Published in final edited form as:

DNA Repair (Amst). 2004 January 5; 3(1): 85–89.

DNA-cytosine Deaminases: From Antibody Maturation to Antiviral Defense

Ashok S. Bhagwat

Department of Chemistry, Wayne State University, Detroit, MI 48202, U.S.A.

Abstract

The deamination of cytosines in DNA to uracil, thought to be initiated by free water within the cells, is a well studied pathway by which C to T mutations occur. Until recently, this conversion was frequently referred to as being spontaneous because of the involvement of cellular water. The recent discovery of a family of enzymes in mammalian cells that catalyze this reaction was unexpected and has created excitement in at least two areas of biology, immunology and virology. One of these enzymes, activation-induced cytidine deaminase, is required for the final steps in the maturation of antibodies. The key features of this process include the introduction of a wide variety of base substitutions in the immunoglobulin genes and the creation of region-specific double-strand breaks. Another member of this family, Apobec3G, is involved in the mutational inactivation and degradation of the human immunodeficiency virus (HIV-1). Among the many intriguing features of these processes is the likely involvement of the enzyme that is thought to “protect” cellular DNA against the accumulation of uracils, uracil-DNA glycosylase (UDG). It appears that in certain situations, the newly discovered DNA-cytosine deaminases can team up with UDG to extensively mutate and degrade DNA. In this article Ashok Bhagwat discusses the many questions raised regarding the role of these enzymes in protecting cells against infections, and about their possible roles in genome evolution and carcinogenesis.

Keywords

Somatic hypermutations; Class-switch recombination; Gene conversion; CEM15; Vif

For over a century there has been a fundamental assumption regarding mutations that in the short term most mutations are harmful and cells try to avoid acquiring them. This is so ingrained in biology and social thought that the extensive efforts made by cells to prevent DNA damage and to repair any damage that escapes the preventive mechanisms, are well publicized. One needs to look no further than the title of this journal to appreciate the importance of this idea. This view of genome stability may have to be significantly altered due to a series of exciting papers that have appeared in the past few months about a new class of human enzymes, DNA-cytosine deaminases.

Activation-induced cytidine deaminase (AID) is the founding member of this class. It is required for the maturation of antibodies through genetic alterations called somatic

*Tel.: +1-313-577-2547; Fax: +1-313-577-8822. axb@chem.wayne.edu.

hypermutations (SHM), class switch recombination (CSR) and gene conversion (GC; in some mammals and chickens). SHM is perhaps the most interesting of these processes because it introduces point mutations in the variable segments of immunoglobulin (Ig) genes at a rate that is 10^5 – 10^6 -fold higher than the rest of the genome. While the key role of AID in SHM has been known for three years, the underlying mechanism has been unclear.

Several recent papers show that AID converts cytosines in DNA to uracil in a transcription-dependent, strand-biased fashion and that this is the likely starting point of genetic alterations that result in SHM, CSR and GC. The ability of AID to attack cytosines in single-stranded (SS) DNA was demonstrated biochemically [1–4] or inferred from genetic data in *Escherichia coli* [1,5,6]. The transcription-dependence of this reaction and its strand-bias were demonstrated in *E. coli* [1,6] and using a T7 RNA polymerase-based *in vitro* transcription system [1,3,7]. The ability of AID to deaminate cytosines in DNA and the expected biological role of this reaction in SHM, CSR and GC suggests that AID should be referred to as a DNA-cytosine deaminase instead of its original name, cytidine deaminase.

Of particular significance is the paper by Pham et al [7] which shows that AID targets cytosines within WRC sequences (W is A or T, R is a purine), and deaminates cytosines in single-stranded DNA in a processive fashion. The targeting of these sequences by AID is apparent in the examination of the observed mutational hotspots, where 14 out of 15 are WRC sequences. The processivity of AID on SS DNA is evidenced in the high number of mutations in each clone (10 to 70 per clone in a 230 nt segment) found in the experiment. The authors persuasively argue that the observed multiple mutations in all the clones are unlikely to be due either to multiple binding events of DNA by AID or AID aggregation.

