

**Insights into the length- and location-dependent
deaminase activities of APOBEC3B/F and the
deaminase activity determinants of APOBEC3F**

Wan Li

Table of contents

Chapter 1 General introduction

1.1 Energy and enzymes	1
1.2 HIV and cancers	1
1.3 APOBEC3 proteins	2
1.4 A3 proteins and antiviral activity	4
1.5 A3 proteins and cancers	5
1.6 Structural characteristics of A3 proteins	6
1.7 Binding interface of A3s with oligonucleotides	8
1.8 A3s' movement on single-stranded DNA	12
1.9 Objectives of this study	13

Chapter 2 Observation by real-time NMR, and interpretation of length- and location-dependent deamination activity of APOBEC3B

2.1 Introduction	15
2.2 Results	15
2.2.1 A3B-CTD catalysis depends on the lengths of ssDNA substrates	15
2.2.2 A3B-CTD interacts with ssDNA with weak binding affinity	17
2.2.3 A3B-CTD catalysis depends on the location of the target sequence on ssDNA	18
2.3 Discussion	22
2.4 Conclusions	27
2.5 Experimental procedures	28
2.5.1 Oligonucleotides and Plasmids	28
2.5.2 Proteins expression and purification	28
2.5.3 <i>In-vitro</i> cytosine deamination assay	29
2.5.3 Annealing oligonucleotides to ssDNA	30
2.5.4 Real-time NMR spectroscopy	31
2.5.6 Electrophoretic mobility shift assay	31
2.5.7 Steady-state fluorescence anisotropy measurement	32

Chapter 3 Influence of the DNA sequence and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F

3.1 Introduction	33
3.2 Results.....	33
3.2.1 Preparation of an active A3F-CTD	33
3.2.2 Sequence context-dependent deaminase activity.....	34
3.2.3 Roles of amino acid residues in loop 1	35
3.2.4 Roles of amino acid residues in loop 3	38
3.2.5 Roles of amino acid residues in or near loop 7.....	39
3.2.6 Deaminase activity of phosphomimetic A3F-CTD mutant.....	41
3.2.7 The effect of pH on A3F-CTD catalysis	42
3.3 Discussion	44
3.4 Conclusions	47
3.5 Experimental procedures	48
3.5.1 Oligonucleotides and plasmids	48
3.5.2 Protein expression and purification.....	49
3.5.3 <i>In-vitro</i> cytosine deamination assay.....	50
3.5.4 Steady-state fluorescence anisotropy measurement.....	50
3.5.5 Circular dichroism (CD) spectroscopy.....	51

Chapter 4 Influence of the ssDNA concentration on the length-dependent deaminase activity of APOBEC3F

4.1 Introduction	52
4.2 Results.....	52
4.2.1 The ssDNA-length-dependent deaminase activity of A3F-CTD.....	52
4.2.2 N214H mutant does not affect the length-dependent deaminase activity of A3F-CTD	53
4.2.3 The concentration of ssDNA affects the length-dependent deamination by A3F	55
4.3 Discussion	55
4.4 Conclusions	57
4.5 Experimental procedures	57
4.5.1 Oligonucleotides	57

4.5.2 Proteins expression and purification	57
4.5.3 <i>In-vitro</i> cytosine deamination assay.....	58
4.5.4 Real-time NMR spectroscopy.....	58
4.5.5 Steady-state fluorescence anisotropy measurement.....	58
 Chapter 5 The inhibition of the deaminase activity of A3F-CTD in presence of Vif complex	
5.1 Introduction	59
5.2 Results.....	60
5.2.1 Purification of Vif complex	60
5.2.2 Vif complex inhibits the deaminase activity of A3F-CTD	60
5.3 Discussion	61
5.4 Conclusions	62
5.5 Experimental procedures	62
5.5.1 Expression and purification of Vif complex.....	62
5.5.2 <i>In-vitro</i> cytosine deamination assay.....	63
 Chapter 6	
General conclusions.....	64
 References.....	66
 List of publications	78
List of presentations	79
List of honors.....	80
Acknowledgements.....	81

Chapter 1

General Introduction

1.1 Energy and enzymes

Energy takes in many different forms, such as light, heat, wind, and electricity. Here, we focus on another kind of energy, bioenergy. Generally, bioenergy can be used to produce heat, transport fuels, *et al.*[1]. In the human body, kinetic and thermal energy provides us with enough power to perform social activities, in other words, it can be defined as the fundamental resource for work. Bioenergy arises from food, water, sunshine *et al.*, during the process of conversion of the chemical energy from these resources into bioenergy, enzymes being the essential proteins that catalyze this conversion [2, 3].

The factors that affect enzymatic efficiency mainly result from the structures of enzymes and the binding affinities with their substrates [3]. Therefore, active enzymes and the binding interfaces between enzymes and their substrates are determinants of enzymatic efficiency. Once the structures of enzymes are destroyed or the binding affinity is affected by other factors, the catalytic activities of the enzymes will also be affected. Only the active structure and proper binding affinity can give effective catalytic activity.

1.2 HIV and cancers

In the world, nearly all people are fearful when they hear the word "HIV (human immunodeficiency virus)" or "cancer". Since the beginning of the epidemic, more than 70 million people have been infected with the HIV virus and about 35 million people have died of HIV. In 2015, about 36.7 million people (all ages) were living with HIV and 2.1 million (all ages) people were nearly HIV infected [4]. As for cancer, according to the statistical results of the World Health Organization, there were 8.8 million people worldwide who die from various cancers in 2015, which is nearly 1 in 6 of all global deaths. Why are the death rates caused by HIV and cancer so high? What are HIV and cancer?

HIV is a lentivirus that infects human immune cells and causes acquired immunodeficiency syndrome (AIDS) [5]. There are two subtypes of HIV, HIV-1 and HIV-2 [6]. HIV-1 is the most common virus in the world, while HIV-2 is limited to western Africa. HIV destroys human T-lymphocytes, blocks the process of cell immunity and humoral immunity, and paralyzes the immunity system, finally resulting in various diseases, such as AIDS. What is more, HIV adopts several pathways to avoid the human immune defense, and mutations often occur in the HIV genome resulting in drug-resistance during the drug-therapeutic process [7, 8]. Because of its tricky properties, so far, there is no effective strategy to eliminate HIV completely from human cells.

Similar to HIV, cancer is another incurable disease and it also weakens the human immune system [9]. Cancer is the name given to a collection of related diseases. Actually, cancers are abnormal cells, which can divide in an uncontrolled way, and grow without limitations. There are numerous cancers and some of them can spread to other parts of the body, which may cause pathological changes in body tissues or organs.

There are multiple treatments to prevent cells from being infected by HIV or developing into cancer [8], however, no therapeutic methods have been developed for HIV infections and cancers up to now, because of the limited knowledge about HIV and cancers. Therefore, insights into the mechanisms of HIV infection and tumorigenesis are greatly needed. They will provide the theoretical bases for developing efficient drugs or treatment methods once the mechanisms have been elucidated.

1.3 APOBEC3 proteins

In the human antiviral defense system, there is a kind of protein named the apolipoprotein B mRNA-editing catalytic polypeptide-like 3 (APOBEC3, denoted after as A3) family proteins, which are encoded by a gene cluster located on human chromosome 22 [10]. A3 proteins belong to the cytosine deaminase family, which catalyze cytosine-to-uracil conversion in single-stranded (ss) DNA and possess the ability to stop HIV-1 from replicating [11]. A3 proteins comprise seven subfamily

members, three of them (A3A, A3C, and A3H) contain only a single zinc-coordinating (Z) DNA cytosine deaminase domain, while the other four proteins (A3B, A3D, A3F, and A3G) contain double Z domains [10-12]. In the latter, only the C-terminal domains (CTDs) are conferred with deaminase activity, the N-terminal domains (NTDs) being reportedly inactive (Figure 1-1) [11-14]. Both types of the domains contain a zinc-signature motif (HXEX₂₃₋₂₈PCX₂₋₄C). On the basis of the sequence similarity, these domains are categorized into Z1 (A3A, CTDs of A3B, and A3G), Z2 (A3C and NTDs of A3B and A3G, as well as both domains of A3D and A3F), and Z3 (A3H) [12].

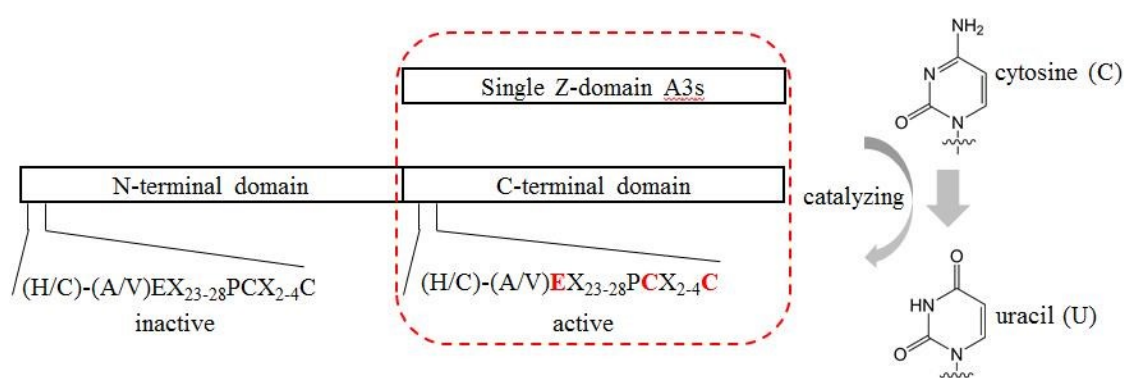


Figure 1-1. Schematic representation of A3 proteins. The active domains containing A3A, A3C, and A3H, as well as the CTDs of A3B, A3D, A3F, and A3G, which are enclosed by a red dashed line, catalyze the conversion of cytosines into uracils. Active residues in the signature motif are colored red.

The mechanism of cytosine deamination catalyzed by A3 proteins has been elucidated on the basis of the crystal structures of cytosine deaminases[15]. The zinc ion, which is indispensable for the catalytic activity, is coordinated by a histidine and two cysteine residues. Another highly conserved amino acid residue, glutamic acid residue, contributes to produce the hydroxide ion, which is required to remove the amine group from pyrimidine ring (Figure 1-2)[15].

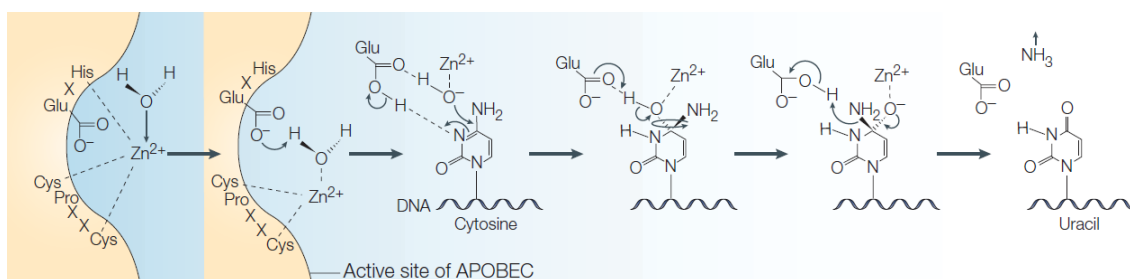


Figure 1-2. The mechanism of cytosine deamination by A3 proteins. First, the cysteine (Cys) and histidine (His) residues of the active site coordinate a single zinc ion (Zn^{2+}); an incoming water molecule (H_2O) is shown. Second, a hydroxide ion (OH^-) is produced when water reacts with glutamic acid (Glu) and zinc. Only the zinc ion and the side chain of the catalytic glutamic acid are illustrated beyond this step. Third, glutamic acid protonates N3 of the cytosine pyrimidine ring, which destabilizes the N3–C4 double bond and renders C4 susceptible to attack by the hydroxide ion. Fourth, this results in the formation of a tetrahedral (transition state) intermediate, as the proton (H^+) of the water molecule is sequestered by glutamic acid. Fifth, the amino group (NH_2) receives this proton in a reaction that causes cleavage of the C–N bond and formation of a C4–O double bond. Sixth, uracil and ammonia (NH_3) are released from the active site. This figure is Figure 2C in R.S.Harris [15].

1.4 A3 proteins and antiviral activity

A3 proteins play biological roles in the intrinsic immune system to prevent potential damage from retroviruses and endogenous retroelements [11, 16-19]. Four of the seven APOBEC3 proteins (A3D, A3F, A3G, and A3H) have been implicated as HIV-1 host restriction factors [20-26]. In the absence of HIV-1-Vif (viral infectivity factor), A3 proteins (A3D, A3F, and A3G) are packed into progeny viral particles during HIV-1- Δvif infection, and then interfere with reverse transcription and mutate cytosines to uracils in nascent cDNA (deaminase-dependent mechanism), generating G-to-A hypermutations in the proviral DNA plus strand. Such kinds of mutations in ssDNA lead to double-stranded breaks, resulting in genome DNA damage, and finally blocking viral replication [27-29]. A3G is also able to inhibit the replication by binding HIV-1 RNA, resulting in physical blocking of reverse transcription (deaminase-independent mechanism) [30].

On the other hand, HIV-1 has developed a countermeasure to neutralize A3 antiviral functions in infected cells by encoding accessory protein Vif. Vif recruits an E3 ubiquitin ligase complex comprising Cullin5 (Cul5), Elongin B (EloB), Elongin C

(EloC), core binding factor β (CBF- β), and a RING-box unit (Rbx2) to polyubiquitinated A3 for ubiquitin-mediated proteasomal degradation [31]. The CTD domains of A3D and A3F, as well as the NTD domain of A3G are the binding domains for Vif. To date, many studies have mainly focused on finding inhibitors or strategies to disrupt this interaction between A3 proteins and Vif, which is promising for the development of the novel anti-HIV-1 drugs.

1.5 A3 proteins and cancers

Recently, a role was proposed for A3 proteins in the generation of particular genome-wide mutational signatures [32]. A3 proteins were reported to contribute to genomic DNA damage as whole exome sequencing, and the whole genome sequencing showed that an A3 mutation signature, cytosine-to-thymine conversion within a TCW sequence, is present in more than seven thousand human cancers [33]. Moreover, another group revealed that A3B mRNA is up-regulated in most primary breast tumors and breast cancer cell lines, and that endogenous A3B correlates with increased genomic mutation frequencies [34]. Zik-Zainal and coworkers demonstrated that the A3A-A3B germline deletion allele confers cancer susceptibility through increased activity of APOBEC-dependent mutational processes [35]. Furthermore, the copy number of A3B is related to the A3B expression level in breast cancer cell lines [36]. When A3A and A3B are heterologously expressed in yeast, genome-wide mutation patterns of A3A- and A3B-mediated deamination showed the strong similarity of mutation signatures found in breast cancer, further supporting the role of A3B, and possibly A3A, in breast cancer hypermutation [37]. A3B has been shown to be up-regulated in human lymphoma cell lines, lung cancers, ovarian carcinomas, diverse ovarian cancer cell lines and high grade primary ovarian cancers [38-40].

Besides that of A3B, the expression level of A3G is also found to be high in colorectal tumors and hepatic metastasis [41]. It was also reported that A3A is biologically up-regulated in skin lesions, suggesting that A3A-mediated deamination can be considered as an alternative molecular basis for the initiation events in skin cancer [42]. Therefore, A3 proteins are vital in cancer-associated somatic mutation

processes, and affect cancer/tumor development and progression.

1.6 Structural characteristics of A3 proteins

After the initial discovery and determination of the structure of A3G [20, 43, 44], more and more A3 protein structures have been solved. Now, the structures of the catalytically active domains of A3A, A3B, A3C, A3F, and A3G, as well as the structures of the inactive domains of A3B and A3G in free forms have been determined, using X-ray crystallography or the NMR method (Table 1.1). The overall structure of A3 proteins comprises a central five-stranded β -sheet surrounded by six α -helices. All of them have a highly similar characteristic architecture ($\alpha 1$ - $\beta 1$ - $\beta 2$ - $\alpha 2$ - $\beta 3$ - $\alpha 3$ - $\beta 4$ - $\alpha 4$ - $\beta 5$ - $\alpha 5$ - $\alpha 6$) (Figure 1-3), which is further supported by the high sequence similarity [43, 45]. It was also noticed that in the catalytic center, Zn^{2+} is coordinated by one histidine and two cysteines, and interacts with the catalytic glutamate through a water molecule [46]. Despite the highly conserved structures among A3 proteins, there are still subtle differences. In the cases of A3A, A3B, and A3G, the $\beta 2$ is interrupted ($\beta 2$ -bugle- $\beta 2'$), while the other A3 proteins have a continuous $\beta 2$. This discontinuous $\beta 2$ is supposed to mediate the interaction with nucleotide acids or proteins, and plays an important role in the conformation and oligomeric state [43, 47, 48].

Over all of the molecular surfaces of A3 proteins, there is a single, partial groove formed by loops 1, 3, and 7. The conformations and orientations of these loops play critical roles in ssDNA-binding, accessibility, and positioning. Compared with other A3 proteins, which have a large and open catalytic center, A3B shows a closed conformation of the catalytic center, which leaves a limited opening for a substrate to access the active center. This structural property of A3B is supposed to cause rearrangement of the surrounding residues from the closed conformation after binding to the substrate [49, 50].

Table 1.1 Structures of A3 proteins

A3 protein	PDB	Method	Reference
A3A	4XXO	X-ray	[51]
A3A	2M65	X-ray	[52]
A3B-CTD	5CQD	X-ray	[49]
A3B-CTD	2NBQ	NMR	[50]
A3C	3VOW	X-ray	[53]
A3F-CTD	4IOU	X-ray	[54]
A3F-CTD	4J4J	X-ray	[55]
A3F-CTD	3WUS	X-ray	[56]
A3G-CTD	2JYW	NMR	[43]
A3G-CTD	3E1U	X-ray	[44]
A3G-CTD	2KBO	NMR	[48]
A3G-CTD	3IR2	X-ray	[57]
A3G-CTD	3V4K	X-ray	[58]
A3G-CTD	4ROW	X-ray	[59]
A3G-NTD	2MZZ	NMR	[60]
A3G-NTD	5K81	X-ray	[61]
A3B-NTD	5TKM	X-ray	[62]

Despite that A3 proteins exhibit similar global structures, there are several differences in the structure and conformation of the loops, especially loops 1, 3, and 7, which are around the catalytic center. Due to the differences in these loops, A3 proteins are bestowed with different functions. For example, A3A and A3B-CTD exhibit a sequence identity of 82.7%, but the deaminase activity of A3A is much higher than that of A3B-CTD, which is due to the shorter loop 1 in A3A than that in A3B-CTD [13]. Loop 3 has the potential to regulate the deaminase activity of A3A or A3G by binding a secondary zinc [63]. The aromatic rich amino acid residues in loop 7 determine the substrate specificity of A3 proteins, as it was revealed that on loop 7 replacement between different A3 proteins, the target sequence preference was also swapped [55, 64].

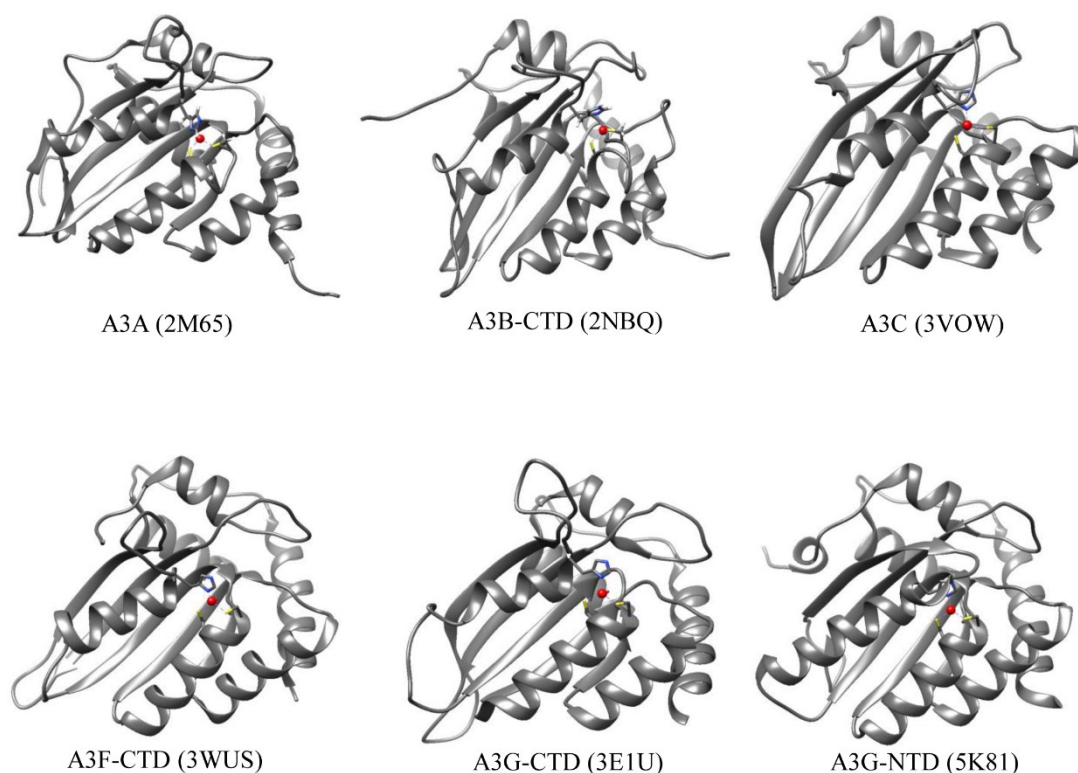


Figure 1-3. Representative structures of A3 proteins. All the structures of A3 proteins that have been solved comprise a central five-stranded β -sheet surrounded by six α -helices. The active zinc is colored red.

1.7 Binding interface of A3s with oligonucleotides

A3 proteins deaminate a target cytosine in an ssDNA substrate. For the double domains of A3 proteins, the deaminase activity is only conferred by the CTD domain, with the NTD domain assisting in ssDNA binding. The ssDNA binding and specific deamination are the basis for their antiviral property and cancer-related characteristics. Therefore, more and more attention is been paid to clarification of how A3 proteins recognize and bind an ssDNA substrate. Numerous research has been performed to try to elucidate this mechanism. When the structure of A3G-CTD was firstly determined by two individual groups, both of them postulated the binding of A3G-CTD with ssDNA oligonucleotides, which have two potential ssDNA-binding interfaces. The mode revealed by Chen *et al.* based on the NMR structure showed that the ssDNA binding residues are located in a straight DNA binding groove [43], which was further

confirmed by Furukawa and coworkers [48]. Unlike other nucleic acid-binding proteins that have a positively charged surface for contact with a negatively charged nucleic acid, on the electrostatic potential surface of A3G-CTD, a largely negative region together with a few positively charged residues form an apparent brim surrounding the concave active-site region (Figure 1-4a). On the basis of this binding mode, it is notable that in order to access the catalytic glutamate, the target cytosine is flipped out from the phosphodiester backbone of the ssDNA [43]. The C-terminal regions of $\alpha 1$ and $\alpha 2$, as well as the N-terminal regions of $\beta 3$ and $\beta 4$ constitute an area that is closed to the main chain of an ssDNA substrate [48]. A mutagenesis study on A3F-CTD from Siu and coworkers showed that the residues of A3F-CTD responding to ssDNA binding are similar to the case of A3G revealed by Chen *et al.* (Figure 1-4b) [55].

While another mode revealed by Holden *et al.* based on the X-ray structure of A3G-CTD revealed that ssDNA-binding residues are located along a kinked groove (Figure 1-4c). In this structure, active loops 1 and 3 together with the region near the active site form a deep, spacious groove that is not present in the NMR structures revealed by Chen *et al.* [43] and Furukawa *et al.* [48]. The ssDNA-binding interface starts from the groove between loop 1 and 3, leads to a deep active pocket, and extends to the left side over $\alpha 6$. It is hard to judge which ssDNA-binding mode is reasonable because different methods were used. The strict ssDNA-binding mode cannot be clarified unless the co-structures of A3s-ssDNA are solved.

Recently, the crystal structures of A3A and chimera A3B-CTD in complexes with ssDNA substrates have become available [65]. Both the co-crystal structures showed a U-shaped DNA conformation (Figure 1-4d, e). At the bottom of the "U", the thymine at position -1 flips out toward the protein and the target cytosine is deeply inserted into the zinc active center. The nucleic acid residues flanking the TC sequence are bent away from the active site, which is between loops 1 and 7, and the sugar-phosphate backbone is rotated in response to those of the flanking sequences. Another co-crystal structure of A3A-ssDNA reported by Kouno group [66], and a solution-state model of A3A interacting with its ssDNA substrate obtained with the "method of small changes" and the NMR method [67] showed the similarity with the co-structure of the A3A-ssDNA complex. One difference regarding these three structures is the length of the ssDNA

substrate. In the first two crystal structures, short ssDNAs (15 nucleotides) were used and co-crystallized, while for the NMR-based structure, a long ssDNA substrate was used for modelling. However, both the two A3A-ssDNA structures showed a well overlapping shape (Figure 1-4f) [67], even though there are some differences outside the TCA sequence, which might be due to the different lengths and compositions of the ssDNA substrates used [67].

The U-shaped ssDNA binding conformations of A3A and A3B-CTD resemble the co-structure of TadA-RNA, which is the distantly related tRNA adenosine deaminase [68] (Figure 1-4g), suggesting that it seems like most deaminases bind a substrate nucleic acid with a U-shape. However, the conformations of oligo(T) co-crystallized with A3G-NTD (Figure 1-4h) and the double-stranded RNA bound with adenosine deaminase ADAR2 [69] are in contrast to the models of A3A-ssDNA and chimera A3B-CTD-ssDNA. The U-shaped DNA/RNA conformation is not observed in these two structures, which could be attributed to the different oligonucleotides used. It is also undeniable that the various conformations for oligonucleotide-binding co-exist in deaminases.

In summary, the reported co-structures of A3A, chimera A3B-CTD, and A3G-CTD with different oligonucleotides provide a basis for clearly understanding the interaction between A3 proteins and their substrates, and for designing mechanistic-based therapeutics for viruses and cancers.

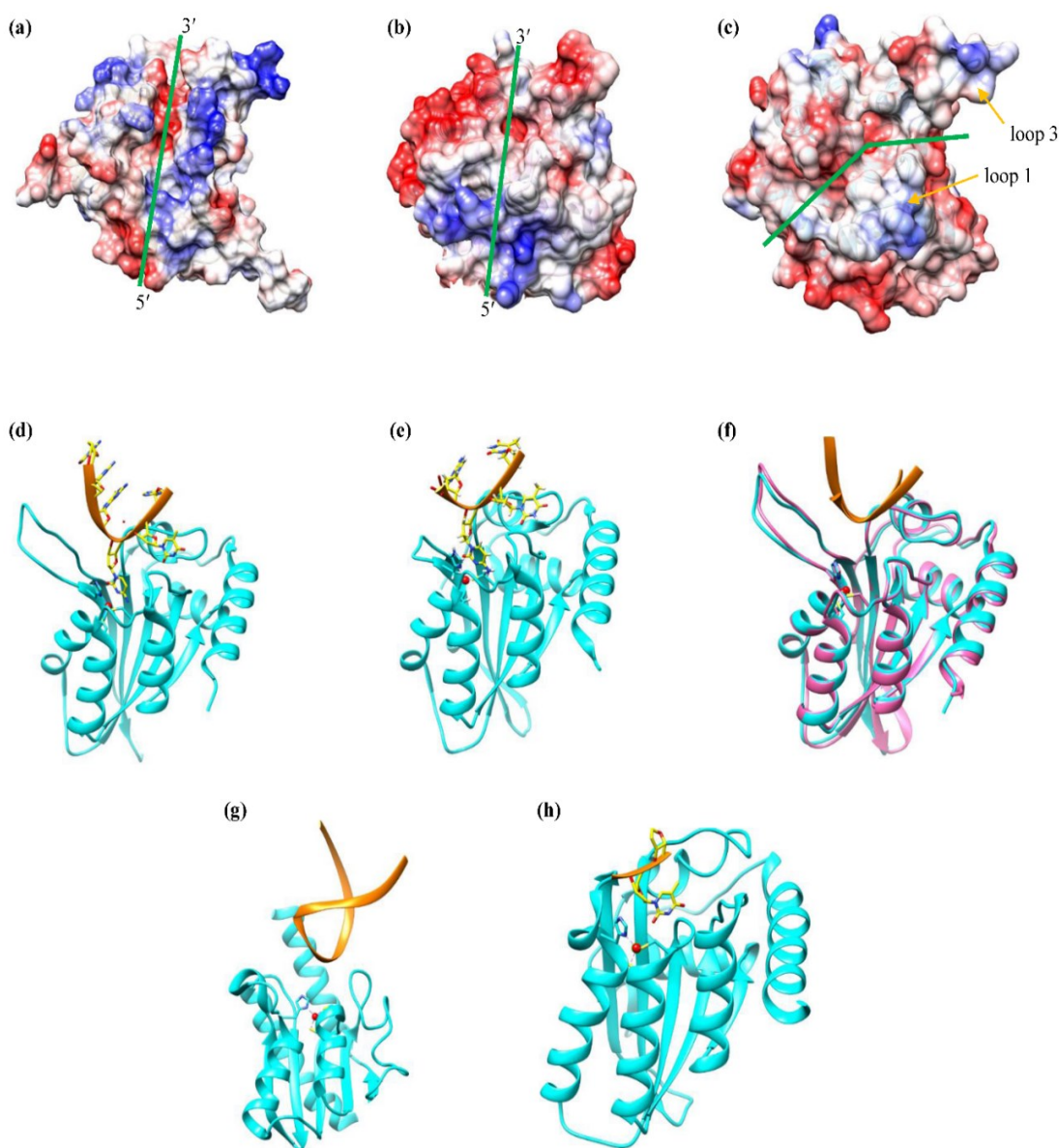


Figure 1-4. A3 proteins and nucleic acid interaction models. Surface representations of A3G-CTD-2K3A (PDB: 2JYW) **(a)** and A3F-CTD (PDB: 3WUS) **(b)**, in which ssDNA-binding residues were found to line a straight DNA-binding groove, and A3G-CTD (PDB: 3E1U) **(c)**, which has the ssDNA-binding residues along a kinked groove. The positions of positive (blue), negative (red), and neutral (white) charge are highlighted. The hypothesized position and polarity of ssDNA are indicated (green line). Ribbon schematics of A3A-ssDNA (PDB: 5SWW) **(d)** and the chimera A3B-ssDNA (PDB: 5TD5) **(e)** complex showing flipped-out target C and -1 nucleotides, as well as the overall U-shaped binding conformation. **(f)** Superposition of **(d)** (cyan) onto another A3A-ssDNA complex (hot pink, PDB: 5KEG), showing the similar U-shaped binding conformation. Ribbon schematics of TadaA-tRNA (PDB: 2B3J) **(g)** and A3G-NTD-poly dT (PDB: 5K83) **(h)**. The active zinc is colored red.

1.8 A3s' movement on single-stranded DNA

Biologically, proteins recognize essentially functional DNA by interacting with specific DNA sequences [70]. The mechanism by which how a protein searches for its target sequence has been studied by numerous groups. On the basis of the theoretical and experimental data, several diffusion-based models, including sliding, hopping/jumping, and intersegmental transfer motions, to explain the movement of a protein on the DNA have been established (Figure 1-5). On sliding, a protein moves along DNA through one-dimensional diffusion with continuous contact with the DNA [71, 72]. Hopping/jumping are used to describe the micro-dissociation of a protein from the DNA with re-association with the same DNA substrate without diffusion into the bulk solution [72]. Most re-bindings are short range "hops" back or near the original dissociation site in a three-dimensionally manner [73]. Occasionally, a protein could jump a long distance via dissociation and then rebind at a distal location on the same DNA molecule [72]. Intersegmental transfer is a term that describes the movement of a protein with two DNA-binding domains that interacts with two distal sites at the same time, and then dissociates from one site [71, 72]. The distance between the two binding sites has a mean step size of ~400 bp, as shorter separation is supposed to hinder the bending/twisting of the intervening DNA [74, 75].

Regarding A3 proteins, the movement of A3G on ssDNA has been well studied. It has been reported that the ssDNA scanning mode of A3G plays a vital role in inducing mutagenesis of minus DNA formed during reverse transcription [76]. One group has found that A3G uses intersegmental transfer to search its target sites [77]. However, several individual groups have reported that A3G searches for its target sequence via combined sliding and jumping [76, 78-80]. The scanning mechanism of another anti-HIV factor, A3F, has also been elucidated by Ara and coworkers [81]. Both A3F and A3G are processive enzymes and can deaminate two target sites on the same ssDNA substrate, but the scanning mode of A3F is different from that of A3G. A3F searches for the target cytosine by jumping, which is determined by a ¹⁹⁰NPM¹⁹² sequence. The scanning movements of other A3 proteins have not been revealed yet.

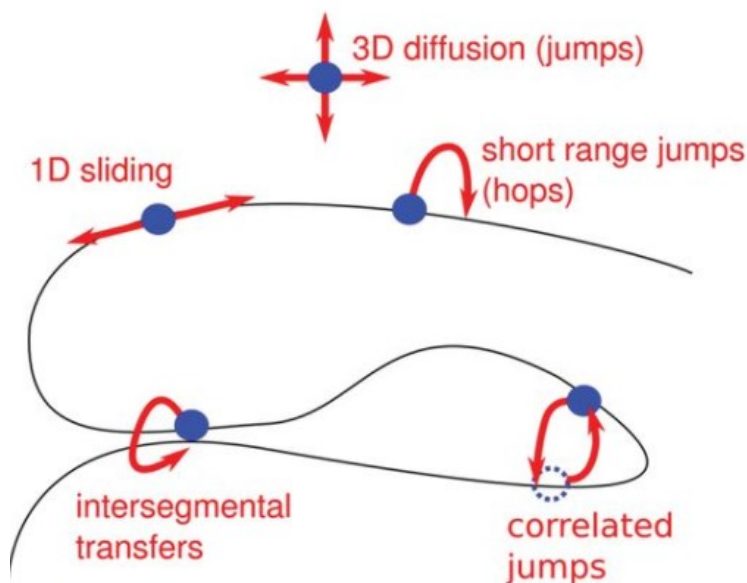


Figure 1-5. Schematic diagram of the different kinds of motion for a protein undergoing facilitated diffusion via a non-specific interaction with the DNA, the protein can randomly walk along the DNA backbone, effectively diffusing in one dimension (sliding). When not interacting with the DNA, the protein performs three-dimensional diffusion (hopping/jumping and intersegmental transfer). This figure is Figure 1A in C.A.Brackley [82].

1.9 Objectives of this study

Among A3 proteins, the relationship of A3B with cancers is mostly noticeable. Given the important biological roles of A3B in cancers, it is important to understand the relationship between A3B and cancers, and how A3B causes tumorigenesis. In the case of A3G, it has been reported that it slides and jumps along ssDNA, and deaminates the target sequence with catalytic orientational specificity (3'→5' polarity) [78], which provides a possible mechanistic explanation for the mutational distribution *in vivo*. However, studies on A3B's movement remain limited. Further biochemical characterization of A3B's deamination properties are needed to reveal the link of A3B with cancers. This thesis describes the deamination behavior of A3B-CTD toward ssDNA substrates of different lengths and the different locations of the target sequences in ssDNAs. On the basis of the experimental results, this thesis aims to reveal how A3B-CTD searches for its target sequence in ssDNA.

A3F is regarded as the second most efficient restrictor of HIV replication following A3G. Previous studies were highly centered on the interaction between A3F and Vif, as well as the development of a method or the finding of inhibitors to abolish this interaction. However, information on the deamination characteristics of A3F remains elusive. This thesis deeply focuses on the deamination properties of A3F-CTD and provides a comprehensive view of the interaction between A3F-CTD with its substrate ssDNA. Besides, the effect of the Vif complex *in vitro* on the deaminase activity of A3F has also been studied in this work.

Chapter 2

Observation by real-time NMR, and interpretation of length- and location-dependent deamination activity of APOBEC3B

2.1 Introduction

Previously, our lab developed the real-time NMR method to monitor the deamination reaction of A3G [48]. This method can directly detect a site-specific deamination reaction and is very sensitive to weak interactions because high concentration of ssDNA (mM order) is used for the NMR experiment. The NMR method could provide the high spectral and time resolution to distinguish the multiple reactions and it could also address the dynamical aspects of the reaction. The NMR method has been applied to study the location- and sequence-dependence of the deamination by A3G and analyze the 3'→5' polarity of A3G's deamination [83]. According to various kinetic parameters obtained by the real-time NMR method, a kinetic model of A3G has been established, which rationally explains the 3'→5' polarity of A3G.

In this chapter, similarly, we used the real-time NMR method, combining the biochemical methods, to study the deamination reaction of A3B-CTD. The length- and location-dependent deaminase activity of A3B-CTD were measured. Different from the case of A3G, A3B-CTD deaminates a target sequence in short ssDNA and the one located close to the center of ssDNA. On the basis of the weak ssDNA-binding affinity of A3B-CTD, these results were rationally interpreted.

2.2 Results

2.2.1 A3B-CTD catalysis depends on the lengths of ssDNA substrates

We firstly used the uracil DNA glycosylase (UDG) assay to analyze the ssDNA-length-dependent activity of A3B-CTD toward ssDNA substrates ranging in size from 10 to 85 nucleotides (nt), each ssDNA containing a single 5'-TC-3' target sequence located near the center (S_{FITC_10-85} , Table 2.1). The ratio of A3B-CTD to ssDNA was 1:4. The results were shown in Figure 3-1a and b. A3B-CTD activity is high for short ssDNA,

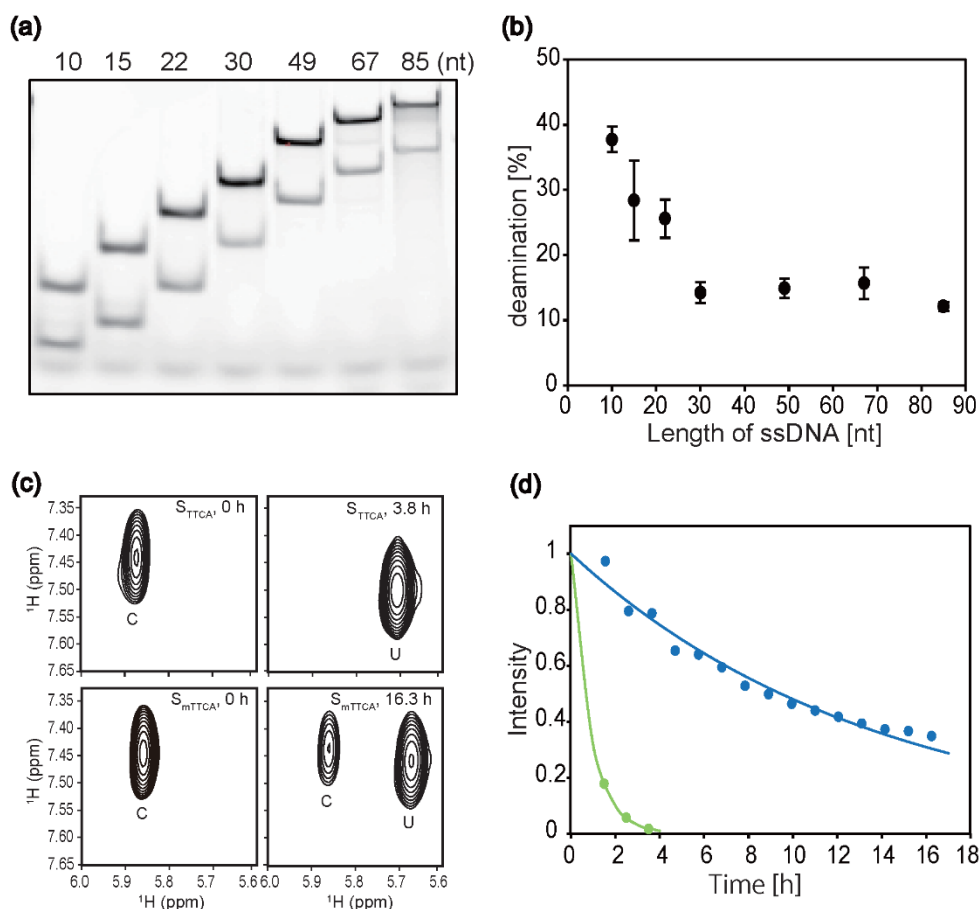


Figure 2-1. Length-dependent deamination of ssDNA by A3B-CTD. (a) Deamination activities of A3B-CTD toward ssDNAs with different lengths were examined by UDG assay. The length of each ssDNA is indicated at the top of the gel. The upper and lower bands in each lane represent unreacted substrate and deamination product, respectively. (b) Efficiency of the deamination for each ssDNA revealed by UDG assay. (c) Deamination activities of A3B-CTD monitored by the real-time NMR method. H5-H6 TOCSY correlation peaks for 10 nt S_{TTCA} (top) at 0 and 3.8 h, and 46 nt S_{mTTCA} (bottom) at 0 and 16.3 h after addition of A3B-CTD are shown. Correlation peaks for a cytosine and a produced uracil are labeled with C and U, respectively. (d) Time-courses of the change in the intensity of a cytosine peak for S_{TTCA} (green) and S_{mTTCA} (blue).

as the Figure 2-1b showed that the deaminase activity decreases monotonically between 10 nt and 30 nt and then keeps constant with longer ssDNA(>30 nt). We also applied the real-time NMR method to monitor the time-course of the intensity of the H5-H6 TOCSY correlation peak of the cytosine of 10 nt (S_{TTCA} , Table 2.1) and 49 nt (S_{mTTCA} , Table 2.1) ssDNA (Figure 2-1c and d). It was indicated that the target sequence in S_{TTCA} is deaminated completely within 4 h, while in case of S_{mTTCA} , about 40 % of the

substrate still remained unreacted after 16 h. These results clearly indicate that A3B-CTD has a tendency to efficiently deaminate a target sequence in short ssDNA.

2.2.2 A3B-CTD interacts with ssDNA with weak binding affinity

To better understand the mechanism of the ssDNA length-dependency of A3B-CTD, we investigate the binding characteristics of A3B-CTD to ssDNA. The binding affinity of A3B-CTD with ssDNA was initially measured by electrophoretic mobility shift assay (EMSA). For comparison, A3F-CTD, which shares the same target sequence with A3B-CTD, was used. Different lengths of ssDNA substrates ($S_{\text{FITC}_{10-85}}$, Table 2.1) were used. The EMSA results showed the band of the A3F-CTD:ssDNA complex (Figure 2-2a), suggesting A3F-CTD efficiently binds ssDNA with the length of 10 - 85 nt. However, in case of A3B-CTD, no band of A3B-CTD:ssDNA complex was observed (Figure 2-2b), indicating the low binding affinity of A3B-CTD with ssDNA.

We also carried out fluorescence anisotropy (FA) measurement which is much more sensitive than EMSA method, to detect the binding affinity. 49 nt FITC-labeled ssDNA ($S_{\text{FITC}_{49}}$, Table 2.1) was used as the substrate (Figure 2-2c). The FA value increased upon addition of A3F-CTD, indicating formation of a complex, which consists with the EMSA result. A3B-CTD titration, however, caused no change in the FA value. These results indicate that the binding affinity of A3B-CTD to ssDNA is too weak to be detected. In our previous study on A3G-CTD, the band of the complex with ssDNA was detected by EMSA analysis, and the dissociation constant, K_d , was determined to be ca. 0.1 mM [48]. For A3B-CTD, the band of the complex with ssDNA was not detected, which suggested that ssDNA-binding affinity of A3B-CTD is weaker than that of A3G-CTD. Therefore, for the discussion below, we assume the K_d value of 1 mM for the A3B-CTD:ssDNA complex, which is larger than that for A3G-CTD:ssDNA complex by one order.

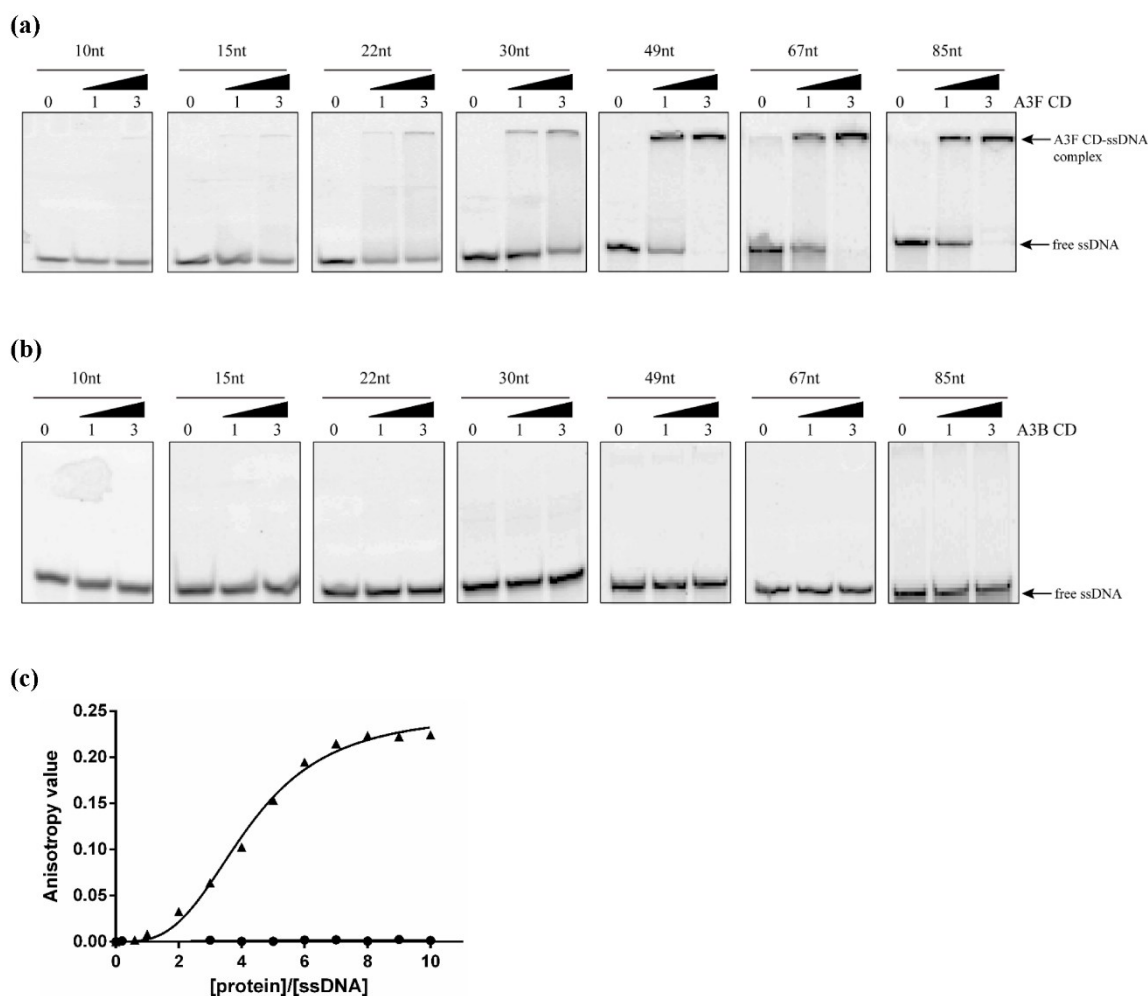


Figure 2-2. Binding assays of A3F-CTD and A3B-CTD with ssDNAs of different lengths. (a) EMSA analysis of A3F-CTD with TC-containing ssDNA substrates (100 nM) of various lengths (Table 2.1): 10 nt (S_{FITC_10}); 15 nt (S_{FITC_15}); 22 nt (S_{FITC_22}); 30 nt (S_{FITC_30}); 49 nt (S_{FITC_49}); 67 nt (S_{FITC_67}); 85 nt (S_{FITC_85}). Lanes: 1, 4, 7, 10, 13, 16, 19, no A3F CTD; 2, 5, 8, 11, 14, 17, 20, 1 μ M A3F CTD; 3, 6, 9, 12, 15, 18, 21, 3 μ M A3F-CTD. The positions of A3F-CTD-bound ssDNA and free ssDNA are indicated. (b) EMSA analysis of A3B-CTD with TC-containing ssDNA substrates of various lengths, as described in (a). (c) Fluorescence anisotropy analysis of binding of either A3F-CTD or A3B-CTD to ssDNA (S_{FITC_49} , Table 2.1). The fluorescence anisotropy is plotted against the ratio of either A3F-CTD (black triangles) or A3B-CTD (black circles) to S_{FITC_49} .

2.2.3 A3B-CTD catalysis depends on the location of the target sequence on ssDNA

Since A3G shows 3'→5' polarity, we examined whether the deamination of A3B-CTD has polarity or not. Three different ssDNA substrates were used, $S_{5'TTCA}$,

S_{mTTCA} , and $S_{3'TTCA}$ (Table 2.1), having a single target sequence at a region close to the 5'-end, center, and 3'-end, respectively. The intensity of the peak of cytosine in the 5'-TC-3' target sequence was plotted for each ssDNA against time. The substrate S_{mTTCA} , which has a target sequence near the center, is deaminated most efficiently by A3B (Figure 2-3). This is quite different from the case of A3G, in which a target sequence located closer to the 5'-end is deaminated more effectively.

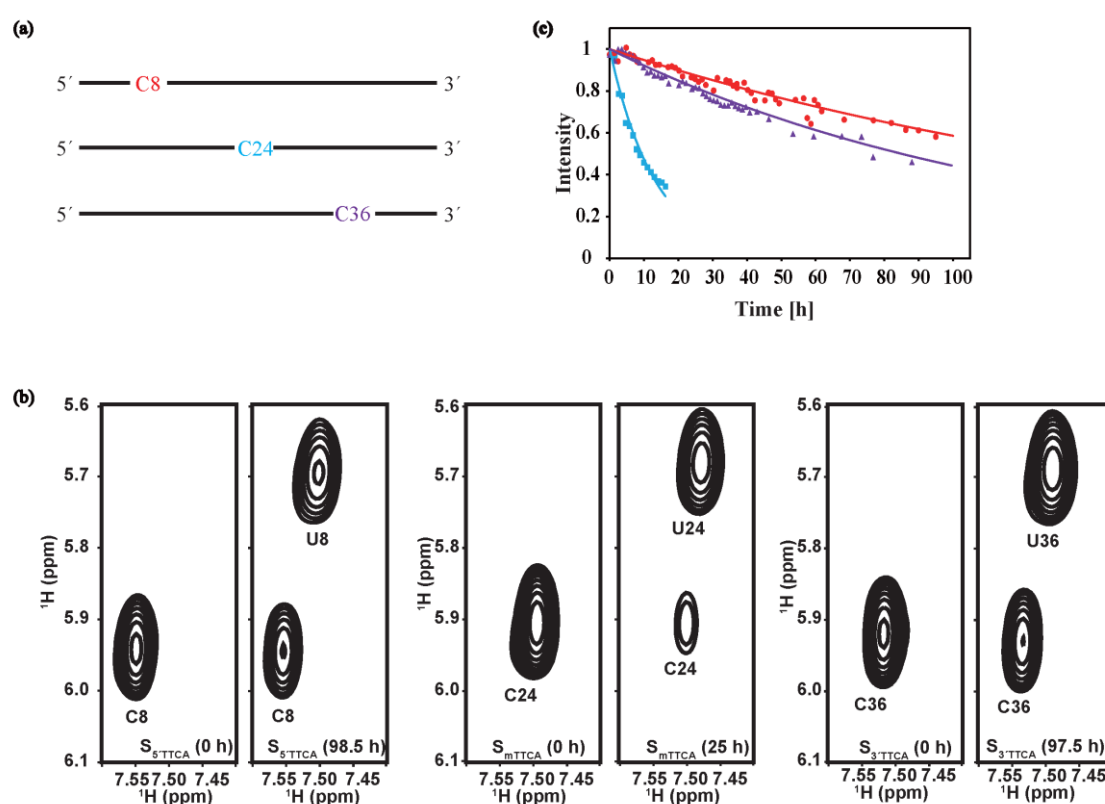


Figure 2-3. Real-time NMR monitoring of deamination reactions catalyzed by A3B-CTD. (a) Three ssDNA substrates containing a single target sequence at a region close to the 5'-end (red), center (cyan), and 3'-end (purple) were shown. (b) Deamination activities of A3B-CTD toward these three ssDNAs were monitored by the real-time NMR method. H5-H6 TOCSY correlation peaks for 46 nt $S_{5'TTCA}$ at 0 and 98.5 h (left), 46 nt S_{mTTCA} at 0 and 25 h (middle), and 46 nt $S_{3'TTCA}$ at 0 and 97.5 h (right) after addition of A3B CTD are shown as representatives, respectively. Correlation peaks for cytosines and produced uracils are labeled. (c) Time courses of the intensity of the target cytosine during deamination for C8 (●) of $S_{5'TTCA}$, C24 (■) of S_{mTTCA} , and C36 (▲) of $S_{3'TTCA}$, respectively.

To further validate the location-dependent deaminase activity of A3B-CTD, we subsequently performed real-time NMR analysis using an ssDNA containing two target sequence, S_{ss} (Table 2.1). The target sequences in S_{ss} are located near the 5'-end and center, respectively. We first confirmed by UDG assay using S_{FITC_1} and S_{FITC_2} (Table 2.1) that the difference in the sequences of the flanking regions does not affect the deaminase activity (Figure 2-4). It was also confirmed that no peak is observed in the amino proton region of 10 - 13 ppm for S_{ss} (Figure 2-5). This indicated that no base pair is formed for S_{ss} , which excluded the possibility that S_{ss} folded into structures. Then, real-time NMR analysis showed that C36 in the target sequence near the center is deaminated more efficiently than C8 in the target sequence near the 5'-end (Figure 2-6a and b), which is the same observation as Figure 2-3a. Next, we introduced a twenty base-pair double-stranded DNA (dsDNA) region to downstream of the central target sequence in S_{ss} and obtained S_{ds} (Table 2.1), and examined the deaminase activity. It turned out that the location-dependence of deaminase activity was significantly lost on the introduction of the dsDNA (Figure 2-6c and d).

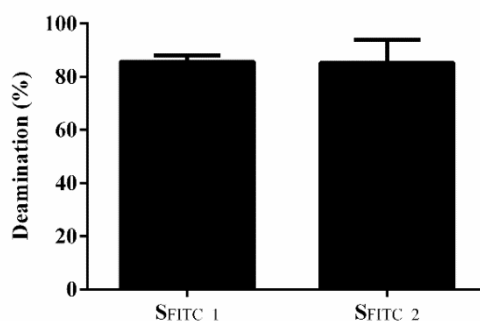


Figure 2-4. Deaminase activities of A3B-CTD toward two units of S_{ss} . Deaminase activities of A3B-CTD toward S_{FITC_1} and S_{FITC_2} (Table 2.1), which are two units of S_{ss} , were measured by UDG assay.

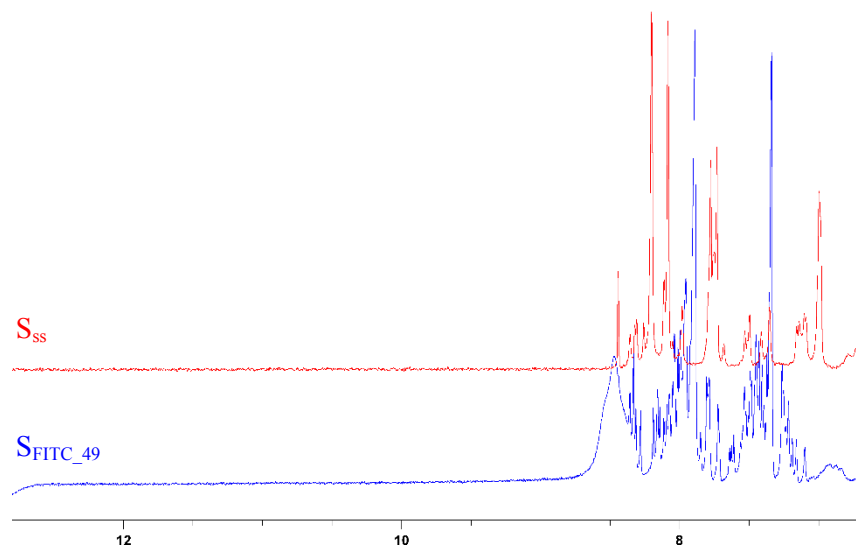


Figure 2-5. ^1H NMR spectrum of long ssDNA. 150 μM S_{ss} (red) and 8 μM S_{FITC_49} (blue) were measured for 1D NMR, respectively. There is no peak in range of 10 - 13 ppm.

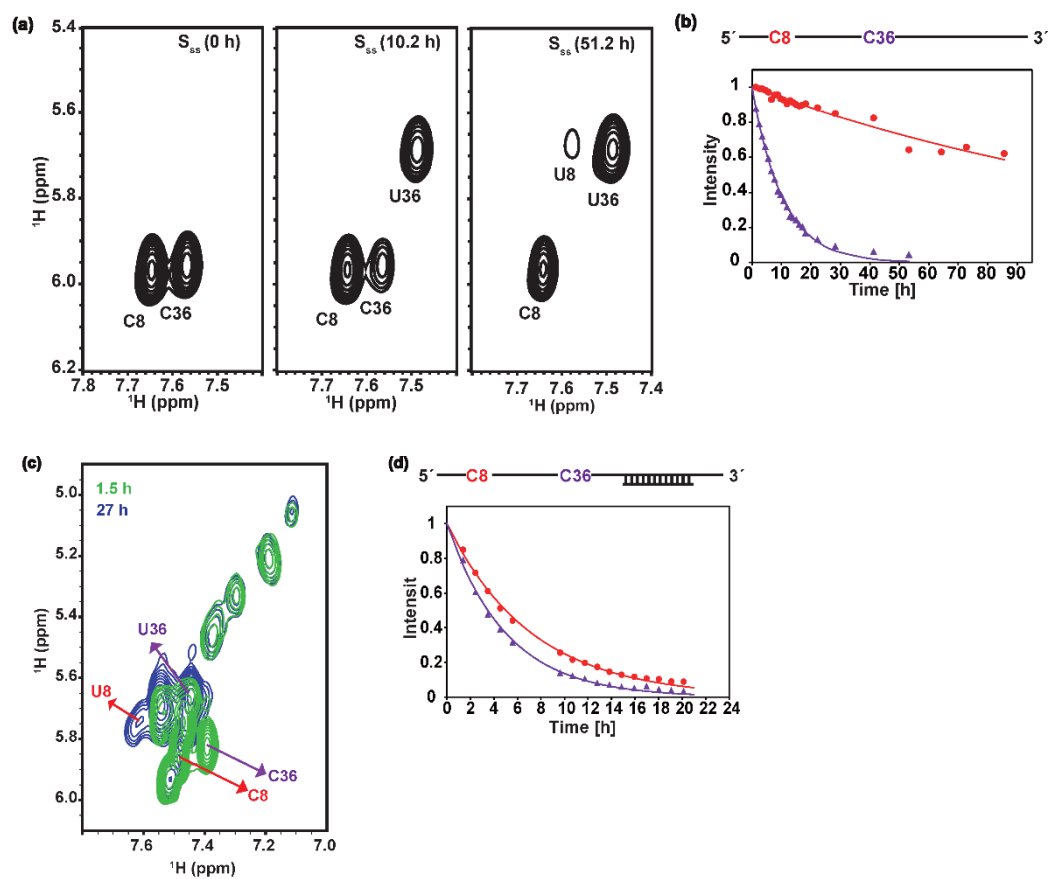


Figure 2-6. Deaminase activities of A3B-CTD toward S_{ss} and S_{ds} by real-time NMR method. (a) H5-H6 TOCSY correlation peaks for S_{ss} at 0 h, 10.2 h, and 51.2 h after addition of A3B-CTD are shown as representatives, respectively. (b) Time courses of the intensity for C8 (●) and C36 (▲) of S_{ss} . (c) Superposition of H5-H6 TOCSY correlation peaks for S_{ds} at 1.5 h (green) and 27 h (blue) after addition of A3B-CTD. (d) Time courses of the intensity for C8 (●) and C36 (▲) of S_{ds} . Correlation peaks for cytosine and produced uracil are labeled, respectively.

2.3 Discussion

Generally, protein scans on ssDNA via three manners, sliding, hopping/jumping, and intersegmental transfer [71, 72]. In this study, we propose that A3B-CTD binds ssDNA nonspecifically as in the case of A3G [79], then A3B-CTD scans ssDNA via either sliding, hopping/jumping, or intersegmental transfer. Here, we take sliding mode as an example to interpret our results.

To interpret the ssDNA-length-dependent deamination activity of A3B-CTD, we consider two substrates, 15 nt ssDNA (Figure 2-7a) and 45 nt ssDNA (Figure 2-7 b - d), both having a single 5'-TC-3' target sequence at the center. 45 nt ssDNA can be considered to comprise three sections, D1, D2, and D3. Considering the weak affinity of A3B-CTD, the K_d value for the A3B-CTD:15 nt ssDNA complex is tentatively assumed to be 1 mM. The total concentrations of A3B-CTD and ssDNA applied for real-time NMR analysis (Figure 2-1c and d) were 1 μ M and 200 μ M, respectively. Under these conditions, the concentrations of A3B-CTD bound to ssDNA were calculated to be 0.165 μ M for 15 nt ssDNA and 0.375 μ M for 45 nt ssDNA, where $200 \mu\text{M} \times (45 \text{ nt} / 15 \text{ nt}) = 600 \mu\text{M}$ was used as the effective concentration of 45 nt ssDNA. Accordingly, A3B-CTD bound at each 15 nt section (D1, D2, or D3) of 45 nt ssDNA could be deduced to be 0.125 μ M (one-third of 0.375 μ M). We assume, as mentioned above, for simplicity, that all A3B-CTD that bound to 15 nt ssDNA (D0) can reach the target sequence in D0 through sliding (Figure 2-7a). This means that A3B-CTD can slide along ssDNA by 7 nt without dissociation. The number of 7 nt was taken as one of the realistic numbers on the basis of the result of Figure 2-1b. It is similarly supposed that all A3B-CTD that bound to the D2 section of 45 nt ssDNA (0.125 μ M) can reach the target sequence in D2. It is supposed that A3B-CTD that bound to D1 (or D3) section slides along ssDNA, but that it mostly dissociates from the ssDNA before

reaching the D2 section, because of the weak affinity (Figure 2-7 b or d). We assume, for example, that only 5% of A3B-CTD that bound to the D1 (or D3) section can reach the D2 section ("survival"). In this case, the concentration of A3B-CTD that bound to D1 (or D3) and slides over to D2 to reach the target sequence is estimated to be $0.125 \mu\text{M} \times 0.05 = 0.00625 \mu\text{M}$. Therefore, the expected concentration of the A3B-CTD that can reach the target sequence in 45 nt ssDNA for possible deamination is $0.125 \mu\text{M} + 2 \times (0.125 \mu\text{M} \times 0.05) = 0.138 \mu\text{M}$. This value is smaller than the concentration of A3B-CTD that reaches the target sequence of 15 nt ssDNA ($0.165 \mu\text{M}$). This coincides with the fact that 15 nt ssDNA is deaminated with higher activity than 45 nt ssDNA. The essence is that although D1 and D3 of 45 nt ssDNA serve as landing points to recruit A3B-CTD to the target sequence in D2, they act more as competitors for binding of A3B-CTD to the target sequence in D2 and thereby reduce the concentration of A3B-CTD that reaches the target sequence in D2. This is due to the relatively short distance of sliding of A3B-CTD as a result of weak affinity. Thus the ssDNA-length-dependent deaminase activity of A3B-CTD can be rationally interpreted.

Besides sliding, hopping/jumping may also be used for A3B CTD to search for the target sequence. This possibility is not excluded. Even in that case, the ssDNA-length-dependent deaminase activity can be explained basically in the same way. It was noticed in Figure 2-1b that the deamination efficiency reached the plateau level at the length of 30 nt, the efficiency being basically the same for longer ssDNAs. One possible interpretation is that the intersegmental transfer of A3B CTD occurs and contributes to the target search for longer ssDNAs. Intersegmental transfer is known as a phenomenon that the protein moves between two sites widely separated along the DNA contour but closely spaced in three dimensions [72, 84]. For longer ssDNA, two sites widely separated along the ssDNA contour can be closely spaced in three dimensions because of easiness of looping for longer ssDNA. Thus, the intersegmental transfer of A3B CTD could occur for longer ssDNAs, resulting in the efficient target search and deamination. This may compensate for the reduction in deamination efficiency caused by the competitive effect of longer ssDNAs. The contribution of the intersegmental transfer to the deamination was suggested for A3G [77] and activation-induced deoxycytidine deaminase [85].

To interpret the location-dependent ability of deamination by A3B-CTD, we consider three cases, in which 45 nt ssDNA has a single target sequence in either the D1, D2, or D3 section (Figure 2-7 e - g). First, the case in which the target sequence is located in D2 (Figure 2-7f) is considered. It is supposed again, for simplicity, that all A3B-CTD that bound to D2 could reach the target sequence through sliding (Figure 2-7f middle). It is supposed, however, that most A3B-CTD that bound to D1 (or D3) cannot reach the target sequence due to dissociation during sliding (Figure 2-7f top or bottom). Secondly, the case in which the target sequence is located in D1 (Figure 2-7e) is considered. It is supposed similarly that all A3B-CTD that bound to D1 can reach the target sequence (Figure 2-7e top). Most A3B-CTD that bound D2, however, is supposed not to be able to reach the target sequence due to dissociation (Figure 2-7e middle). It should be noted that it is even more difficult for A3B-CTD that bound to D3 to reach the target sequence, because the distance between the target sequence in D1 and A3B-CTD bound at D3 is greater and thus the possibility of dissociation during sliding is higher (Figure 2-7e bottom). In total, the concentration of A3B-CTD that reaches the target sequence is higher for the case in which the target sequence is in the central section, D2, than for the case in which the target sequence is in the left section, D1. Thirdly, the case in which the target sequence is located in D3 (Figure 2-7 g) should exhibit a similar situation as for the case in which the target sequence is located in D1. In summary, a target sequence located at the centre of ssDNA has more chance to meet A3B-CTD for possible deamination than one located close to either the 5'-end or 3'-end. The essence is when the target sequence is at the center, it has two DNA regions, one on both sides that serve to recruit A3B-CTD, while when the target sequence is close to either 5'-end or 3'-end, it has only one DNA region on one side that serves for recruitment. In this way, the location dependent deamination activity of A3B-CTD can be rationally interpreted.

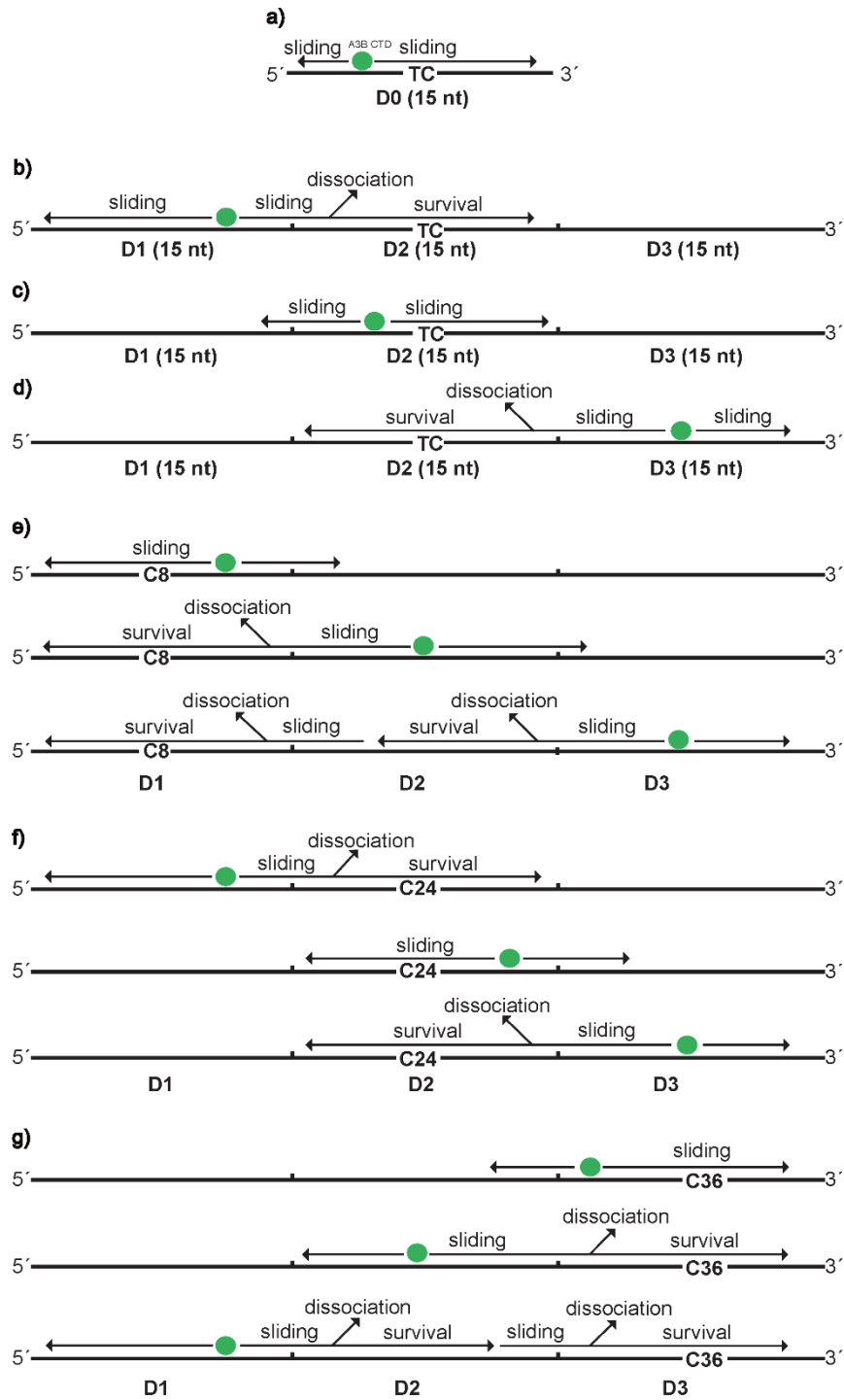


Figure 2-7. Description of the target search by A3B-CTD with binding, sliding, dissociation, and survival of sliding. (a) A case of 15 nt ssDNA (D0). (b - d) Cases of 45 nt ssDNA that comprises three 15 nt sections (D1, D2, and D3), having a target sequence, 5'-TC-3', in D2. A3B CTD binds to either D1 (b), D2 (c) or D3 (d), respectively. (e - g) Cases of 45 nt ssDNA which has the target sequence in either D1 (e), D2 (f), or D3 (g), respectively. A3B CTD binds to either D1, D2 or D3 in each case.

Furthermore, Figure 2-6b and d confirmed our interpretation. A3B reported does not interact with dsDNA [86], therefore, we assume that there is no landing point downstream of C36 for seven adenosine residues. When the dsDNA region and the following ssDNA region in a 3'-end are ignored, C8 and C36 are symmetrically located as to each other with respect to the center of the remaining ssDNA region of S_{ds} . Thus, the concentration of A3B-CTD that reaches either C8 or C36 of S_{ds} should be basically the same, which rationally explains the results. It should be added that the deaminase activity increased for both C8 and C36, when the duplex was introduced (Figure 2-6d). The introduced duplex may repress the flexibility of a whole DNA including the single-stranded regions to some extent. This may enhance the deaminase activity through the stabilization of the complex. In fact, the similar enhancement of the activity at two different sites on introduction of the duplex was observed for A3G as well [83].

Observation of similar deaminase activity of A3B-CTD for a target sequence located close to either the 5'- or 3'-end revealed that A3B does not exhibit 3'→5' polarity. We previously observed 3'→5' polarity for A3G-CTD[83]. We discussed that a bump hanging over one side of the catalytic pocket may be responsible for the 3'→5' polarity (Figure 2-8a). The surface around the catalytic pocket of A3B-CTD does not have a bump (Figure 2-8b), having no bump as an obstacle, no matter from which direction A3B-CTD approaches the target motifs on ssDNA, the cytosine within the target motif may always be accommodated into the catalytic pocket. The bump around the catalytic pocket is not present for the chimera A3F-CTD bound to ssDNA[65]. Therefore, we suppose that this structural feature of A3B-CTD causes the lack of the 3'→5' polarity.

Recently, it has been reported that A3B and other A3 family proteins preferentially deaminate the lagging strand template during DNA replication [87-89], although it is supposed that single-stranded DNA binding proteins cover the DNA during the replication and that so the accessibility of ssDNA segments to A3B and other enzymes is rather blocked. In the course of DNA replication, for the lagging strand template, there transiently exist multiple short ssDNA regions that are not base-paired by Okazaki fragments and exposed to the solvent [90]. The target sequence present in these short ssDNA regions of the lagging strand template may be effectively deaminated by A3B that shows higher activity toward its target sequence in short ssDNA, resulting in the

genomic DNA lesion related to cancers. The unique properties of A3B may have such a biological relevance.

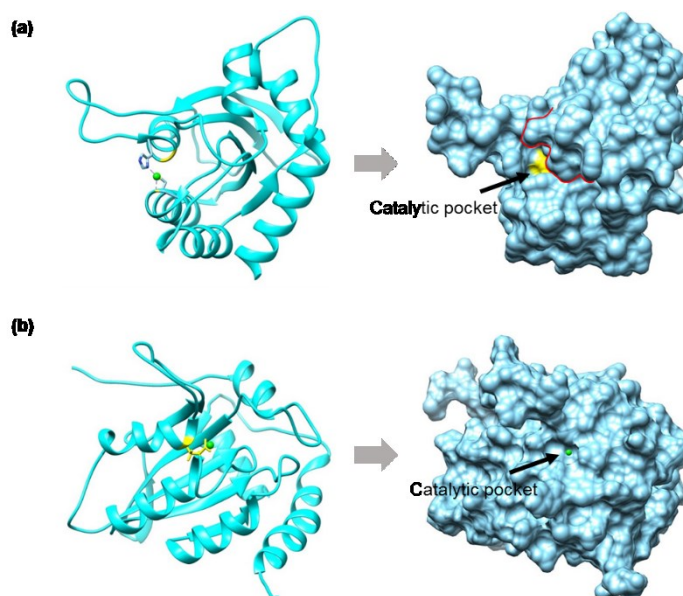


Figure 2-8 Structural comparison of A3G-CTD and A3B-CTD. (a) The catalytic pocket of A3G-CTD (PDB: 3E1U) includes a zinc ion (green) and Glu259 (yellow), which are important for the activity. There is a putative bump near the catalytic pocket of A3G-CTD (highlighted by a red line). (b) The catalytic pocket of A3B-CTD (PDB: 2NBQ) includes a zinc (green) and Glu255 (yellow). Glu255 cannot be seen on the surface due to the smaller catalytic pocket. The surface near the catalytic pocket is flat and the putative bump observed in A3G-CTD is not present.

2.4 Conclusions

In this chapter, we investigated the characteristics of the enzymatic reaction of A3B-CTD by combining a real-time NMR method, UDG assay, EMSA assay, and FA measurement. It turned out that the length- and location-dependence of the deaminase activity of A3B-CTD are quite different from those of A3G-CTD. On the basis of the weak affinity of A3B-CTD to ssDNA, these characteristics were rationally interpreted. The biological relevance of the unique characters of A3B-CTD is implied. Finally, our real-time NMR method should be applicable to other enzyme systems to provide us with new insights into the behavior of enzymes.

temperature was set to 20°C and a final concentration of 0.3 mM isopropyl β -D-1-thiogalactopyranoside was added at $OD_{600} = 0.6$. Then the cell culture was incubated for another 20 h and harvested. The precipitated cells were re-suspended in lysis buffer (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 5 mM 2-mercaptoethanol, 10 % (v/v) glycerol, 4 mM $MgCl_2$, and 20 mM imidazole) supplemented with 5 units mL^{-1} DNase I, 1 $\mu g\ mL^{-1}$ RNase H, and 0.2 mg mL^{-1} lysozyme. Subsequently, the cell suspension was sonicated and the insoluble fraction was removed by centrifugation at 50,000 g for 30 min at 4°C. The soluble fraction was then applied to an Ni-Sepharose resin column (GE healthcare), washed with 50 mM Tris-HCl buffer (pH 8.0) containing 500 mM NaCl, 5 mM 2-mercaptoethanol, 10 % (v/v) glycerol, and 20 mM imidazole, and then eluted with 50 mM Tris-HCl buffer (pH 8.0) containing 500 mM NaCl, 5 mM 2-mercaptoethanol, 2 % (v/v) glycerol, and 250 mM imidazole. The fractions containing A3B-CTD were pooled and concentrated with Amicon Ultra-15 centrifugal filter units (MWCO, 3500 kDa; Merck Millipore), and further purified on a HiLoad 16/60 Superdex 75 PG column (GE healthcare), equilibrated with 20 mM Tris-HCl buffer (pH 8.0) containing 100 mM NaCl and 5 mM DTT. The expression and purification of A3F-CTD were performed as previously described with modifications [56]. The purities of A3B-CTD and A3F-CTD were confirmed by sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE), both of which reached approximately 98%.

2.5.3 *In-vitro* cytosine deamination assay

The deamination assay of A3B-CTD against ssDNA substrates was performed as described previously[91], described hereinafter as UDG assay. The ssDNA substrates were dissolved in deamination buffer (15 mM Tris-HCl (pH 7.5), 50 mM NaCl and 5 mM DTT) to give a final concentration of 2 μM , and A3B was added to make the molar ratio of A3B-CTD : ssDNA = 1:4. The deamination reaction was undertaken at 37°C for 30 min. The reaction was terminated at 95°C for 5 min. Then, two units of UDG (New England Biolabs) were added to the solution, followed by further incubation at 37°C for 1 h. Subsequently, a final concentration of 0.2 M NaOH was added, followed by incubation at 37°C for 20 min. Finally, the equal volume of 2 $\times$ loading buffer (42 % (w/v) urea, 0.25 % (w/v) bromophenol blue) was added to the reaction and incubated at 95°C for 5 min. The reaction products were resolved on a 20 % urea-polyacrylamide

(Urea-PAGE) gel (Table 2.2 and note), and the obtained bands were visualized using a Pharos FXTM Molecular Imager (BIO-RAD) and quantified with Quantity One 1D analysis software. The fraction of a deaminated product as a percentage was calculated as the ratio of product intensity over total intensity. Error bars were obtained through three independent experiments.

Table 2.2 Competent of 20% Urea-PAGE (10 mL)

competent	amount
urea	4.2 g
40 % acryamide (19:1)	5 mL
10 × TBE*	1 mL
10% ammonium persulphate	100 µL
TEMED	3 µL

* **100 mL 10 ×TBE:** 10.9 g tris (hydroxymethy) aminotongue, 5.56 g boric acid, 0.04 M EDTA-3Na

Note:

1. According to different length of nucleic acid, different concentration of gels are used: 20 % for <10 mer, 20 - 16 % for 10 - 20 mer, 16 - 12 % for 20 - 30 mer, 12 - 10 % for 20 - 40 mer, 10 - 8 % for 40 - 70 mer.
2. Before loading the samples into the wells of gels, pre-running the gel at 15 mA for 30 min (for one gel), then 20 mA for 30 - 60 min after loading the samples.

2.5.3 Annealing oligonucleotides to ssDNA

Equal concentrations of S_{ss} (Table 2.1) and 20 nt oligonucleotides (5'-TACACCACCATAACCATACA-3') (15 µM) were mixed in buffer of 10 mM TrisHCl (pH 8.0), 50 mM NaCl, and 1 mM EDTA (pH 8.0). The procedures of annealing was shown in Figure 2-9a. After annealing, the product was dialyzed against ddH₂O using GeBaFex-tube (1 kDa MWCO, Gene Bio-Application L. T. D, Israel), and then concentrated. The concentrated product was checked using native PAGE as described in Table 2.2 without adding urea. The native PAGE result was shown in Figure 2-9b. From Figure 2-9b, the band of S_{ds} immigrates slower than S_{2-1} , indicating that the 20 nt oligonucleotide was annealed into S_{ss} successfully.

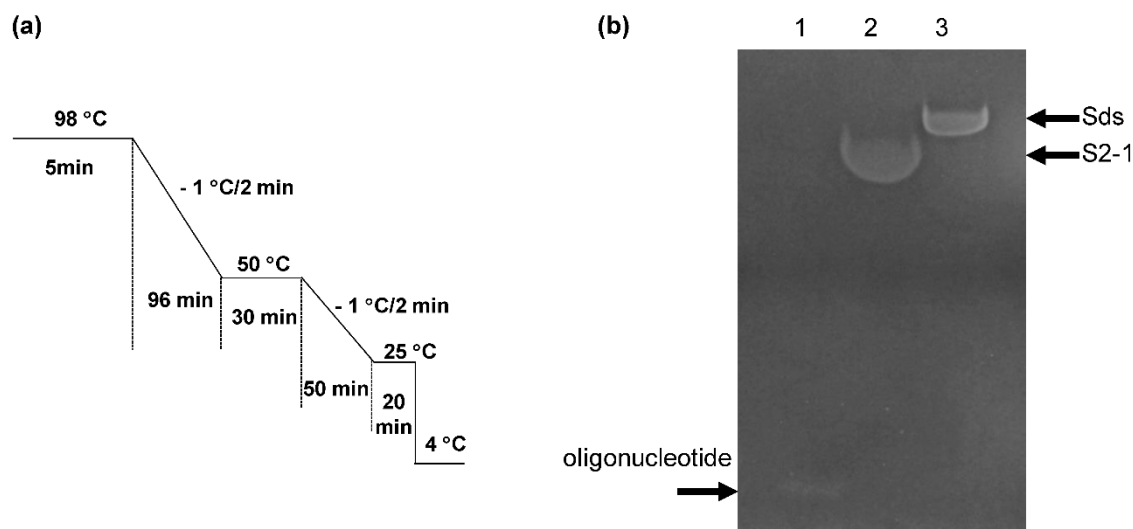


Figure 2-9. Annealing oligonucleotide to ssDNA. (a) The procedure for annealing. (b) Checking the product using native PAGE (20%). Lane 1, oligonucleotide; Lane 2, ssDNA S_{2-1} ; Lane 3, the product S_{ds} .

2.5.4 Real-time NMR spectroscopy

All the ssDNA substrates used for NMR experiments (Table 2.1) were synthesized, purified, and desalted by FASMAC Co., Ltd. (Kanagawa, Japan). S_{ds} , in which a 20 nt-dsDNA segment was introduced within the 3'-region of S_{ss} , was prepared by 2.5.3. Each ssDNA substrate was dissolved in 15 mM Tris-HCl buffer (pH 7.5) containing 50 mM NaCl and 5 mM DTT to give a final concentration of 200 μ M. The deamination reaction was started by adding A3B-CTD to this solution in the ratio of A3B-CTD:ssDNA = 1: 200. All NMR data were recorded using a Bruker AVANCE III 600, equipped with a Z-axis gradient cryoprobe (Bruker Biospin). Sequential measurement of 1H - 1H two-dimensional Total Correlation Spectroscopy (2D TOCSY) spectra was performed at 298 K with a mixing time of 20 ms to monitor the time course of the reaction. The signal intensities of the H5-H6 correlation peaks of the cytosines were quantified with Sparky [92]. The real-time NMR data were fitted to $I(t) = I_0 \exp(-kt)$, where I_0 is the initial intensity and k is the deamination rate constant [48]. I_0 , k , and error bars were obtained by nonlinear least-squares analysis [93].

2.5.6 Electrophoretic mobility shift assay

Each of the FITC-labeled ssDNA substrates of various lengths (Table 2.1) was

dissolved individually in 15 mM Tris-HCl buffer (pH 7.5) containing 50 mM NaCl and either 1 mM TCEP (for A3F-CTD) or 5 mM DTT (A3B-CTD). Either A3B-CTD or A3F-CTD was added to each of these solutions. The concentration of the proteins was either 1 μ M or 3 μ M and that of FITC-labeled ssDNA substrates was 100 nM in a 10 μ L volume. These reaction solutions were incubated at 25°C for 1 h. Aliquots of each solution were loaded onto a 20% native polyacrylamide gel comprising 95 mM Tris-borate and 4 mM EDTA. Pre-running the gel at 4 °C (15 mA for one gel) without loading the samples, and then the gels were run at 4°C (15 mA) until the bromphenol blue dye had migrated about 2/3 through the gel. The obtained bands were visualized using a Pharos FXTM Molecular Imager (BIO-RAD) and quantified with Quantity One 1D analysis software.

2.5.7 Steady-state fluorescence anisotropy measurement

Formation of a complex between either A3B-CTD or A3F-CTD mutants with 49 nt FITC-labeled ssDNA ($S_{\text{FITC}_{49}}$, Table 2.1) was monitored by means of steady-state fluorescence depolarization using an FP-8500 spectrofluorometer (JASCO Asia Portal). 400 nM $S_{\text{FITC}_{49}}$ was dissolved in 15 mM Tris-HCl buffer (pH 7.5) containing 50 mM NaCl, and 5 mM DTT. These DNA solutions were titrated with either A3B-CTD or A3F-CTD mutants at 25°C. A sample at each titration point was excited with vertically polarized light at 490 nm, and both vertical and horizontal emission were monitored at 520 nm. Fluorescence anisotropy (r) is defined as: $r = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2I_{\perp})$, where $I_{\parallel}$ is the intensity of emission parallel to the plane of the excitation light and $I_{\perp}$ is the intensity of emission perpendicular to the plane of excitation. The r value was plotted against the ratio of protein to $S_{\text{TTCA}_{49}}$.

Chapter 3

Influence of the DNA sequence and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F

3.1 Introduction

As one of the double-domain proteins, A3F is recognized to have efficient anti-HIV-1 activity as A3G. A3 proteins show their anti-HIV-1 activities via deaminase activities. A3 proteins act on ssDNA and deaminate cytosine into uracil, leading to the hypermutation in ssDNA, which disrupts the genome. Therefore, it is significant to study the interaction between A3 proteins and nucleic acid. Up to now, several groups have reported the interface of A3A or A3G with ssDNA using NMR titration or site-mutagenesis analysis [43, 44, 48, 52]. And there are also studies on the binding abilities of A3B and A3F to ssDNA [14, 49]. The co-structures of A3A:ssDNA[65-67], chimera A3B-CTD:ssDNA[65], as well as A3G-NTD:ssDNA[61], have also become available, such as, however, little information on deamination characteristics of A3F. Since it is extremely difficult to express and obtain soluble and stable full-length A3F or A3F-CTD, it becomes challengeable to solve the structure of the A3F:ssDNA complex by X-ray or NMR method. To clearly elucidate the deamination characteristics of A3F, in this study, we have systematically analyzed the properties of the deaminase activity of A3F-CTD, together with the roles of the amino acid residues located close to the catalytic center in deaminase activity and ssDNA binding. The current study allows the basic characterization of the deamination by A3F-CTD, together with a deeper insight into the roles of key amino acid residues in A3F in deamination.

3.2 Results

3.2.1 Preparation of an active A3F-CTD

A soluble and stable A3F-CTD is difficult to express and purify using an *Escherichia coli* expression system. Previously, L-Arginine (300 mM) was used to improve the solubility and stability of an A3F-CTD, which had an N-terminal GST tag attached [56]. In this study, we simplified the purification method as described in 3.5.2. We expressed the wt A3F-CTD connected with an N-terminal hexahistidine-tag in *Escherichia coli*,

and purification was performed using a high concentration of L-Arginine (300 mM). Although the yield of the A3F-CTD was low, the obtained A3F-CTD was active, as evaluated by *in vitro* cytosine deamination assay (Figure 3.1a). Following the same method, the A3F-CTD mutants used for our studies were expressed and purified.

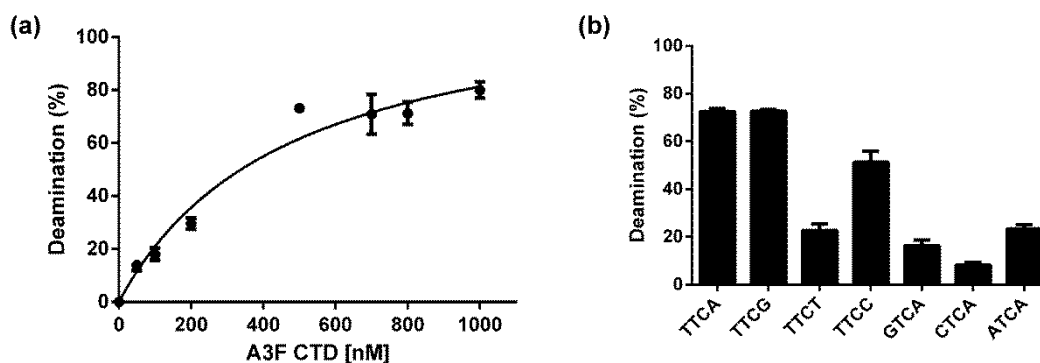


Figure 3-1. The dependence of A3F-CTD deaminase activity on the sequence of the substrate. (a) The deamination percentage (%) with A3F CTD at 1 h was obtained as described under Experimental Procedures and plotted against increasing amounts of A3F-CTD. (b) The deamination percentage with A3F-CTD at 1 h for 49 nt ssDNA substrates containing the following sequences at the center: TTCA, TTCG, TTCT, TTCC, GTCA, CTCA, and ATCA. 1,000 nM A3F-CTD and 100 nM ssDNA were used as substrates.

3.2.2 Sequence context-dependent deaminase activity

A3F has a strong preference for for 5'-TC-3' dinucleotides (TC), but not all TC sites can be deaminated by A3F with equal efficiency [94, 95]. To identify the most preferred sequence for A3F-CTD's deamination, we synthesized seven 49 nt ssDNAs containing a centrally located TC motif flanked by various sequences (Table 3.1). Firstly, the nucleotide on 5' side of TC was fixed as T, and that on 3' side were changed to A, G, T or C, respectively. As shown in Figure 3-1b, TTCA and TTCG exhibited higher activities than other two sequences. Replacement of the nucleotide on the 3' side with T/C resulted in reduction of deamination, suggesting A/G are preferred over T/C on the 3' side. Then the nucleotide on 3' side of TC was fixed as A, and that on 5' side was changed to A, G, T or C, respectively. Compared with TTCA, the replacement of the nucleotide on the 5' side with G, C or A resulted in ~4.5, ~9 and ~3-fold decreases in deamination, respectively. These results indicate that A3F-CTD most prefers TTCA/G

as the deamination target sequences.

3.2.3 Roles of amino acid residues in loop 1

To evaluate the roles of the amino acid residues that are close to the catalytic center (Figure 3-2), we examined the deaminase activities and ssDNA-binding affinities of the mutant A3F-CTD proteins by means of UDG-based cytosine deamination assay and steady-state fluorescence anisotropy measurement. In A3 proteins, the conformation of loop 1 is reportedly important for substrate recognition [11, 96, 97]. Here, we assumed that the three positively charged amino acid residues in loop 1, R208, K209 and R213 (Figures 3-2 and 3-3a), may be responsible for binding to ssDNA, which has negatively charged phosphate groups. We replaced each of these positively charged amino acid residues with a negatively charged glutamate residue, to obtain mutants, R208E, K209E and R213E (Figure 3-3b), and examined the deaminase activity of each mutant (Figure 3-3c). Compared to the wt A3F-CTD, the R208E and K209E mutants exhibited ~10-fold and ~2-fold reduction of the deaminase activity, respectively. The R213E mutant, in which the arginine that is well conserved throughout A3 family proteins was replaced, exhibited the minimal detectable deaminase activity. The ssDNA-binding assay showed that all three mutants exhibit reduced ssDNA-binding affinities compared to the wt (Figure 3-3d). The deaminase activity is determined by two factors, ssDNA-binding affinity and catalytic efficiency. Thus, the reduced deaminase activities of these three mutants are supposed to be due to their reduced ssDNA-binding affinities.

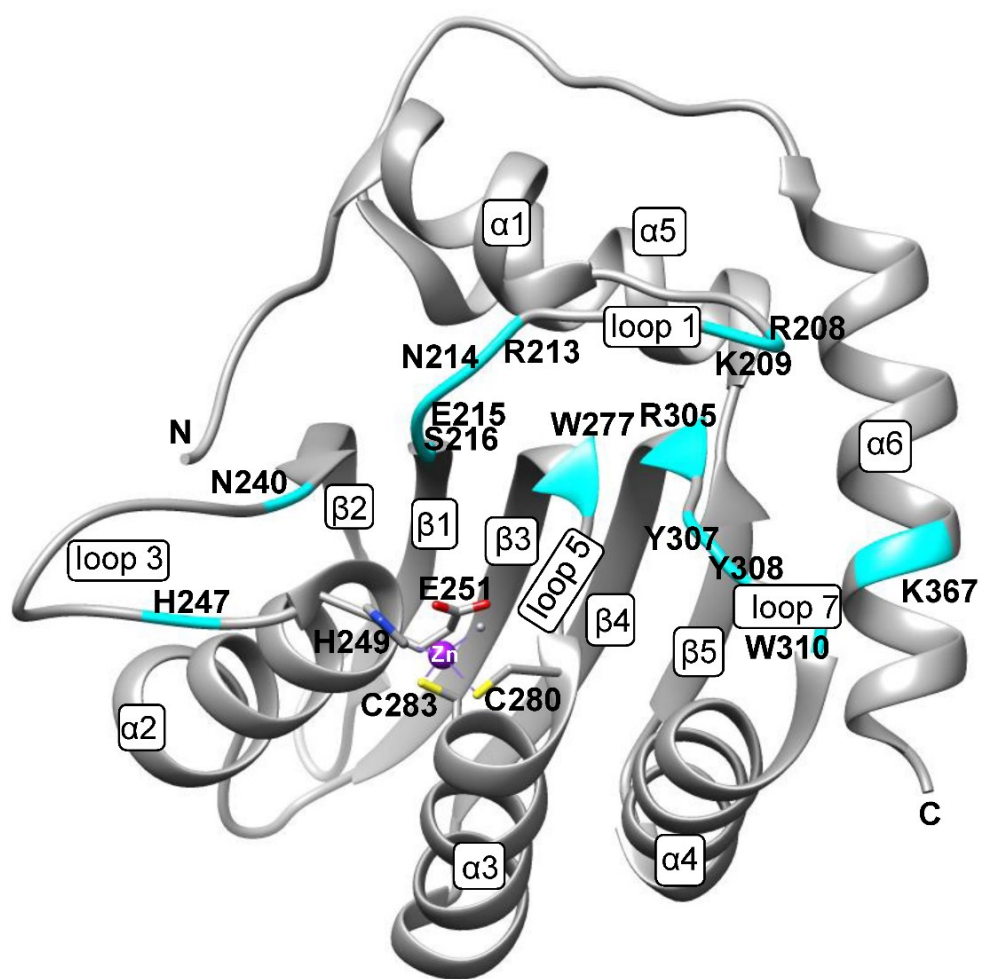


Figure 3-2. The residues of A3F-CTD that were mutated in this study are shown in cyan on the crystal structure of A3F-CTD (PDB: 3WUS). A purple sphere indicates the catalytic zinc coordinated by H249, C280, C283, and a water molecule. The catalytic residue, E251, is also shown.

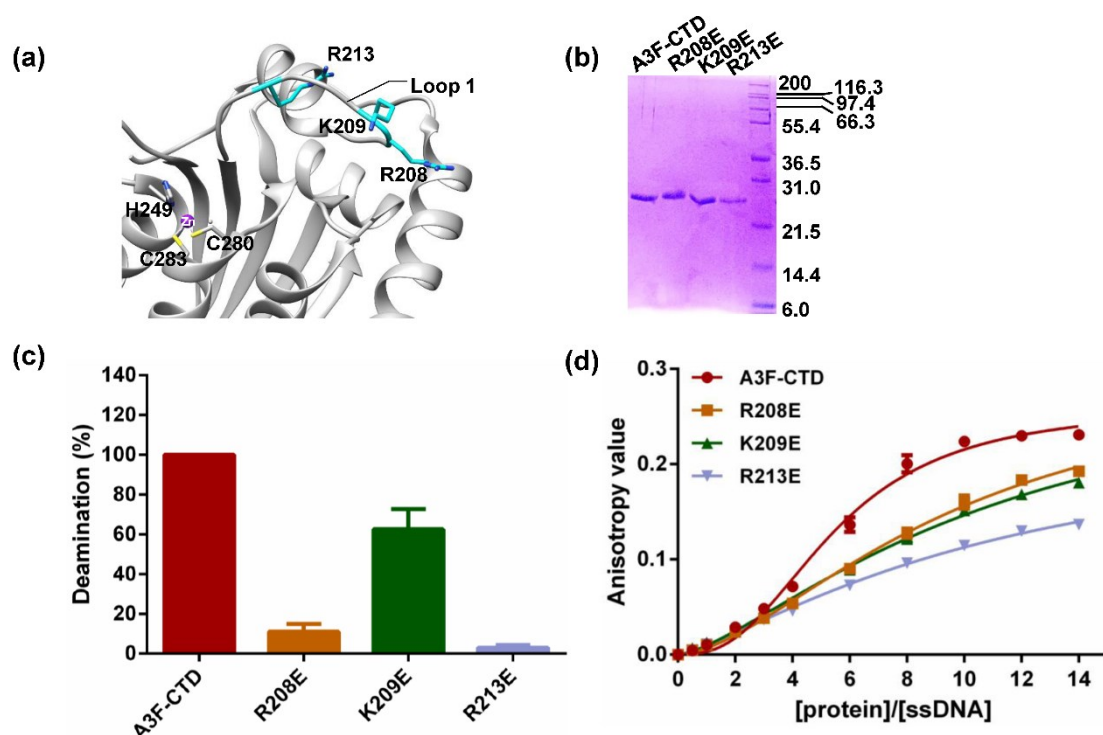


Figure 3-3. The roles of the three amino acid residues in loop 1 in deaminase activity and binding affinity. (a) A close-up view of the A3F-CTD structure (PDB ID: 3WUS). The positions of the three amino acid residues in loop 1 mutated in this study are colored cyan. (b) SDS-PAGE analysis of the purified wt A3F-CTD, and the R208E, K209E and R213E mutants are shown. Normalized deamination percentage at 1 h(c) and binding (d) of the A3F-CTD mutants in loop 1 with 49-TTCA substrate. Error bars were obtained through three independent experiments.

CD spectra of R208E and R213E for which the significant decrease in the deaminase activity was found were recorded (Figure 3-4). These CD spectra were basically similar to that of wt A3F-CTD. This suggested that overall folding was preserved even for these mutants and that the decrease in the activity was not due to destabilization of overall folding by mutation. Similarly, CD spectra were also recorded for mutant A3F-CTDs for which the significant decrease in the deaminase activity was found in the following sections (Figure 3-4). For all of them, their CD spectra were basically similar to that of wt A3F-CTD. This suggested that the decrease in the activity for these mutants was not due to destabilization of overall folding by mutation, either.

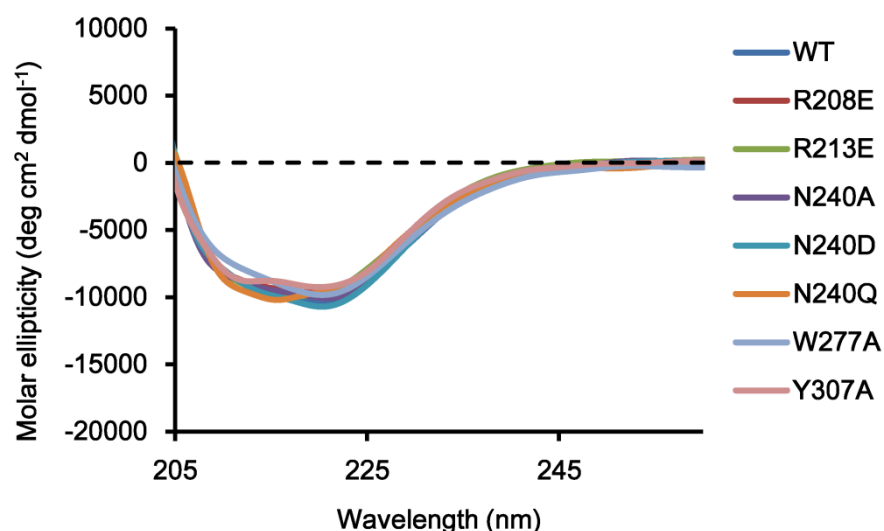


Figure 3-4. CD spectra of wild type and mutant A3F-CTDs whose deaminase activities were significantly low.

3.2.4 Roles of amino acid residues in loop 3

Loop 3 in A3 proteins reportedly plays critical roles in deaminase activity[31, 44, 52]. In this study, we focused on two special amino acid residues located in loop 3 of A3F-CTD, N240 and H247 (Figures 3-2 and 3-5a). N240, located at the N-terminal end of loop 3 is highly conserved throughout the zinc-dependent cytosine deaminase family (Figure 3-5b, left panel). In other zinc-deaminases, the equivalent residues N244 of A3G-CTD, N42 of TadA, and N54 of CDA, are in direct contact the target substrate [44, 68, 98]. To investigate the role of N240 in A3F, three mutants, N240A, N240D and N240Q, were prepared, respectively (Figure 3-5b, right panel). All three mutants showed very weak deaminase activity: N240D, which had lost the amino group, and N240Q, which had gained a methylene group, retained only ~4.5% of the wt A3F-CTD's activity; N240A, which had lost the functional groups, lost the activity completely (Figure 3-5c). These results strongly indicate that the side chain of the N240 residue is critically important for the deaminase activity

As shown in Figures 3-2 and 3-5a, H247 is close to the zinc-binding site, and the deaminase activity increased on substitution of this histidine into arginine (Figure 3-5c).

The deaminase activity of another mutant, H247A (Figure 3-5b, right panel), was ~70% of that of the wt A3F-CTD (Figure 3-5c). All the mutants, as to the N240 and H247 residues showed slight changes in the ssDNA-binding affinity (Figure 3-5d). This indicates that the deficient deaminase activity of these mutants is caused by defective catalytic efficiency.

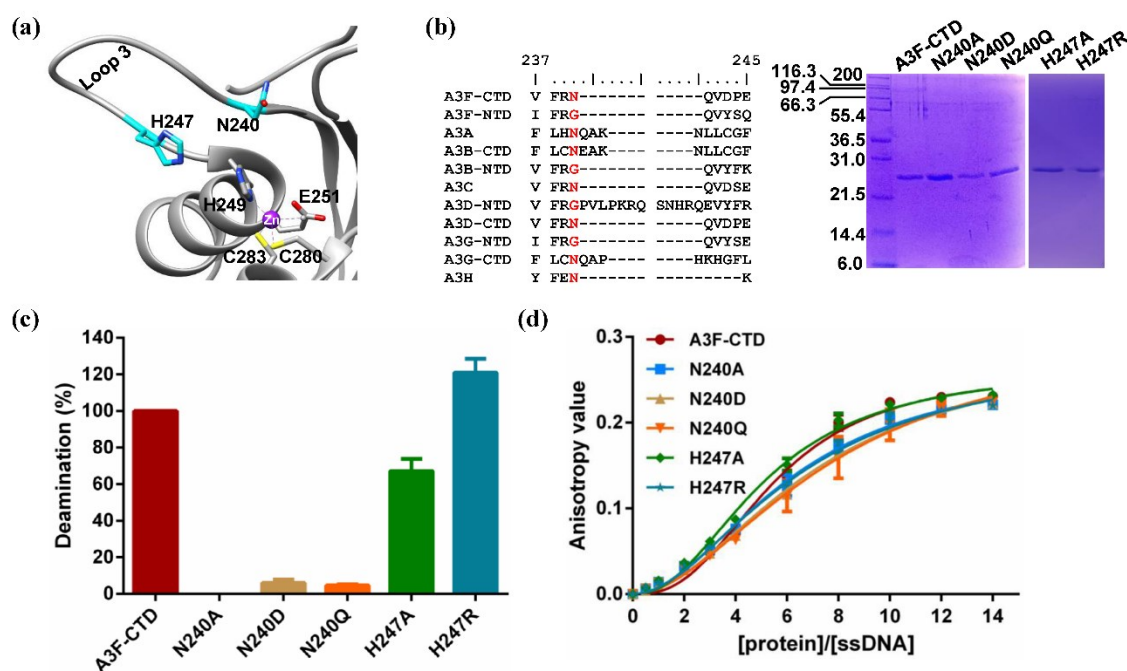


Figure 3-5. The roles of the two amino acid residues in loop 3 in deaminase activity and binding affinity. (a) A close-up view of the A3F-CTD structure. N240 and H247 in loop 3, mutated in this study, are colored cyan. (b) Sequence alignment of the residues in loop 3 of the human A3 proteins is shown on the left: A3F numbering is shown at the top, the conserved Asn residue is shown in red. SDS-PAGE analysis of the purified proteins is shown on the right. Normalized deamination percentage at 1 h(c) and binding (d) of the A3F-CTD mutants in loop 3 with 49-TTCA substrate. Error bars were obtained through three independent experiments.

3.2.5 Roles of amino acid residues in or near loop 7

Loop 7 has been suggested to be the primary determinant of specific recognition of a DNA substrate by A3 proteins. When loop 7 was swapped between two different A3 proteins, the DNA sequence specificity was also exchanged [55, 64]. Here, we focused on five amino acid residues in or near loop 7, W277, Y307, Y308, W310 and K367 (Figures 3-2 and 3-6a and b). Alanine mutagenesis of W277, Y307 and Y308 resulted in

no detectable or minimal deaminase activity (Figure 3-6c), suggesting the side chains of these amino acid residues are critical for the deaminase activity. The ssDNA-binding affinities of Y307A and Y308A were the same as that of the wt A3F-CTD, while that of W277A was just slightly weaker (Figure 3-6d), This indicates that the decreased deaminase activities of these mutants are due to the defect in the catalytic efficiency.

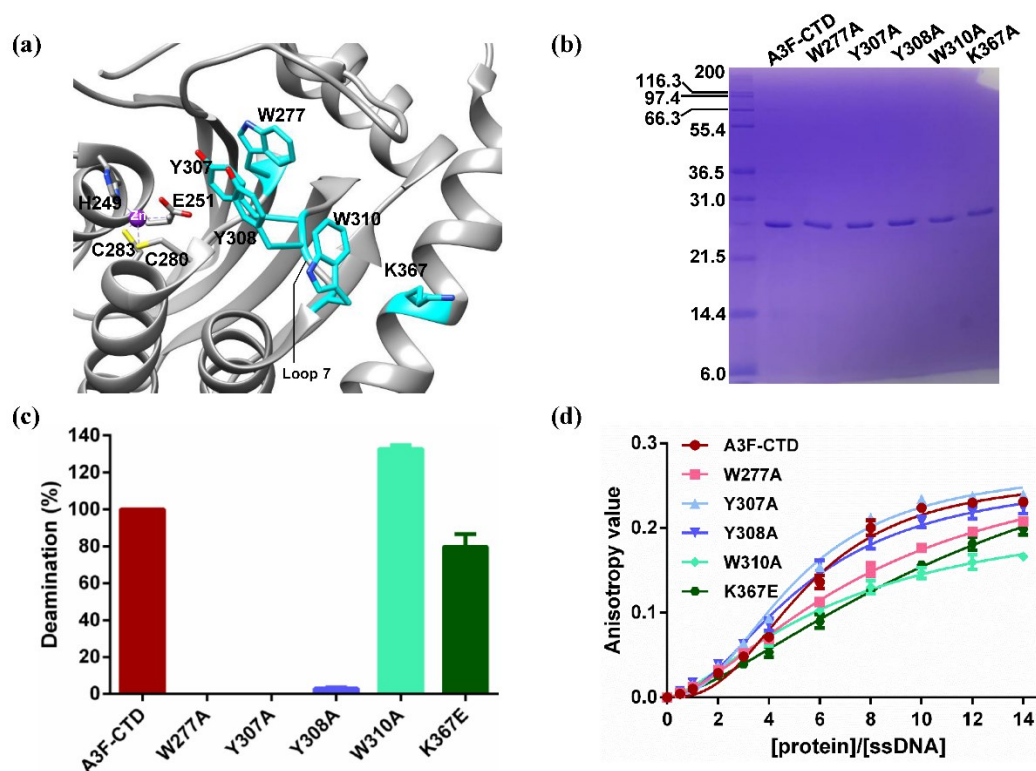


Figure 3-6. The roles of the five amino acid residues in or near the loop 7 in deaminase activity and binding affinity. (a) A close-up view of the A3F-CTD structure. W277 in $\beta 3$, three amino acid residues in loop 7 (Y307, Y308, and W310), and K367 in $\alpha 6$, colored cyan, were mutated in this study. (b) SDS-PAGE analysis of the purified proteins. Normalized deamination percentage at 1 h(c) and binding (d) of the A3F-CTD mutants in or near loop 7 with 49-TTCA substrate. Error bars were obtained through three independent experiments.

W310, located close to the active site of A3F-CTD, also plays an important role in the deaminase activity, as it was shown that the W310A mutant exhibits higher activity than the wt A3F-CTD (Figure 3-6c). This increased activity of the W310A mutant is in contrast with previous finding that the W310A mutant exhibited decreased deaminase activity [14]. This discrepancy could be due to the tag attached to the A3F-CTD. In our study, the A3F-CTD W310A mutant possessed an N-terminal hexahistidine-tag,

whereas in Chen *et al.*'s study, A3F-CTD W310A mutant possessed a large MBP tag. In both cases, ssDNA-binding affinity decreased (Figure 3-6d). This suggests that in our case, the increase in catalytic efficiency overcame the decrease in ssDNA-binding affinity, resulting in higher deaminase activity.

We also found that a positively charged K367 residue, located slightly far away from the core structure, plays a role in the function of A3F-CTD. Replacement of K367 with a negatively charged glutamate resulted in reduction of both deaminase activity and ssDNA-binding affinity (Figure 3-6c, d). This suggests that the K367 residue contributes to the deaminase activity through binding to ssDNA.

3.2.6 Deaminase activity of phosphomimetic A3F-CTD mutant

Activation -induced deaminase (AID) and nearly all the A3 proteins have a position in loop 1 at which either threonine or serine is highly conserved (Figure 3-7a, left panel). The threonine/serine at this position is important for these proteins to be regulated under physiological conditions [99]. AID is reportedly phosphorylated at T27 in loop 1 by protein kinase (PKA), by which its deaminase activity is directly suppressed [99]. Similarly, PKA-induced negative regulation is also known for A3G, in which case T218 in loop 1 is the corresponding site [99]. The PKA phosphorylation site is highly conserved, which is RxxT/S [100], and found also in A3F-CTD (RNES), in which S216 is the putative phosphorylation site. To address the possibility that A3F-CTD's deaminase activity could be similarly regulated through phosphorylation of S216 residue, two A3F-CTD mutants, S216A and S216E, were prepared (Figure 3-7a, right panel). The S216A mutant showed slightly higher activity than the wt A3F-CTD (Figure 3-7b). The phosphomimetic mutant, S216E, exhibited low deaminase activity (Figure 3-7b). This would be partly due to the decreased ssDNA-binding affinity (Figure 3-7c).

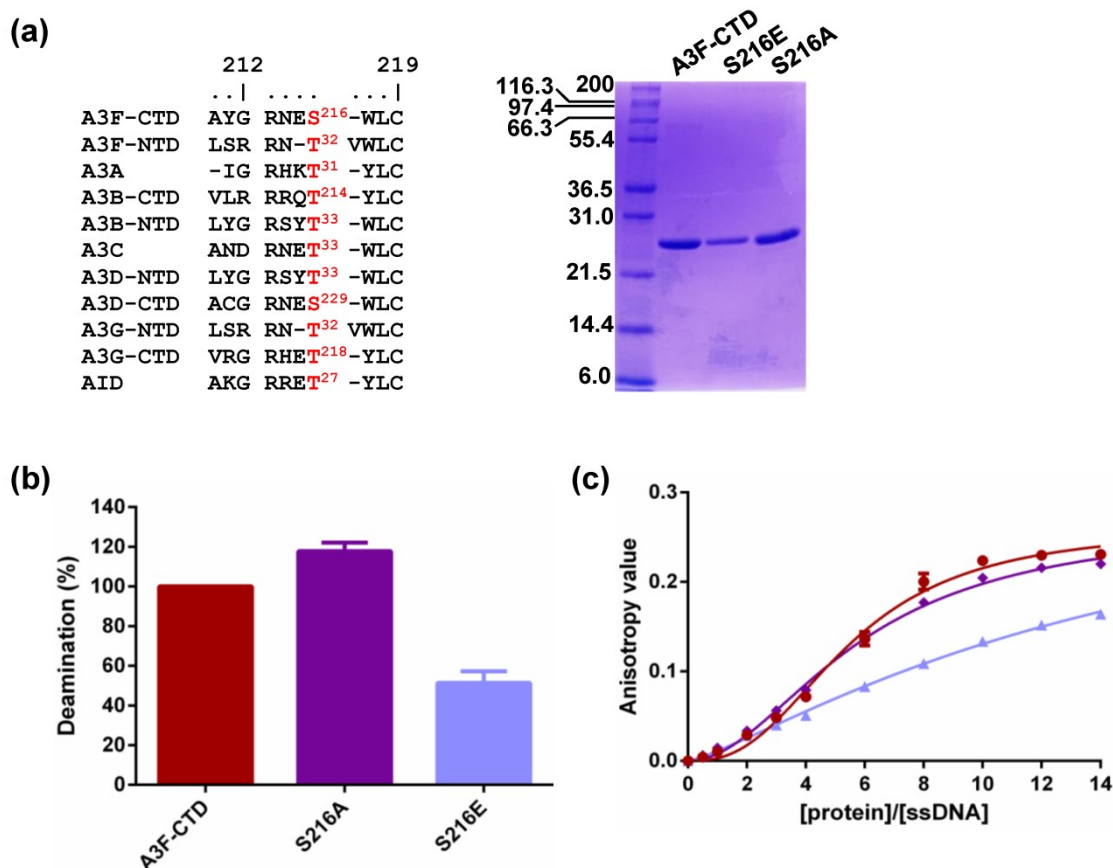


Figure 3-7. The effects of A3F-CTD-S216 mutants on deaminase activity and ssDNA-binding. (a) Sequence alignment of a part of loop 1 of the human A3 proteins and AID (left). A3F numbering is shown at the top. The conserved T/S residue is highlighted in red. SDS-PAGE analysis of the purified mutants is also shown (right). Normalized deamination percentage at 1 h (b) and binding (c) of the S216 mutants with 49-TTCA substrate. Error bars were obtained through three independent experiments.

3.2.7 The effect of pH on A3F-CTD catalysis

A3 proteins have different subcellular localizations playing various biological functions in different physiological phases [101], which suggests that the biological functions of A3 proteins are closely linked to the physiological pH. To test the effect of pH on A3F deamination, a 49 nt Alexa 488-labeled ssDNA (49-TTCA-A488, Table 3.1) containing a centrally located TTCA sequence was used as a substrate to measure A3F-CTD activity over the pH range of 5.0 to 9.5. The deamination rate of A3F-CTD increased nearly 3-fold when the pH was increased from 5.0 to 5.5, and then stayed constant from pH 5.5 to 9.5 (Figure 3-9). Considering the theoretical isoelectric point of

A3F-CTD is 6.25, as calculated with ProtParam [102], it is supposed that the binding affinity of A3F-CTD for negatively charged ssDNA is unaffected by the charge of A3F-CTD. The decrease in the deamination rate at pH 5.0, where A3F-CTD is positively charged, may be caused by the partial denaturation of A3F-CTD.

Protonation/deprotonation of the side chains of the aspartate, glutamate, histidine, lysine and arginine residues are reportedly responsible for the pH-dependent deaminase activities of A3 proteins [103]. Among these five amino acid residues, only E215 and H247 are spatially close to the catalytic center (Figure 3-2). Then, we performed site-directed mutagenesis at these positions and obtained two A3F-CTD mutants, E215Q and H247R. The deamination rates of both the mutants at pH 5.0 were the same as that of the wt A3F-CTD. But at pHs higher than 5.0, the rates increased 2.9 - 3.6 fold and 1.5 - 2.0 fold for E215Q and H247R, respectively. However, there was no drastic change in pH dependency for them (Figure 3-8). Thus, manipulation of either E215 or H247 was not effective to change pH dependency of the wt A3F-CTD.

In A3G, H216 plays a pivotal role in the regulation of the deaminase activity; the imidazole ring of the H216 residue is protonated and deprotonated through a pH change, by which the deaminase activity changes [103]. Here, we mutated the corresponding residue of A3F-CTD, N214, to a histidine, obtaining N214H. The deamination rate of N214H was ~27-fold higher than that of the wt A3F-CTD (Figure 3-8). Since pH 5.5 is close to the pKa of the imidazole ring of histidine, we assume that this pH dependency of the N214H mutant is related to protonation/deprotonation of histidine; the protonation of the N214H mutant might have had a positive effect on the catalytic efficiency.

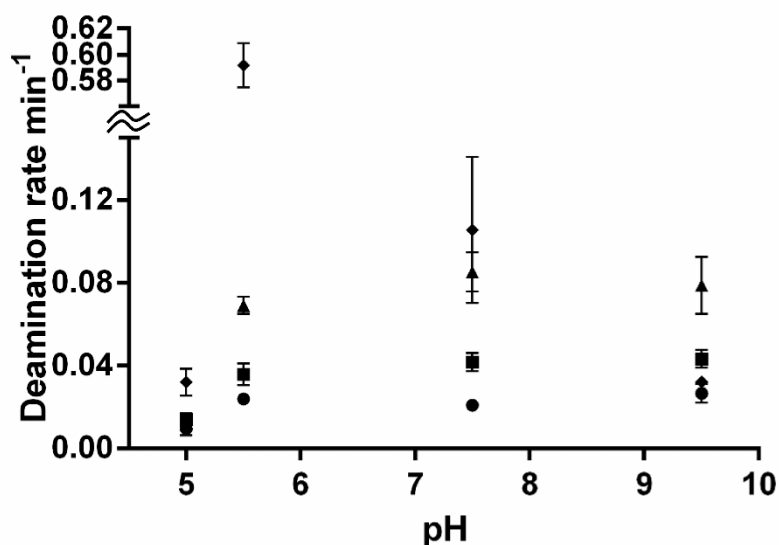


Figure 3-8. pH dependence of the rate of deamination of A3F-CTD. 49 nt Alexa 488-labeled ssDNA (100 nM) was incubated with 1,000 nM A3F-CTD (circles), E215Q (triangles), H247R (squares), and N214H (diamonds), respectively.

3.3 Discussion

Although the structure of ssDNA free A3F-CTD is available, that of an A3F-CTD:ssDNA complex has not yet been solved. Currently, the co-crystal structures of the A3A:ssDNA and A3B-CTD:ssDNA complexes are available [65, 66], both of which may provide a structural basis to rationally understand the interaction between A3F-CTD and ssDNA. In the co-crystal structure of A3B-CTD:ssDNA, however, loop 3 of A3B-CTD was deleted. Therefore, this complex is not suitable for use for the interpretation of our results. It turned out that A3F-CTD can be superimposed well on A3A of the A3A:ssDNA complex (Figure 3-9a). Thereby, after the superposition of A3F-CTD onto the A3A:ssDNA complex, the coordinate of A3A was deleted, by which the structural model of the A3F:ssDNA complex was constructed (Figure 3-9b).

Several A3F-CTD mutants, R208E, K209E and R213E, of loop 1, W277A of β 3, W310 of loop 7, and K367 of α 6, exhibited decreased ssDNA-binding affinities compared to the wt A3F-CTD, which could be one of the reasons why the deaminase activity is affected. The model of the A3F-CTD:ssDNA (Figure 3-9b) shows that the nucleotide at the -1 position is inserted into the groove formed by loop 1 and loop 7.

The R208, K209, R213, W277 and W310 residues are located in the proximity of ssDNA (Figure 3-9b), and it seems reasonable that the mutations at their positions resulted in decreases in ssDNA-binding. The position of the K367 residue seems far away from the position of ssDNA in the current model. However, recently, another structural model of an A3A:ssDNA complex was constructed on the basis of the NMR method and binding data[67], which showed that the nucleotide at the -2 position could approach $\alpha 6$ of A3A. This suggests that K367 from $\alpha 6$ of A3F-CTD may also be involved in ssDNA-binding, and thus the mutation at K367 could result in decreased ssDNA-binding affinity, as observed in Figure 3-6d.

The N240A/D/Q, H247A/R, Y307A, and Y308A mutants exhibited alternation of deaminase activity but no change in ssDNA-binding affinity, which indicates that the catalytic efficiency is affected by these mutations. The position of the side chain of N240 is very close to the sugar and 3' phosphate group at position 0 of ssDNA; the side chain of N240 may play a role in positioning of the target DNA sequence in a proper orientation for efficient deamination. H247 may restrict the conformation of loop 3 to guarantee the access of ssDNA to the catalytic center. The Y307 and Y308 residues, as well as the R305 and W277, are near the catalytic center. The side chains of the tyrosine and tryptophan residues form a hydrophobic patch. The aromatic side chains of Y307 and W277 are stacked on top of each other, forming a stacked ring. To this stacked ring, the backbone moieties of the nucleic acids at the -1 and 0 positions (the sugar moiety and the 3'-phosphate of the nucleotide at the -1 position, and the sugar moiety of the nucleotide at the 0 position) wrap around with the van der Waals interaction. On the other hand, the base of the nucleotide at the 0 position stacks onto the aromatic ring of Y307. Thus, these tyrosine and tryptophan residues play important roles in the proper positioning of the target cytosine (0 position) in the catalytic center (Figure 3-9b). Therefore, the mutations at the N240, H247, Y307 and Y308 positions, together with that at the W277 position that is responsible also for ssDNA binding, as described above, probably affected the roles of the residues at each position described above, and thereby resulted in the decrease of the catalytic efficiency.

Figure 3-9c summarizes the amino acid residues responsible for deaminase activity through either ssDNA binding or catalytic efficiency.

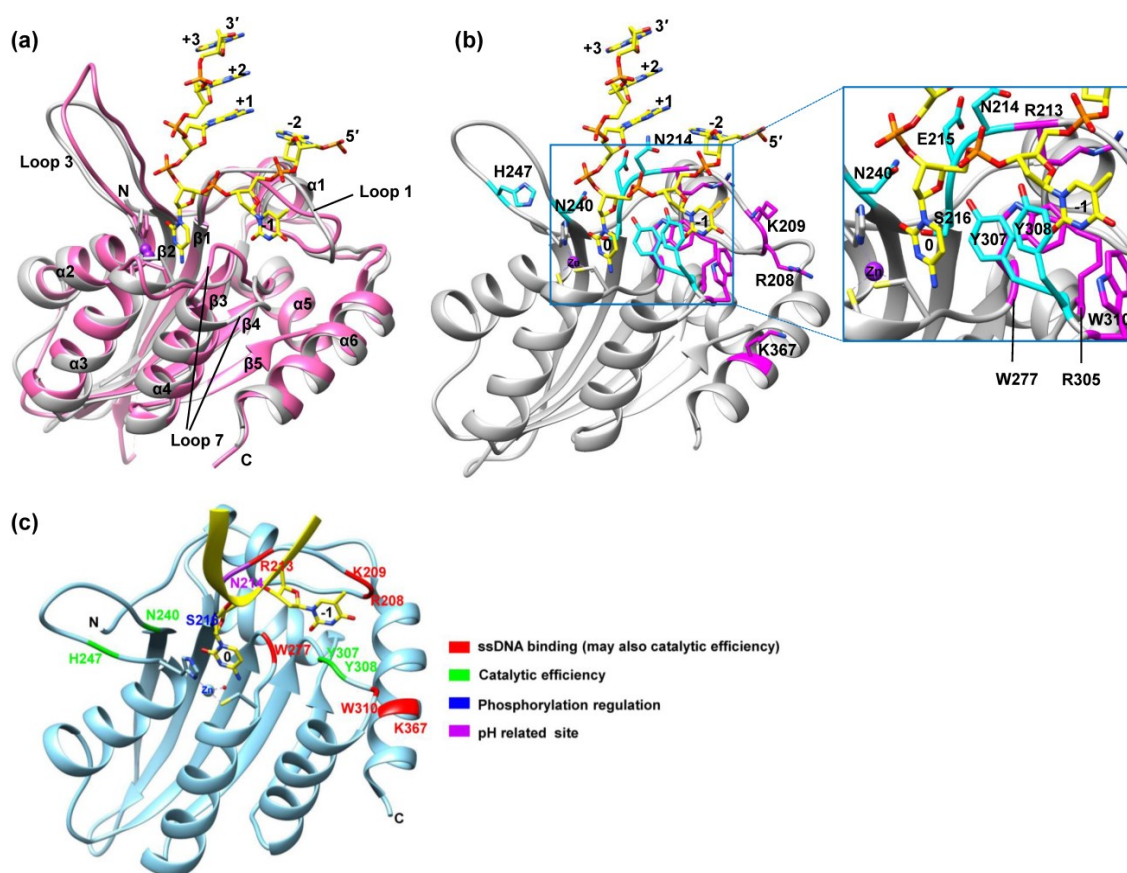


Figure 3-9. A model of A3F-CTD in a complex with ssDNA. (a) Superposition of the crystal structure of A3F-CTD (PDB: 3WUS, gray) onto that of the A3A:ssDNA complex (PDB: 5SWW, hot pink). (b) A model of A3F-CTD in a complex with ssDNA derived from the A3A:ssDNA complex structure. See the main text for detail. The amino acid residues responsible for both deamination activity and ssDNA binding are colored magenta, and those that are responsible only for deaminase activity are colored cyan. (c) Summary of the amino acid residues responsible for deaminase activity through either ssDNA binding (red) or catalytic efficiency (green), together with one involved potentially in regulation by phosphorylation (blue) and the other related to the pH issue (magenta). Some of red residues may also be responsible for deaminase activity through catalytic efficiency.

We revealed in the previous section that A3F-CTD requires the ssDNA length of 30 nt for maximum activity. The length of ssDNA in contact with A3F-CTD in the model of Figure 3-9 is much less than 30 nt. This is consistent with the proposed idea that ssDNA regions that are apart from the target TC serve as landing points of A3F-CTD. After landing, A3F-CTD would search for TC by jumping, as discussed above. Thus, ssDNA regions are supposed to contribute to the recruitment of A3F-CTD to TC,

resulting in increase in the deaminase activity. The total length of 30 nt seems to be required to provide the suitable landing points for A3F-CTD.

For S216E mutant, the decrease in the catalytic efficiency may also be partly a reason for the decreased deaminase activity. Since the location of S216 is very close to the base moiety at nucleotide position 0 (Figure 3-9b), the negative charge and/or bulkiness of the glutamate may repel the cytosine base, resulting in the mis-positioning of the target DNA and thus a decrease in the catalytic efficiency. This repulsive effect should also occur when S216 is phosphorylated.

In the A3A:ssDNA complex structure, the protonated H29 is supposed to interact with ssDNA and thereby to act as a scaffold to stabilize ssDNA [65]. The side chain of H29 of A3A could stack on the +1 nucleotide and come into contact with the -2 nucleotide through van der Waals. Therefore, it is likely that the protonated histidine interacts with ssDNA, causing a positive effect on the deaminase activity. The model in Figure 3-9b suggests that similar interactions with ssDNA are possible for the N214H mutant of A3F-CTD. This may rationalize the unique pH dependency of the N214H mutant.

3.4 Conclusions

In this chapter, we examined the effects of the 5'- and 3'-flanking sequences of the target TC on A3F-CTD deaminase activity, and found that TTCA/G are the most preferred target sequences of A3F-CTD. We also examined the ssDNA length-dependent deaminase activity and found that A3F-CTD shows higher activity toward longer ssDNA. Then, we mutated the residues that are close to the catalytic center and found that some amino acid residues are responsible for the binding to ssDNA, while others are responsible for the catalytic efficiency. The distinct role of each amino acid residue can be rationally understood on the basis of the model structure of the A3F-CTD:ssDNA complex derived from the A3A:ssDNA complex structure. Additionally, we revealed that the deaminase activity of A3F may be modulated through phosphorylation and de-phosphorylation of the S216 residue, possibly in response to innate immune stimulation. Lastly, we found that the N214H mutant exhibits

significantly high deaminase activity at pH 5.0. Our data suggest that the protonation of histidine at this position in A3 proteins has a positive effect on the deaminase activity.

3.5 Experimental procedures

3.5.1 Oligonucleotides and plasmids

All of the ssDNA substrates used in this chapter including FITC- and Alexa488-labeled ssDNA substrates that were obtained from FASMAC Co., Ltd. (Kanagawa, Japan) are listed in Table 3.1.

Table 3.1 Oligonucleotides used in this study

Name	Sequence^a (5'→3')
S _{TTCA} _49	*ATAATAATAATAATAATAATAATATT <u>C</u> ATTTATAATAATAATAATAATA
S _{TTCG} _49	*ATAATAATAATAATAATAATAATAATATT <u>C</u> GTTTATAATAATAATAATAATA
S _{TTCT} _49	*ATAATAATAATAATAATAATAATAATATT <u>C</u> TTTTATAATAATAATAATAATA
S _{TTCC} _49	*ATAATAATAATAATAATAATAATAATATT <u>C</u> CTTTATAATAATAATAATAATA
S _{GTCA} _49	*ATAATAATAATAATAATAATAATAATAGT <u>C</u> ATTTATAATAATAATAATAATA
S _{CTCA} _49	*ATAATAATAATAATAATAATAATAACT <u>C</u> ATTTATAATAATAATAATAATA
S _{ATCA} _49	*ATAATAATAATAATAATAATAATAAT <u>C</u> ATTTATAATAATAATAATAATA
S _{49-TCA-A488}	(Alexa488)ATAATAATAATAATAATAATAATAATATT <u>C</u> ATTTATAATAATAATAATAATA

*sequences that are labeled with FITC at 5' end.

^a the target cytosine of the TC sequence is in bold and underlined.

Mutants of A3F-CTD in pET28b were constructed using KOD-Plus-mutagenesis kit (TOYOBO, Japan). The oligonucleotides design (Table 3.2) and mutagenic procedures were performed following the manufacturers' protocol. All of the plasmids were sequenced to confirm the correct sequences before transferring into BL21 (DE3) Gold.

Table 3.2 Primers used in this study

Primers	Sequence(5'→3')
A3F-CTD(187-373)_F	GGAATTCCATATGATTCTGCGTAACCCGATGGAGG
A3F-CTD(187-373)_R	CGGAATTCTTATTCCAGGATTTCTTGCAG
A3F-CTD_R208E_F	AAACCTGGAAAAAGCATACGGTCGTAACG
A3F-CTD_R208E_R	CTTTTCCAGGTTTTTGAAATGAAAGTAGA
A3F-CTD_K209E_F	CTGCGCGAAGCATACGGTCGTAACGAAA
A3F-CTD_K209E_R	GTATGCTTCGCGCAGGTTTTTGAAATG
A3F-CTD_R213E_F	TACGGTGAAAACGAAAGCTGGCTGTGT
A3F-CTD_R213E_R	TTCGTTTTACCGTATGCTTTGCGCA
A3F-CTD_N214H_F	CGGTCGTCATGAAAGCTGGCTGTGTTTTAC
A3F-CTD_N214H_R	AGCTTTCATGACGACCGTATGCTTTGC
A3F-CTD_E215Q_F	TCGTAACCAGAGCTGGCTGTGTTTTACAATG
A3F-CTD_E215Q_R	GCCAGCTCTGGTTACGACCGTATGCTTTGC
A3F-CTD_S216A_F	AACGAAGCATGGCTGTGTTTTACAATGGA
A3F-CTD_S216A_R	CAGCCATGCTTCGTTACGACCGTATGCT
A3F-CTD_S216E_F	AACGAAGAATGGCTGTGTTTTACAATGGA
A3F-CTD_S216E_R	CAGCCATTCTTCGTTACGACCGTATGCT
A3F-CTD_N240A_F	TTTCGTGCACAGGTTGATCCGGAAACC
A3F-CTD_N240A_R	AACCTGTGCACGAAAAACACCACGTTTCC
A3F-CTD_N240D_F	TTTCGTGATCAGGTTGATCCGGAAACC
A3F-CTD_N240D_R	AACCTGATCACGAAAAACACCACGTTTCC
A3F-CTD_N240Q_F	TTTCGTCAACAGGTTGATCCGGAAACC
A3F-CTD_N240Q_R	AACCTGTTGACGAAAAACACCACGTTTCC
A3F-CTD_H247A_F	GAAACCGCATGTCATGCAGAACGTTGTTT
A3F-CTD_H247A_R	ATGACATGCGGTTTCCGGATCAACCTG
A3F-CTD_H247R_F	GAAACCCGTTGTCATGCAGAACGTTGTTT
A3F-CTD_H247R_R	ATGACAACGGGTTTCCGGATCAACCTG
A3F-CTD_W277A_F	ATACCTCTGCAAGCCCGTGTCCGGAAT
A3F-CTD_W277A_R	GGCTTGCAGAGGTATACCAGGTAACCTCATAG
A3F-CTD_Y307A_F	CCTGGCATACTTTTGGGATACCGATTATC
A3F-CTD_Y307A_R	CAAAAGTATGCCAGGCGTGCGGTAAAA
A3F-CTD_Y308A_F	GTATGCATTTTGGGATACCGATTATCAAG
A3F-CTD_Y308A_R	CCCAAATGCATACAGGCGTGCGGTAAA
A3F-CTD_W310A_F	ACTTTGCAGATACCGATTATCAAGAAGGTC
A3F-CTD_W310A_R	GGTATCTGCAAAGTAATACAGGCGTGCG
A3F-CTD_K367E_F	GATAGCGAACTGCAAGAAATCCTGGAATAA
A3F-CTD_K367E_R	TGCAGTTCGCTATCCAGAAACAGGAAGTTA

3.5.2 Protein expression and purification

Expression and purification of the wt A3F-CTD and its mutants were performed as

previously described with slight modifications[56]. BL21 (DE3) Gold was transformed with the prepared pET28b (+) expression vectors individually and grown at 37°C until the OD₆₀₀ reached 0.6. Protein expression was induced by adding 0.3 mM isopropyl β -D-1-thiogalactopyranoside and subsequent incubation at 20°C for 20 h. Cells were harvested and lysed by sonication in lysis buffer (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10 % (v/v) glycerol, 4 mM MgCl₂, and 20 mM imidazole) supplemented with 5 units mL⁻¹ DNase I, 1 μ g mL⁻¹ RNase, and 0.2 mg mL⁻¹ lysozyme. A cell-free lysate was obtained by centrifugation at 50,000 g for 30 min and purified by loading on an Ni-Sepharose resin column (GE Healthcare), followed by extensive washing with washing buffer (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 % (v/v) glycerol, and 20 mM imidazole). Proteins eluted with elution buffer (50 mM Tris-HCl buffer (pH 8.0), 500 mM NaCl, 2 % (v/v) glycerol, and 250 mM imidazole) were concentrated and further purified by loading onto a HiLoad 16/60 Superdex 75 PG column (GE Healthcare) with equilibration buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 300 mM L-arginine HCl, 5 mM 2-mercaptoethanol). Target fractions were identified by SDS-PAGE, pooled, concentrated and stored at 4°C on ice.

3.5.3 *In-vitro* cytosine deamination assay

The deaminase activity measurements for A3F-CTD and its mutants were carried out as described in chapter 2. For pH-dependent reactions, 100 nM 49 nt Alexa 488-labeled ssDNA (S_{49-TTCA-A488}, Table 3.1) was incubated with 1,000 nM A3F-CTD or A3F-CTD mutants at each pH at 37°C for 1, 10, 20, 30, 60, 90 and 120 min, respectively. The deamination percentage obtained at each time point was individually fitted to $I(t) = I_0 \exp(-kt)$, where I_0 is the initial intensity and k is the deamination rate constant [48]. I_0 , k , and error bars were obtained by nonlinear least-squares analysis [93].

3.5.4 Steady-state fluorescence anisotropy measurement

The binding of A3F-CTD to ssDNA was monitored as changes in steady-state fluorescence depolarization using an FP-8500 spectrofluorometer (JASCO Asia Portal), as described in chapter 2.

3.5.5 Circular dichroism (CD) spectroscopy

Wild type and mutant A3F-CTDs (2-6 μM) were individually dissolved in the 20 mM Tris-HCl (pH 7.5) solution containing 300 mM NaCl, 0.05 % CHAPS and 1 mM DTT. CD spectra were recorded with a Jasco J-720A spectrometer (Japan Spectroscopic Co., Japan) at 25°C using a cell of the optical path of 1 mm. Each spectrum was background-corrected, smoothed, and converted to molar ellipticity $[\theta]$ ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$).

Chapter 4

Influence of the ssDNA concentration on the length-dependent deaminase activity of APOBEC3F

4.1 Introduction

In chapter 2, we studied the length- and location-dependent deaminase activity of the A3B using a real-time NMR method, as well as biochemical methods, and interpreted the results according to the fact that A3B moves along ssDNA for a relatively short distance and tends to dissociate from the ssDNA. Previous research showed that A3F scans ssDNA via jumping, and that the intersegmental transfer only occurs under crowded reaction conditions [81], which is totally different from the case of A3B. Similar to the studies of A3B, in this chapter, we investigated the length-dependent deaminase activity of A3F. We showed that A3F deaminates ssDNA in a length-dependent manner, which is affected by the concentration of the ssDNA substrate. On the basis of the binding affinities of A3F with different lengths of ssDNA, the results were rationally interpreted.

4.2 Results

4.2.1 The ssDNA-length-dependent deaminase activity of wt A3F-CTD

As described previously, A3 proteins show different deaminase activities toward ssDNA substrates of various lengths [78, 91]. In chapter 2, we have identified that A3B-CTD shows higher activity toward the target sequence in short ssDNA than the one in long ssDNA. It would be of interest to determine how the length of ssDNA affects the deamination by wt A3F-CTD. We analyzed the length-dependent deaminase activities of A3F-CTD using seven ssDNA substrates of different lengths, ranging from 10 to 85 nt, each of which contains a single TTCA sequence near the center (S_{FITC}_{10-85} , Table 2.1). The deaminase activity of A3F-CTD increased monotonically from 10 to 30 nt and reached a plateau level (Figure 4-1). At the reaction time of 1 h, the deamination percentage for 30 - 85 nt ssDNA substrate was ~3.5-fold higher than that for 10 nt ssDNA substrate, indicating that A3F-CTD deaminates the target sequence in long

ssDNA more efficiently. We suppose that the non-specific binding sites of a long ssDNA substrate serve as landing points for the recruitment of much more A3F-CTD, which searches for the target sequence by jumping as described by Ara *et al*[81]. This would be the reason why the target sequence in long ssDNA is deaminated more efficiently.

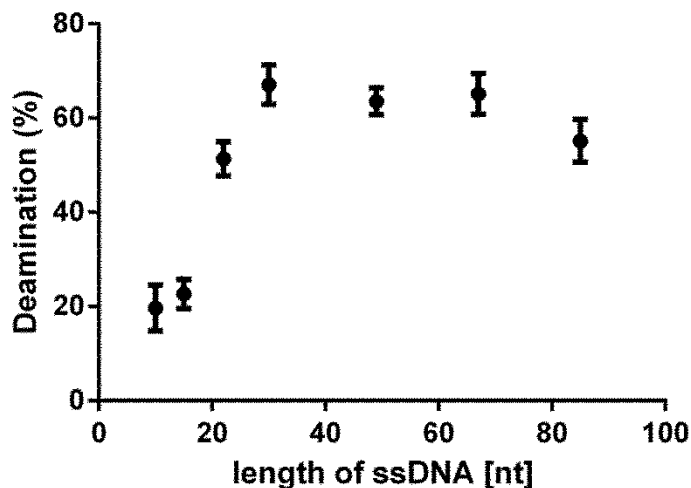


Figure 4-1 The length-dependence of the deamination percentage with wt A3F-CTD. The length of the ssDNA substrate was varied from 10 to 85 nt. The deamination percentage (%) with A3F-CTD was plotted against the length of the ssDNA substrate.

4.2.2 N214H mutant does not affect the length-dependent deaminase activity of A3F-CTD

Figure 3-8 shows that the N214H mutant exhibits much higher deaminase activity toward 46 nt ssDNA at pH 5.5 and pH 7.5 than wt A3F-CTD. This finding was further confirmed by the real-time NMR method, as described in Figure 4-2, in which 150 μ M 10 nt ssDNA (S_{TTCA} , Table 2.1), and 1 μ M wt A3F-CTD and the N214H mutant were used, respectively. To determine whether or not the N214H mutant affects the essential deaminase property of wt A3F-CTD, we applied the UDG assay to study the length-dependent deaminase activity of the N214H mutant. 100 nM 10 nt and 49 nt ssDNA ($S_{\text{FITC}_{10}}$ and $S_{\text{FITC}_{49}}$, Table 2.1) were used. As shown in Figure 4-3a and b, the N214H mutant showed higher deaminase activity toward 49 nt ssDNA than 10 nt ssDNA, which was similar to the case of wt A3F-CTD (Figure 4-1), indicating that the N214H mutant has no effect on the length-dependence of A3F-CTD. Considering the

high deaminase activity, the N214H mutant was used in the following experiments.

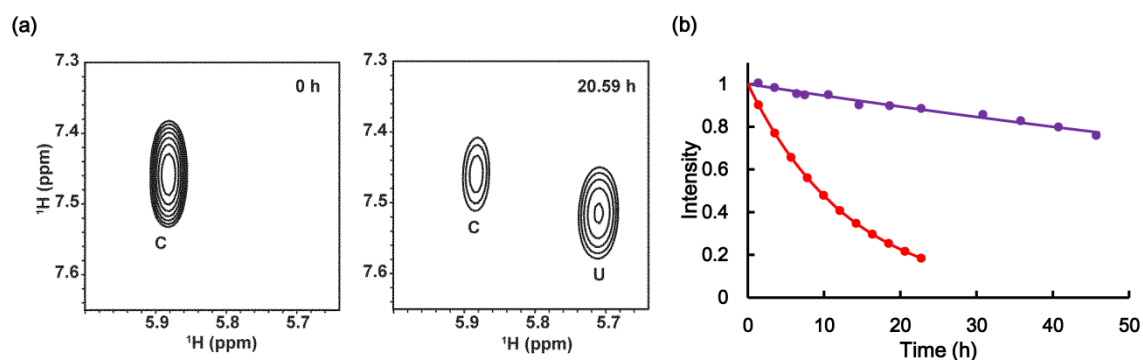


Figure 4-2. Deaminase activities of wt A3F-CTD and the N214H mutant. (a) Deaminase activities of the N214H mutant monitored by the real-time NMR method. H5-H6 TOCSY correlation peaks for 10 nt S_{TTCA} at 0 and 20.59 h after addition of the N214H mutant are shown. (b) Time-courses of the change in the intensity of the cytosine peak for S_{TTCA} for wt A3F-CTD (purple) and the N214H mutant (red).

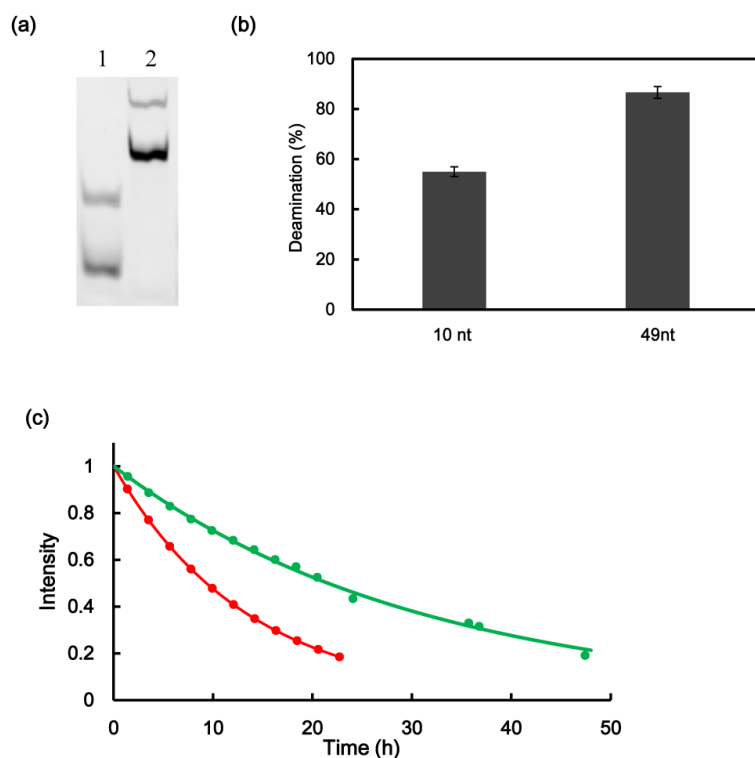


Figure 4-3. ssDNA-length-dependent deaminase activity of the N214H mutant. (a) Deaminase activities of the N214H mutant toward 10 nt ssDNA ($S_{\text{FITC}_{10}}$, lane 1) and 46 nt ssDNA ($S_{\text{FITC}_{49}}$, lane 2) were examined by means of the UDG assay. The upper and lower bands in each lane represent the unreacted substrate and deamination product, respectively. (b) Efficiencies of the deamination for each ssDNA revealed by the UDG assay. (c) Time-courses of the change in the intensity of the cytosine peak for S_{TTCA} (red) and S_{mTTCA} (green).

4.2.3 The concentration of ssDNA affects the length-dependent deamination by A3F

Figure 4-3b shows that N214H efficiently deaminates a target sequence in 49 nt ssDNA under a low ssDNA concentration (100 nM). However, the real-time result showed that N214H prefers to deaminate a target sequence in 10 nt ssDNA (S_{TTCA} , Table 2.1) rather than the same one in 46 nt ssDNA (S_{mTTCA} , Table 2.1, Figure 4-3c), which is opposite to the case of UDG. It should be noted that in a real-time NMR experiment, a high ssDNA concentration was used (150 μ M), while in the UDG assay, a low ssDNA concentration was used (100 nM). Different ssDNA concentrations cause a change of the length-dependent deaminase activity of A3F, indicating that the ssDNA-length dependent deamination by A3F-CTD could be determined by the ssDNA substrate concentration.

4.3 Discussion

In chapter 2, we interpreted the length- and location-dependent deaminase activity of A3B-CTD on the basis of its weak binding affinity. One obvious difference between A3B and A3F is that the binding affinity of A3F to ssDNA is higher than that of A3B (Figure 2-2). To interpret the results obtained with the N214H mutant, firstly we measured the dissociation constant (K_d) of the N214H mutant for binding to 10 nt and 49 nt ssDNA, respectively. As shown in Figure 4-5, the calculated K_d values for 10 nt and 49 nt ssDNA are 7.733 μ M and 0.931 μ M, respectively.

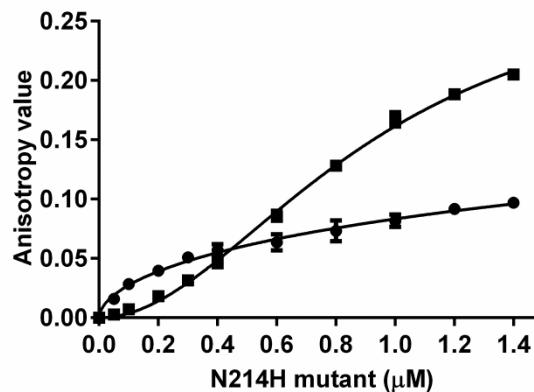


Figure 4-5. The binding of the N214H mutant with 10 nt ssDNA (circle) and 49 nt ssDNA (square).

When attempting to interpret the ssDNA-length-dependence of deamination by A3F in the presence of a low ssDNA concentration, we calculated the concentration of N214H bound to 10 nt and 49 nt ssDNA to be 0.011 μM and 0.051 μM , respectively, indicating that part of A3F binds to 10 nt ssDNA while much more A3F binds to 49 nt ssDNA. In the case of 49 nt ssDNA, even if A3F partially dissociates from the ssDNA during the process of searching for the target sequence, the final concentration of A3F that can reach the target section in 49 nt ssDNA will always be higher than in the case of the 10 nt ssDNA (Figure 4-6a), resulting in higher deaminase activity toward the target sequence in 49 ssDNA than that in 10 nt ssDNA (Figure 4-3b).

In presence of a high ssDNA concentration, the concentrations of N214H bound to 10 nt and 46 nt ssDNA were calculated to be 0.951 μM and 0.994 μM , respectively, indicating that almost all A3F binds to either 10 nt or 46 nt ssDNA. (Figure 4-6b). In the case of 46 nt ssDNA, although all A3F binds to the ssDNA, some A3F cannot reach the target section, which arises from the dissociation of A3F from the ssDNA during the process of searching for the target sequence (Figure 4-6b). Therefore, the final concentration of A3F that reaches the target section in 46 nt ssDNA will be lower than in the case of the 10 nt ssDNA, resulting in higher deaminase activity toward the target sequence in 10 nt ssDNA than that in 46 nt ssDNA (Figure 4-3c).

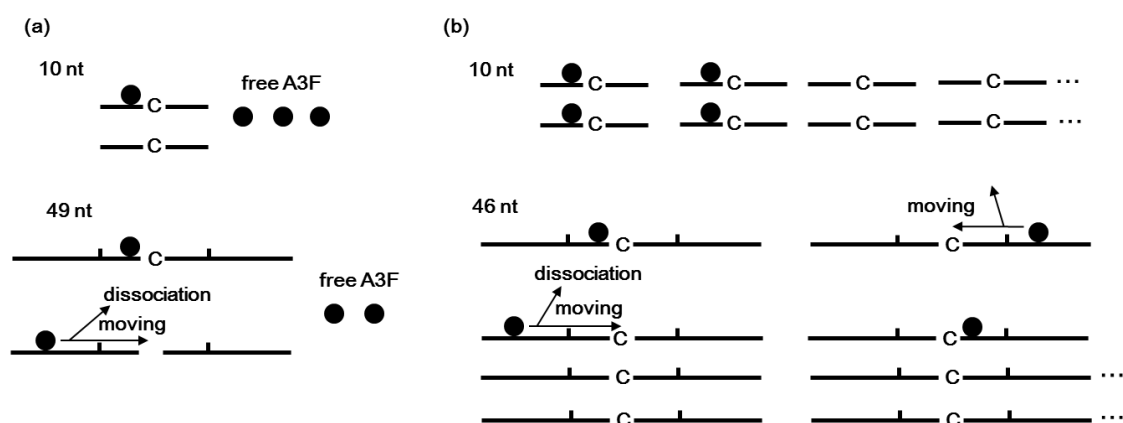


Figure 4-6. Description of the target search by A3F. (a) In the presence of a low ssDNA concentration, much more protein binds the 49 nt ssDNA, contributing to the high deaminase activity. (b) In presence of a high ssDNA concentration, all protein binds to either 10 nt or 46 nt ssDNA. For 46 nt ssDNA, the dissociation of A3F from the ssDNA during moving process results in low deaminase activity.

4.4 Conclusions

In this chapter, we analyzed the ssDNA-length-dependent deaminase activity of A3F. In the presence of a low ssDNA concentration, A3F turned out to efficiently deaminate a target sequence in long ssDNA. However, in the presence of a high ssDNA concentration, A3F prefers to deaminate a target sequence in short ssDNA. On the basis of the different dissociation constants of N214H toward 10 nt and 49 nt ssDNA, these unique characteristics were rationally interpreted. Our data provides a new insight into the specific biochemical properties for efficient restriction of HIV by A3 proteins.

4.5 Experimental procedures

4.5.1 Oligonucleotides

All of the ssDNA substrates used in this chapter that were obtained from FASMAC Co., Ltd. (Kanagawa, Japan) are listed in Table 2.1.

4.5.2 Proteins expression and purification

The expression and purification of A3F-CTD and A3F-CTD N214H mutant were described as 3.5.2 in chapter 3.

4.5.3 *In-vitro* cytosine deamination assay

The *in-vitro* cytosine deamination assay followed the method as described in 2.5.3 in chapter 2.

4.5.4 Real-time NMR spectroscopy

The real-time NMR measurement was performed as described in 2.5.4 in chapter 2.

4.5.5 Steady-state fluorescence anisotropy measurement

The binding affinities of N214H mutant with 10 nt ssDNA and 49 nt ssDNA was performed using steady-state fluorescence anisotropy measurement, as described in chapter 2.

Chapter 5

The inhibition of the deaminase activity of A3F-CTD in the presence of the Vif complex

5.1 Introduction

In infected cells, HIV-1 inactivates the antiviral functions of A3 proteins by expressing the Vif protein. The Vif protein recruits an E3 ubiquitin ligase complex (Cul5, CBF β , ElonB, ElonC, and Rbx-2) to bind and target A3 proteins for proteasomal degradation [31, 104]. Study on the destroying of the binding between Vif and A3 proteins is a promising area for finding antiretroviral drugs. Recently, the crystal structure of the HIV-1 Vif-Cul5-CBF β -EloB-EloC was determined (Figure 5-1)[105]. However, there is not any A3 substrate in this structure. In A3F, the CTD domain is responsible for the DNA cytosine deamination and the binding with HIV-1 Vif. Previous biochemical analysis identified several motifs in HIV-1 Vif involved in A3F binding, as well as the binding domains or motifs in A3F responding to HIV-1 Vif binding [54, 56, 108]. In this chapter, we investigate the effects of the Vif complex on the deaminase activity of A3F-CTD.

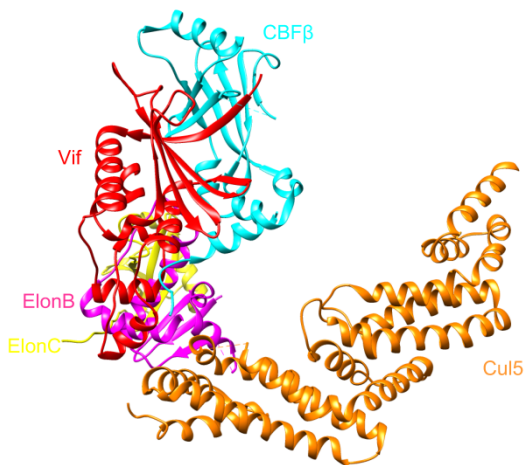


Figure 5-1. Structure of the Vif complex. Structure of Vif (red) in a complex with Cul5 (orange), CBF β (cyan), EloB (magenta), and EloC (yellow) (PDB: 4N9F[105]) was elucidated using UCSF Chimera [107].

5.2 Results

5.2.1 Purification of the Vif complex

The crystal structure of the Vif complex allowed us to construct the Vif complex for this study. Here, we used Vif (1-192), Cul5 (12-386), CBF β (1-170), EloB (1-118), and EloC (17-112). Firstly, Vif (1-192), CBF β (1-170), EloB (1-118), and EloC (17-112) were co-expressed in *E.coli* and then purified with a Nickel column, and then mixed with Cul5 (12-386), which was expressed separately. The mixture was sequentially purified by anion exchange chromatography and size exclusion chromatography. The purified Vif complex was confirmed by SDS-APGE, as shown in Figure 5-2. More details are given under Experimental procedures.

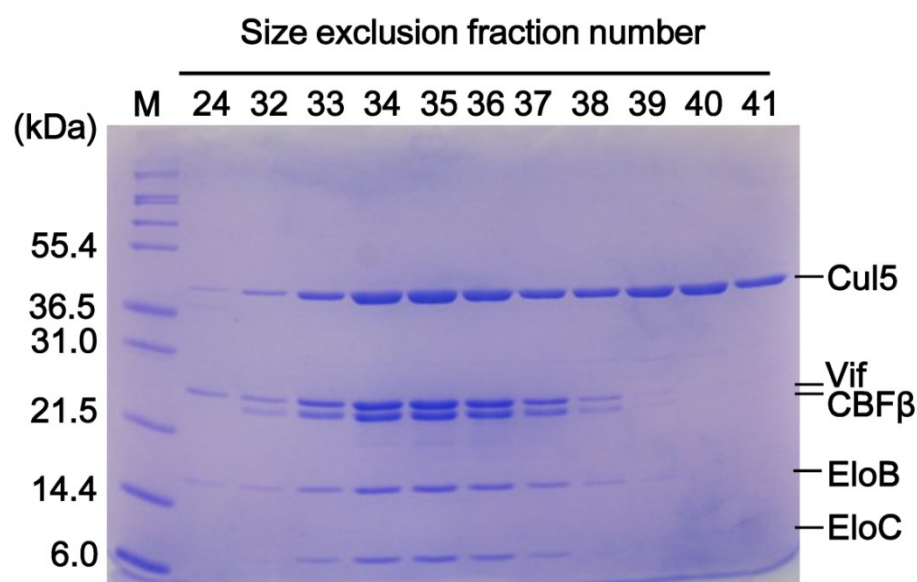


Figure 5-2. SDS-PAGE and Coomassie blue staining of the Vif complex in the peak fractions on size exclusion chromatography.

5.2.2 The Vif complex inhibits the deaminase activity of A3F-CTD

After obtaining purified Vif complex, we examined the effect of the Vif complex on the deaminase activity of wt A3F-CTD by using the UDG assay. S_{FITC_49} (Table 2.1) was used as the substrate. The results showed that with increasing concentrations of Vif complex added to the mixture, the deaminase activity of A3F-CTD decreased (Figure

5-3), suggesting that the Vif complex causes inhibition of the deaminase activity of A3F-CTD.

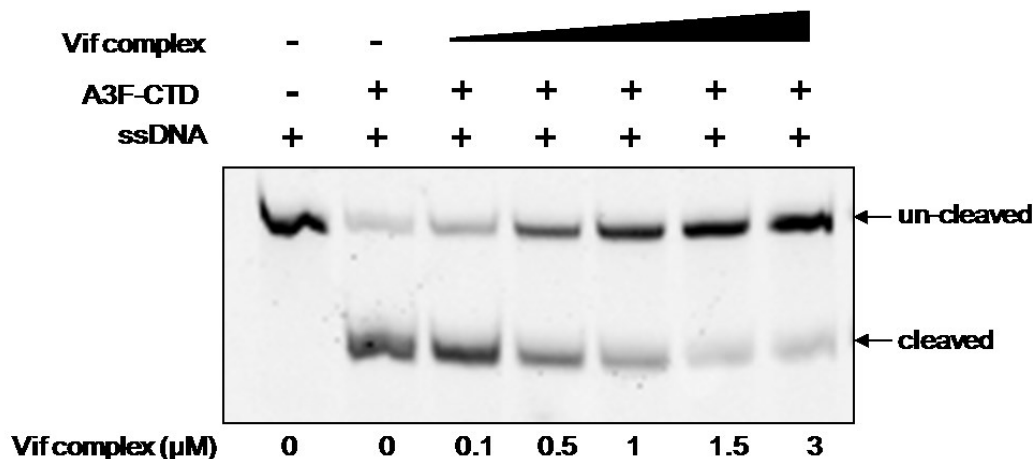


Figure 5-3. Deaminase activity of A3F-CTD in the presence of the Vif complex. Measurement of the deaminase activity of A3F-CTD was carried out by UDG assay. S_{FITC_49} was used as the substrate. The un-cleaved substrate and cleaved product are indicated. The various concentrations of the Vif complex are shown on the bottom of the gel.

5.3 Discussion

The deaminase activity of A3F-CTD in the presence of the Vif complex studied in this chapter indicates that the Vif complex has a negative effect on the deamination by A3F-CTD. There are two possible mechanisms. One is that the Vif complex targets and binds to A3F-CTD, decreasing the amount of ssDNA-bound A3F-CTD. Another possibility is that the positively charged Vif complex binds to the negatively charged ssDNA, reducing the free ssDNA in the mixture. Actually, a number of studies are in agreement with the first possibility. Albin *et al.* stated that Glu324, which is located on the surface of $\alpha 4$ in human A3F, is a critical determinant of A3F susceptibility to HIV-1 Vif-mediated degradation [108]. Land *et al* found that the $\alpha 3$ and $\alpha 4$ helices of human A3F are important determinants of the interaction with HIV-1 Vif [109]. Recently, Richards *et al* proposed a wobble model to explain how HIV-1 Vif has evolved to bind A3F[106]. In this model, the $\alpha 2$, $\alpha 3$, and $\alpha 4$ helices of A3F constitute the interface with HIV-1 Vif. In this interface, a persistent electrostatic interaction between A3F E289 and

Vif R15 was best highlighted. On the other hand, one study supported the second possibility that Vif complex competitively binds ssDNA, as revealed by Kamba[110]. In general, different from previous report that Vif destroys A3F-CTD by ubiquitin-mediated proteasomal degradation [31, 104], our result shown that the Vif complex also reduce the deamination of A3F *in vitro*.

5.4 Conclusions

In this chapter, we applied *in vitro* deaminase activity assay to study the effects of the Vif complex on the deamination by A3F-CTD. With increasing Vif complex, the deaminase activity of A3F-CTD decreased. This finding indicates that the Vif complex inhibits the deaminase activity of A3F-CTD. This assay constitutes another possible meaner of of finding to find the drugs that can destroy the interaction between Vif and A3F.

5.5 Experimental procedures

5.5.1 Expression and purification of Vif complex

The expression and purification followed the method described by Guo *et al* [107] with some modifications. Briefly, the gene fragments of Vif (1-192), CBF β (1-170), were inserted into *NcoI/BamHI* sites of pRSF Duet, *NdeI/BamHI* sites of pET15b, respectively. EloB (1-118) and EloC (17-112) were insert into the vector of pCDF Duet1 via the sites of *NcoI/BamHI* and *NcoI/KpnI*, respectively. Cul5 (12-386) was inserted into the *BamHI/EcoRI* sites of pGEX6p-1. The Vif, CBF β , and EloB/C constructs were sequentially transferred into BL21 (DE3) Gold. The Cul5 construct was transferred into BL21 (DE3) Gold separately. Expression of the Vif-CBF β -EloB-EloC was induced by adding 0.1 mM IPTG at 20°C for 20 h. Then the cells were harvested by centrifugation and re-suspended in lysis buffer (25 mM Tris-HCl (pH 7.5), 500 mM NaCl, 20 mM imidazole), supplemented with 5 U mL⁻¹ DNase I, 1 μ g mL⁻¹ RNase H, 0.2 mg mL⁻¹ lysozyme. A cell-free lysate was obtained by centrifugation at 50,000 g for 30 min and purified by loading onto an Ni-Sepharose resin column (GE Healthcare), followed by extensive washing with lysis buffer. Proteins was eluted with elution buffer

(25 mM Tris-HCl buffer (pH 7.5), 500 mM NaCl, and 250 mM imidazole). The cells expressing Cul5 were lysated with the buffer of 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM benzamidine, and 250 μ L MgCl₂, supplemented with 5 U mL⁻¹ DNase I, 1 μ g mL⁻¹ RNase H, 0.2 mg mL⁻¹ lysozyme. After centrifuging, the cell-free lysate was loaded onto glutathione sepharose 4B column (GE healthcare). The GST tag was removed on the column digested with prescission protease in the cleavage buffer (25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 1 mM DTT) at 4°C overnight. Cul5 was washed out by the cleavage buffer. Collected Cul5 and Vif-CBF β - EloB-EloC were mixed, concentrated and further fractionated by anion exchange chromatography with buffer A (25 mM Tris-HCl (pH 7.5), 3 mM DTT) and buffer B (25 mM Tris-HCl (pH 7.5), 1 M NaCl, and 3 mM DTT). Finally, the fraction was by loading onto a HiLoad 16/60 Superdex 200 PG column (GE Healthcare) with equilibration buffer (20 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 5 mM DTT). Target fractions were identified by SDS-PAGE, pooled, concentrated and stored at 4°C on ice.

5.5.2 *In-vitro* cytosine deamination assay

The deamination assay of A3B-CTD toward S_{FITC_49} in presence of Vif complex was performed as described in chapter 2. Briefly, 1 μ M A3F-CTD and 100 nM S_{FITC_49} were incubated in the deamination buffer (15 mM Tris-HCl (pH 7.5), 50 mM NaCl and 5 mM DTT) in presence of 0, 0.1, 0.5, 1, 1.5 and 3 μ M Vif complex, respectively. The deamination reaction was undertaken at 37°C for 2 h. Then the reaction was stopped. The following steps were the same to the 2.5.3 in chapter 2.

Chapter 6

General conclusions

Human A3 proteins play important biological roles in the intrinsic immune system by defending against endogenous and exogenous DNA pathogens. Specifically, A3F exhibits efficient anti-HIV-1 activity by being packed into the progeny viral particles, and then interferes with reverse transcription, and mutates cytosines into uracils in nascent cDNA, resulting in blocking of viral replication. Different from A3F, A3B is involved in various cancer types, and it was revealed that aberrant expression of A3B causes a genomic DNA lesion. Here, we used the real-time NMR method, together with biochemical methods, to study the deamination characteristics of A3B and A3F, and identified the mechanism of deamination by A3B and A3F.

In chapter 1, a general introduction was given, including energy and enzymes, HIV and cancers, and the structural characteristics of A3 proteins, together with its antiviral activity and relationship with cancers, the binding interface of A3 proteins with oligonucleotides, and the movement on ssDNA.

In chapter 2, how the target sequences are searched for and deaminated by A3B-CTD was analyzed. The real-time NMR method, as well as the UDG assay, were applied to measure the length- and location-dependent deaminase activity of A3B-CTD. We found that A3B-CTD shows higher activity toward its target sequence in short ssDNA and efficiently deaminates a target sequence located near the center of the ssDNA. These properties are quite different from those of well-studied A3G, which shows higher activity toward its target sequence in long ssDNA and one located close to the 5'-end. The unique properties of the A3B-CTD can be rationally interpreted by considering that after non-specific binding to ssDNA, A3B-CTD only slides for a relatively short distance to search for the target sequence and tends to dissociate from the ssDNA before reaching the target sequence. The possible biological significance of this was also discussed.

In chapter 3, the deaminase activity of A3F-CTD, including the influence of the DNA sequence and pH on deaminase activity, together with the roles of the amino acid residues located close to the catalytic center in deaminase activity and ssDNA binding

were systematically characterized. Firstly, we optimized the target sequence and the length of the ssDNA substrate, and found that A3F-CTD efficiently deaminates 5'-TTCA/G-3' in longer ssDNA(>30 nt). Secondly, on the basis of the result of mutagenesis analysis of the amino acid residues that are close to the catalytic center, combining the structure model of the A3F-CTD:ssDNA complex, we found that some amino acid residues are responsible for the binding to ssDNA, while others are responsible for the catalytic efficiency. Thirdly, we revealed that the deaminase activity of A3F-CTD may be regulated through phosphorylation and de-phosphorylation of the S216 residue. Lastly, we investigated the pH effects on the deaminase activity of A3F-CTD and noticed that the N214H mutant shows obviously high deaminase activity at pH 5.0 and pH 7.5.

In chapter 4, the length-dependence of the deaminase activity of A3F-CTD was studied. We identified the length-dependent deamination by wt A3F-CTD and the deaminase activity of the N214H mutant. Comparing the results obtained with the UDG assay and real-time NMR method, we found that the concentration of the ssDNA substrate could affect the length-dependent deamination by A3F. In presence of a low ssDNA concentration, A3F prefers to deaminate a target sequence in long ssDNA, while in presence of high ssDNA concentration, A3F efficiently deaminates a target sequence in short ssDNA. On the basis of the different binding affinities of N214H with different lengths of ssDNA, these unique characteristics were rationally interpreted.

In chapter 5, the deaminase activity of A3F-CTD in presence of the Vif complex was investigated using the UDG assay. The results suggested that the Vif complex efficiently inhibits the deamination by A3F-CTD *in vitro*. This finding indicates that the Vif complex neutralizes the function of A3F not only through E3 ubiquitin mediated proteasomal degradation, but also by direct interference with the deamination.

In chapter 6, a general conclusion was given.

In this study, we mainly investigated the length- and location-dependent deaminase activities of A3B and A3F, respectively, as well as the deamination properties of A3F. The results will help us to better understand the overall molecular mechanism of DNA editing by cytosine deaminases.

References

1. Robledo-Abad C., Althaus H. J., Berndes G., Bolwig S., Corbera E., Creutzig F., Garcia-Ulloa J., Geddes A., Gregg J. S., Haberl H., Hanger S., Harper R. J., Hunsberger C., Larsen R. K., Lauk C., Leitner S., Lilliestam J., Lotze-Campen H., Muys B., Nordborg M., Olund M., Orlowsky B., Popp A., Portugal-Pereira J., Reinhard J., Scheiffle L. and Smith P., Bioenergy production and sustainable development: science base for policymaking remains limited. *Global Change Biology Bioenergy*, 2017. **9**(3): p. 541-556.
2. Kamerlin S. C. L. and Warshel A., At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 2010. **78**(6): p. 1339-1375.
3. Robinson P. K., Enzymes: principles and biotechnological applications. Understanding Biochemistry: *Enzymes and Membranes*, 2015. **59**: p. 1-41.
4. *UNAIDS World AIDS Day Report*. 2016.
5. Weiss R. A., How Does Hiv Cause Aids. *Science*, 1993. **260**(5112): p. 1273-1279.
6. Nyamweya S., Hegedus A., Jaye A., Rowland-Jones S., Flanagan K. L. and Macallan D. C., Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Rev. Med. Virol.*, 2013. **23**(4): p. 221-40.
7. Rong R., Li B., Lynch R. M., Haaland R. E., Murphy M. K., Mulenga J., Allen S. A., Pinter A., Shaw G. M., Hunter E., Robinson J. E., Gnanakaran S. and Derdeyn C. A., Escape from Autologous Neutralizing Antibodies in Acute/Early Subtype C HIV-1 Infection Requires Multiple Pathways. *PloS Pathog.*, 2009. **5**(9).
8. Arts E. J. and Hazuda D. J., HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.*, 2012. **2**(4): p. a007161.
9. Reiche E. M., Nunes S. O. and Morimoto H. K., Stress, depression, the immune system, and cancer. *Lancet Oncol.*, 2004. **5**(10): p. 617-25.
10. Jarmuz A., Chester A., Bayliss J., Gisbourne J., Dunham I., Scott J. and Navaratnam N., An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*, 2002. **79**(3): p. 285-96.

11. Conticello S. G., The AID/APOBEC family of nucleic acid mutators. *Genome Biol.*, 2008. **9**(6): p. 229.
12. LaRue R. S., Andresdottir V., Blanchard Y., Conticello S. G., Derse D., Emerman M., Greene W. C., Jonsson S. R., Landau N. R., Lochelt M., Malik H. S., Malim M. H., Munk C., O'Brien S. J., Pathak V. K., Strebel K., Wain-Hobson S., Yu X. F., Yuhki N. and Harris R. S., Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.*, 2009. **83**(2): p. 494-7.
13. Siriwardena S. U., Guruge T. A. and Bhagwat A. S., Characterization of the Catalytic Domain of Human APOBEC3B and the Critical Structural Role for a Conserved Methionine. *J. Mol. Biol.*, 2015. **427**(19): p. 3042-55.
14. Chen Q., Xiao X., Wolfe A. and Chen X. S., The in vitro Biochemical Characterization of an HIV-1 Restriction Factor APOBEC3F: Importance of Loop 7 on Both CD1 and CD2 for DNA Binding and Deamination. *J. Mol. Biol.*, 2016. **428**(13): p. 2661-70.
15. Reuben S. Harris, Mark T. Liddament. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.*, 2004. **4**(11): p. 868-77
16. Esnault C., Heidmann O., Delebecque F., Dewannieux M., Ribet D., Hance A. J., Heidmann T. and Schwartz O., APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, 2005. **433**(7024): p. 430-3.
17. Bogerd H. P., Wiegand H. L., Doehle B. P., Lueders K. K. and Cullen B. R., APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.*, 2006. **34**(1): p. 89-95.
18. Jonsson S. R., Hache G., Stenglein M. D., Fahrenkrug S. C., Andresdottir V. and Harris R. S., Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. *Nucleic Acids Res.*, 2006. **34**(19): p. 5683-94.
19. Moris A., Murray S. and Cardinaud S., AID and APOBECs span the gap between innate and adaptive immunity. *Front. Microbiol.*, 2014. **5**.
20. Sheehy A. M., Gaddis N. C., Choi J. D. and Malim M. H., Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 2002. **418**(6898): p. 646-650.

21. Sheehy A. M., Gaddis N. C. and Malim M. H., The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.*, 2003. **9**(11): p. 1404-1407.
22. Malim M. H. and Emerman M., HIV-1 accessory proteins - Ensuring viral survival in a hostile environment. *Cell Host & Microbe*, 2008. **3**(6): p. 388-398.
23. Malim M. H., APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos. Trans. R. Soc.Lond. B. Biol. Sci.*, 2009. **364**(1517): p. 675-687.
24. Berger G., Durand S., Fargier G., Nguyen X. N., Cordeil S., Bouaziz S., Muriaux D., Darlix J. L. and Cimarrelli A., APOBEC3A Is a Specific Inhibitor of the Early Phases of HIV-1 Infection in Myeloid Cells. *Plos Pathog.* , 2011. **7**(9).
25. Koning F. A., Goujon C., Bauby H. and Malim M. H., Target Cell-Mediated Editing of HIV-1 cDNA by APOBEC3 Proteins in Human Macrophages. *J.Virol.*, 2011. **85**(24): p. 13448-13452.
26. Harris R. S., Hultquist J. F. and Evans D. T., The Restriction Factors of Human Immunodeficiency Virus. *J. Biol.Chem.*, 2012. **287**(49): p. 40875-40883.
27. Harris R. S., Bishop K. N., Sheehy A. M., Craig H. M., Petersen-Mahrt S. K., Watt I. N., Neuberger M. S. and Malim M. H., DNA deamination mediates innate immunity to retroviral infection. *Cell*, 2003. **113**(6): p. 803-9.
28. Khan M. A., Goila-Gaur R., Opi S., Miyagi E., Takeuchi H., Kao S. and Strebel K., Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. *Retrovirology*, 2007. **4**: p. 48.
29. Goila-Gaur R. and Strebel K., HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology*, 2008. **5**: p. 51.
30. Bishop K. N., Holmes R. K. and Malim M. H., Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J.Virol.*, 2006. **80**(17): p. 8450-8458.
31. Conticello S. G., Harris R. S. and Neuberger M. S., The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.*, 2003. **13**(22): p. 2009-13.
32. Nowarski R. and Kotler M., APOBEC3 Cytidine Deaminases in Double-Strand DNA Break Repair and Cancer Promotion. *Cancer Res.* , 2013. **73**(12): p.

3494-3498.

33. Alexandrov L. B., Nik-Zainal S., Wedge D. C., Aparicio S. A., Behjati S., Biankin A. V., Bignell G. R., Bolli N., Borg A., Borresen-Dale A. L., Boyault S., Burkhardt B., Butler A. P., Caldas C., Davies H. R., Desmedt C., Eils R., Eyfjord J. E., Foekens J. A., Greaves M., Hosoda F., Hutter B., Ilcic T., Imbeaud S., Imielinski M., Jager N., Jones D. T., Jones D., Knappskog S., Kool M., Lakhani S. R., Lopez-Otin C., Martin S., Munshi N. C., Nakamura H., Northcott P. A., Pajic M., Papaemmanuil E., Paradiso A., Pearson J. V., Puente X. S., Raine K., Ramakrishna M., Richardson A. L., Richter J., Rosenstiel P., Schlesner M., Schumacher T. N., Span P. N., Teague J. W., Totoki Y., Tutt A. N., Valdes-Mas R., van Buuren M. M., van 't Veer L., Vincent-Salomon A., Waddell N., Yates L. R., Australian Pancreatic Cancer Genome I., Consortium I. B. C., Consortium I. M.-S., PedBrain I., Zucman-Rossi J., Futreal P. A., McDermott U., Lichter P., Meyerson M., Grimmond S. M., Siebert R., Campo E., Shibata T., Pfister S. M., Campbell P. J. and Stratton M. R., Signatures of mutational processes in human cancer. *Nature*, 2013. **500**(7463): p. 415-21.
34. Burns M. B., Lackey L., Carpenter M. A., Rathore A., Land A. M., Leonard B., Refsland E. W., Kotandeniya D., Tretyakova N., Nikas J. B., Yee D., Temiz N. A., Donohue D. E., McDougale R. M., Brown W. L., Law E. K. and Harris R. S., APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, 2013. **494**(7437): p. 366-70.
35. Nik-Zainal S., Wedge D. C., Alexandrov L. B., Petljak M., Butler A. P., Bolli N., Davies H. R., Knappskog S., Martin S., Papaemmanuil E., Ramakrishna M., Shlien A., Simonic I., Xue Y. L., Tyler-Smith C., Campbell P. J. and Stratton M. R., Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.*, 2014. **46**(5): p. 487-491.
36. Komatsu A., Nagasaki K., Fujimori M., Amano J. and Miki Y., Identification of novel deletion polymorphisms in breast cancer. *Int. J. Oncol.*, 2008. **33**(2): p. 261-270.
37. Taylor B. J. M., Nik-Zainal S., Wu Y. L., Stebbings L. A., Raine K., Campbell P. J., Rada C., Stratton M. R. and Neuberger M. S., DNA deaminases induce

- break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife*, 2013. **2**.
38. Shinohara M., Io K., Shindo K., Matsui M., Sakamoto T., Tada K., Kobayashi M., Kadowaki N. and Takaori-Kondo A., APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci. Rep.* , 2012. **2**: p. 806.
 39. Leonard B., Hart S. N., Burns M. B., Carpenter M. A., Temiz N. A., Rathore A., Vogel R. I., Nikas J. B., Law E. K., Brown W. L., Li Y., Zhang Y. J., Maurer M. J., Oberg A. L., Cunningham J. M., Shridhar V., Bell D. A., April C., Bentley D., Bibikova M., Cheetham R. K., Fan J. B., Grocock R., Humphray S., Kingsbury Z., Peden J., Chien J., Swisher E. M., Hartmann L. C., Kalli K. R., Goode E. L., Sicotte H., Kaufmann S. H. and Harris R. S., APOBEC3B Upregulation and Genomic Mutation Patterns in Serous Ovarian Carcinoma. *Cancer Res.*, 2013. **73**(24): p. 7222-7231.
 40. Sasaki H. S. A., Tatematsu T., Shitara M., Hikosaka Y., Okuda K., Moriyama S., Yano M., Fujii Y., APOBEC3B gene overexpression in non-small-cell lung cancer. *Biomed Rep.* , 2014. **2**: p. 392-395.
 41. Ding Q., Chang C. J., Xie X., Xia W., Yang J. Y., Wang S. C., Wang Y., Xia J., Chen L., Cai C., Li H., Yen C. J., Kuo H. P., Lee D. F., Lang J., Huo L., Cheng X., Chen Y. J., Li C. W., Jeng L. B., Hsu J. L., Li L. Y., Tan A., Curley S. A., Ellis L. M., Dubois R. N. and Hung M. C., APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. *J. Clin. Invest.*, 2011. **121**(11): p. 4526-36.
 42. Pham P., Landolph A., Mendez C., Li N. and Goodman M. F., A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *J. Biol.Chem.*, 2013. **288**(41): p. 29294-304.
 43. Chen K. M., Harjes E., Gross P. J., Fahmy A., Lu Y., Shindo K., Harris R. S. and Matsuo H., Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*, 2008. **452**(7183): p. 116-9.
 44. Holden L. G., Prochnow C., Chang Y. P., Bransteitter R., Chelico L., Sen U., Stevens R. C., Goodman M. F. and Chen X. S., Crystal structure of the anti-viral

- APOBEC3G catalytic domain and functional implications. *Nature*, 2008. **456**(7218): p. 121-4.
45. Aydin H., Taylor M. W. and Lee J. E., Structure-guided analysis of the human APOBEC3-HIV restrictome. *Structure*, 2014. **22**(5): p. 668-84.
 46. Neuberger M. S., Harris R. S., Di Noia J. and Petersen-Mahrt S. K., Immunity through DNA deamination. *Trends Biochem. Sci.*, 2003. **28**(6): p. 305-312.
 47. Bransteitter R., Prochnow C. and Chen X. S., The current structural and functional understanding of APOBEC deaminases. *Cell Mol. Life Sci.*, 2009. **66**(19): p. 3137-47.
 48. Furukawa A., Nagata T., Matsugami A., Habu Y., Sugiyama R., Hayashi F., Kobayashi N., Yokoyama S., Takaku H. and Katahira M., Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.*, 2009. **28**(4): p. 440-51.
 49. Shi K., Carpenter M. A., Kurahashi K., Harris R. S. and Aihara H., Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. *J. Biol. Chem.*, 2015. **290**(47): p. 28120-30.
 50. Byeon I. J., Byeon C. H., Wu T., Mitra M., Singer D., Levin J. G. and Gronenborn A. M., Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity. *Biochemistry*, 2016. **55**(21): p. 2944-59.
 51. Bohn M. F., Shandilya S. M., Silvas T. V., Nalivaika E. A., Kouno T., Kelch B. A., Ryder S. P., Kurt-Yilmaz N., Somasundaran M. and Schiffer C. A., The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. *Structure*, 2015. **23**(5): p. 903-11.
 52. Byeon I. J., Ahn J., Mitra M., Byeon C. H., Hercik K., Hritz J., Charlton L. M., Levin J. G. and Gronenborn A. M., NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.*, 2013. **4**: p. 1890.
 53. Kitamura S., Ode H., Nakashima M., Imahashi M., Naganawa Y., Kurosawa T., Yokomaku Y., Yamane T., Watanabe N., Suzuki A., Sugiura W. and Iwatani Y., The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat.*

- Struct. Mol. Biol.*, 2012. **19**(10): p. 1005-10.
54. Bohn M. F., Shandilya S. M., Albin J. S., Kouno T., Anderson B. D., McDougale R. M., Carpenter M. A., Rathore A., Evans L., Davis A. N., Zhang J., Lu Y., Somasundaran M., Matsuo H., Harris R. S. and Schiffer C. A., Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure*, 2013. **21**(6): p. 1042-50.
 55. Siu K. K., Sultana A., Azimi F. C. and Lee J. E., Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat. Commun.*, 2013. **4**: p. 2593.
 56. Nakashima M., Ode H., Kawamura T., Kitamura S., Naganawa Y., Awazu H., Tsuzuki S., Matsuoka K., Nemoto M., Hachiya A., Sugiura W., Yokomaku Y., Watanabe N. and Iwatani Y., Structural Insights into HIV-1 Vif-APOBEC3F Interaction. *J. Virol.*, 2015. **90**(2): p. 1034-47.
 57. Shandilya S. M., Nalam M. N., Nalivaika E. A., Gross P. J., Valesano J. C., Shindo K., Li M., Munson M., Royer W. E., Harjes E., Kono T., Matsuo H., Harris R. S., Somasundaran M. and Schiffer C. A., Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure*, 2010. **18**(1): p. 28-38.
 58. Li M., Shandilya S. M., Carpenter M. A., Rathore A., Brown W. L., Perkins A. L., Harki D. A., Solberg J., Hook D. J., Pandey K. K., Parniak M. A., Johnson J. R., Krogan N. J., Somasundaran M., Ali A., Schiffer C. A. and Harris R. S., First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.*, 2012. **7**(3): p. 506-17.
 59. Lu X., Zhang T., Xu Z., Liu S., Zhao B., Lan W., Wang C., Ding J. and Cao C., Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *J. Biol. Chem.*, 2015. **290**(7): p. 4010-21.
 60. Kouno T., Luengas E. M., Shigematsu M., Shandilya S. M. D., Zhang J. Y., Chen L., Hara M., Schiffer C. A., Harris R. S. and Matsuo H., Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat. Struct. Mol. Biol.*, 2015. **22**(6): p. 485-U79.
 61. Xiao X., Li S. X., Yang H. and Chen X. S., Crystal structures of APOBEC3G

- N-domain alone and its complex with DNA. *Nat. Commun.*, 2016. **7**: p. 12193.
62. Xiao X., Yang H., Arutiunian V., Fang Y., Besse G., Morimoto C., Zirkle B. and Chen X. S., Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. *Nucleic Acids Res.*, 2017.
 63. Marx A., Galilee M. and Alian A., Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes. *Sci. Rep.*, 2015. **5**: p. 18191.
 64. Rathore A., Carpenter M. A., Demir O., Ikeda T., Li M., Shaban N. M., Law E. K., Anokhin D., Brown W. L., Amaro R. E. and Harris R. S., The Local Dinucleotide Preference of APOBEC3G Can Be Altered from 5 '-CC to 5 '-TC by a Single Amino Acid Substitution. *J. Mol. Biol.*, 2013. **425**(22): p. 4442-4454.
 65. Shi K., Carpenter M. A., Banerjee S., Shaban N. M., Kurahashi K., Salamango D. J., McCann J. L., Starrett G. J., Duffy J. V., Demir O., Amaro R. E., Harki D. A., Harris R. S. and Aihara H., Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.*, 2017. **24**(2): p. 131-139.
 66. Kouno T., Silvas T. V., Hilbert B. J., Shandilya S. M. D., Bohn M. F., Kelch B. A., Royer W. E., Somasundaran M., Kurt Yilmaz N., Matsuo H. and Schiffer C. A., Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.*, 2017. **8**: p. 15024.
 67. Harjes S., Jameson G. B., Filichev V. V., Edwards P. J. and Harjes E., NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate. *Nucleic Acids Res.*, 2017. **45**(9): p. 5602-5613.
 68. Losey H. C., Ruthenburg A. J. and Verdine G. L., Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.*, 2006. **13**(2): p. 153-9.
 69. Matthews M. M., Thomas J. M., Zheng Y., Tran K., Phelps K. J., Scott A. I., Havel J., Fisher A. J. and Beal P. A., Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct.*

- Mol. Biol.*, 2016. **23**(5): p. 426-33.
70. Otashne M. and Gann A., Genes and Signals. 2001, *Cold Spring Harbor Laboratory Press*: Cold Spring Harbor, NY.
 71. Vonhippel P. H. and Berg O. G., Facilitated Target Location in Biological-Systems. *J. Biol. Chem.*, 1989. **264**(2): p. 675-678.
 72. Halford S. E. and Marko J. F., How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, 2004. **32**(10): p. 3040-3052.
 73. Berg O. G., On Diffusion-Controlled Dissociation. *Chem. Phys.*, 1978. **31**(1): p. 47-57.
 74. Bellomy G. R. and Record M. T., Jr., Stable DNA loops in vivo and in vitro: roles in gene regulation at a distance and in biophysical characterization of DNA. *Prog. Nucleic Acid Res. Mol. Biol.*, 1990. **39**: p. 81-128.
 75. Vologodskii A. and Cozzarelli N. R., Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophys. J.*, 1996. **70**(6): p. 2548-56.
 76. Feng Y. and Chelico L., Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. *J. Biol. Chem.*, 2011. **286**(13): p. 11415-26.
 77. Nowarski R., Britan-Rosich E., Shiloach T. and Kotler M., Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat. Struct. Mol. Biol.*, 2008. **15**(10): p. 1059-66.
 78. Chelico L., Pham P., Calabrese P. and Goodman M. F., APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.*, 2006. **13**(5): p. 392-9.
 79. Senavirathne G., Jaszczur M., Auerbach P. A., Upton T. G., Chelico L., Goodman M. F. and Rueda D., Single-stranded DNA Scanning and Deamination by APOBEC3G Cytidine Deaminase at Single Molecule Resolution. *J. Biol. Chem.*, 2012. **287**(19): p. 15826-15835.
 80. Shlyakhtenko L. S., Lushnikov A. Y., Miyagi A., Li M., Harris R. S. and Lyubchenko Y. L., Nanoscale Structure and Dynamics of ABOBEC3G Complexes with Single-Stranded DNA. *Biochemistry*, 2012. **51**(32): p. 6432-6440.
 81. Ara A., Love R. P. and Chelico L., Different mutagenic potential of HIV-1

- restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms. *PLoS Pathog.*, 2014. **10**(3): p. e1004024.
82. Brackley C. A., Cates M. E. and Marenduzzo D., Effect of DNA conformation on facilitated diffusion. *Biochem. Soc. Trans.*, 2013. **41**(2): p. 582-8.
 83. Furukawa A., Sugase K., Morishita R., Nagata T., Kodaki T., Takaori-Kondo A., Ryo A. and Katahira M., Quantitative analysis of location- and sequence-dependent deamination by APOBEC3G using real-time NMR spectroscopy. *Angew. Chem. Int. Ed.*, 2014. **53**(9): p. 2349-52.
 84. Mechetin G. V. and Zharkov D. O., Mechanisms of diffusional search for specific targets by DNA-dependent proteins. *Biochemistry (Mosc)*, 2014. **79**(6): p. 496-505.
 85. Mak C. H., Pham P., Afif S. A. and Goodman M. F., A Mathematical Model for Scanning and Catalysis on Single-stranded DNA, Illustrated with Activation-induced Deoxycytidine Deaminase. *J. Biol. Chem.*, 2013. **288**(41): p. 29786-29795.
 86. Yu Q., Chen D., Konig R., Mariani R., Unutmaz D. and Landau N. R., APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *J. Biol. Chem.*, 2004. **279**(51): p. 53379-86.
 87. Bhagwat A. S., Hao W., Townes J. P., Lee H., Tang H. and Foster P. L., Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in *Escherichia coli*. *Proc. Natl. Acad. Sci. U S A*, 2016. **113**(8): p. 2176-81.
 88. Haradhvala N. J., Polak P., Stojanov P., Covington K. R., Shinbrot E., Hess J. M., Rheinbay E., Kim J., Maruvka Y. E., Braunstein L. Z., Kamburov A., Hanawalt P. C., Wheeler D. A., Koren A., Lawrence M. S. and Getz G., Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell*, 2016. **164**(3): p. 538-49.
 89. Hoopes J. I., Cortez L. M., Mertz T. M., Malc E. P., Mieczkowski P. A. and Roberts S. A., APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep.*, 2016. **14**(6): p. 1273-82.

90. D V. and G V. J., Biochemistry. 4 ed. 2010, Chichester: *John Wiley & Sons*, p.1173-1180.
91. Mitra M., Hercik K., Byeon I. J., Ahn J., Hill S., Hincbee-Rodriguez K., Singer D., Byeon C. H., Charlton L. M., Nam G., Heidecker G., Gronenborn A. M. and Levin J. G., Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. *Nucleic Acids Res.*, 2014. **42**(2): p. 1095-110.
92. T. D. Goddard D. G. K., SPARKY 3. *University of California, San Francisco*, 2006.
93. Harris D. C., Nonlinear least-squares curve fitting with Microsoft excel solver. *J. Chem. Edu.*, 1998. **75**(1): p. 119-121.
94. Bishop K. N., Holmes R. K., Sheehy A. M., Davidson N. O., Cho S. J. and Malim M. H., Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.*, 2004. **14**(15): p. 1392-6.
95. Liddament M. T., Brown W. L., Schumacher A. J. and Harris R. S., APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr. Biol.*, 2004. **14**(15): p. 1385-91.
96. Refsland E. W. and Harris R. S., The APOBEC3 family of retroelement restriction factors. *Curr. Top Microbiol. Immunol.*, 2013. **371**: p. 1-27.
97. Vasudevan A. A., Smits S. H., Hoppner A., Haussinger D., Koenig B. W. and Munk C., Structural features of antiviral DNA cytidine deaminases. *Biol. Chem.*, 2013. **394**(11): p. 1357-70.
98. Chung S. J., Fromme J. C. and Verdine G. L., Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.*, 2005. **48**(3): p. 658-60.
99. Demorest Z. L., Li M. and Harris R. S., Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. *J. Biol. Chem.*, 2011. **286**(30): p. 26568-75.
100. Amanchy R., Periaswamy B., Mathivanan S., Reddy R., Tattikota S. G. and Pandey A., A curated compendium of phosphorylation motifs. *Nat. Biotechnol.*, 2007. **25**(3): p. 285-6.
101. Lackey L., Law E. K., Brown W. L. and Harris R. S., Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA

- deamination. *Cell Cycle*, 2013. **12**(5): p. 762-772.
102. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. and Bairoch A., ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, 2003. **31**(13): p. 3784-8.
 103. Harjes S., Solomon W. C., Li M., Chen K. M., Harjes E., Harris R. S. and Matsuo H., Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J. Virol.*, 2013. **87**(12): p. 7008-14.
 104. Harris R. S. and Dudley J. P., APOBECs and virus restriction. *Virology*, 2015. **479-480**: p. 131-45.
 105. Guo Y., Dong L., Qiu X., Wang Y., Zhang B., Liu H., Yu Y., Zang Y., Yang M. and Huang Z., Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature*, 2014. **505**(7482): p. 229-33.
 106. Richards C., Albin J. S., Demir O., Shaban N. M., Luengas E. M., Land A. M., Anderson B. D., Holten J. R., Anderson J. S., Harki D. A., Amaro R. E. and Harris R. S., The Binding Interface between Human APOBEC3F and HIV-1 Vif Elucidated by Genetic and Computational Approaches. *Cell Rep.*, 2015. **13**(9): p. 1781-1788.
 107. Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. and Ferrin T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 2004. **25**(13): p. 1605-12.
 108. Albin J. S., LaRue R. S., Weaver J. A., Brown W. L., Shindo K., Harjes E., Matsuo H. and Harris R. S., A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. *J. Biol. Chem.*, 2010. **285**(52): p. 40785-92.
 109. Land A. M., Shaban N. M., Evans L., Hultquist J. F., Albin J. S. and Harris R. S., APOBEC3F determinants of HIV-1 Vif sensitivity. *J. Virol.*, 2014. **88**(21): p. 12923-7.
 110. Kamba K., Development of real-time NMR monitoring method and elucidation of the deamination mechanism of APOBEC3G. *Kyoto University*, 2016.

List of publications

Publications that constitute this thesis

1. Observation by real-time NMR and interpretation of length- and location-dependency of deamination by APOBEC3B.

Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
ACS Chemical Biology, 2017, DOI: 10.1021/acscchembio.7b00662

(chapter 2)

2. Influence of the DNA sequence/length and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F.

Li Wan, Takashi Nagata, Masato Katahira
RSC Physical Chemistry Chemical Physics, 2017, DOI: 10.1039/C7CP04477A

(chapter 3)

3. Influence of the ssDNA concentration on the length- and location-dependent deaminase activity of APOBEC3F

Li Wan, Takashi Nagata, Masato Katahira

In preparation.

(chapter 4)

4. Investigation of the effects of Vif complex on the deaminase activity of APOBEC3F

Li Wan, Takashi Nagata, Masato Katahira

In preparation.

(chapter 5)

List of presentations

1. Deamination mechanism of human antiviral factor APOBEC3G by Real-Time NMR
Li Wan, Keisuke Kamba, Takashi Nagata, Masato Katahira
The 16th NMR Symposium, for Young Researchers, September 27-29, 2015, Mura, Japan
2. Analysis of the biochemical characteristics of APOBEC3F by real-time NMR
Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
The 54th Annual Meeting of the Biophysical Society of Japan, November 25-27, 2016, Tsukuba, Japan
3. Characterization of the deamination activity of APOBEC3B by real-time NMR, which is distinct from that of APOBEC3G
Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
The 54th Annual Meeting of the Biophysical Society of Japan, November 25-27, 2016, Tsukuba, Japan
4. Characterization of the deamination activity of APOBEC3B by real-time NMR, which is distinct from that of APOBEC3G
Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
The 7th Asia-Pacific NMR Symposium, February 16-19, 2017, Bangalore, India
5. Influences of length and concentration of the DNA substrate, as well as the location of the target sequence, on deamination by APOBEC3F
Li Wan, Takashi Nagata, Masato Katahira
The 55th Annual Meeting of the Biophysical Society of Japan, September 19-21, 2017, Kumamoto, Japan

List of honors

1. The 54th Annual Meeting of the Biophysical Society of Japan, Student Presentation Award

Characterization of the deamination activity of APOBEC3B by real-time NMR, which is distinct from that of APOBEC3G

Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
November 25-27, 2016, Tsukuba, Japan

2. The 7th Asia-Pacific NMR Symposium, The Best Student Oral/Poster Award

Characterization of the deamination activity of APOBEC3B by real-time NMR, which is distinct from that of APOBEC3G

Li Wan, Takashi Nagata, Ryo Morishita, Akifumi Takaori-Kondo, Masato Katahira
February 16-19, 2017, Bangalore, India

Acknowledgements

I have spent more than three years in Japan without noticing it and I have acquired so much new knowledge and skills during my Ph.D. period. Here, I want first and foremost to express my appreciation and respect to my supervisor, Prof. Masato Katahira. I have never experienced such a nice professor from who I learned a lot about how to conduct a research topic and treat it with heart. You inspired me with your kind smile and positive attitude when I felt disappointed and confused. You also provided me many opportunities to attend kinds of the domestic and international conferences, by which my abilities of introducing and explaining my research to others were well trained. I was deeply impressed on your serious attitude to researches, as well as the amiable feeling you gave me in the daily communication. I am eternally grateful for everything that you has done for me.

I would also like to thank Associate Prof. Takashi Nagata who guided me a lot about my experiments and also my paper writing. Once I met some problems and troubles with my experiments, I just knocked at your door and you always listened and explained my problems patiently. I realized that you possesses a quite wide range of knowledge. I often set a goal to aim at you but I know there is still a long way forward for me to improve myself. The spirit I learned from you is that you treat nearly all things seriously. When I prepared my first manuscript, I thought I did well. After checking by you, I was shocked that I made so many mistakes in the manuscript. From that time on, I tried my best to learn this kind of serious attitude to research. I think it will be a lifetime of treasure for me.

I appreciated it very much that I had so many excellent colleagues in this lab. Keisuke Kamba guided me to enter this research topic. Keiko Kondo and Yudai Yamaoki often helped me to solve my problems that I met in experiments and gave me plenary suggestions. Lin Meng-I taught me how to use Nickel column and AKTA purification system. And the other members, Wan Hasnida WAN OSMAN, Nesreen Hamad, Mamiko Iida, Masayuki Miyake and Akiho Hiyama, I enjoyed every moment with them and I would never forget the happy time we have been together.

Here, I would also like to thank my friends in Uji called "Uji group", Kai Xue,

Yanyan Liu, Shaotang Song, Chen Chen, Xinhao He, and Xiangxun Lu. We experienced a lot together. No matter it was stressful, happy or sad, we just called and went outside to have fun, then released. I appreciated everyone and I was so happy that your guys kept me company within these oversea-years. It must be yearning when I recall my Ph.D. study in the future and this unforgettable experience will be memorable forever.

Meanwhile, I would like to thank my friends in China. They helped me with spirits or financial support on my way of pursuing my Ph.D. degree. I treat your guys with my grateful heart. Without your support, I could not keep going ahead. Your guys have already become one part of my life.

Finally, I am grateful to my family, the most important people for me in this world. You always support me silently. Without your support and care, I would not have been able to achieve my goals and pursue my dreams. You always give me a kind of power which encourages and supports me on my way to the future. Thank you that you have been there for me over past 30 years. I will try my best to be someone who is deserving of your support and love.

Li Wan

Laboratory of Bioenergy

Department of Fundamental Energy Science

Graduate School of Energy Science

Kyoto University

September, 2017