cytosine deamination catalyzed by A3Gctd

As shown in **Fig. 2.1a**, the progress of A3Gctd-catalyzed deamination of 5′-ATTCCCAATT to 5′-ATTCCUAATT was monitored by using a series of 1H NMR

spectra. The rising product signal at 5.6 ppm is sufficiently resolved to allow quantitative analysis. By using appropriate enzyme and substrate concentrations, the speed of the reaction can be adjusted so that we can monitor the reaction course on a time scale of several hours. As shown in **Fig. 2.1b**, the speed of the catalytic reaction is constant for the first 25 data points. Thus, we can determine the initial speed of the reaction under the given conditions. This determination allows us to characterize the enzyme kinetics of A3Gctd by measurements at several different substrate concentrations. Our first attempt was performed at a physiological pH of 7.4. Although we used very high substrate concentrations of up to 10 mM, we were not able to saturate the enzyme. Instead, the initial speed of the reaction increased linearly with substrate concentration (**Fig. 2.2a**). We tested both wild-type A3Gctd and a soluble variant containing five substitutions, L234K, C243A, F310K, C321A, and C356A, called A3Gctd-2K3A [29] and found that both proteins exhibited the same linear dependence of reaction speed on substrate concentration. This finding suggested that either the reaction cannot be described by a Michaelis-Menten-type mechanism or saturation of the enzyme occurs at concentrations higher than 10 mM under the conditions used for these experiments.

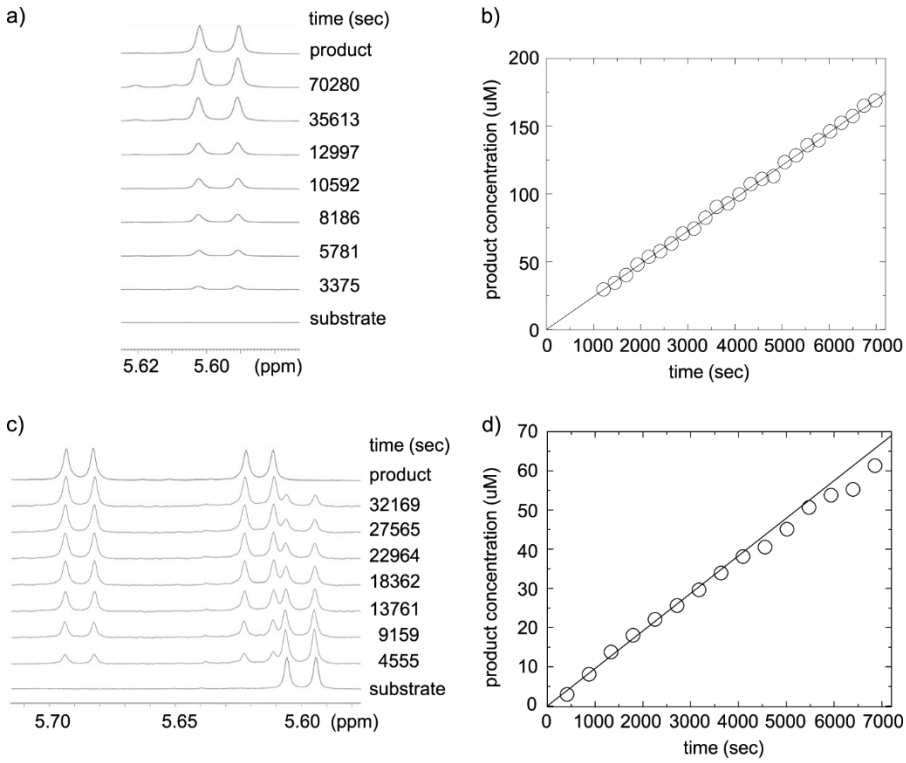

Figure 2.1. Time-resolved NMR analysis of A3Gctd-catalyzed ssDNA cytidine deamination. (a) An enlarged region of a 1H spectrum of 5′-ATTCCCAATT-3′ is compared at several time points during a deamination reaction to illustrate the increase in the H5 proton signal of uridine due to formation of the product 5′-ATTCCUAATT-3′. The spectrum of a synthesized decanucleotide containing the product sequence is provided at the top, and that of the substrate is shown at the bottom. An 818 μM concentration of oligonucleotide was mixed with a 7.7 μM concentration of A3Gctd in 50 mM phosphate buffer (pH 7.4) with 10% D2O at 25°C. After addition of enzyme and preparation of the spectrometer, a 1H spectrum was recorded every 220 s. (b) The H5 proton signal of uridine at 5.60 ppm was integrated, calibrated, and plotted versus time. The straight line is a linear fit to the data. (c) Same as panel a, but the substrate is 5′-ATTCCUAATT-3′, and the product is 5′-ATTCUUAATT-3′. A 237 μM concentration of oligonucleotide was incubated with 10 μM A3Gctd in 75 mM citrate buffer (pH 5.5) with 10% D2O at 25°C. A 1H spectrum was recorded every 460 s. (d) The H5 proton signal of uridine at 5.69 ppm was integrated, calibrated, and plotted versus time. The straight line is a linear fit to the data.

In order to find better conditions for kinetic analysis, we tested the influence of pH, salt concentration, and the presence of Mg2+, Ca2+, and Zn2+ in the buffer solution on the reaction speed of A3Gctd deamination. Only in the case of varied pH did we observe a strong influence on the reaction speed. At low substrate concentrations, the

reaction speed increased significantly with more acidic pH. As shown in **Fig. 2.2b**, wild-type A3Gctd reaches a maximum enzymatic activity at pH 5.5, where the speed of reaction is ~30 times higher than that at pH 7.5 (**Fig. 2.2b**, circles). Catalytic activity of the 2K3A variant peaked at pH 6.0 (**Fig. 2.2b**, stars).

We therefore monitored deamination reactions of wild-type A3Gctd for several different substrate concentrations at pH 5.5. As shown in **Fig. 2.2c**, at pH 5.5, we were able to saturate the enzymes. The speed of enzymatic catalysis first increases with substrate concentration but then reaches a plateau (**Fig. 2.2c**). Accordingly, we fit the double-reciprocal data with a straight line to determine a wild-type Km of 0.57 ± 0.09 mM and a kcat of 5.80 ± 2.20 deamination events per minute for the 5′-ATTCCCAATT-to-5′-ATTCCUAATT reaction (**Table 2.1**). We also observed Michaelis-Menten-type behavior for A3Gctd-2K3A and obtained similar constants (Km = 1.4 ± 0.7 mM; kcat = 4.6 ± 2.9 min−1).

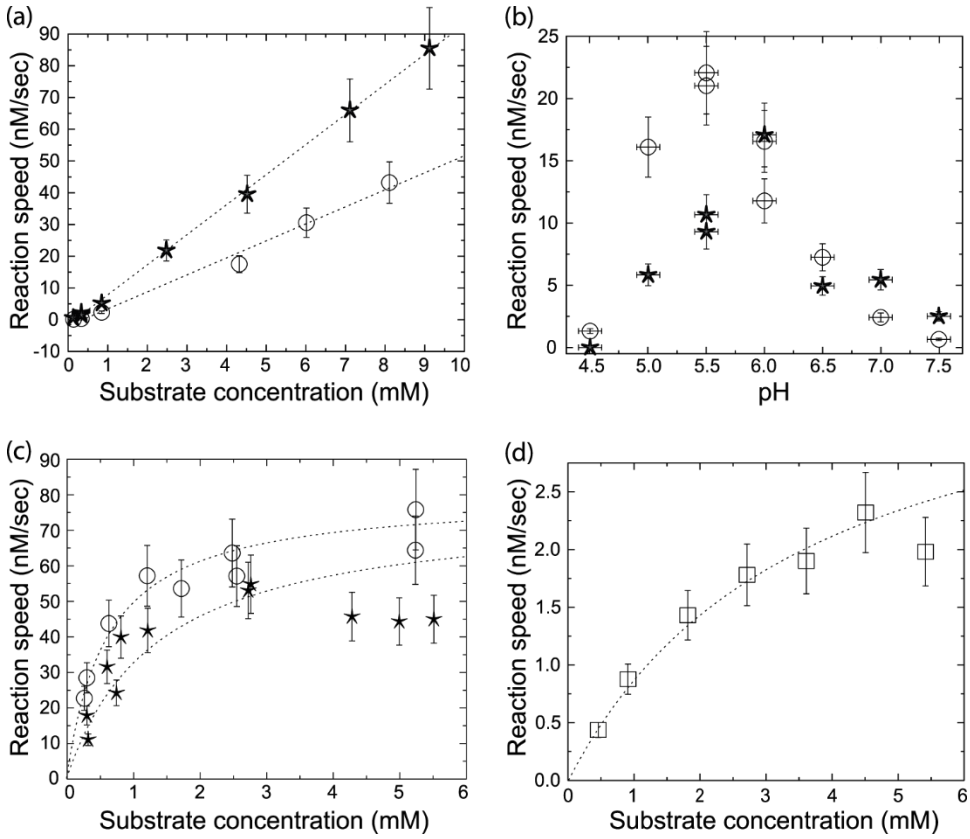

Figure 2.2. Speed of A3Gctd-catalyzed deaminations. (a) Speed of A3Gctd-catalyzed 5′-CCC-to-5′-CCU deamination at pH 7.4. The initial speed of the deamination was determined for different substrate concentrations for wild-type A3Gctd (circles) or the 2K3A variant (stars). (b) pH dependence of the speed of A3Gctd-catalyzed deamination. 5′-ATTCCCAATT-3′ oligonucleotides (300 μM) were incubated with wild-type A3Gctd (circles) or the 2K3A variant (stars), and the initial speed of the reaction at the indicated pH was determined. The ionic strength of the different buffers was adjusted to similar levels by addition of sodium chloride. (c) Substrate concentration dependence of A3Gctd-catalyzed 5′-CCC-to-5′-CCU deamination at pH 5.5. Initial speed of the deamination at 25°C was determined for several substrate concentrations using wild-type A3Gctd (circles) and the 2K3A variant (stars). For panels a to c, the data were collected from four independent enzyme preparations and calibrated to the same enzyme concentration of 820 nM. (d) Substrate dependence of wild-type A3Gctd-catalyzed 5′-CCU-to-5′-CUU deamination at pH 5.5. Initial speed of catalysis was determined by using wild-type A3Gctd at a 2.4 μM concentration. To compare these data with those for the 5′-CCC-to-5′-CCU reaction in panel c, the speed of the reaction was calibrated to an 820 nM enzyme concentration. For panels c and d, dashed lines show reaction curves generated by double-reciprocal (Lineweaver-Burk) analysis of the data.

The product of the initial deamination reaction, 5′-ATTCCUAATT, was further deaminated to 5′-ATTCUUAATT. This second product exhibited a distinct doublet NMR

signal at 5.69 ppm (**Fig. 2.1c and d**). We monitored wild-type A3Gctd's catalytic reactions with different 5′-ATTCCUAATT substrate concentrations to determine the initial speed of each reaction, which was plotted against the corresponding substrate concentration in **Fig. 2.2d**. The speed of this secondary deamination initially increases with increasing substrate concentrations but then plateaus. We determined Michaelis-Menten constants by using the linear fit of double-reciprocal data to find that this reaction is significantly slower, with a kcat of less than 1 deamination event per minute (**Table 2.1**). In addition, a higher Km was measured, which indicates that 5′-ATTCCUAATT binds A3Gctd with lower affinity than 5′-ATTCCCAATT. Moreover, the value of kcat/Km can be regarded as a substrate specificity constant [67]. This value was calculated to be 10 M−1 s−1 for 5′-ATTCCCAATT and 0.22 M−1 s−1 for 5′-ATTCCUAATT. Therefore, A3Gctd exhibited a 45-fold greater preference for the 5′-CCC substrate over the 5′-CCU substrate. It should be noted that we have previously shown that A3Gctd produces undetectable levels of 5′-CUC or 5′-UCC products from the 5′-CCC substrate using an electrochemical deamination assay [68].

	Mean value for reaction ± SD	
Parameter	CCC to CCU	CCU to CUU
K_m (mM)	0.57 ± 0.09	3.60 ± 0.72
k_{cat} (1/min)	5.80 ± 2.20	0.80 ± 0.29

[a] Wild-type A3Gctd deaminase reactions included 5′-ATTCCCAATT-to-5′-ATTCCUAATT and 5′-ATTCCUAATT-to-5′-ATTCUUAATT reactions. The constants originate from a linear fit of the function $y = a \times x + b$ to the data, with y as the reciprocal of the rate of production and x as the reciprocal of the substrate concentration, where k_{cat} equals $1/b$ and K_m equals a/b.

Table 2.1. Linear regression analysis of wild-type A3Gctd deaminase reactions[a]

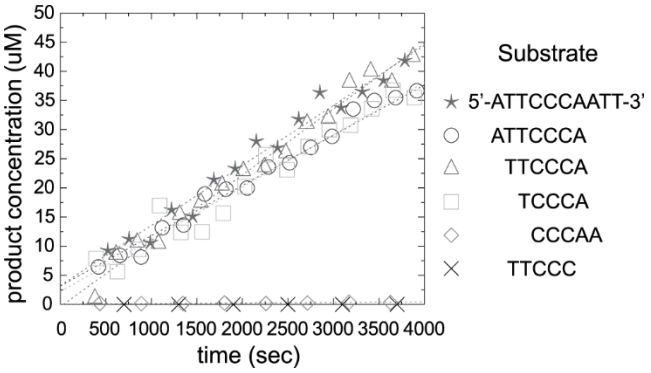

Figure 2.3. Speed of deamination catalysis as a function of substrate sequence. Initial speed was determined for the conversion of 600 μM substrate of the indicated sequence and A3Gctd-2K3A at pH 5.5. The enzyme concentration for all experiments was 820 nM, except for the 5′-CCCAA-3′ determination, for which 8.2 μM enzyme was used. The product concentration of 5′-CCCAA-3′ is calibrated to the 820 nM enzyme concentration.

2.3.2 Determination of a minimum substrate for A3Gctd

To test whether H216 is responsible for the pH dependency of A3Gctd deamination, we mutated H216 to alanine (A3Gctd-2K3A-H216A) and to arginine (A3Gctd-2K3A-H216R) and measured reaction speeds at pH 7.3 and pH 6.0. A3Gctd-2K3A-H216A was catalytically dead at pH 7.3 (data not shown), which is consistent with previous E. coli assay results [29]. In comparison, A3Gctd-2K3A-H216R showed a reaction speed similar to that of A3Gctd-2K3A at pH 7.3 (**Fig. 2.4b**). The difference was revealed at pH 6.0, as A3Gctd-2K3A's reaction speed was 8-fold higher than that at pH 7.3, while A3Gctd-2K3A-H216R showed only a modest 1.4-fold increase in the reaction speed (**Fig. 2.4b**).

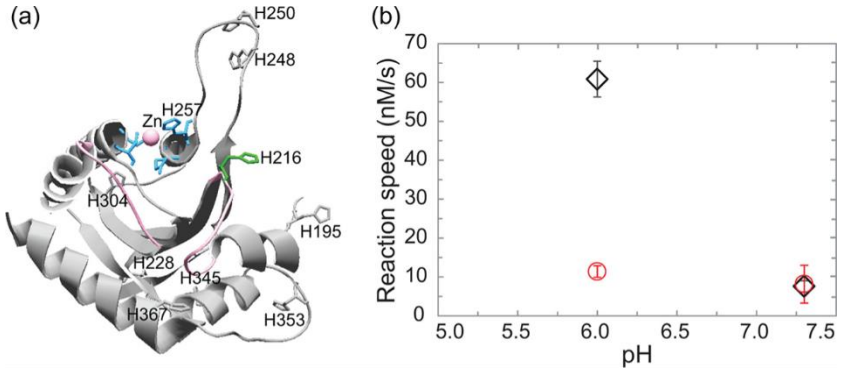

Figure 2.4. Role of H216 in A3Gctd's catalytic reaction. (a) Locations of 10 histidine residues in the three-dimensional structure of A3Gctd (Protein Data Bank [PDB] accession number 3IR2). A Zn atom (pink) is coordinated by H257, C288, and C291 (blue). H216 is shown in green, whereas all other histidines are gray. Two loops spanning residues 207 to 217 and 313 to 320 are shown in pink. (b) pH dependence of the speed of A3Gctd-2K3A-catalyzed deamination. 5′-ATTCCCAATT-3′ oligonucleotides (150 μM concentration) were mixed with a 1.5 μM concentration of the A3Gctd-2K3A (black diamonds) or A3Gctd-2K3A-H216R mutant (red circles) in sodium-phosphate buffer at pH 7.3 or pH 6.0, and the initial speed of the reaction was determined. Three independent experiments were conducted for each pH, and error bars show standard errors.

2.3.3 pH dependency of full-length A3G deamination

We next used a FRET-based deamination assay to ask whether wild-type full-length A3G deamination is similarly modulated by pH. Figure 2.5 shows that A3G has 4-fold-higher deamination activity at pH 5.5 than at pH 7.4 (values were compared at a 33.7 nM protein concentration). Since we observed increased deamination activity at lower pH, we generated H216A (A3G-H216A) and H216R (A3G-H216R) variants of full-length A3G in order to find whether H216 is responsible for the pH dependence of deamination activity. Figure 2.5 shows that A3G-H216A was catalytically dead at pH 7.4 as well as at pH 5.5, but A3G-H216R had deamination activity similar to that of wild-type A3G at pH 7.4. Interestingly, A3G-H216R showed only 1.5-fold-higher deamination activity at pH 5.5 than at pH 7.4 (values were compared at a 33.7 nM protein

concentration) (**Fig. 2.5**), further indicating that protonation of H216 is a key part of the mechanism by which A3G shows increased deamination activity at pH 5.5.

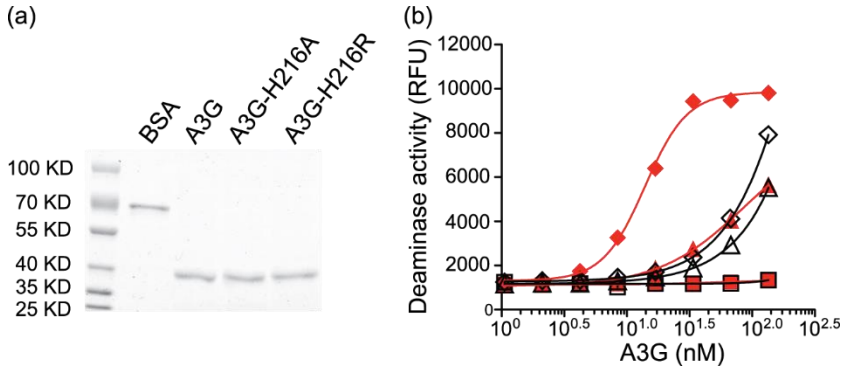

Figure 2.5. pH dependence of the deamination activities of A3G, A3G-H216A, and A3G-H216R. (a) Purities of A3G, A3G-H216A, and A3G-H216R, as shown by SDS-PAGE and Coomassie blue R250 staining. These purified proteins were used for deaminase activity assays shown in panel b. BSA, bovine serum albumin. (b) pH dependence of deaminase activities of A3G at pH 5.5 (red diamonds), A3G-H216A at pH 5.5 (red squares), A3G-H216R at pH 5.5 (red triangles), A3G at pH 7.4 (white diamonds), A3G-H216A at pH 7.4 (white squares), and A3G-H216R at pH 7.4 (white triangles). Deaminase activities are shown by using fluorescence intensity emitted by products in relative fluorescence units (RFU).

2.3.4 Impact of H216 mutations on HIV-1 restriction activity

Since H216 is important for A3G's deaminase activity, A3G-H216R and A3G-H216A were tested for their capability to restrict HIV infection. A3G-H216A showed almost no restriction capability, but A3G-H216R maintained restriction of HIV infection (**Fig. 2.6**). The restriction activity of A3G-H216R was slightly weaker than that of wild-type A3G, consistent with biochemical data presented above and prior studies demonstrating the importance of DNA deaminase activity for full levels of HIV-1 restriction [69], [5], [70].

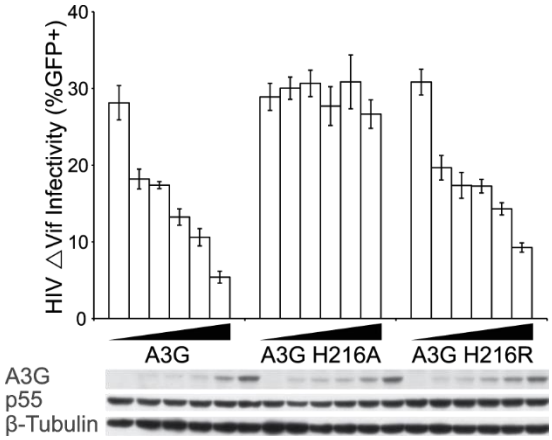

Figure 2.6. HIV-1 restriction activities of A3G-H216A and A3G-H216R. HIV-1 ΔVif infectivity was determined by measuring the ratio of GFP-expressing target 293T cells by flow cytometry. Each data point represents the average of three independent reactions. The standard deviations are shown as error bars. The expression of A3G-mycHis and mutant proteins in producer 293T cells was detected by Western blotting using an A3G monoclonal antibody (catalog number 7105; ImmunoDiagnostics). Gag (p55) was detected by using a p24 monoclonal antibody, and β-tubulin is shown as a loading control.

2.4 Discussion

A3G predominantly deaminates 3′ cytosine in a 5′-CCC sequence ([2], [3], [52], [64], [65]). The middle cytosine can be deaminated, but its deamination speed is much lower than that of the primary 3′ cytosine. Previous studies failed to distinguish A3G's preference for the primary 3′ cytosine. We used time-resolved NMR, which was originally proposed by Furukawa and coworkers [64], to determine kinetic parameters for A3Gctd-catalyzed deamination. The reaction speed was maximal at pH 5.5, and it followed Michaelis-Menten-type kinetics. This pH dependency may imply involvement of histidine residues for substrate binding and/or the catalytic reaction, since the pKa value of the conjugated acid of imidazole is around pH 6.0. A3Gctd contains 10 histidine residues, including a Zn2+-coordinating histidine, H257. Protonation of a Zn2+-

coordinating histidine is likely to destabilize the protein structure, and this is probably the reason for the decline of catalytic activity observed at pH 4.5. E. coli cytidine deaminase, which has a Zn2+-coordinating histidine, did not show a maximum reaction speed at pH 5.5 to 6.0 [71]. Therefore, H257 is not likely the cause of the observed pH dependency of A3Gctd deamination. Among the remaining histidines, alanine mutants of H216 and H345 have shown reductions in mutation frequency of 30-fold and 10-fold, respectively, whereas substitution of the others maintained >25% of wild-type catalytic activity [29], [72]. H216 was one of the five most perturbed residues upon titration of a 21-mer single-stranded DNA containing a 5′-CCC hot spot, identified by monitoring main-chain 1H-15N chemical shifts [73]. Figure 2.4a shows positions of all histidine residues in a three-dimensional structure of A3Gctd. Loops spanning residues 207 to 217 and 313 to 320 are also shown in color, since they have been suggested to play important roles in DNA binding [61], [66], [29], [30], [73], [74]. H216 is located in the loop at residues 207 to 217, and it is spatially close to the catalytic site (Fig. 4a). Indeed, amino acid substitutions of H216 modulated catalytic activity of A3Gctd-2K3A, as A3Gctd-2K3A-H216A was catalytically dead, whereas A3Gctd-2K3A-H216R, as well as A3Gctd-2K3A, was active at pH 7.3, and it showed only a modest increase in catalytic speed at pH 6.0 (Fig. 4b). Therefore, it is plausible that H216 interacts with substrate DNA using a mechanism which is enhanced by protonation of its imidazole ring. Importantly, wild-type full-length A3G also showed a significant increase in deamination activity at pH 5.5 (Fig. 5), indicating that the results obtained with A3Gctd reflect the DNA binding and deamination mechanism of the full-length enzyme. It should also be noted that full-length A3G binds 10-mer ssDNA as strongly as 69-mer ssDNA, with apparent Kd (dissociation

constant) values of 50 nM and 52 nM, respectively [61], suggesting that shorter substrates make all essential contacts required for binding and catalysis. Furthermore, our data indicate that the observed pH dependency is independent of substrate length, as shorter substrates (decamer in NMR experiments) and longer substrates (24-mers in FRET-based experiments) behaved similarly. There were small but significant decreases in viral restriction with A3G-H216R compared to wild-type A3G, although A3G-H216R's deamination activity was very similar to wild-type A3G activity at pH 7.4 (**Fig. 2.5 and 2.6**). These data imply that H216 may be protonated to increase A3G's catalytic activity inside the viral capsid during reverse transcription.

The Km value for 5′-CCC-to-5′-CCU deamination by A3Gctd is on the same order of magnitude as mononucleoside deamination by E. coli cytidine deaminase (0.12 mM), whereas A3Gctd's kcat value is much lower than the value of 299 s−1 determined for E. coli cytidine deaminase [75]. A3Gctd's kcat value is similar to that determined for human RNA-editing adenosine deaminase 2 (ADAR2) (0.88 min−1) [76]. A common feature between ADAR2 and A3G is that both require binding to polynucleotide substrates prior to deamination, as ADAR2's substrate is duplex RNA [77], [78] and A3G's substrate is single-stranded DNA, whereas E. coli cytidine deaminase deaminates the mononucleoside cytidine. Therefore, A3Gctd and ADAR2 may take additional time to position their substrates before catalysis can occur, which may cause their speed of deamination to be lower than that of E. coli cytidine deaminase. The kcat value of A3Gctd is on the same order of magnitude as the reported kcat value of wild-type full-length A3G (9.78 ± 2.4 min−1) [52], while the Km value of wild-type A3G was reported to be 0.82 ± 0.06 nM [52], which is 105 times lower than that of A3Gctd. The

nonspecific ssDNA binding of wild-type A3G, with a dissociation constant of 52 to 238 nM, is much stronger than that of A3Gctd, which has a dissociation constant of 130 to 400 μM [29], [52], [61], [62], [64], [66], [79]. Thus, the large difference in Km values may be partly due to the stronger binding capability of the N-terminal domain of A3G, which was not present in our NMR studies because full-length A3G aggregates at higher protein concentrations. It should be noted that the above-mentioned Km value was determined by using an 80-mer oligonucleotide substrate containing one 5′-CCC hot spot ([52]; see also references [61] and [66]), suggesting that longer substrates are more efficiently deaminated than shorter substrates, since deamination predominantly occurs when A3G approaches a 5′-CCC target with a 3′→5′ orientation. Therefore, it is plausible that 3′→5′ sliding enhances the likelihood of binding in a catalytically active orientation of substrate DNA, which could explain the lower Km value determined by using a longer ssDNA substrate [52].

Although RNA- and ssDNA-dependent oligomerization of A3G has been observed, atomic force microscopy studies showed that A3G was able to bind and slide over ssDNA as a monomer [61], [66], [69], [80]–[84]. In addition, a double mutant (F126A/W127A) that was predominantly monomeric in solution bound ssDNA and catalyzed deamination with 3′→5′ directionality similarly to wild-type A3G [41]. Therefore, monomeric A3Gctd is a useful model for studying the kinetic parameters of the full-length enzyme.

The minimum substrate length for A3Gctd was found to be a pentamer containing a 5′-CCC hot spot flanked on both sides by one nucleotide, suggesting that key interactions for substrate specificity may involve these five nucleotides. This result is

consistent with substrate specificity studies using nucleoside analogues, which revealed that A3G was sensitive to nucleotides at positions −3 to +1 of the primary target cytosine [65]. The present study provides fundamental knowledge of A3G-catalyzed cytosine deamination and may aid in the design of ssDNA for high-resolution structure determination of an A3G-substrate complex.

2.5 Methods summary

2.5.1 Purification of A3Gctd

The APOBEC3G C-terminal domain (A3Gctd), comprising amino acids 191 to 384, was expressed and purified as previously described ([25]). Briefly, the glutathione S-transferase (GST)-fused A3Gctd was expressed in Escherichia coli BL21(DE3) cells overnight at 17°C. After harvesting, the cells were resuspended in 50 mM sodium phosphate buffer (pH 7.4) and lysed by sonication. After ultracentrifugation at 25,000 × g for 10 min, the supernatant was added to glutathione (GSH)-Sepharose, which was subsequently washed. For kinetic analysis, the GST fusion protein was eluted from the Sepharose matrix with 100 mM GSH in phosphate buffer. By using filtration at 4,000 × g, the buffer was changed to a solution containing 50 mM phosphate (pH 7.4), 50 mM NaCl, 50 μM ZnCl2, and 10% glycerol, and the protein was concentrated and frozen for storage. For pH analysis and measurements at pH 5.5, the buffer was composed of 75 mM sodium phosphate and 75 mM citrate, and the pH was adjusted accordingly. The ionic strength of different buffers was adjusted by using sodium chloride.

2.5.2 Kinetic analysis of A3Gctd's deaminase activity

Kinetic measurements were conducted in the same buffer at a glycerol concentration of 0.6% and with 10% deuterium oxide. All nuclear magnetic resonance (NMR) data were acquired on a 700-MHz Bruker NMR spectrometer equipped with a 1.7-mm cryoprobe. A series of 1H spectra of the oligonucleotide substrates including 5′-ATTCCCAATT-3′, 5′-ATTCCUCAATT-3′, 5′-ATTCUUAATT-3′, 5′-ATTCCCA-3′, 5′-TTCCCA-3′, 5′-TCCCA-3′, 5′-CCCAA-3′, and 5′-TTCCC-3′ at concentrations ranging from 100 μM to 10 mM was recorded. The spectra were baseline corrected, and the H5 proton signals of the 3′ and middle uracils from the 5′-CCU and 5′-CUU products, respectively, were integrated; these appeared at 5.60 ppm and 5.69 ppm, respectively. Two groups of signals were used as an internal standard in order to determine the product concentration during the reaction, including oligonucleotide signals at 8.18 to 7.97 ppm and a doublet of doublets at 2.57 to 2.39 ppm originating from citrate buffer. The product signal area was converted to product concentration and plotted versus time of reaction. To determine the initial speed, this plot was fitted with linear regression, and the slope of the initial linear part was recorded as initial speed. Enzyme concentrations were determined by SDS-acrylamide gel electrophoresis. A dilution series by a factor of 5 of the protein of interest was prepared, and usually, three samples were loaded onto an 18% gel. To calibrate the measurement, commercial carbonic anhydrase (CA) was weighted to 10 mg/ml. A dilution series by a factor of 1.5 of the 1-mg/ml CA solution was loaded next to the protein of interest. After electrophoresis and staining, the gel was photographed, and the image was analyzed by using ImageJ [85].

2.5.3 Purification of full-length A3G

Full-length A3G-mycHis, A3G H216A-mycHis, and A3G H216R-mycHis proteins were purified from transiently transfected 293T cells as described previously [26], [69]. Briefly, 293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) with 10% fetal bovine serum (FBS) (Gibco), 50 units/ml penicillin, and 50 μg/ml streptomycin (Invitrogen) at 37°C in 5% CO2. Forty-eight hours after transfection with TransIT-LTI (Mirus Bio), cells were harvested and lysed in a solution containing 25 mM HEPES (pH 7.4), 150 mM NaCl, 1 mM MgCl2, 1 mM ZnCl2, 1 mM EDTA, 0.5% Triton X-100, and 10% glycerol. The soluble cell lysates were harvested by centrifugation (14,000 rpm for 10 min), and A3G-mycHis proteins were purified by affinity chromatography with Ni-nitrilotriacetic acid (NTA) agarose (Qiagen). The purified proteins were subjected to SDS-PAGE/Coomassie blue R250 staining for purity and concentration determination.

2.5.4 FRET-based deamination assay

Fluorescence resonance energy transfer (FRET)-based deamination assays were performed as described previously [26], [69], [86], [87]. Purified A3G-mycHis and mutant proteins were incubated with a dually labeled single-stranded DNA (5′–6-carboxyfluorescein [FAM]–AAATATCCCAAAGAGAGAATGTGA–6-carboxytetramethylrhodamine [TAMRA]–3′) (Biosearch Technologies, Inc.) and E. coli uracil-DNA glycosylase (NEB) in a 384-well black plate (Nunc) at 37°C for 2 h. The products with abasic sites were cleaved by adding 3 μl of 4 N NaOH. The pH of the reaction solution was regulated with 3 μl of 4 N HCl and 37 μl of 2 M Tris-Cl (pH.9), and

the fluorescence was read in a Synergy Mx monochromator-based multimode microplate reader (BioTek).

Single-cycle HIV-1 infectivity assay.Single-cycle HIV-1 infectivity assays were performed as described previously [88]. 293T cells were plated into a 6-well plate and transfected with 0.22 μg pCS-CG, 0.14 μg pRK5/Pack1 (Gag-Pol), 0.07 μg pRK5/Rev, and 0.07 μg pMDG (vesicular stomatitis virus G protein) along with 4.4, 5.5, 7.3, 11, or 22 ng of pcDNA3.1-A3G-mycHis plasmid or mutant A3G plasmids by using TransIT-LTI (Mirus Bio). After 48 h of incubation, virus-containing supernatants were harvested to infect target 293T cells. The transduced target 293T cells were harvested after an additional 24-h incubation and subjected to flow cytometry (BD FACSCanto II) to measure the ratio of green fluorescent protein (GFP)-expressing cells.

Chapter 3

Structural and Mechanistic Insight into APOBEC3G Substrate Binding and Selection

William C. Solomon *and Hiroshi Matsuo*

Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, USA

3.1 Synopsis

APOBEC3G is capable of finding and deaminating cytidines in ssDNA located in specific motifs while excluding oligonucleotides containing a single RNA base at the target position. In this chapter we examine the structural interactions involved in catalytic substrate binding, discrimination between target motifs, and the mechanism for exclusion of non-target nucleotides. A3Gctd binds ssDNA with an extended binding interface that encompasses a large portion of the catalytic face of the protein. A3Gctd undergoes significant conformational flexing during catalytic binding and deamination resulting in the perturbation of multiple residues located in the core of the protein. A3Gctd binds ssDNA non-specifically through a small subset of the binding residues and selectively engages substrates containing the target motif through a more extensive set of adjacent residues. We show that A3Gctd fails to engage the more extensive set of residues on substrates containing target nucleotides that preferentially exhibit the C3'-endo conformation typical of RNA. We also identify a novel substrate for A3G containing a 2'-F-ANA cytidine at the target position, which adopts the C2'-endo conformation seen in the canonical substrate ssDNA.

3.2 Introduction

APOBEC3G has an important role in human defense against retroviral pathogens, including HIV-1. Its single-stranded DNA cytosine deaminase activity, located in its C-terminal domain (A3Gctd), can mutate viral cDNA. Targeted deamination of the viral genome effects viral genome stability manifesting in reduced integration and non-functional gene products [3], [48], [53]. This deamination occurs in a context dependent

manner, selectively targeting a minimal di-cytidine motif [16], [17], [64], [65], [68], [74], [89], [90]. We previously showed that the C-terminal domain of APOBEC3G preferentially deaminates the 3' cytidine in the hotspot 5'-CCC motif with a 45-fold greater preference than the di-cytidine substrate motif 5'-CCU [17]. In the same study we determined that A3Gctd demonstrates a considerable increase in the catalytic rate at reduced pH with a maximal increase at pH 6.0. We identified the histidine at position 216 to be responsible for the pH sensitivity as arginine and alanine substitutions in this position resulted in a loss of pH sensitivity or a complete loss of deamination activity, respectively. H216 is located in a loop adjacent to the active site, but does not interact with the active site residues. Since H216 does not directly participate in deamination, we postulated that it must be involved in substrate binding and functions to increase A3Gctd affinity for substrate oligomers.

Several APOBEC3 proteins have had their structures determined [18], [20], [22], [23], [25], yet to date no structures of an APOBEC protein has been solved in complex with DNA. NMR studies have been pursued with APOBEC3A [20] and APOBEC3G [29], both studies suffered from low substrate affinities (~55μM for A3A, ~450μM for A3Gctd) resulting in small chemical shift perturbations. Structural interactions of ssDNA with APOBEC3A indicate an extended binding interface involving residues in loop 7, between β4 and α4, and loop 3, between β2 and α2, along with residues immediately surrounding the catalytic site [20]. Molecular dynamics simulations of the related protein AID with substrate DNA display a similar binding interface[91].

A3G interacts extensively with both endogenous and viral RNAs through its N-terminal domain. This interaction is involved in subcellular localization to mRNA

processing bodies and incorporation into viral particles [59], [92], [93]. A3G strongly binds RNA, which contains a plethora of di-cytidine motifs, but displays no catalytic activity on RNA nucleotides [89]. The lack of activity on RNA cytidines must be the result of a yet to be identified gating interaction between A3Gctd and RNA. Kohli and coworkers proposed a mechanism for AID's lack of deamination of RNA substrates involving the pucker of the ribose sugar [94]. DNA predominantly assumes the C2'-endo conformation, while RNA assumes the C3'-endo conformation. These two conformations position the nucleotide base in different orientations with respect to the sugar. AID displayed activity on substrates containing 2'-F-ANA and ANA substrates, both of which assume the canonical DNA C2'-endo conformations [94].

Given the lack of significant structural data on the binding of APOBEC3G to substrate DNA, an unknown mechanism of RNA exclusion, and APOBEC3G's important role in HIV-1 restriction; we have undertaken this study to determine the A3Gctd substrate interface with DNA, RNA, and conformational mimics. Taking advantage of the increased substrate affinity of A3Gctd at low pH we have identified the relevant interactions of A3Gctd with various oligonucleotides.

3.3 Results

3.3.1 pH 6.0 assignment of A3Gctd-2K3A and A3Gctd-2K3A E259A mutant

As previously reported [17], A3Gctd-2K3A binds ssDNA with a higher affinity at low pH with a maximal increase in catalytic rate at pH 6.0. In order to assess the effects of reduced pH on the ssDNA binding of A3Gctd we first had to assign the NMR chemical shifts at pH 6.0. Initial pH titration data for A3Gctd-2K3A indicated a

substantial number of shifts in NMR peak positions that correlated exclusively to solvent exposed residues. Since the surface residues exhibited fast exchange dynamics, we were able to follow peak shifts through a range of titration points from the previously reported [25] chemical shifts to assign the pH 6.0 spectrum (**Fig. 3.1**).

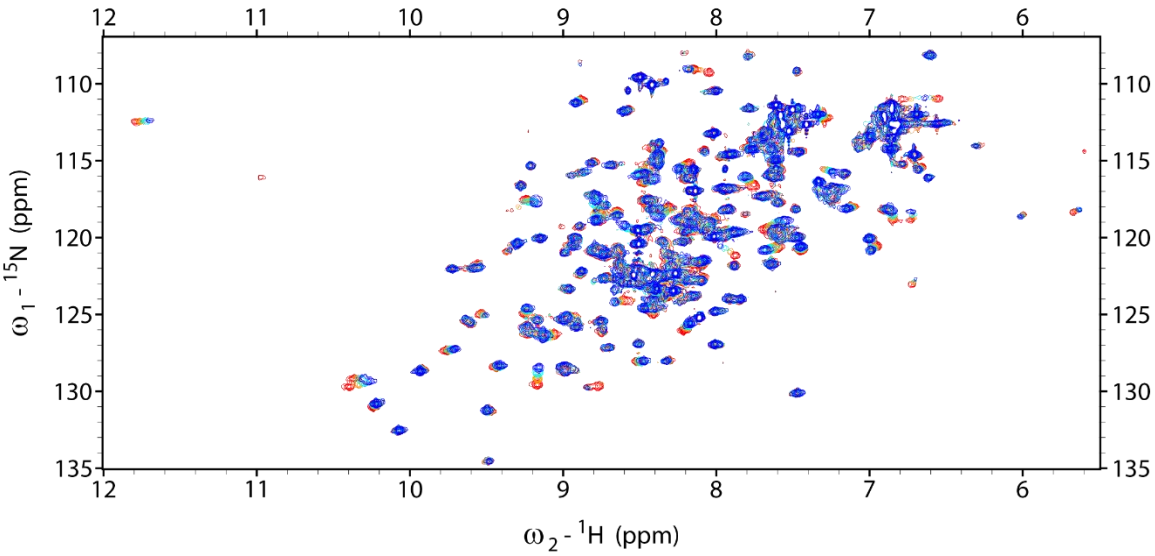

Fig. 3.1. pH titration of A3Gctd-2K3A. ^{1}H-^{15}N HSQC spectra overlay of A3Gctd-2K3A at pH 7.0 (red), pH 6.6 (orange), pH 6.3 (cyan), pH 6.0 (blue).

To determine specific binding differences between potential substrates, a single alanine point mutation was introduced at the catalytic glutamate (E259A) to produce a catalytically inactive mutant A3Gctd-2K3A E259A protein. The resulting A3Gctd-2K3A E259A spectrum displayed multiple chemical shift perturbations compared to the A3Gctd-2K3A reference, requiring *de novo* assignment of A3Gctd-2K3A E259A. Due to increased protein stability, the initial assignment was done at pH 7.3 and the assignments were transferred to the pH 6.0 spectrum in the same manner as A3Gctd-2K3A. We were able to assign 76% of non-proline main chain residues at pH 6.0, representative of a

majority of the surface exposed binding area. A small number of residues were absent at pH 6.0 that were assigned at pH 7.3, indicating that dynamics of those residues may change at low pH.

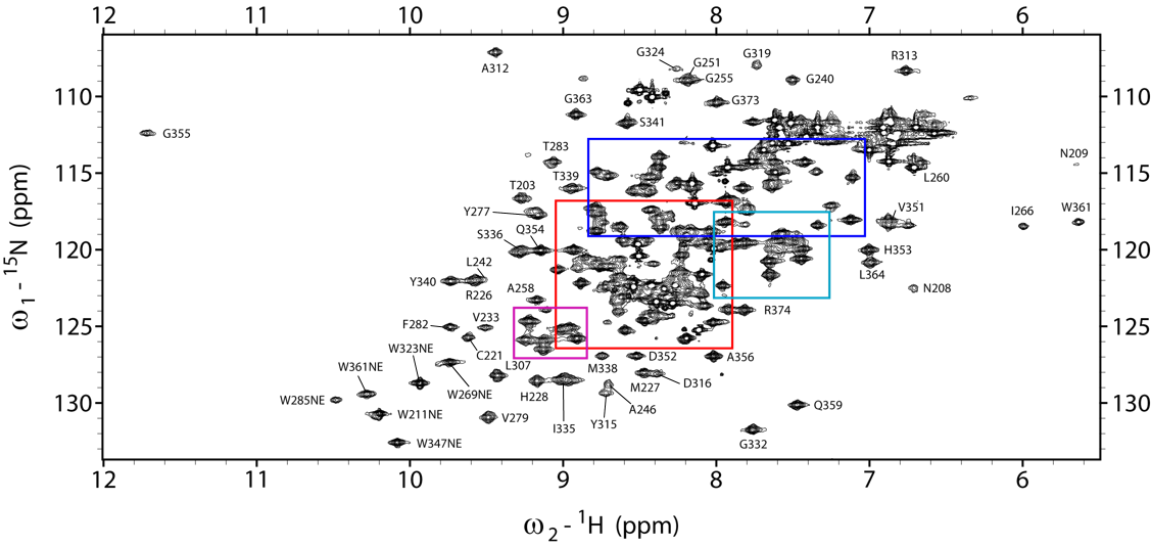

Fig. 3.2.a. Assignment of A3Gctd-2K3A E259A at pH 6.0. ^{1}H-^{15}N HSQC spectrum of A3Gctd-2K3A E259A. Peak assignments determined using TROSY versions of HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO, HN(CA)CO triple resonance experiments. (Insets in **Fig. 3.2.b** below)

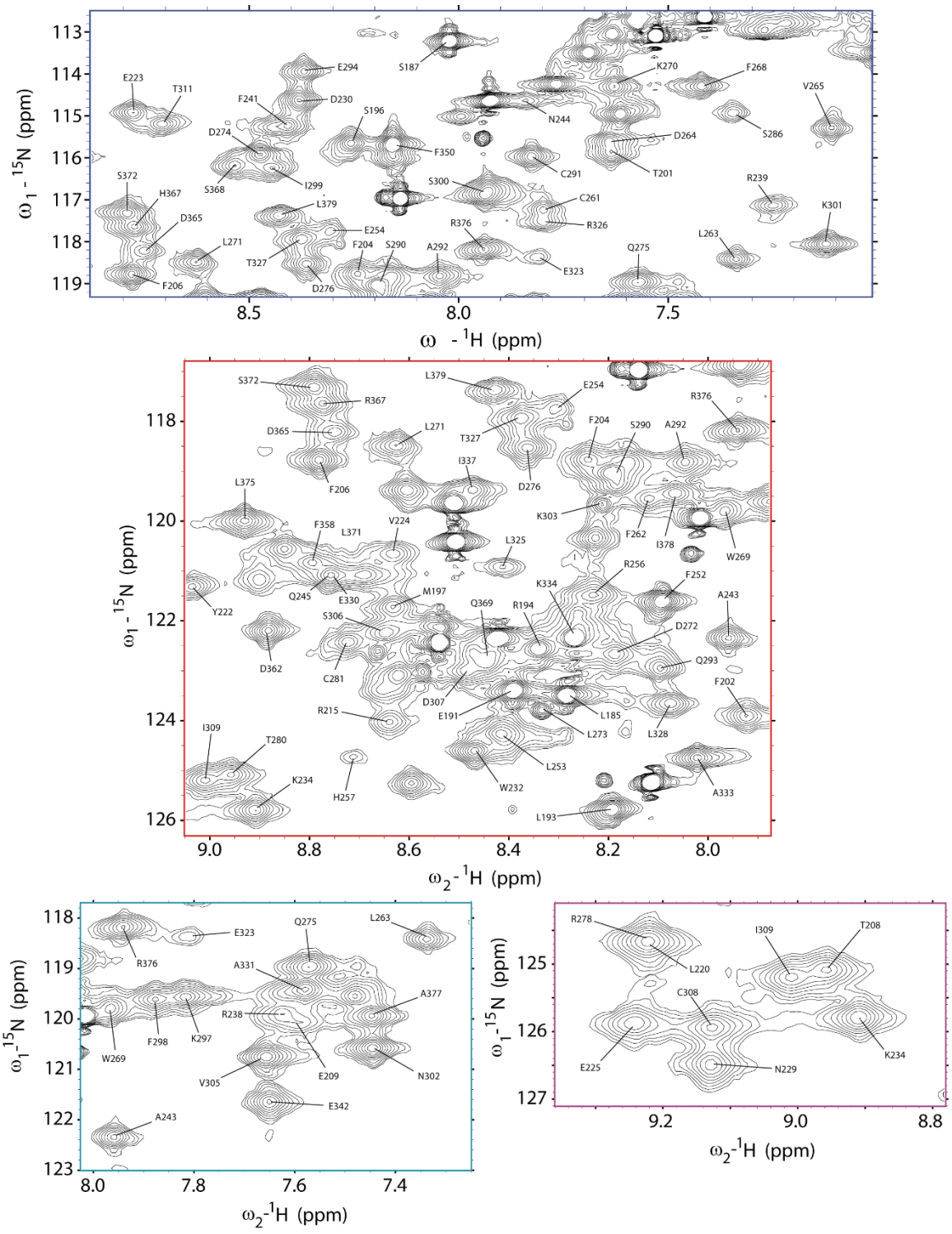

Fig. 3.2.b. Assignment of A3Gctd-2K3A E259A at pH 6.0. Inset locations indicated with colored boxes (**Fig. 3.2.a)** corresponding to expanded inset spectrum borders.

3.3.2 Single-stranded DNA binding of A3Gctd-2K3A undergoing active catalysis

A3Gctd-2K3A rapidly deaminates substrates that include the hotspot target sequence 5'-CCC motif. To capture A3Gctd-2K3A during active catalysis, we acquired spectra with a 20-fold excess of substrate ssDNA 5'-AATCCCAAA at pH 6.0 in a time window that ensured catalysis was actively proceeding. We confirmed that we were acquiring data during the appropriate time period by monitoring the 1D ^{1}H substrate spectrum at the end of the HSQC acquisition and verifying that non-deaminated substrate is still present (**Fig. 3.3**). Following the acquisition period the 3' cytidine was completely deaminated and approximately half of the middle cytidine had been deaminated resulting in a spectrum that contained a mixture of 5'-AATCCCAAA and 5'-AATCCUAAA binding.

A3Gctd 2K3A shows substantial perturbation upon ssDNA binding. The bound spectrum displayed both chemical shift perturbations and intensity changes. Significantly perturbed residues were W211NE, R215, T218, R239, L242, L253, E254, A258, D274, D276, R278, W285NE, E294, I314, G319, D365, L375, and H376. Significant intensity reductions were seen for residues S196, E209, W211NE, A243, N244, A246, F252, G255, R256, H257, L260, F262, L263, W269NE, D274, S284, W285, C288, A292, M295, F298, A312, I314, D316, L328, I337, M338, F350, and L375.

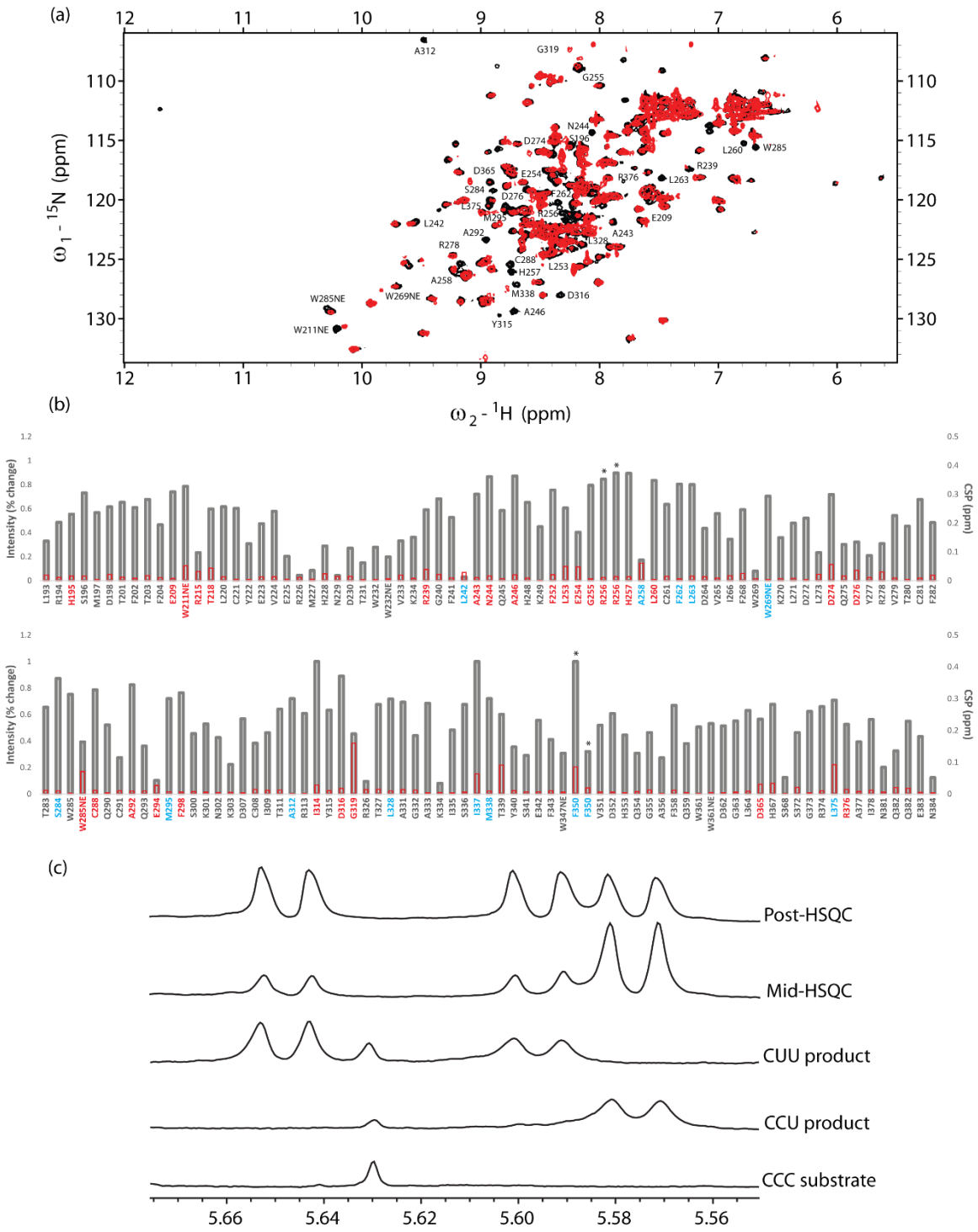

Fig. 3.3. A3Gctd-2K3A perturbations upon AATCCCAAA binding. (a) 1H-15N HSQC spectrum of 4mM 5'-AATCCCAAA : 0.2mM A3Gctd-2K3A pH 6.0 (red) overlaid onto A3Gctd-2K3A (black) at pH 6.0 (potential target cytidines underlined). Spectra were acquired with 8 scans and 64 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray

bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue. Asterisks indicated residues that have two distinct peaks at pH 6.0. (c) 1D 1H spectra of substrate and product oligomers (bottom three spectra) and 1D 1H spectra of 20:1 AATCCCAAA:A3Gctd-2K3A pH 6.0 at mid-HSQC and post-HSQC time points.

3.3.3 Substrate oligomer binding to A3Gctd-2K3A E259A

Not all target 5'-CC (deamination target cytidine underlined) sites are deaminated with equal efficiency. In order to determine binding changes that are responsible for these disparities we titrated A3Gctd-2K3A E259A with different substrate and non-substrate oligomers. We first bound A3Gctd-2K3A E259A with the 5'-AATCCCAAA (5'-CCC) substrate containing a centrally located 5'-CCC hotspot motif (**Fig. 3.4**). A3Gctd-2K3A E259A showed substantial perturbation upon saturation with the substrate. Perturbations were localized to the predicted binding surface of the protein with additional perturbations to the neighboring loops and helices along with core beta sheets of the protein indicative of structural rearrangement. Binding perturbed four primary regions of the protein. Binding region 1 (BR1) spans residues T201-L220, with significant perturbation of residues F202, F206, E209, W211NE, R215, and L220. BR1 includes residues in helix 1 spanning 198-208. Binding region 2 (BR2) spans residues R238-K270, with significant perturbation of residues R238, R239, L242, N244, Q245, and A246, in the non-continuous β2 sheet. Residues G251, E254, R256, H257 (Zn^{2+} coordinating His residue), and A258 in the loop leading to the catalytic site show extensive perturbation upon DNA binding. F262, L263, V265, and W269, all located in the α2 helix following the catalytic residue with their side chains directed into the core, display slow exchange

dynamics upon binding indicative of core structure rearrangement. W285 constitutes the bottom of the catalytic pocket [29], displays slow exchange dynamics and exhibits the single greatest change in chemical shift upon DNA binding. Binding region 3 (BR3) spans non-consecutive residues W285, T311-E330, and I335-T339 containing the previously identified substrate recognition loop T311-G319 [16], [74] and form the left side of the binding surface. Residues A312, R313, and Y315 display substantial perturbation and slow exchange dynamics. Residues in the α4 helix including E323, G324, L325, R326, T327, L328, and E330 are undergoing intermediate exchange dynamics with reductions in intensity greater than 60 percent. Residues F358, G363, L364, and R374 show reduced intensity and D365 displays extensive perturbation. D365 has been previously implicated in dimerization of C-terminal domains [72].

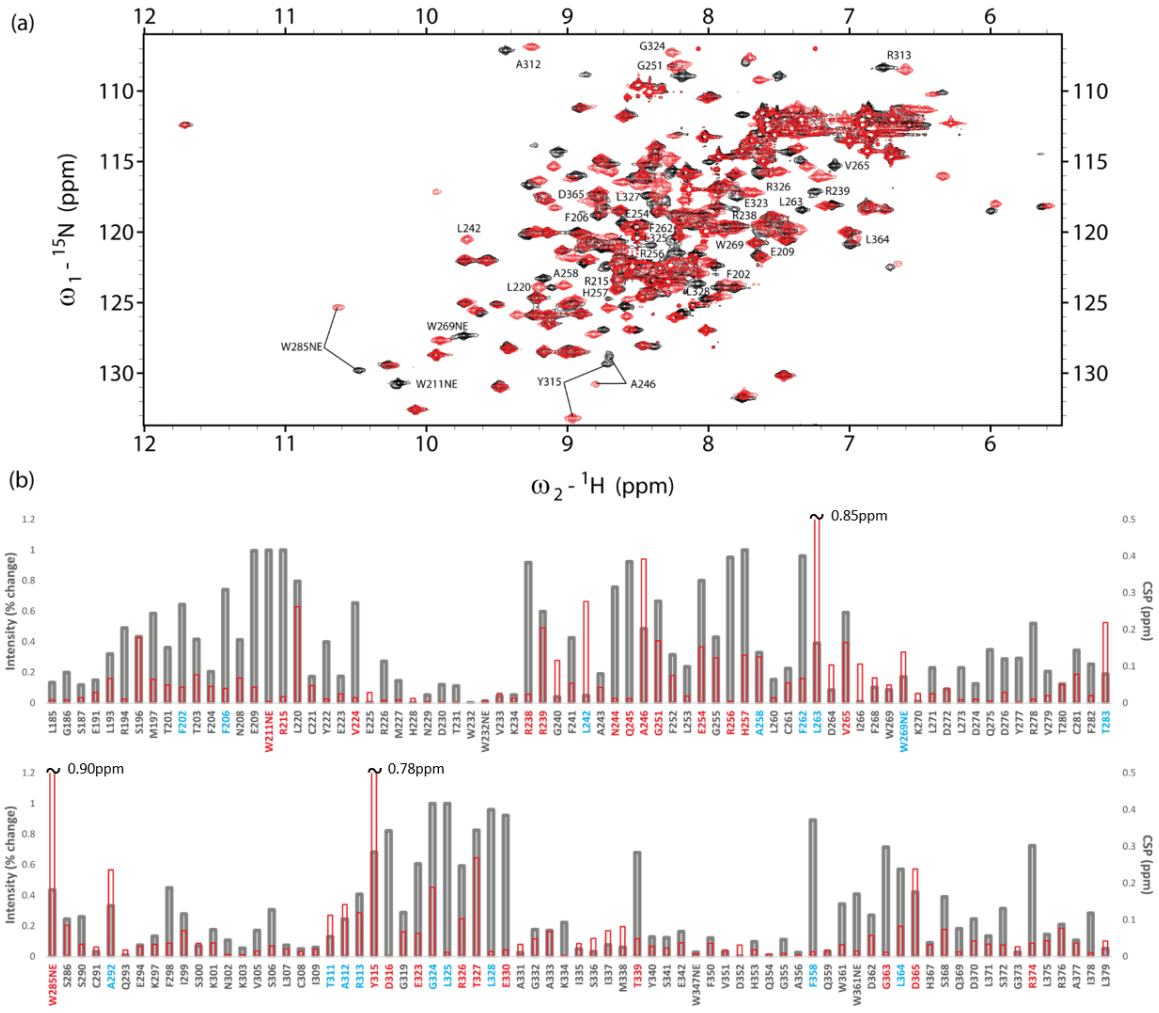

Fig. 3.4. A3Gctd-2K3A E259A binding to hotspot 5'-CCC motif. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCCCAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0 (target cytidine underlined). Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue.

We then bound A3Gctd-2K3A E259A with 5'-AATCCUAAA to determine the binding profile of a substrate containing a di-cytidine motif representative of the minimal A3G target (**Fig 3.5**). Utilizing the 5'-CCU target site also helps to elucidate the factors

that limit sequential deamination of the 5'-CCC target that has been observed by several groups [17], [64], [69], [89]. The 5'-AATCCUAAA (5'-CCU) substrate oligomer displays a similar binding profile to the 5'-AATCCCAAA substrate and engages all of the same binding regions. However, A3Gctd-2K3A E259A shows no slow exchange dynamics with the 5'-CCU substrate as was seen in the 5'-CCC substrate indicative of a difference in the binding kinetics between the two substrates. BR1 overall shows a limited changes in peak intensity with W211NE and R215 showing significant changes in intensity. BR2 displays significant intensity changes at residues R238, R239, G240, F241, N244, Q245, A246, G251, E254, G255, R256, H257, A258, F262, L263, D264, V265, and W269NE. BR3 displays significant intensity changes at residues W285NE, S286, A312, R313, Y315, D316, G319, E323, G324, L325, R326, T327, L328, E330, A333, M338, and T339. C-terminal residues D365, H367, and R376 also display significant intensity changes.

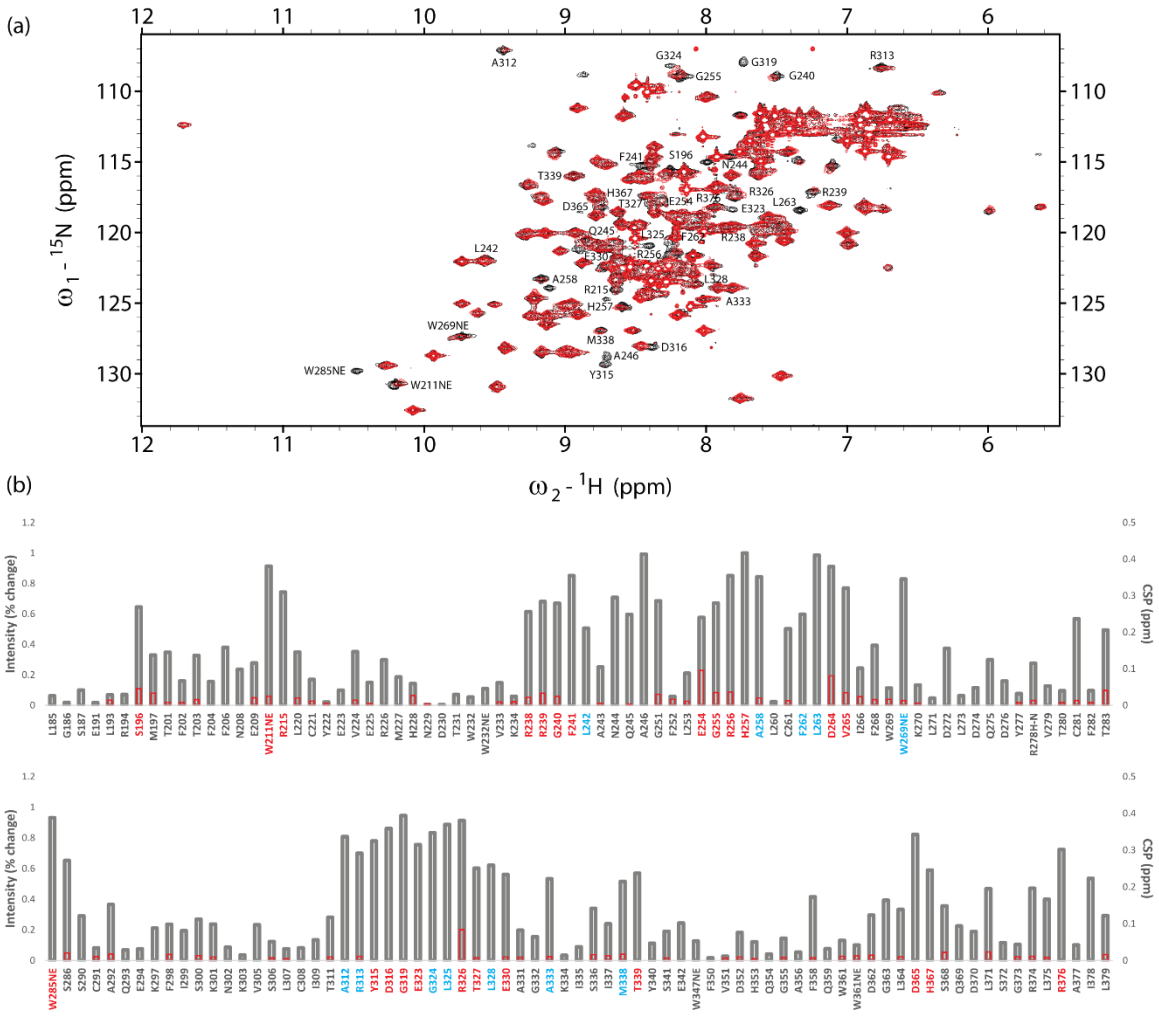

Fig. 3.5. A3Gctd-2K3A E259A binding to minimal di-cytidine target motif. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCCUAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0 (target cytidine underlined). Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue.

3.3.4 Non-substrate oligomer binding to A3Gctd-2K3A E259A

In order to determine non-specific DNA binding effects on A3Gctd-2K3A E259A we then bound the protein to a non-substrate oligomer 5'-AATCUUAAA (5'-CUU),

which lacks a 5’-CC target site (**Fig. 3.6**). The overall binding profile of the 5’-CUU oligomer shows no chemical shift perturbation upon binding and limited localized changes in peak intensities. Interaction with BR1 is almost completely abrogated with the exception of W211NE and R215. BR2 displays non-consecutive changes in peak intensity limited to N244, A246, E254, R256, H257, and L263. BR3 is almost completely unperturbed with the exception of D316. Addition of five-fold excess 5’-AAAAAAAAA oligomer to A3Gctd-2K3A E259A (data not shown) displayed the same reductions in intensity confirming that 5’-CUU behaves the same as ssDNA lacking cytidine.

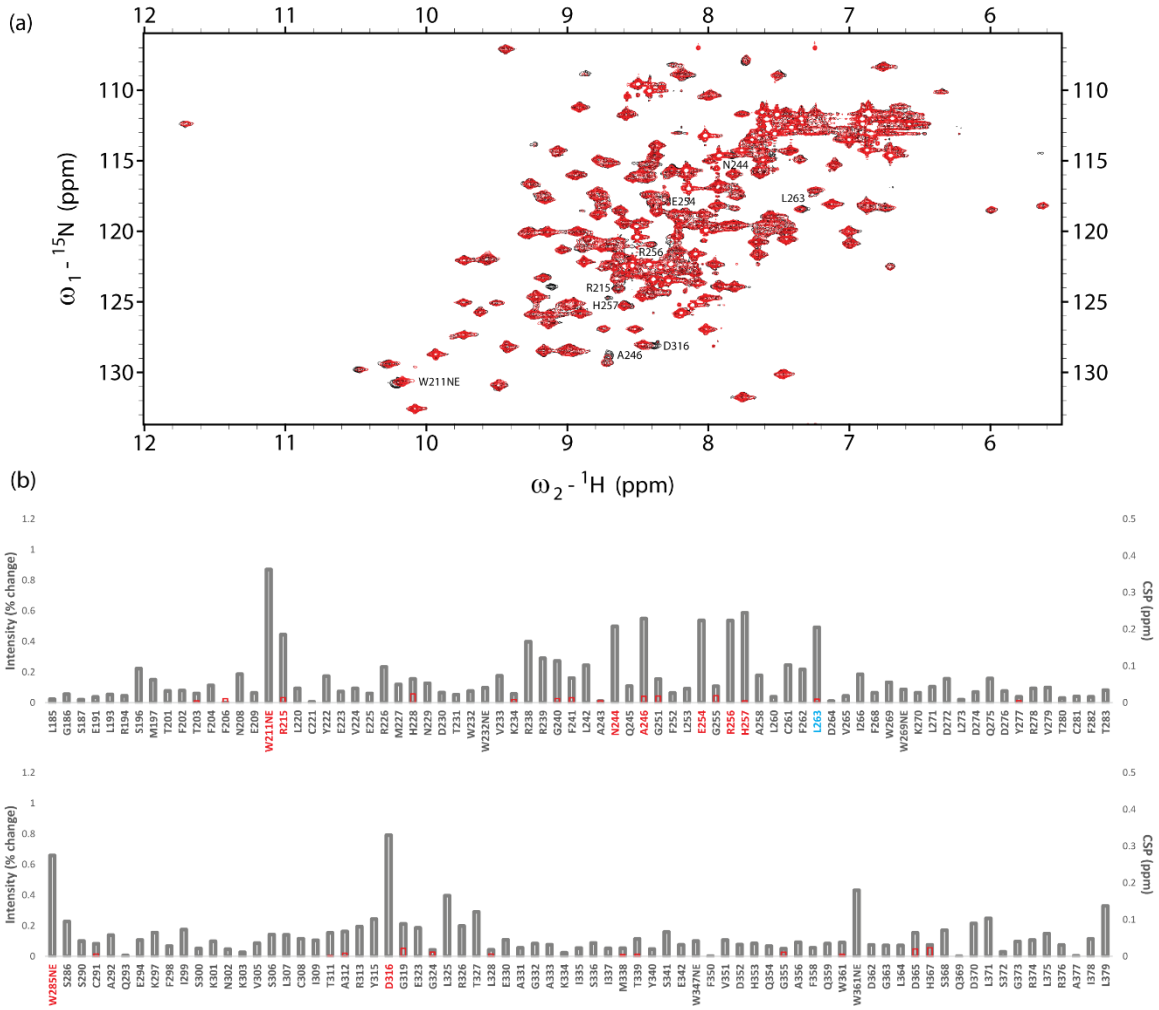

Fig. 3.6. A3Gctd-2K3A E259A binding to non-substrate oligomer. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCUUAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0. Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue.

3.3.5 Ribose containing oligomer binding to A3Gctd-2K3A E259A

The introduction of a single RNA base at the 3' position of the 5'-CCC hotspot motif has been shown to eliminate A3G catalytic activity on the ribose modified substrate [17], [65], [89]. In order to test the effects of this ribose substitution on the ssDNA binding of A3Gctd-2K3A E259A, we added excess 5'-ATTCCrCAATT (5'-CCrC) to the enzyme and observed the ^{15}N-HSQC spectrum of the protein (**Fig. 3.7**). BR1 interaction is limited to W211NE and R215. BR2 shows significant intensity changes at residues R238, N244, A246, E254, R256, H257, F262, and L263. Interestingly, residue Q275 which is outside of the established binding regions also exhibits a significant change in intensity. BR3 interaction is limited to W285NE, D316, G319, G324, and L325. Residues G355, D365, and R367 also show intensity changes consistent with those seen in other binding regions.

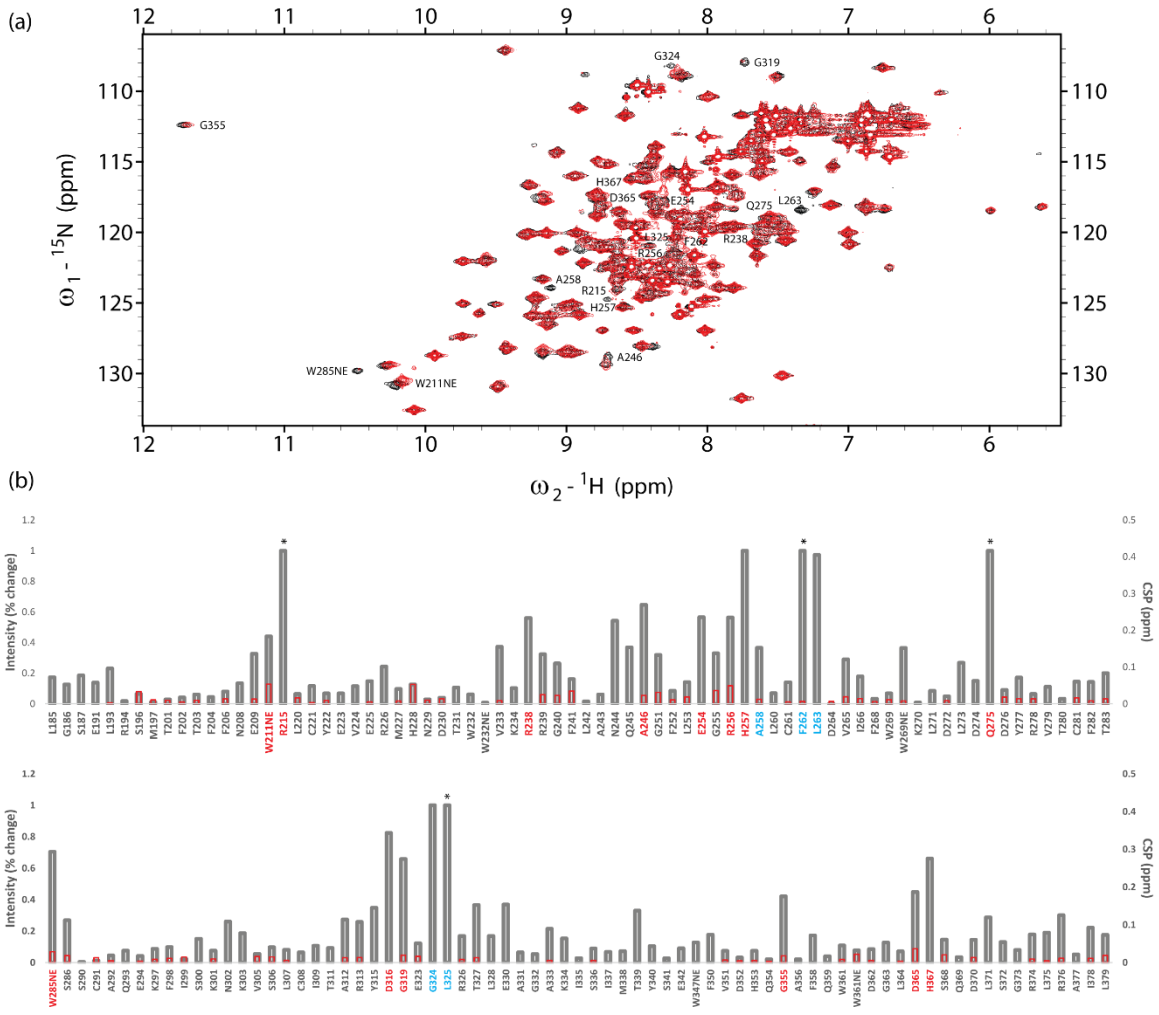

Fig. 3.7. A3Gctd-2K3A E259A binding to ribose target cytidine. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCCrCAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0. Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue. Asterisks indicate increases in peak intensity.

3.3.6 2’-F-RNA and 2’-F-ANA containing oligonucleotide binding to A3Gctd-2K3A E259A

Two potential mechanisms could be responsible for the exclusion of ribose bases from catalysis by A3G, the presence of the hydroxyl moiety at the C2’ position of the target sugar or the base conformation of the target sugar with ribose bases preferring the C3’-endo conformation and deoxyribose bases preferring the C2’-endo conformation [90]. To discriminate between these two possible mechanisms we assayed two fluorinated cytidine substrates, the first containing a fluorine substituted for the C2’ hydroxyl of the ribose (2’-deoxy-2’-fluororibonucleic acid, 2’-F-RNA) and the second containing an arabinose sugar with the C2’ hydroxyl substituted for fluorine (2’-deoxy-2’-fluoroarabonucleic acid, 2’-F-ANA) (**Fig. 3.8**).

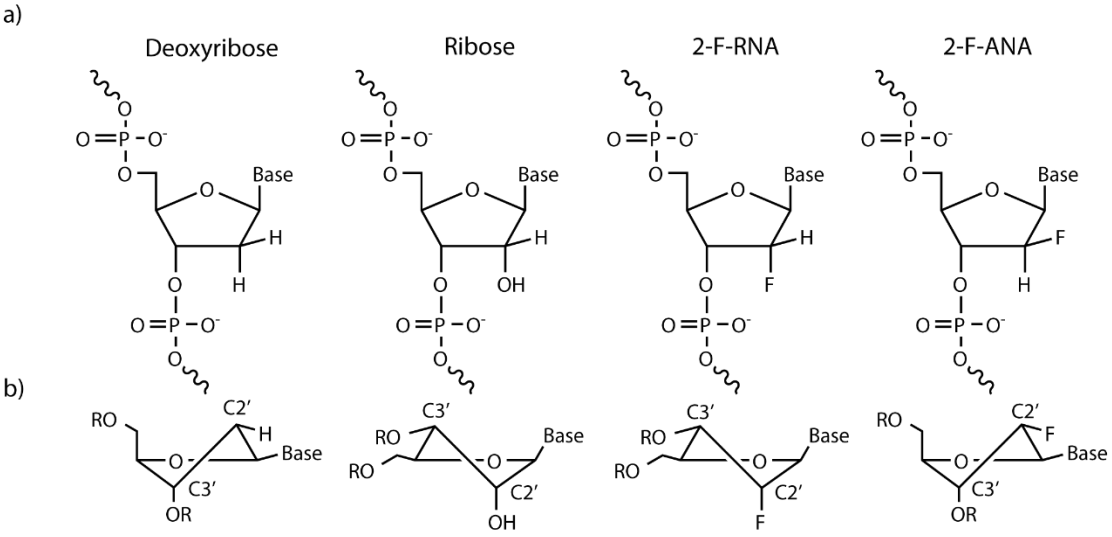

Fig. 3.8. Comparison of nucleotide sugar conformation. Varying the atom composition and stereochemistry at the C2’ position of the ribose ring significantly impacts the ribose ring conformation. (a) Stereochemistry at the C2’ position of the nucleotides used in this study. (b) Preferred conformations of the ribose ring containing specified substitutions in polynucleotide contexts. Exo-conformations indicated by vertical lines connecting

functional groups at the indicated stereo-centers. Endo-conformations indicated by horizontal lines connecting functional groups.

We assayed the binding of five-fold excess 2'-F-RNA containing substrate with A3Gctd-2K3A E259A (**Fig. 3.9**). Significant intensity changes were seen at residues R215, R238, N244, A246, L253, E254, R256, H257, L263, Q275, D276, W285NE, Y315, D316, G324, L325, T327, G355, and H367. The overall profile of the 2'-F-RNA containing oligomer displayed a binding profile intermediate to the RNA containing oligomer and the substrate 5'-CCU oligomer.

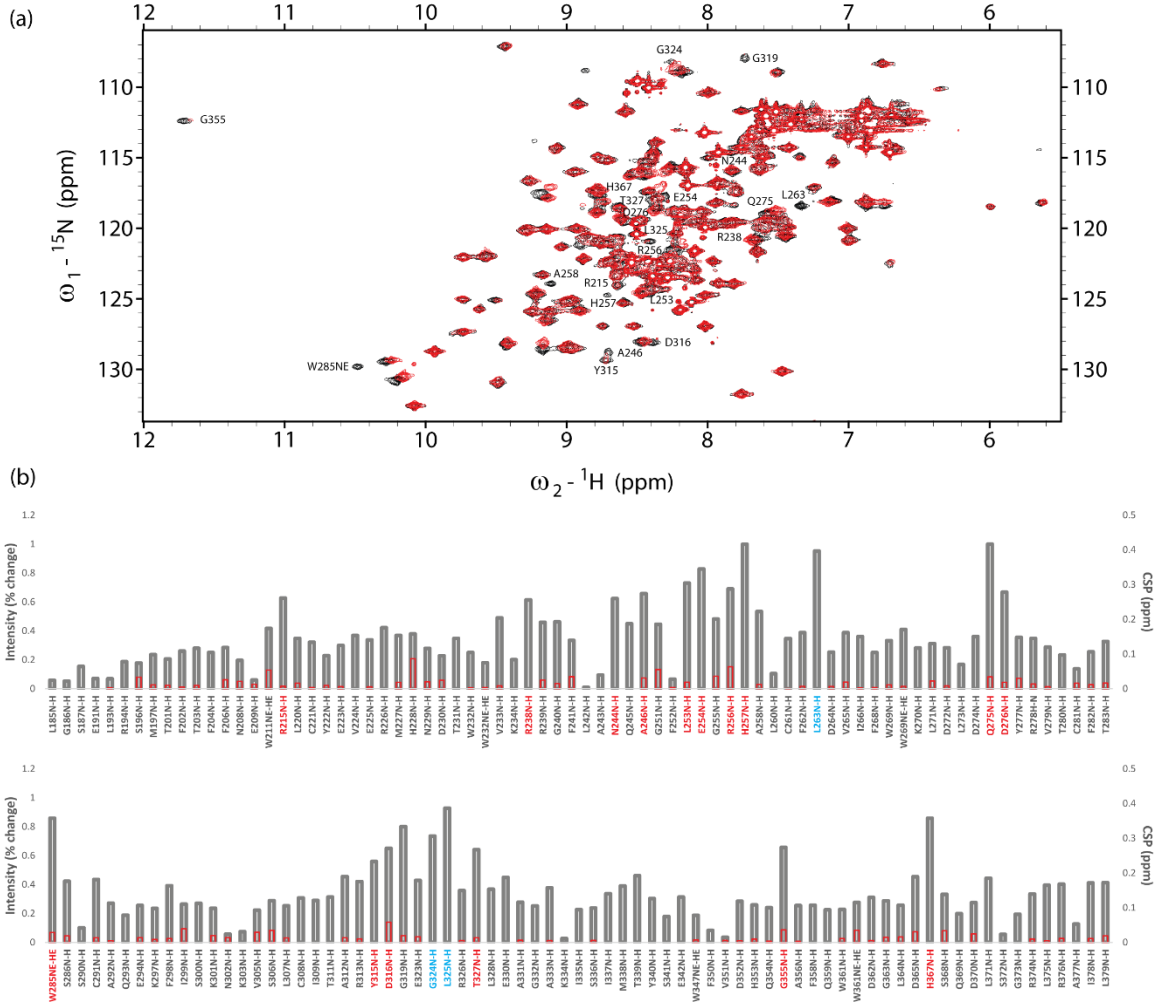

Fig. 3.9. A3Gctd-2K3A E259A binding fluorinated RNA mimic. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCC(2'-F-RNA)CAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0. Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue.

We then assayed the binding of five-fold excess 2'-F-ANA containing substrate with A3Gctd-2K3A E259A (**Fig. 3.10**). Binding of the 2'-F-ANA containing oligomer showed a combination of chemical shift perturbations and intensity changes similar to substrate oligomers with substantial reductions in intensity across the majority of the protein. Significant chemical shift perturbations were seen for residues H228, G251, G255, D264, W269NE, R313, and R326. Significant intensity changes were seen for residues N208, W211NE, R238, R239, G240, A246, L253, E254, R256, H257, A258, L263, D264, V265, W269NE, T283, W285NE, S306, A312, R313, Y315, G319, E323, L325, T327, L328, M338, T339, G355, D365, H367, and R376.

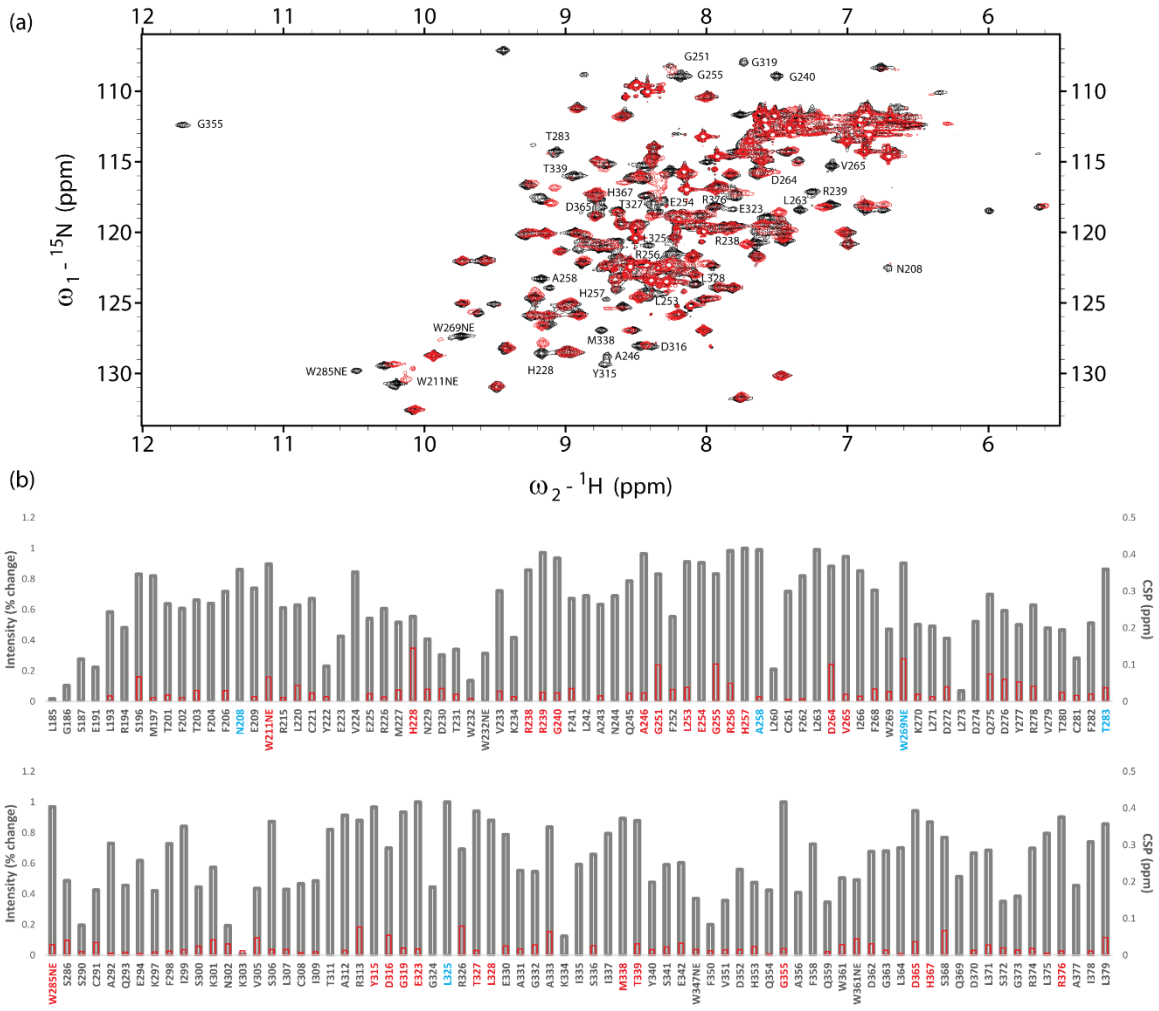

Fig. 3.10. A3Gctd-2K3A E259A binding to the 2'-F-ANA target cytidine substrate. (a) ^{1}H-^{15}N HSQC spectrum of 1mM 5'-AATCC(2'-F-ANA)CAAA : 0.2mM A3Gctd-2K3A E259A (red) overlaid onto 0.2mM A3Gctd-2K3A E259A (black) at pH 6.0. Spectra were acquired with 128 scans and 100 real points in the indirect dimension. Significantly shifted peaks labeled. (b) Quantification of peak intensity changes (gray bars, left axis) and chemical shifts (red bars, right axis). Significant changes in either peak intensity or chemical shift are indicated by colored residue names, cyan indicates core residue, red indicates surface residue.

3.3.7 Deamination of 2-F-RNA and 2-F-ANA containing substrates

Following confirmation of 2'-F-RNA containing oligomer binding to A3Gctd-2K3A E259A we then assayed the ability of the catalytically active A3Gctd-2K3A to

deaminate the potential substrate. We monitored deamination using real-time 1D ^{1}H NMR, as previously described in [17], with a 30-fold excess of the fluorinated substrate looking for the presence of the characteristic H5 signal of uracil between 5.50 and 5.75 ppm over the course of 8 hours (**Fig. 3.11c**). During the course of the experiment no uracil signals were observed confirming that the 2'-F-RNA cytidine base was effectively excluded from the catalytic site.

We then performed the same 1D ^{1}H deamination assay at 30-fold excess over A3Gctd-2K3A to determine if the 2'-F-ANA containing oligomer was a substrate. Over the course of 8 hours we observed the appearance of the H5 signal from the deaminated 2'-F-ANA uracil product 5'-ATTCC(2'-F-ANA)UAATT at 5.58ppm followed by the appearance of the deamination product of the middle cytidine in the 5'-ATTCU(2'-F-ANA)UAATT oligomer (**Fig. 3.11d**). The rate of deamination was dramatically reduced in the case of the 2'-F-ANA oligomer compared to the fully deoxyribose substrate (**Fig. 3.11a)**.

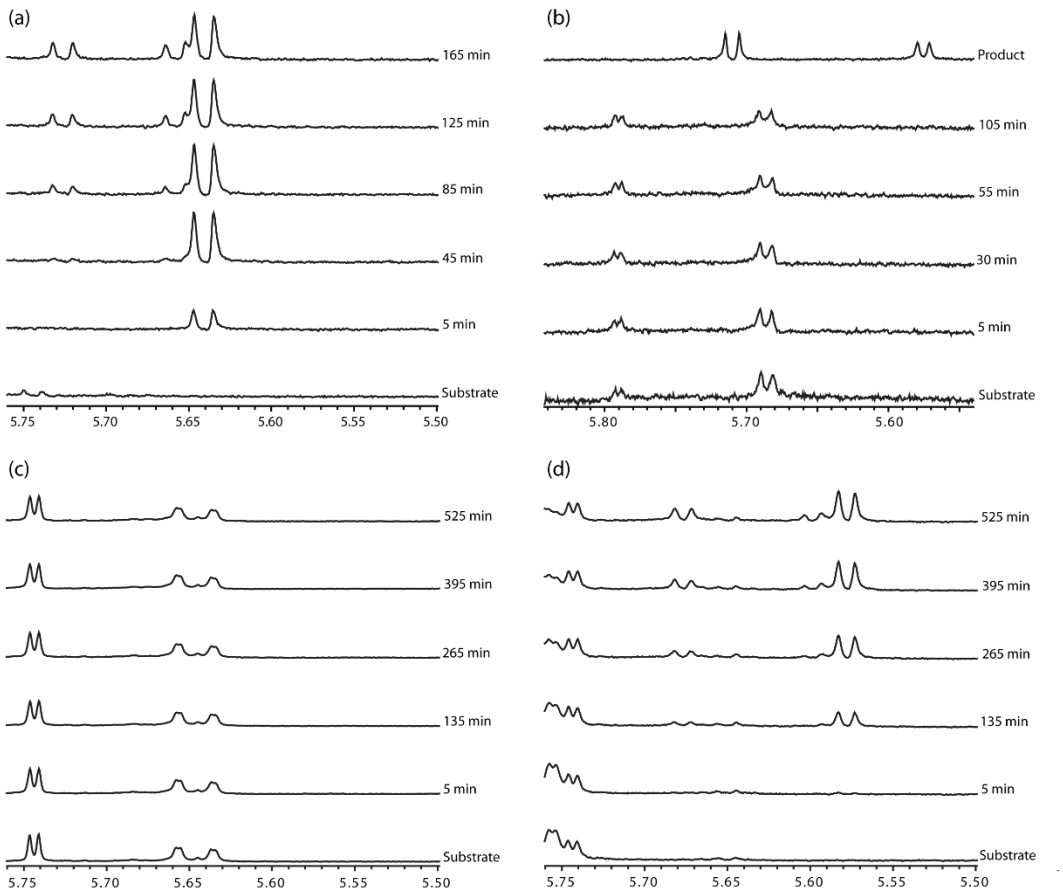

Fig. 3.11. Real-time NMR monitoring of substrate and potential substrate oligomer deamination. (a) 1D ^{1}H spectra series of 0.15mM 5'-ATTCCCAATT : 0.0015mM A3Gctd-2K3A (deaminated cytidines underlined). 5'-CCU product doublet appears at 5.64ppm, 5'-CUU product doublet appears at 5.735ppm with concurrent shifting of the 3' uracil doublet to 5.66ppm. (b) 1D ^{1}H spectra series of 0.15mM 5'-ATTCCrCAATT : 0.05mM A3Gctd-2K3A. Spectrum of 5'-ATTCCrUAATT product (top) for reference. No deamination products were observed. (c) 1D ^{1}H spectra series of 0.15mM 5'-ATTCCrCAATT : 0.03mM A3Gctd-2K3A. No shifts in full 1D ^{1}H spectrum were observed. (d) 1D ^{1}H spectra series of 0.15mM 5'-ATTCC(2'-F-ANA)CAATT : 0.03mM A3Gctd-2K3A (deaminated cytidines underlined). 5'-CC(2'-F-ANA)U product doublet observed at 5.58ppm, 5'-CU(2'-F-ANA)U product doublet appears at 5.68ppm with concurrent shifting of the 3' (2'-F-ANA) uracil to 5.60ppm.

3.4 Discussion

The C-terminal domain of APOBEC3G has been shown to deaminate cytidines in single stranded DNA with differing rates depending on target site context. APOBEC3G also possess the ability to selectively exclude ribonucleotides from deamination. The data presented in this study represent a comprehensive comparison of the structural interactions involved in discriminating between substrates as well as potential mechanisms for the rejection of ribonucleotides from catalytic binding. We have also presented the first data showing the full extent of ssDNA binding to both catalytically active A3Gctd-2K3A during active catalysis and inactive A3Gctd-2K3A E259A. The ssDNA binding interface of A3Gctd extends across the catalytic face of the protein engaging both the loops adjacent to the catalytic site as well as internal residues indicative of structural rearrangement upon binding. The structural rearrangement appears to require interactions distal from the active site in order to position the target base for catalysis consistent with APOBEC3G's inability to deaminate single nucleosides.

A3Gctd-2K3A shows extensive perturbation across the protein upon DNA binding when compared to the more localized interactions of A3Gctd-2K3A E259A. As we have shown previously [17], A3Gctd completes deamination of the available high affinity 5'-CCC targets before beginning to deaminate the low affinity 5'-CCU targets. The likely cause of the observed protein-wide perturbation is conformational flexing between bound and unbound states present as the protein binds the initial 5'-CCC substrate and ejecting the deaminated 5'-CCU product to search for another high affinity 5'-CCC substrate. This is consistent with A3G's lack of processive deamination within

target sites containing multiple 5'-CC minimal substrate motifs. While A3Gctd-2K3A shows increased protein-wide perturbations the most significantly perturbed residues remain consistent with those seen in substrate binding to A3Gctd-2K3A E259A.

The generation of A3Gctd-2K3A E259A allowed us to identify the differences in substrate binding among differing target cytidine contexts without the convoluting effects of product release and subsequent product binding. Binding of the 5'-CCC substrate showed the highest apparent affinity of the substrates assayed with the most significant chemical shift perturbations across the binding interface. Similarly to the binding of A3Gctd-2K3A, A3Gctd-2K3A E259A bound to 5'-CCC displayed perturbations outside of the ssDNA binding interface indicative of structural rearrangement of the α2 and α3 helices. Interestingly, perturbations are also present at the C-terminal end of the non-sheet region A234-R238 between the discontinuous β2 and β2' sheet implying rearrangement of this region that persists into binding region 2 (BR2) containing the loop from residues A243-H257. These interactions likely help position the target cytidine in the active site as these perturbations are only present in the 5'-CCC and 5'-CCU substrates that are actively deaminated. The binding of 5'-CUU shows slightly higher than average perturbations at these positions but that lack of a target cytidine limits the extent. The primary region that differs between substrate and non-substrate oligomers is binding region 1 (BR1), specifically the alpha1 helix (T201-N207) and the following loop (N208-T218). With the exceptions of W211NE and R215, which appear to be involved in all ssDNA interactions, these residues are perturbed extensively only in substrate oligomers. This is consistent with single alanine substitutions in this region preventing substrate deamination [29]. Perturbations of BR3 also appear to be intimately involved in the

discrimination between substrate and non-substrate oligomers. Residues W285NE and D316, similarly to W211NE and R215, are consistently perturbed upon all ssDNA binding. However, neighboring residues in binding region 3 (BR3) are only perturbed upon binding of target 5'-CC motifs. These perturbations extend to neighboring loop residues T311-G319 and also into the α4 helix. BR3 is structurally adjacent to BR1, further supporting this regions role in substrate discrimination. Interestingly, D365 which is conserved in 8 of the 11 APOBEC members [72] was only perturbed in oligomers containing a 5'-CC motif (including fluorinated substrates discussed later) indicating that this interaction is not exclusive to substrate binding but rather a more generalized di-cytidine recognition. Taken together, residue W211NE, R215, W285NE, and D316 constitute the minimal interaction interface required for non-specific ssDNA binding. Secondary interactions around these residues are responsible for high affinity ssDNA binding, substrate discrimination, and positioning of the target cytidine for catalysis.

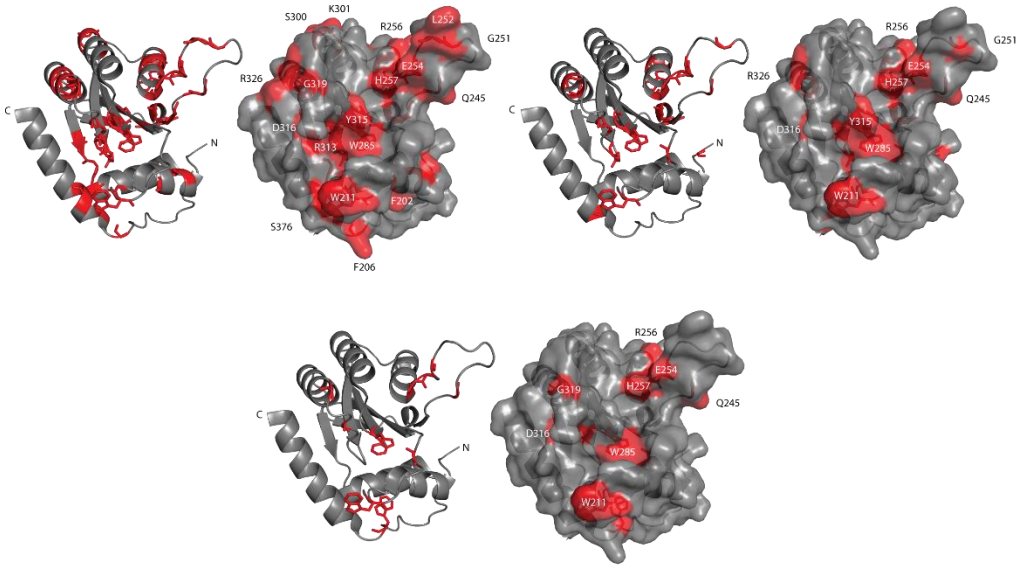

Fig. 3.12. Structural comparison of substrate and non-substrate binding interfaces. Significant perturbations (red) from substrate binding assays mapped onto A3Gctd (PDB

ID: 3V4K); 5’-AATCCCAAA (top left), 5’-AATCCUAAA (top right), and 5’-AATCUUAAA (bottom) with surface residues labeled.

APOBEC3G binds RNA to enable incorporation into budding HIV-1 virions [80], [95]–[97], yet excludes oligomers containing only a single RNA base at the 3’ target site from deamination [89]. To determine structural factors involved in DNA/RNA differentiation, we bound A3Gctd-2K3A E259A to oligomers containing RNA, 2’-F-RNA, and 2’-F-ANA bases. Binding of 5’-CCrC to A3Gctd-2K3A E259A resulted in a pattern of perturbations similar to that seen in the 5’-CUU oligomer. We observed a slightly more extensive pattern of interaction with BR3 than seen in the non-substrate oligomer which is likely due to interactions between this region and the 5’-CC motif present in the 5’-CCrC pseudo target motif. 5’-CCrC binding lacked the characteristic BR1 interactions seen in substrate oligomers indicating that the presence of the ribose base inhibits access to the catalytic pocket and prevents the oligomer from assuming full catalytic binding.

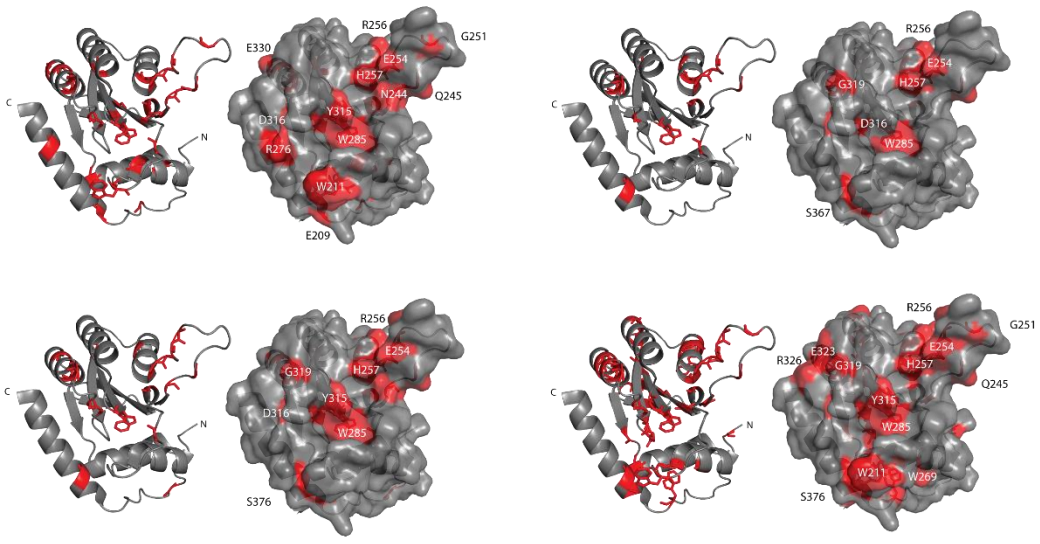

Fig. 3.13. Structural comparison of ribose and fluorine modified oligomer binding. Significant perturbations (red) from substrate binding assays mapped onto A3Gctd (PDB ID: 3V4K); 5'-ATTCCCATT (top left), 5'-ATTCCrCAATT (top right), 5'-ATTCC(2'-F-RNA)CAATT (bottom left), and 5'-ATTCC(2'-F-ANA)CAATT (bottom right) with surface residues labeled.

Fluorinated nucleosides have been shown to have anti-viral and anti-cancer properties (reviewed in [98]). Fluorine serves as an isopolar and isosteric mimic of the native hydroxyl moiety in ribonucleosides, retaining similar interbond distances and the ability to be a hydrogen bond acceptor. For the purposes of this study we utilized 2'-F-RNA and 2'-F-ANA to distinguish the mechanism of nucleotide exclusion from the catalytic site. The 5'-ATTCC(2'-F-RNA)CAATT oligomer retains the C3'-endo conformation of the un-substituted ribose base without being a hydrogen bond donor (**Fig. 3.8**). The 5'-ATTCC(2'-F-ANA)CAATT oligomer biases the base pucker towards the C2'-endo conformation typically seen in the ssDNA substrate while still retaining the strongly electronegative atom at the C2' position to mimic the presence of an oxygen (**Fig. 3.8**). We observed an overall increase in perturbation upon binding to both fluorinated oligomers indicative of an increase in non-specific binding interactions. Binding to the non-substrate 2'-F-RNA oligomer displayed perturbations similar to 5'-CCrC including minimal interaction with BR1. These results are consistent with the target 2'-F-RNA cytidine base assuming the C3'-endo conformation with little impact due to the loss of hydrogen bond donation as the mechanism of non-substrate gating. Comparatively, the substrate containing the 2-F'-ANA nucleotide engaged BR1 allowing for deamination of both the 2'-F-ANA cytidine as well as the subsequent deamination of middle deoxycytidine. While both of these substrates contain the highly electronegative

fluorine at their C2', the propensity for 2'-F-ANA to assume the C2'-endo conformation of the native ssDNA allows for extensive interaction with the secondary binding residues and catalytic binding. Kohli and co-workers have suggested the importance of the C2'-endo conformation for AID-catalyzed deamination using a substrate containing the 2'-F-ANA modification in the target cytidine [90], and our result strongly support extending this hypothesis to APOBEC3G as the 2'-F-ANA did not prevent the catalytic binding, but 2'-F-RNA did. Therefore, we concluded that the ribose ring pucker of the target cytidine either allows, C2'-endo conformation, or excludes, C3'-endo conformation, the base from entering the catalytic pocket and/or assuming the catalytically active orientation.

3.5 Methods summary

3.5.1 Substrates

Oligonucleotides containing standard DNA and RNA bases were synthesized by Integrated DNA Technologies. Oligonucleotides containing 2'-deoxy-2'-fluororibonucleic acid and 2'-deoxy-2'-fluoroarabonucleic acid at the underlined position in the 5'-ATTCCCAATT oligonucleotide were synthesized by Boston Open Labs.

3.5.2 Plasmid generation and protein purification

A pGEX-6p-1 vector (GE Healthcare) containing the C-terminal catalytic domain of APOBEC3G, residues 191-384, with the previously reported 2K3A mutations (L234K, C243A, F310K, C321A, C356A) was used as the template for Quikchange mutagenesis

(Stratagene/Agilent Technologies). The resulting plasmid contained the catalytic domain with the addition of the E259A mutation. *E. coli* were transformed with plasmids containing either A3Gctd-2K3A or A3Gctd-2K3A E259A, grown to OD 0.5 at 37°C followed by a reduction in temperature to 17°C for 30 minutes, and protein expression was induced using a 0.1mM final concentration of IPTG. The cells were lysed using sonication into buffer containing 50mM sodium phosphate pH 7.3, 100mM NaCl, 2mM DTT, 0.002% Tween 20. Following high speed centrifugation the supernatant was bound to glutathione sepharose resin (GenScript) and washed under high salt/high detergent conditions, 400mM NaCl and 0.06% Tween 20, followed by two washes in low salt/low detergent conditions, 30mM NaCl and 0.002% Tween 20. The GST-tag was removed using PreScission protease (GE Healthcare) in 50mM sodium phosphate buffer at pH 7.3 with 100mM NaCl, 2mM DTT, and 0.002% Tween 20. Following cleavage, the protein was dialyzed into sample buffer containing 50mM sodium phosphate pH 6.0, 100mM NaCl, 2mM DTT, 0.002% Tween 20, and 50µM $ZnCl_2$.

3.5.3 NMR assignment of A3Gctd-2K3A and A3Gctd-2K3A E259A

All spectra were acquired on an 850MHz Bruker Ascend spectrometer equipped with a 5mm Z-gradient TCI cryoprobe. Samples contained a final volume of 300µL (97% H_2O/3% D_2O, v/v), and spectra were taken at 293K. Backbone resonance assignments for the A3Gctd 2K3A E259A mutant were derived using TROSY versions of a standard set of triple resonance spectra (HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO, HN(CA)CO) on uniformly $^{15}N/^{13}C$ labeled protein with 85% random deuteration at pH

7.3. Assignments were transferred to pH 6.0 HSQC spectra by titrating pH to identify relevant peak shifts. All spectra were processed using NMRPipe [99] and analyzed using SPARKY [100].

3.5.4 Chemical shift perturbation analysis

^{15}N-HSQC substrate DNA titrations were collected on 0.2mM ^{15}N-labelled A3Gctd-2K3A or A3Gctd-2K3A E259A samples at pH 6.0 with unlabeled substrate ratios indicated. Each titration point was collected with 128 scans and 100 real data points in the indirect ^{15}N dimension. Chemical shift and intensity changes were monitored through a series of spectra at varying relative concentration ratios of 1:1, 1:2, 1:5 (A3Gctd-2K3A E259A:ssDNA). Chemical shift changes were calculated using the equation:

$$\Delta\delta_{ppm} = \sqrt{(\delta H_x - \delta H_0)^2 + \left(\frac{\delta N_x - \delta N_0}{5}\right)^2}$$

Intensity changes were calculated using the difference in peak height at the center of the ^{15}N-HSQC peak between the unbound and bound spectra divided by the unbound peak height.

Chapter 4

Summary and Future Directions

In this thesis, I have described a mechanism for the observed pH dependent effects on A3G's affinity for substrate DNA. I have identified the A3G single stranded DNA binding interface during active catalysis. Furthermore, I have identified specific differences in the interaction interface between different deamination targets, substrate and non-substrate oligomers. By thoroughly characterizing the optimal conditions and substrates for specific binding to A3G, I have helped establish a foundation for further structural studies. Future experiments will be concentrated around solving the complex structure. Since the A3Gctd-2K3A E259A mutant displays strong, specific binding to substrate DNA, it is a good candidate for co-crystallization with single stranded DNA substrates. Incorporating this catalytically inactive C-terminal domain into a full length protein containing the recently solubilized N-terminal domain chimera protein [27] could allow for the first full-length A3G structure in complex with ssDNA. Utilizing the E59A mutant combined with the loop swap mutants previously characterized [16], might allow for the examination of TC target site binding to identify the residues involved in excluding this substrate from catalysis by A3G.

I have also identified a novel substrate for A3G in 2'-F-ANA cytidine and identified the ribose sugar pucker as the likely mechanism for RNA exclusion. Based on these results, further experiments utilizing ribose bases that exhibit the C2'-endo conformation should be conducted. Specifically, the introduction of an arabinose sugar containing the native hydroxyl at C2' would solidify the sugar pucker hypothesis and exclude the hydroxyl gating hypothesis. Other non-native sugars, such as locked nucleic acids, might also yield interesting results and prove useful in formulating therapeutics targeting A3G.

References Cited

[1] B. Mangeat, P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono, "Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.," *Nature*, vol. 424, no. 6944, pp. 99–103, Jul. 2003.

[2] H. Zhang, B. Yang, R. J. Pomerantz, C. Zhang, S. C. Arunachalam, and L. Gao, "The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.," *Nature*, vol. 424, no. 6944, pp. 94–8, Jul. 2003.

[3] Q. Yu, R. König, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, and N. R. Landau, "Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome.," *Nat. Struct. Mol. Biol.*, vol. 11, no. 5, pp. 435–42, May 2004.

[4] R. S. Harris and M. T. Liddament, "Retroviral restriction by APOBEC proteins.," *Nat. Rev. Immunol.*, vol. 4, no. 11, pp. 868–77, Nov. 2004.

[5] E. P. Browne, C. Allers, and N. R. Landau, "Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent.," *Virology*, vol. 387, no. 2, pp. 313–21, May 2009.

[6] F. Guo, S. Cen, M. Niu, Y. Yang, R. J. Gorelick, and L. Kleiman, "The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA.," *J. Virol.*, vol. 81, no. 20, pp. 11322–31, Oct. 2007.

[7] R. S. Harris, S. K. Petersen-Mahrt, and M. S. Neuberger, "RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators," *Mol. Cell*, vol. 10, no. 5, pp. 1247–1253, 2002.

[8] A. Jarmuz, A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott, and N. Navaratnam, "An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.," *Genomics*, vol. 79, no. 3, pp. 285–296, 2002.

[9] N. K. Duggal, W. Fu, J. M. Akey, and M. Emerman, "Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations," *Virology*, vol. 443, no. 2, pp. 329–337, 2013.

[10] M. Henry, C. Terzian, M. Peeters, S. Wain-Hobson, and J. P. Vartanian, "Evolution of the primate APOBEC3A cytidine deaminase gene and identification of related coding regions," *PLoS One*, vol. 7, no. 1, pp. 1–7, 2012.

[11] R. S. LaRue, V. Andrésdóttir, Y. Blanchard, S. G. Conticello, D. Derse, M. Emerman, W. C. Greene, S. R. Jónsson, N. R. Landau, M. Löchelt, H. S. Malik, M. H. Malim, C. Münk, S. J. O'Brien, V. K. Pathak, K. Strebel, S. Wain-Hobson,

X.-F. Yu, N. Yuhki, and R. S. Harris, “Guidelines for naming nonprimate APOBEC3 genes and proteins.,” *J. Virol.*, vol. 83, no. 2, pp. 494–497, 2009.

[12] R. S. LaRue, S. R. Jónsson, K. a T. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hötzel, V. Andrésdóttir, T. P. L. Smith, and R. S. Harris, “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.*, vol. 9, p. 104, 2008.

[13] C. Münk, T. Beck, J. Zielonka, A. Hotz-Wagenblatt, S. Chareza, M. Battenberg, J. Thielebein, K. Cichutek, I. G. Bravo, S. J. O’Brien, M. Löchelt, and N. Yuhki, “Functions, structure, and read-through alternative splicing of feline APOBEC3 genes.,” *Genome Biol.*, vol. 9, no. 3, p. R48, 2008.

[14] S. L. Sawyer, M. Emerman, and H. S. Malik, “Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G,” *PLoS Biol.*, vol. 2, no. 9, 2004.

[15] J. Zhang and D. M. Webb, “Rapid evolution of primate antiviral enzyme APOBEC3G,” *Hum. Mol. Genet.*, vol. 13, no. 16, pp. 1785–1791, 2004.

[16] A. Rathore, M. a. Carpenter, Ö. Demir, T. Ikeda, M. Li, N. M. Shaban, E. K. Law, D. Anokhin, W. L. Brown, R. E. Amaro, and R. S. Harris, “The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution,” *J. Mol. Biol.*, vol. 425, no. 22, pp. 4442–4454, 2013.

[17] S. Harjes, W. C. Solomon, M. Li, K.-M. Chen, E. Harjes, R. S. Harris, and H. Matsuo, “Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G.,” *J. Virol.*, vol. 87, no. 12, pp. 7008–14, 2013.

[18] C. Prochnow, R. Bransteitter, M. G. Klein, M. F. Goodman, and X. S. Chen, “The APOBEC-2 crystal structure and functional implications for the deaminase AID.,” *Nature*, vol. 445, no. 7126, pp. 447–451, 2007.

[19] S. Hayashi, F., Umehara, T., Kigawa, T., Yokoyama, “Solution structure of the monomeric form of mouse APOBEC2,” *To be Publ.*

[20] I.-J. L. Byeon, J. Ahn, M. Mitra, C.-H. Byeon, K. Hercík, J. Hritz, L. M. Charlton, J. G. Levin, and A. M. Gronenborn, “NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity.,” *Nat. Commun.*, vol. 4, p. 1890, 2013.

[21] M.-F. Bohn, S. M. D. Shandilya, T. V. Silvas, E. A. Nalivaika, T. Kouno, B. A. Kelch, S. P. Ryder, N. Kurt-Yilmaz, M. Somasundaran, and C. A. Schiffer, “The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization,” *Structure*, vol. 23, no. 5, pp. 903–911, 2015.

[22] S. Kitamura, H. Ode, M. Nakashima, M. Imahashi, Y. Naganawa, T. Kurosawa, Y. Yokomaku, T. Yamane, N. Watanabe, A. Suzuki, W. Sugiura, and Y. Iwatani, "The APOBEC3C crystal structure and the interface for HIV-1 Vif binding," *Nature Structural & Molecular Biology*, vol. 19, no. 10. pp. 1005–1010, 2012.

[23] M. F. Bohn, S. M. D. Shandilya, J. S. Albin, T. Kouno, B. D. Anderson, R. M. McDougle, M. A. Carpenter, A. Rathore, L. Evans, A. N. Davis, J. Zhang, Y. Lu, M. Somasundaran, H. Matsuo, R. S. Harris, and C. A. Schiffer, "Crystal structure of the DNA cytosine deaminase APOBEC3F: The catalytically active and HIV-1 Vif-binding domain," *Structure*, vol. 21, no. 6, pp. 1042–1050, 2013.

[24] A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, and M. Katahira, "Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G.," *EMBO J.*, vol. 28, no. 4, pp. 440–451, 2009.

[25] E. Harjes, P. J. Gross, K.-M. Chen, Y. Lu, K. Shindo, R. Nowarski, J. D. Gross, M. Kotler, R. S. Harris, and H. Matsuo, "An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model.," *J. Mol. Biol.*, vol. 389, no. 5, pp. 819–32, Jun. 2009.

[26] M. Li, S. M. D. Shandilya, M. a Carpenter, A. Rathore, W. L. Brown, A. L. Perkins, D. a Harki, J. Solberg, D. J. Hook, K. K. Pandey, M. a Parniak, J. R. Johnson, N. J. Krogan, M. Somasundaran, A. Ali, C. a Schiffer, and R. S. Harris, "First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G.," *ACS Chem. Biol.*, vol. 7, no. 3, pp. 506–17, Mar. 2012.

[27] R. S. H. & H. M. Takahide Kouno, Elizabeth M Luengas, Megumi Shigematsu, Shivender M D Shandilya, JingYing Zhang, Luan Chen, Mayuko Hara, Celia A Schiffer, "Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G," *Nat. Struct. Mol. Biol.*, 2015.

[28] A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, and M. Katahira, "Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G.," *EMBO J.*, vol. 28, no. 4, pp. 440–51, Feb. 2009.

[29] K.-M. Chen, E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris, and H. Matsuo, "Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G.," *Nature*, vol. 452, no. 7183, pp. 116–9, Mar. 2008.

[30] L. G. Holden, C. Prochnow, Y. P. Chang, R. Bransteitter, L. Chelico, U. Sen, R. C. Stevens, M. F. Goodman, and X. S. Chen, "Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications.," *Nature*, vol. 456, no. 7218, pp. 121–4, Nov. 2008.

[31] H. C. Losey, A. J. Ruthenburg, and G. L. Verdine, “Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA.,” *Nat. Struct. Mol. Biol.*, vol. 13, no. 2, pp. 153–9, Feb. 2006.

[32] L. Chelico, P. Pham, P. Calabrese, and M. F. Goodman, “APOBEC3G DNA deaminase acts processively 3’ --> 5' on single-stranded DNA.,” *Nat. Struct. Mol. Biol.*, vol. 13, no. 5, pp. 392–9, May 2006.

[33] J. W. Rausch, L. Chelico, M. F. Goodman, and S. F. J. Le Grice, “Dissecting APOBEC3G substrate specificity by nucleoside analog interference.,” *J. Biol. Chem.*, vol. 284, no. 11, pp. 7047–58, Mar. 2009.

[34] R. M. Kohli, S. R. Abrams, K. S. Gajula, R. W. Maul, P. J. Gearhart, and J. T. Stivers, “A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase.,” *J. Biol. Chem.*, vol. 284, no. 34, pp. 22898–904, Aug. 2009.

[35] M. a Khan, R. Goila-Gaur, S. Opi, E. Miyagi, H. Takeuchi, S. Kao, and K. Strebel, “Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions.,” *Retrovirology*, vol. 4, p. 48, Jan. 2007.

[36] M. A. Khan, S. Kao, E. Miyagi, R. Goila-gaur, S. Opi, C. L. Gipson, T. G. Parslow, H. Ly, K. Strebel, and H. Takeuchi, “Viral RNA Is Required for the Association of APOBEC3G with Human Immunodeficiency Virus Type 1 Nucleoprotein Complexes Viral RNA Is Required for the Association of APOBEC3G with Human Immunodeficiency Virus Type 1 Nucleoprotein Complexes,” pp. 1–6, 2005.

[37] E. S. Svarovskaia, H. Xu, J. L. Mbisa, R. Barr, R. J. Gorelick, A. Ono, E. O. Freed, W.-S. Hu, and V. K. Pathak, “Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs.,” *J. Biol. Chem.*, vol. 279, no. 34, pp. 35822–8, Aug. 2004.

[38] V. B. Soros, W. Yonemoto, and W. C. Greene, “Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H.,” *PLoS Pathog.*, vol. 3, no. 2, p. e15, Feb. 2007.

[39] T. Wang, C. Tian, W. Zhang, K. Luo, P. T. N. Sarkis, L. Yu, B. Liu, Y. Yu, and X.-F. Yu, “7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G.,” *J. Virol.*, vol. 81, no. 23, pp. 13112–24, Dec. 2007.

[40] L. Chelico, E. J. Sacho, D. A. Erie, and M. F. Goodman, “A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV.,” *J. Biol. Chem.*, vol. 283, no. 20, pp. 13780–91, May 2008.

[41] L. Chelico, C. Prochnow, D. a Erie, X. S. Chen, and M. F. Goodman, “Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G.,” *J. Biol. Chem.*, vol. 285, no. 21, pp. 16195–205, May 2010.

[42] L. S. Shlyakhtenko, A. Y. Lushnikov, M. Li, L. Lackey, R. S. Harris, and Y. L. Lyubchenko, “Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G.,” *J. Biol. Chem.*, vol. 286, no. 5, pp. 3387–95, Feb. 2011.

[43] C. Prochnow, R. Bransteitter, M. G. Klein, M. F. Goodman, and X. S. Chen, “The APOBEC-2 crystal structure and functional implications for the deaminase AID.,” *Nature*, vol. 445, no. 7126, pp. 447–51, Jan. 2007.

[44] T. C. Krzysiak, J. Jung, J. Thompson, D. Baker, and A. M. Gronenborn, “APOBEC2 is a monomer in solution: implications for APOBEC3G models.,” *Biochemistry*, vol. 51, no. 9, pp. 2008–17, Mar. 2012.

[45] R. P. Bennett, J. D. Salter, X. Liu, J. E. Wedekind, and H. C. Smith, “APOBEC3G subunits self-associate via the C-terminal deaminase domain.,” *J. Biol. Chem.*, vol. 283, no. 48, pp. 33329–36, Nov. 2008.

[46] W. M. McDougall, C. Okany, and H. C. Smith, “Deaminase activity on single-stranded DNA (ssDNA) occurs in vitro when APOBEC3G cytidine deaminase forms homotetramers and higher-order complexes.,” *J. Biol. Chem.*, vol. 286, no. 35, pp. 30655–61, Sep. 2011.

[47] L. S. Shlyakhtenko, A. Y. Lushnikov, A. Miyagi, M. Li, R. S. Harris, and Y. L. Lyubchenko, “Nanoscale Structure and Dynamics of ABOBEC3G Complexes with Single-Stranded DNA.,” *Biochemistry*, Jul. 2012.

[48] R. Harris, K. Bishop, and A. Sheehy, “DNA deamination mediates innate immunity to retroviral infection,” *Cell*, vol. 113, pp. 803–809, 2003.

[49] H. Xu, E. Chertova, J. Chen, D. E. Ott, J. D. Roser, W.-S. Hu, and V. K. Pathak, “Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions.,” *Virology*, vol. 360, no. 2, pp. 247–56, Apr. 2007.

[50] A. E. Armitage, K. Deforche, C.-H. Chang, E. Wee, B. Kramer, J. J. Welch, J. Gerstoft, L. Fugger, A. McMichael, A. Rambaut, and A. K. N. Iversen, “APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete ‘all or nothing’ phenomenon.,” *PLoS Genet.*, vol. 8, no. 3, p. e1002550, Mar. 2012.

[51] G. Senavirathne, M. Jaszczur, P. a Auerbach, T. G. Upton, L. Chelico, M. F. Goodman, and D. Rueda, “Single-stranded DNA scanning and deamination by

APOBEC3G cytidine deaminase at single molecule resolution.," *J. Biol. Chem.*, vol. 287, no. 19, pp. 15826–35, May 2012.

[52] R. Nowarski, E. Britan-Rosich, T. Shiloach, and M. Kotler, "Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase.," *Nat. Struct. Mol. Biol.*, vol. 15, no. 10, pp. 1059–66, Oct. 2008.

[53] S. G. Conticello, "The AID/APOBEC family of nucleic acid mutators.," *Genome Biol.*, vol. 9, no. 6, p. 229, 2008.

[54] R. Goila-Gaur and K. Strebel, "HIV-1 Vif, APOBEC, and intrinsic immunity.," *Retrovirology*, vol. 5, p. 51, Jan. 2008.

[55] R. S. Harris, J. F. Hultquist, and D. T. Evans, "The restriction factors of human immunodeficiency virus," *J. Biol. Chem.*, vol. 287, no. 49, pp. 40875–40883, 2012.

[56] M. H. Malim and M. Emerman, "HIV-1 Accessory Proteins-Ensuring Viral Survival in a Hostile Environment," *Cell Host Microbe*, vol. 3, no. 6, pp. 388–398, 2008.

[57] J. L. Smith, W. Bu, R. C. Burdick, and V. K. Pathak, "Multiple ways of targeting APOBEC3-virion infectivity factor interactions for anti-HIV-1 drug development," *Trends Pharmacol. Sci.*, vol. 30, no. 12, pp. 638–646, 2009.

[58] G. Haché, M. T. Liddament, and R. S. Harris, "The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain," *J. Biol. Chem.*, vol. 280, no. 12, pp. 10920–10924, 2005.

[59] F. Navarro, B. Bollman, H. Chen, R. König, Q. Yu, K. Chiles, and N. R. Landau, "Complementary function of the two catalytic domains of APOBEC3G," *Virology*, vol. 333, no. 2, pp. 374–386, 2005.

[60] E. N. C. Newman, R. K. Holmes, H. M. Craig, K. C. Klein, J. R. Lingappa, M. H. Malim, and A. M. Sheehy, "Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity," *Curr. Biol.*, vol. 15, no. 2, pp. 166–170, 2005.

[61] L. Chelico, P. Pham, P. Calabrese, and M. F. Goodman, "APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA.," *Nat. Struct. Mol. Biol.*, vol. 13, no. 5, pp. 392–399, 2006.

[62] Y. Iwatani, H. Takeuchi, K. Strebel, and J. G. Levin, "Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect.," *J. Virol.*, vol. 80, no. 12, pp. 5992–6002, 2006.

[63] H. Xu, E. S. Svarovskaia, R. Barr, Y. Zhang, M. a Khan, K. Strebel, and V. K. Pathak, "A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 15, pp. 5652–7, Apr. 2004.

[64] A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, and M. Katahira, "Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G.," *EMBO J.*, vol. 28, no. 4, pp. 440–451, 2009.

[65] J. W. Rausch, L. Chelico, M. F. Goodman, and S. F. J. Le Grice, "Dissecting APOBEC3G substrate specificity by nucleoside analog interference," *J. Biol. Chem.*, vol. 284, no. 11, pp. 7047–7058, 2009.

[66] L. Chelico, E. J. Sacho, D. a. Erie, and M. F. Goodman, "A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV," *J. Biol. Chem.*, vol. 283, no. 20, pp. 13780–13791, 2008.

[67] R. Eisenthal, M. J. Danson, and D. W. Hough, "Catalytic efficiency and kcat/KM: a useful comparator?," *Trends Biotechnol.*, vol. 25, no. 6, pp. 247–249, 2007.

[68] J. Chiba, T. Kouno, S. Aoki, H. Sato, J. Zhang, H. Matsuo, and M. Inouye, "Electrochemical direct detection of DNA deamination catalyzed by APOBEC3G," *Chem. Commun.*, pp. 12115–12117, 2012.

[69] L. S. Shlyakhtenko, A. Y. Lushnikov, M. Li, L. Lackey, R. S. Harris, and Y. L. Lyubchenko, "Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G," *J. Biol. Chem.*, vol. 286, no. 5, pp. 3387–3395, 2011.

[70] E. Miyagi, S. Opi, H. Takeuchi, M. Khan, R. Goila-Gaur, S. Kao, and K. Strebel, "Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1.," *J. Virol.*, vol. 81, no. 24, pp. 13346–13353, 2007.

[71] D. C. Carlow, a a Smith, C. C. Yang, S. a Short, and R. Wolfenden, "Major contribution of a carboxymethyl group to transition-state stabilization by cytidine deaminase: mutation and rescue.," *Biochemistry*, vol. 34, no. 13, pp. 4220–4224, 1995.

[72] S. M. D. Shandilya, M. N. L. Nalam, E. a Nalivaika, P. J. Gross, J. C. Valesano, K. Shindo, M. Li, M. Munson, W. E. Royer, E. Harjes, T. Kono, H. Matsuo, R. S. Harris, M. Somasundaran, and C. a Schiffer, "Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces.," *Structure*, vol. 18, no. 1, pp. 28–38, Jan. 2010.

[73] K. Shindo, M. Li, P. J. Gross, W. L. Brown, E. Harjes, Y. Lu, H. Matsuo, and R. S. Harris, "A Comparison of Two Single-Stranded DNA Binding Models by Mutational Analysis of APOBEC3G," *Biology (Basel).*, vol. 1, no. 2, pp. 260–276, 2012.

[74] R. M. Kohli, S. R. Abrams, K. S. Gajula, R. W. Maul, P. J. Gearhart, and J. T. Stivers, "A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase," *J. Biol. Chem.*, vol. 284, no. 34, pp. 22898–22904, 2009.

[75] D. C. Carlow, S. a. Short, and R. Wolfenden, "Complementary truncations of a hydrogen bond to ribose involved in transition-state stabilization by cytidine deaminase," *Biochemistry*, vol. 37, no. 5, pp. 1199–1203, 1998.

[76] E. a. Véliz, L. M. Easterwood, and P. a. Beal, "Substrate analogues for an RNA-editing adenosine deaminase: Mechanistic investigation and inhibitor design," *J. Am. Chem. Soc.*, vol. 125, no. 36, pp. 10867–10876, 2003.

[77] T. Melcher, S. Maas, a Herb, R. Sprengel, P. H. Seeburg, and M. Higuchi, "A mammalian RNA editing enzyme.," *Nature*, vol. 379, no. 6564. pp. 460–464, 1996.

[78] M. a. O'Connell, A. Gerber, and W. Keller, "Purification of human double-stranded RNA-specific editase 1 (HRED1) involved in editing of brain glutamate receptor B pre-mRNA," *J. Biol. Chem.*, vol. 272, no. 1, pp. 473–478, 1997.

[79] Y. Iwatani, D. S. B. Chan, F. Wang, K. S. Maynard, W. Sugiura, A. M. Gronenborn, I. Rouzina, M. C. Williams, K. Musier-Forsyth, and J. G. Levin, "Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G," *Nucleic Acids Res.*, vol. 35, no. 21, pp. 7096–7108, 2007.

[80] A. Burnett and P. Spearman, "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.," *J. Virol.*, vol. 81, no. 10, pp. 5000–5013, 2007.

[81] Y.-L. Chiu, V. B. Soros, J. F. Kreisberg, K. Stopak, W. Yonemoto, and W. C. Greene, "Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells.," *Nature*, vol. 435, no. 7038, pp. 108–114, 2005.

[82] H. Huthoff, F. Autore, S. Gallois-Montbrun, F. Fraternali, and M. H. Malim, "RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1.," *PLoS Pathog.*, vol. 5, no. 3, p. e1000330, Mar. 2009.

[83] L. S. Shlyakhtenko, A. Y. Lushnikov, A. Miyagi, M. Li, R. S. Harris, and Y. L. Lyubchenko, "Nanoscale structure and dynamics of ABOBEC3G complexes with single-stranded DNA," *Biochemistry*, vol. 51, no. 32, pp. 6432–6440, 2012.

[84] J. E. Wedekind, R. Gillilan, A. Janda, J. Krucinska, J. D. Salter, R. P. Bennett, J. Raina, and H. C. Smith, "Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits," *J. Biol. Chem.*, vol. 281, no. 50, pp. 38122–38126, 2006.

[85] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imageJ," *Biophotonics Int.*, vol. 11, no. 7, pp. 36–41, 2004.

[86] M. D. Stenglein, M. B. Burns, M. Li, J. Lengyel, and R. S. Harris, "APOBEC3 proteins mediate the clearance of foreign DNA from human cells.," *Nat. Struct. Mol. Biol.*, vol. 17, no. 2, pp. 222–229, 2010.

[87] B. K. Thielen, K. C. Klein, L. W. Walker, M. Rieck, J. H. Buckner, G. W. Tomblingson, and J. R. Lingappa, "T cells contain an RNase-insensitive inhibitor of APOBEC3G deaminase activity," *PLoS Pathog.*, vol. 3, no. 9, pp. 1320–1334, 2007.

[88] A. J. Schumacher, G. Haché, D. a Macduff, W. L. Brown, and R. S. Harris, "The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction.," *J. Virol.*, vol. 82, no. 6, pp. 2652–2660, 2008.

[89] K. Kamba, T. Nagata, and M. Katahira, "Catalytic Analysis of APOBEC3G Involving Real-Time NMR Spectroscopy Reveals Nucleic Acid Determinants for Deamination," *PLoS One*, vol. 10, no. 4, p. e0124142, 2015.

[90] R. M. Kohli, R. W. Maul, A. F. Guminski, R. L. McClure, K. S. Gajula, H. Saribasak, M. a. McMahon, R. F. Siliciano, P. J. Gearhart, and J. T. Stivers, "Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification," *J. Biol. Chem.*, vol. 285, no. 52, pp. 40956–40964, 2010.

[91] J. J. King, C. A. Manuel, C. V. Barrett, S. Raber, H. Lucas, P. Sutter, and M. Larijani, "Catalytic Pocket Inaccessibility of Activation-Induced Cytidine Deaminase Is a Safeguard against Excessive Mutagenic Activity," *Structure*, vol. 23, no. 4, pp. 615–627, 2015.

[92] S. Gallois-Montbrun, B. Kramer, C. M. Swanson, H. Byers, S. Lynham, M. Ward, and M. H. Malim, "Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules.," *J. Virol.*, vol. 81, no. 5, pp. 2165–2178, 2007.

[93] L. Lackey, E. K. Law, W. L. Brown, and R. S. Harris, “Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination,” *Cell Cycle*, vol. 12, no. 5, pp. 762–772, 2013.

[94] C. S. Nabel, J. W. Lee, L. C. Wang, and R. M. Kohli, “Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 35, pp. 14225–30, 2013.

[95] M. Douaisi, S. Dussart, M. Courcoul, G. Bessou, R. Vigne, and E. Decroly, “HIV-1 and MLV Gag proteins are sufficient to recruit APOBEC3G into virus-like particles,” *Biochem. Biophys. Res. Commun.*, vol. 321, no. 3, pp. 566–573, 2004.

[96] E. S. Svarovskaia, H. Xu, J. L. Mbisa, R. Barr, R. J. Gorelick, A. Ono, E. O. Freed, W. S. Hu, and V. K. Pathak, “Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs,” *J. Biol. Chem.*, vol. 279, no. 34, pp. 35822–35828, 2004.

[97] V. Zennou, D. Perez-caballero, and P. D. Bieniasz, “APOBEC3G Incorporation into Human Immunodeficiency Virus Type 1 Particles APOBEC3G Incorporation into Human Immunodeficiency Virus Type 1 Particles,” *J. Virol.*, vol. 78, no. 21, pp. 12058–12061, 2004.

[98] P. Liu, A. Sharon, and C. K. Chu, “Fluorinated nucleosides: Synthesis and biological implication,” *J. Fluor. Chem.*, vol. 129, no. 9, pp. 743–766, 2008.

[99] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and a Bax, “NMRPipe: a multidimensional spectral processing system based on UNIX pipes.,” *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–93, Nov. 1995.

[100] T. D. G. and D. G. Kneller, “SPARKY 3.” University of California, San Francisco.