The targeting of WRC (or WRCY, Y is a pyrimidine) sequences during maturation of Ig genes is a well known and much-debated property of SHM. A key question has been whether the targeting occurs in phase I of mutagenesis where AID acts or during phase II where lesion bypass polymerases are presumed to act (Figure 1). The results by Pham et al suggest that the targeting occurs in phase I and not II. It also reinforces the notion that the target of AID is DNA and not mRNA as is sometimes suggested [8]. However, this work is unlikely to be the complete story of how AID targets cytosines in DNA. This is because extended SS DNA without the protection of single-strand DNA-binding protein (RPA in mammalian cells) rarely occurs in cells. Additionally, AID mutagenesis is coupled with transcription [9] and the latter process rarely creates long SS regions in DNA *in vivo*. However, it is possible that AID targets WRC sequences within the transient SS non-transcribed strand within the transcription bubble. This remains to be demonstrated in mammalian cells or in an *in vitro* transcription system involving eukaryotic RNA polymerase.

An additional issue regarding targeting by AID concerns the lack of strand bias in C to T mutations found in SHM. If AID targets the non-transcribed strand of Ig genes and the resulting uracils are not repaired, then there should be an excess of C to T mutations in that strand over G to A changes. The absence of such a bias in SHM even in UDG^{-/-} mice is puzzling [10]. How is this bias “lost” during the processing of uracils? Or, is it possible that

AID somehow targets WRC sequences in both the DNA strands within the transcription bubble?

The reported processivity of AID *in vitro* [7] may explain a subclass of SHM where the mutations are found to be clustered. Clustering of mutations in Ig that have undergone SHM is an interesting observation [11], but it has not been studied extensively. It is easy to see how processive action of AID would create a cluster of mutations in a single encounter, but there are several problems with this interpretation. First, although there are few careful studies of the mutational clusters *per se* [12], a simple analysis would suggest that mutational clusters are unlikely to occur in a majority of B cells undergoing SHM. For example, a recent study found that among the 384 clones of Ig that had undergone SHM, ~59% had suffered three or fewer mutations ([13] and U. Storb, personal communication). Second, a possible alternate explanation for why SHM are sometimes found in clusters is that they are acquired in successive generations of the same B cell clone. “Geneological” analysis of independent clones of Ig support this idea {see [13,14] for example}. Third, if the biologically relevant target of AID is the non-transcribed strand of a transcription bubble, AID is unlikely to act processively. RNA polymerases move along DNA at ~30 nt per sec, while most sequence-specific DNA-binding enzymes such as DNA methyltransferases and restriction endonucleases turn over at a rate of ~1 per min. If AID is similarly slow in its catalytic turnover, it would fall off DNA as soon as the transcription bubble passes it by. One way in which a slow AID could create a cluster of mutations during transcription is if it acted at transcriptional pause sites [9] or within R-loops [3]. At present, there is little experimental evidence to support these possibilities.

To examine the possibility that AID may cause multiple cytosine deaminations in a single encounter with a transcription bubble, we examined the sequence of independent mutations caused by AID in a transcribed *kan* gene [1]. The genetic selection used in this experiment requires that a C to T mutation is introduced in codon 94 of the gene to make the gene *kan*⁺ and the question was whether unselected C to T changes occurred elsewhere in the gene. A careful examination of 15 clones obtained in this experiment found no additional mutations in a 600 bp window, despite the presence of third position C’s in 83 codons of *kan*, 6 of which are within WRC sequence (M. Samaranayake and A. Bhagwat; unpublished results). Thus the relevance of the processivity of AID on SS DNA to SHM remains an intriguing, but untested, possibility.

While the mechanism of action of AID in SHM was being elucidated, a cousin of this protein, Apobec3G (aka CEM15) made its own splash. In another remarkable cluster of papers published in early summer, several groups independently demonstrated that the expression of this protein provided substantial protection for human cells against an HIV-1 strain that was missing a protein called virus infectivity factor (Vif) [15–18]. Moreover, the mechanism by which Apobec3G/CEM15 (hereafter referred to as Apobec3G) seems to restrict HIV growth is through the massive deamination of cytosines in the minus strand DNA copy of this RNA virus. Earlier, it had been shown that Apobec3G was required for the resistance of certain T-cell lines against *vif*⁻ HIV-1 [19] and that it is a mutator in *E. coli* [20]. However, the direct demonstration that Apobec3G prevents virus growth while introducing C to T mutations was a major achievement [15–18]. It has furthermore been

demonstrated that Apobec3G is active against other retroviruses besides HIV-1, including murine leukemia virus, simian immunodeficiency virus, and equine infectious anemia virus [16,18]. Whether Apobec3G and/or other related DNA cytosine deaminases are active against other families of viruses remains to be determined.

These findings raise many other issues which should occupy the attention of biologists for years to come. Some of these concern the protein Vif. How does Vif prevent the action of Apobec3G? Early indications are that Vif binds the latter protein reducing its packaging into the virion [21]. It also appears that some Apobec3G can be packaged despite the action of Vif causing C to T mutations. Do such mutations contribute to the known high mutability of HIV-1? Also, why does HIV-1 Vif not bind the non-human Apobec3G [21]? Another question concerns how widespread this phenomenon may be in nature. Vif is present in the genomes of only lentiviruses, which are not found in many mammals (rodents, for example). Consequently, are there Vif-like activities encoded by other types of retroviruses that may inhibit the non-primate Apobec3G proteins?

Yet another fascinating observation is that uracil-DNA glycosylase (UDG) is packaged into many animal viruses including HIV-1. Understanding the function of this DNA repair protein in viral life cycle becomes important because of the observations mentioned above. Presumably, the enzyme is present in the virions to protect the DNA copy of the virus against cytosine deaminations, but as pointed out by Harris et al. [16], UDG may actually aid the destruction of the viral DNA (Figure 2). If so, should UDG be considered an “antiviral” protein? Like most really good experiments, the recent series of experiments on Apobec3G raise important new questions while answering the older ones.

AID, Apobec1 and Apobec3G are members of a family of about 10 proteins [22]. With the emerging evidence that AID (through antibody maturation) and Apobec3G are involved in protecting cells against infectious agents, it would be interesting to know whether the remaining members of the family perform a similar function. It is already known that one other member of the family, Apobec3C, is a mutator in *E. coli*, while another member, Apobec2, is not [20]. It will be important to determine how many members of this class are DNA-cytosine deaminases and are involved in causing cellular or viral mutations. If these proteins are indeed found to cause mutations in other animal viruses or in cellular genes other than Ig, we may have to reevaluate what we mean by “spontaneous” mutations. It is known that constitutive expression of AID causes tumors-presumably by increasing the mutational load [23] and it is not far-fetched to think that the AID/Apobec3G class of enzymes are a significant source of mutations in many vertebrate cellular and viral genomes. While the controlled targeting of these mutations leads to protective effects, occasional uncontrolled mutagenesis on inappropriate targets by these proteins could have serious deleterious consequences for the cell.

Acknowledgements

I am indebted to F. Yoshimura (Wayne State University) for educating me about the biology of retroviruses and appreciate her comments on the manuscript. I would also like to thank P. Gearhart (National Institute of Aging, Baltimore, MD) and U. Storb (University of Chicago, Chicago, IL) for their comments on the manuscript. Further, I am grateful to Dr. Storb for sharing unpublished results. This work was supported by grants GM57200 and CA97899 from the National Institutes of Health.

Literature Cited

1. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* 2003; 31:2990–2994. [PubMed: 12799424]
2. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med.* 2003; 197:1291–1296. [PubMed: 12756266]
3. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature.* 2003; 422:726–730. [PubMed: 12692563]
4. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A.* 2003; 100:4102–4107. [PubMed: 12651944]
5. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature.* 2002; 418:99–103. [PubMed: 12097915]
6. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol.* 2003; 4:452–456. [PubMed: 12692548]
7. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature.* 2003; 424:103–107. [PubMed: 12819663]
8. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell.* 2000; 102:553–563. [PubMed: 11007474]
9. Storb U. The molecular basis of somatic hypermutation of immunoglobulin genes. *Curr Opin Immunol.* 1996; 8:206–214. [PubMed: 8725944]
10. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice. *Curr Biol.* 2002; 12:1748–1755. [PubMed: 12401169]
11. Gearhart PJ, Bogenhagen DF. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc Natl Acad Sci U S A.* 1983; 80:3439–3443. [PubMed: 6222379]
12. Winter DB, Phung QH, Umar A, Baker SM, Tarone RE, Tanaka K, Liskay RM, Kunkel TA, Bohr VA, Gearhart PJ. Altered spectra of hypermutation in antibodies from mice deficient for the DNA mismatch repair protein PMS2. *Proc Natl Acad Sci U S A.* 1998; 95:6953–6958. [PubMed: 9618520]
13. Michael N, Martin TE, Nicolae D, Kim N, Padjen K, Zhan P, Nguyen H, Pinkert C, Storb U. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity.* 2002; 16:123–134. [PubMed: 11825571]
14. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci U S A.* 1984; 81:3180–3184. [PubMed: 6203114]
15. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature.* 2003; 424:94–98. [PubMed: 12808465]
16. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. DNA deamination mediates innate immunity to retroviral infection. *Cell.* 2003; 113:803–809. [PubMed: 12809610]
17. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science.* 2003; 300:1112. [PubMed: 12750511]
18. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature.* 2003; 424:99–103. [PubMed: 12808466]

19. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*. 2002; 418:646–650. [PubMed: 12167863]
20. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell*. 2002; 10:1247–1253. [PubMed: 12453430]
21. Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, Munk C, Nymark-McMahon H, Landau NR. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*. 2003; 114:21–31. [PubMed: 12859895]
22. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*. 2002; 79:285–296. [PubMed: 11863358]
23. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T. Constitutive expression of AID leads to tumorigenesis. *J Exp Med*. 2003; 197:1173–1181. [PubMed: 12732658]
24. Phung QH, Winter DB, Cranston A, Tarone RE, Bohr VA, Fishel R, Gearhart PJ. Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med*. 1998; 187:1745–1751. [PubMed: 9607916]
25. Rada C, Ehrenstein MR, Neuberger MS, Milstein C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*. 1998; 9:135–141. [PubMed: 9697843]

Figure 1. A Model for Somatic Hypermutations

The stepwise processing of DNA by AID and other enzymes is shown schematically. The first row of steps (Phase I) followed by the filling-in of the one nucleotide gap in DNA by a (relatively) error-free DNA polymerase to restore a C•G pair is the normal base excision repair pathway for DNA repair. This creates no mutations. If no repair occurs (Phase II, Left), the uracil is fixed as thymine during replication causing a C to T mutation. The intermediate in this pathway, U•G, is unlikely to be subject to mismatch correction. If the abasic site created by the removal of uracil is copied by a lesion bypass polymerase (LBP; Phase II, Center), then mutations are introduced in the bottom (G containing) strand. The incorrectly paired bases introduced by LBPs are represented by X, Y and Z and this SHM pathway is labeled as "A". The intermediates in this pathway may be subjected to further processing by the cell including mismatch correction. If the gap created by the actions of an AP endonuclease and a dRPase at the abasic site is filled-in by an LBP (Phase II, Right), mutations are introduced in the top (C containing) strand. This is labeled as SHM pathway "B". The intermediates in this pathway may also be subject to mismatch correction. In some earlier publications [24,25] Phase I was distinguished from Phase II only by the presence of mismatch correction during the latter stage. However, lesion bypass is likely to be coupled with mismatch correction and hence these two processes have been combined together in Phase II in my model. This is a minimal model for SHM because it does not attempt to explain many of the well-documented features of somatic hypermutations in detail- mutations at non-C•G pairs, for example.

Figure 2. A Model for Mutagenesis and Degradation of HIV-1 Genome promoted by Apobec3G
 The possible sequence of steps in the damage and degradation of the HIV-1 genome are schematically shown [16]. The copying of the positive strand RNA as a negative DNA strand is initiated by the reverse transcriptase (RT) using a tRNA primer. Concomitant with this synthesis, parts of RNA in the RNA:DNA hybrid are degraded by RNase H activity associated with the RT. The resulting RNA primers are extended by the RT to make the positive DNA strand. While this is occurring, Apobec3G deaminates cytosines in the exposed single-stranded DNA. If the resulting uracils are copied by RT, G to A transitions will result in the HIV-1 genome. Alternately, if the uracils are excised by UDG and the resulting abasic

sites are processed by AP endonuclease, the synthesis of the positive DNA strand will be terminated and the HIV-1 genome will be fragmented. Many of the steps in the model, including deamination of cytosines in the minus DNA strand, have not been demonstrated biochemically.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript