Structural features of antiviral APOBEC3 proteins are linked to their functional activities

Shingo Kitamura[1,2], Hirotaka Ode[1] and Yasumasa Iwatani[1,3] *

[1] Laboratory of Infectious Diseases, Department of Infectious Diseases and Immunology, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan
[2] Department of Biotechnology, Graduate School of Engineering, Nagoya University, Nagoya, Japan
[3] Department of AIDS Research, Graduate School of Medicine, Nagoya University, Nagoya, Japan

Edited by:
Akio Adachi, The University of
Tokushima Graduate School, Japan

Reviewed by:
Youichi Suzuki, National University of
Singapore, Singapore
Kenzo Tokunaga, National Institute of
Infectious Diseases, Japan

*Correspondence:
Yasumasa Iwatani, Department of
Infectious Diseases and Immunology,
Clinical Research Center, National
Hospital Organization Nagoya Medical
Center, 4-1-1 San-no-Maru, Naka-ku,
Nagoya, Japan.
e-mail: iwataniy@nnh.hosp.go.jp

Human APOBEC3 (A3) proteins are cellular cytidine deaminases that potently restrict the replication of retroviruses by hypermutating viral cDNA and/or inhibiting reverse transcription. There are seven members of this family including A3A, B, C, DE, F, G, and H, all encoded in a tandem array on human chromosome 22. A3F and A3G are the most potent inhibitors of HIV-1, but only in the absence of the virus-encoded protein, Vif. HIV-1 utilizes Vif to abrogate A3 functions in the producer cells. More specifically, Vif, serving as a substrate receptor, facilitates ubiquitination of A3 proteins by forming a Cullin5 (Cul5)-based E3 ubiquitin ligase complex, which targets A3 proteins for rapid proteasomal degradation. The specificity of A3 degradation is determined by the ability of Vif to bind to the target. Several lines of evidence have suggested that three distinct regions of A3 proteins are involved in the interaction with Vif. Here, we review the biological functions of A3 family members with special focus on A3G and base our analysis on the available structural information.

Keywords: APOBEC3, Vif, APOBEC3G, HIV, retrovirus, ubiquitin, cytidine deaminase, structure
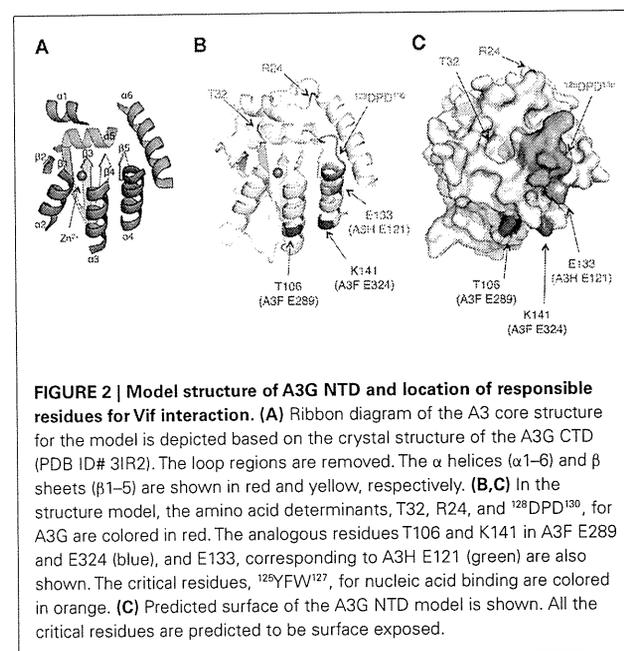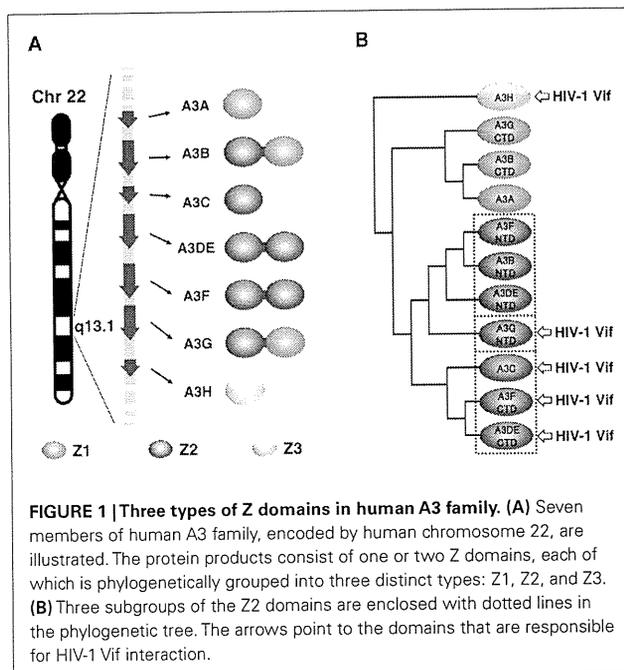
INTRODUCTION

During coevolution of mammalian organisms and retroelements, the hosts have developed defense systems to restrict replication of these elements. The restriction factors include the A3 family of DNA cytidine deaminases, which is characterized by the presence of one or two Zn clusters consisting of $(H/C)XE(X)_{23-28}CXXC$ motifs (reviewed in Wedekind et al., 2003). In humans, there are seven members of the A3 family (A3A, B, C, DE, F, G, and H), each encoded in a tandem array on chromosome 22 (Jarmuz et al., 2002; Figure 1A). Each protein has been found to have different inhibitory effects on various retroelements that are mediated by cytidine deamination and other mechanisms (reviewed in Goila-Gaur and Strebel, 2008; Albin and Harris, 2010). In order to overcome A3 antiviral activity, the viruses have acquired their own strategies (Yu, 2006). HIV-1 encodes the Vif protein to counteract the most potent inhibitors, human A3G (hA3G) and A3F (Sheehy et al., 2002; Wiegand et al., 2004; Zheng et al., 2004; Simon et al., 2005; Miyagi et al., 2010). Vif binds the A3 proteins and serves as a substrate receptor to recruit the Cul5-based E3 ubiquitin (Ub) ligase complex, which facilitates the polyubiquitination and subsequent proteasome-mediated degradation (Yu et al., 2003). The elimination of A3 during virus production prevents its encapsidation into progeny viruses. Thus, given Vif's critical role in eliminating A3 function, it may be viewed as one of the most attractive pharmacologic targets for an anti-HIV drug, which would restore the activity of the intrinsic antiviral factor in the context of HIV infection. Here, we briefly review what is known about A3-Vif interactions and the subsequent

ubiquitination, based on the available biological and structural information.

THE HUMAN A3 FAMILY OF CYTIDINE DEAMINASES

The seven members of the human A3 family have a defining feature: each protein has one or two conserved zinc (Z)-coordinating deaminase domains. Zinc coordination is mediated by a histidine and two cysteines, which form a catalytic center for a cytidine deaminase activity. Based on the phylogenetic analysis, the Z domains fall into three types: Z1 [A3A and the C-terminal half domains (CTD) of A3B and A3G], Z2 [A3C, both halves of A3DE and A3F, and the N-terminal domains (NTD) of A3B and A3G], and Z3 (A3H; LaRue et al., 2009; Figure 1A). Within the Z2 types, the domain can be further subdivided into three subgroups based on the identity of the amino acid sequence: (1) the A3F NTD; (2) the A3G NTD; (3) the A3F CTD (Figure 1B). Each subgroup has highly conserved amino acid sequences. For example, A3F CTD is 77 and 88% identical to A3C and A3DE CTD, respectively, whereas it is 42% identical to A3G NTD. As described in more detail below, only the Z3 type (A3H) and the Z2 type (more specifically, the A3G NTD and the A3F CTD subgroups) contain a critical interface for binding HIV-1 Vif (Figure 1B). It is thought that there is a common structural feature for the organization of cytidine deaminases: all Z domains are believed to have a conserved core structure composed of five β-strands (β1–β5) and six α-helices (α1–α6; Figure 2A). Although, to date, only the three-dimensional structure of the A3G CTD has been determined by NMR (Chen et al., 2008; Furukawa et al., 2009) or by X-ray crystallography

FIGURE 1 | Three types of Z domains in human A3 family. (A) Seven members of human A3 family, encoded by human chromosome 22, are illustrated. The protein products consist of one or two Z domains, each of which is phylogenetically grouped into three distinct types: Z1, Z2, and Z3. **(B)** Three subgroups of the Z2 domains are enclosed with dotted lines in the phylogenetic tree. The arrows point to the domains that are responsible for HIV-1 Vif interaction.

FIGURE 2 | Model structure of A3G NTD and location of responsible residues for Vif interaction. (A) Ribbon diagram of the A3 core structure for the model is depicted based on the crystal structure of the A3G CTD (PDB ID# 3IR2). The loop regions are removed. The α helices (α1–6) and β sheets (β1–5) are shown in red and yellow, respectively. **(B,C)** In the structure model, the amino acid determinants, T32, R24, and ^{128}DPD^{130}, for A3G are colored in red. The analogous residues T106 and K141 in A3F E289 and E324 (blue), and E133, corresponding to A3H E121 (green) are also shown. The critical residues, ^{125}YFW^{127}, for nucleic acid binding are colored in orange. **(C)** Predicted surface of the A3G NTD model is shown. All the critical residues are predicted to be surface exposed.

African green monkeys, determines species specificity by influencing Vif–A3G binding (Bogerd et al., 2004; Mangeat et al., 2004; Schrofelbauer et al., 2004; Xu et al., 2004). In addition, extensive site-direct mutagenesis revealed that the ^{128}DPD^{130} motif of A3G, located at the loop between β4 and α4 shown in red (**Figures 2B,C**), is crucial for direct binding to HIV-1 Vif (Huthoff and Malim, 2007). Furthermore, there is a report that residue T32, which may be potentially phosphorylated by protein kinase A, is also involved in the Vif–A3G interaction by collaborating with R24 (Shirakawa et al., 2008). All of these critical residues are mapped on the variable loop structure in proximity to the nucleic acid binding surface (**Figure 2B**).

In contrast, two independent studies have shown that two C-terminal A3F residues, E289 and E324, located in helices α3 and α4, respectively, are critical for the interaction with HIV-1 Vif (Albin et al., 2010; Smith and Pathak, 2010). Interestingly, in the structural model of the A3F CTD, these two residues are close to some negatively charged surface residues although the location of these residues is separate from ^{128}DPD^{130}. In accord with phylogenetic similarities, A3F E289 and E324 are highly conserved in hA3C (E106 and E141) and hA3DE CTD (E302 and E337). These findings suggest that structural features of the Vif-binding interfaces might be conserved among the A3F CTD, A3C, and A3DE CTD, but different from the A3G NTD. In the case of A3H, the interface for Vif is likely to have a surface area close to that of ^{128}DPD^{130} in the A3G NTD. The A3H gene is polymorphic, with four major haplotypes in humans. The four proteins have different levels of antiviral activity and sensitivity to HIV-1 Vif, in which haplotype II has the highest antiviral activity (Dang et al., 2008a; OhAinle et al., 2008; Harari et al., 2009; Tan et al., 2009; Li et al., 2010). By comparing the A3H variants, Zhen et al. (2010) identified a critical residue, E121, in A3H haplotype II for binding to HIV-1 (**Figures 2B,C**). Because the identity of amino acid sequences between Z2 and Z3 types, particularly the β4–α4 loop and the α4, is quite low, it is assumed that the putative Vif interface structure of A3H might be different from those of the A3G NTD or A3F CTD.

RESIDUES OF HIV-1 VIF THAT ARE CRITICAL FOR BINDING A3 PROTEINS

Extensive mutational analysis of HIV-1 Vif has led to the characterization of several distinct motifs in HIV-1 Vif that are required for formation of the Cul5-based E3 Ub ligase complex and recruitment of human A3 proteins (reviewed in Albin and Harris, 2010; **Figure 3**). The C-terminal half of Vif contains three conserved motifs: (1) the HCCH domain, which chelates zinc mediates the interaction with Cul5; (2) the SLQ motif, which binds elongin C (Elo C); and (3) the PPLP motif, which is important for Vif dimerization and recruitment of A3G, albeit by an unknown mechanism. Meanwhile, the N-terminal half of Vif is involved in the interaction with A3 proteins. **Figure 3** shows the compiled map of the critical residues that have been identified by several groups (Schrofelbauer et al., 2006; Tian et al., 2006; Russell and Pathak, 2007; He et al., 2008; Zhang et al., 2008; Chen et al., 2009; Pery et al., 2009; Zhen et al., 2010; Binka et al., 2011). Overall, it appears that the critical residues for binding to each Z domain type are discontinuous. This suggests that the interfaces in Vif might be determined by conformational constraints.
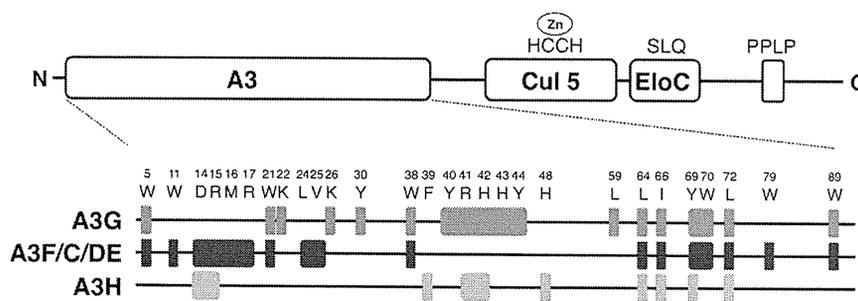
(Holden et al., 2008; Shandilya et al., 2010), no structure of the Z domain having an HIV-1 Vif interface has been solved yet.

THREE DISTINCT INTERFACES OF A3 Z DOMAINS INTERACT WITH HIV-1 VIF

The region in A3G responsible for binding to HIV-1 Vif was initially identified by comparative studies of the species specificity of A3G degradation by Vif. A single amino acid difference in hA3G, aspartic acid at position 128 (D128) versus lysine in the A3G of

FIGURE 3 | Map of critical residues in HIV-1 Vif. Important domains in Vif (HIV-1 HXB2) are illustrated in the schematic diagram. The domains labeled "A3," "Cul5," and "Elo C" are responsible for Vif's functional interactions. The "A3" domain is responsible for binding to A3 family proteins. The amino acid sequence motifs, "HCCH," "SLQ," and "PPLP" are important for binding to Cul5, and Elo C, and Vif dimerization, respectively. The positions and the names of amino acid residues that were identified by several groups are shown. The residues that were identified as important for forming an interface with human A3 proteins are mapped in Vif, i.e., A3G (red), A3F, A3C, and A3DE (blue), and A3H (green).

Interestingly, some residues are unique among the three types of Z domains, suggesting that such residues might determine binding specificity. For example, K22, K26, Y30, Y40, H43, Y44, and L59 are important for interaction only with A3G, whereas the residues W11, M16, R17, L24, V25, and W79 are required for specific interactions with A3F. These residues tend to have two characteristics: they are positively charged or hydrophobic. Interestingly, the electrostatic charge of the Vif residues is mostly positive whereas the key residues that comprise the A3 interfaces are negatively charged. Thus, such electrostatic interactions might be one of the common features for binding between human A3 proteins and HIV-1 Vif. This could also determine the species-specific interactions between Vif and A3 proteins, as demonstrated in the case of Vif–A3G (Bogerd et al., 2004; Mangeat et al., 2004; Schrofelbauer et al., 2004; Xu et al., 2004). In contrast to the aforementioned regions, there are some aromatic or hydrophobic residues, W5, W21, W38, W70, and W89, which are commonly responsible for binding to two (the A3G NTD and the A3F CTD), or L64, I66, Y69, and L72 to all types of Z domains. These regions could play fundamental roles in maintaining the overall conformation of A3-binding interfaces or could form a partly shared structure of the interface for either two or all Z domains.

VIF-MEDIATED UBIQUITINATION/DEGRADATION
Information concerning the specific interaction between Vif and A3G in the E3 Ub ligase complex suggests that there may be a specific mechanism for Ub conjugation of the A3G protein. Previously, our studies using structure-guided mutagenesis demonstrated that four lysine residues (K297, 301, 303, and 334) near the C-terminal end of A3G are critical sites for HIV-1 Vif-mediated A3G ubiquitination and degradation (Iwatani et al., 2009). The data suggested that these sites, which are located on the opposite end of A3G relative to the Vif-binding interface, might be specifically dictated by the rigid structure of the Cul5-based scaffold. In addition, the study clearly demonstrated that the additional residues, particularly lysines near the C-terminal tag of A3G could be potential targets for Vif-induced ubiquitination. However, recently, Wang et al. (2011) showed that a detectable level of Vif-mediated ubiquitination occurs at the N-terminus in addition

to certain lysine residues within 20 lysine residues of A3G. In contrast, another group could detect no N-terminal ubiquitination (Dang et al., 2008b). Although these apparent contradictions might be due to differences in the detection level of ubiquitination, it is important to consider the possibility that different tags attached at the C-terminal end of an A3G protein could create potential ubiquitination sites and/or mask the area where the four lysine residues are located (Iwatani et al., 2009). Moreover, while ubiquitination is associated primarily with the lysines in the CTD, the observation that there is also N-terminal ubiquitination could be of interest. The evidences of two distal ubiquitination sites in A3G may provide important structural insight, which implies two distinct types of structural configuration for Vif–A3G interaction. Further studies are needed to clarify Vif-mediated ubiquitination in the context of the E3 Ub ligase complex and to allow us to answer the following two questions: (1) How can we rationalize the relationship between the configuration of the N-terminal end and the proximal Vif-binding interface? (2) In particular, how can the Ub molecules be conjugated at the N-terminus, which is predicted to be a structurally flexible end?

CONCLUSION
Phylogenetic analyses and genetic studies of A3 and Vif have provided important evidence for three distinct types of interactions between human A3 and HIV-1 Vif proteins, which are determined by the characteristic Z domain types of A3G, A3F/C/DE, and A3H. Further understanding of Vif–A3 interactions could advance efforts to develop novel anti-HIV drugs, which would function as anti-Vif inhibitors. Although presenting a tremendous challenge, complementary studies focusing on the structure of the Vif-interactive A3 domain and Vif are also critical to accelerate future progress in this field.

REFERENCES

Albin, J. S., and Harris, R. S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev. Mol. Med.* 12, e4.

Albin, J. S., LaRue, R. S., Weaver, J. A., Brown, W. L., Shindo, K., Harjes, E., Matsuo, H., and Harris, R. S. (2010). A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. *J. Biol. Chem.* 285, 40785–40792.

Binka, M., Ooms, M., Steward, M., and Simon, V. (2011). The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F and APOBEC3H. *J. Virol.* (in press).

Bogerd, H. P., Doehle, B. P., Wiegand, H. L., and Cullen, B. R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3770–3774.

Chen, G., He, Z., Wang, T., Xu, R., and Yu, X. F. (2009). A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. *J. Virol.* 83, 8674–8682.

Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* 452, 116–119.

Dang, Y., Siew, L. M., Wang, X., Han, Y., Lampen, R., and Zheng, Y. H. (2008a). Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *J. Biol. Chem.* 283, 11606–11614.

Dang, Y., Siew, L. M., and Zheng, Y. H. (2008b). APOBEC3G is degraded by the proteasomal pathway in a Vif-dependent manner without being polyubiquitylated. *J. Biol. Chem.* 283, 13124–13131.

Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H., and Katahira, M. (2009). Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.* 28, 440–451.

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

Harari, A., Ooms, M., Mulder, L. C., and Simon, V. (2009). Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. *J. Virol.* 83, 295–303.

He, Z., Zhang, W., Chen, G., Xu, R., and Yu, X. F. (2008). Characterization of conserved motifs in HIV-1 Vif required for APOBEC3G and APOBEC3F interaction. *J. Mol. Biol.* 381, 1000–1011.

Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008). Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* 456, 121–124.

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

Iwatani, Y., Chan, D. S., Liu, L., Yoshii, H., Shibata, J., Yamamoto, N., Levin, J. G., Gronenborn, A. M., and Sugiura, W. (2009). HIV-1 Vif-mediated ubiquitination/degradation of APOBEC3G involves four critical lysine residues in its C-terminal domain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19539–19544.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 79, 285–296.

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. (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.* 83, 494–497.

Li, M. M., Wu, L. I., and Emerman, M. (2010). The range of human APOBEC3H sensitivity to lentiviral Vif proteins. *J. Virol.* 84, 88–95.

Mangeat, B., Turelli, P., Liao, S., and Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. *J. Biol. Chem.* 279, 14481–14483.

Miyagi, E., Brown, C. R., Opi, S., Khan, M., Goila-Gaur, R., Kao, S., Walker, R. C. Jr., Hirsch, V., and Strebel, K. (2010). Stably expressed APOBEC3F has negligible antiviral activity. *J. Virol.* 84, 11067–11075.

OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S., and Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* 4, 249–259.

Pery, E., Rajendran, K. S., Brazier, A. J., and Gabuzda, D. (2009). Regulation of APOBEC3 proteins by a novel YXXL motif in human immunodeficiency virus type 1 Vif and simian immunodeficiency virus SIVagm Vif. *J. Virol.* 83, 2374–2381.

Russell, R. A., and Pathak, V. K. (2007). Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. *J. Virol.* 81, 8201–8210.

Schrofelbauer, B., Chen, D., and Landau, N. R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc. Natl. Acad. Sci. U.S.A.* 101, 3927–3932.

Schrofelbauer, B., Senger, T., Manning, G., and Landau, N. R. (2006). Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with non-human primate APOBEC3G. *J. Virol.* 80, 5984–5991.

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. (2010). Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* 18, 28–38.

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

Shirakawa, K., Takaori-Kondo, A., Yokoyama, M., Izumi, T., Matsui, M., Io, K., Sato, T., Sato, H., and Uchiyama, T. (2008). Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat. Struct. Mol. Biol.* 15, 1184–1191.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D. D., and Bieniasz, P. D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog.* 1, e6. doi:10.1371/journal.ppat.0010006

Smith, J. L., and Pathak, V. K. (2010). Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. *J. Virol.* 84, 12599–12608.

Tan, L., Sarkis, P. T., Wang, T., Tian, C., and Yu, X. F. (2009). Sole copy of

Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. *FASEB J.* 23, 279–287.

Tian, C., Yu, X., Zhang, W., Wang, T., Xu, R., and Yu, X. F. (2006). Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F. *J. Virol.* 80, 3112–3115.

Wang, Y., Shao, Q., Yu, X., Kong, W., Hildreth, J. E., and Liu, B. (2011). N-terminal hemagglutinin tag renders lysine-deficient APOBEC3G resistant to HIV-1 Vif-induced degradation by reduced polyubiquitination. *J. Virol.* 85, 4510–4519.

Wedekind, J. E., Dance, G. S., Sowden, M. P., and Smith, H. C. (2003). Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.* 19, 207–216.

Wiegand, H. L., Doehle, B. P., Bogerd, H. P., and Cullen, B. R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J.* 23, 2451–2458.

Xu, H., Svarovskaia, E. S., Barr, R., Zhang, Y., Khan, M. A., Strebel, K., and Pathak, V. K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5652–5657.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X. F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 302, 1056–1060.

Yu, X. F. (2006). Innate cellular defenses of APOBEC3 cytidine deaminases and viral counter-defenses. *Curr. Opin. HIV AIDS* 1, 187–193.

Zhang, W., Chen, G., Niewiadomska, A. M., Xu, R., and Yu, X. F. (2008). Distinct determinants in HIV-1 Vif and human APOBEC3 proteins are required for the suppression of diverse host anti-viral proteins. *PLoS ONE* 3, e3963. doi:10.1371/journal.pone.0003963

Zhen, A., Wang, T., Zhao, K., Xiong, Y., and Yu, X. F. (2010). A single amino acid difference in human APOBEC3H variants determines HIV-1 Vif sensitivity. *J. Virol.* 84, 1902–1911.

Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B. M. (2004). Human APOBEC3F is another host

factor that blocks human immunod-eficiency virus type 1 replication. *J. Virol.* 78, 6073–6076.

Two Genetic Determinants Acquired Late in *Mus* Evolution Regulate the Inclusion of Exon 5, which Alters Mouse APOBEC3 Translation Efficiency

Jun Li[1], Yoshiyuki Hakata[1]*, Eri Takeda[1¤], Qingping Liu[2], Yasumasa Iwatani[3], Christine A. Kozak[2], Masaaki Miyazawa[1]*

1 Department of Immunology, Kinki University School of Medicine, Osaka, Japan, 2 Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, 3 Department of Infection and Immunology, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan

Abstract

Mouse apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like editing complex 3 (mA3), an intracellular antiviral factor, has 2 allelic variations that are linked with different susceptibilities to beta- and gammaretrovirus infections among various mouse strains. In virus-resistant C57BL/6 (B6) mice, mA3 transcripts are more abundant than those in susceptible BALB/c mice both in the spleen and bone marrow. These strains of mice also express mA3 transcripts with different splicing patterns: B6 mice preferentially express exon 5-deficient (Δ5) mA3 mRNA, while BALB/c mice produce exon 5-containing full-length mA3 mRNA as the major transcript. Although the protein product of the Δ5 mRNA exerts stronger antiretroviral activities than the full-length protein, how exon 5 affects mA3 antiviral activity, as well as the genetic mechanisms regulating exon 5 inclusion into the mA3 transcripts, remains largely uncharacterized. Here we show that mA3 exon 5 is indeed a functional element that influences protein synthesis at a post-transcriptional level. We further employed *in vitro* splicing assays using genomic DNA clones to identify two critical polymorphisms affecting the inclusion of exon 5 into mA3 transcripts: the number of TCCT repeats upstream of exon 5 and the single nucleotide polymorphism within exon 5 located 12 bases upstream of the exon 5/intron 5 boundary. Distribution of the above polymorphisms among different *Mus* species indicates that the inclusion of exon 5 into mA3 mRNA is a relatively recent event in the evolution of mice. The widespread geographic distribution of this exon 5-including genetic variant suggests that in some *Mus* populations the cost of maintaining an effective but mutagenic enzyme may outweigh its antiviral function.

Introduction

The family of apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like editing complex 3 (APOBEC3) proteins consists of cytidine deaminases that function as cellular restriction factors against various exogenous and endogenous viruses [1–17]. Seven APOBEC3 paralogues have been identified on human chromosome 22, while only a single copy of the *Apobec3* gene is found in the mouse genome [10,18,19]. Among the human APOBEC3 enzymes, APOBEC3G (hA3G) is the best characterized member and is known to inhibit HIV-1 replication when the virus lacks the functional accessory protein, viral infectivity factor (Vif) [reviewed in 20]. In the absence of Vif, hA3G is incorporated into newly generated virions budding from virus-producing cells and exhibits its antiviral effect in subsequently infected cells. Thus, during reverse transcription in the target cells, the virion-incorporated hA3G catalyzes C-to-U deamination on the minus strand of

nascent viral DNA, resulting in G-to-A mutations on the plus strand of the double-stranded viral DNA, which can be detrimental to viral replication [7,9,10,21,22]. In addition, a deaminase-independent antiviral mechanism exerted by hA3G has also been reported [23,24].

In contrast to its human counterparts, mouse APOBEC3 (mA3) restricts HIV-1 regardless of the presence of Vif, as well as mouse mammary tumor virus (MMTV), ecotropic murine leukemia viruses (MuLVs), Friend MuLV (F-MuLV) and Moloney MuLV (M-MuLV), along with endogenous mouse retroviruses including the AKR ecotropic virus (AKV) [5,25–30]. This suggests that APOBEC3 enzymes protect host genomes from the retroviruses they commonly encounter, although some retroviruses, like HIV-1, have evolved to counter the intracellular restriction mechanisms of their natural hosts.

Friend virus (FV) is an acutely leukemogenic retroviral complex composed of replication-competent F-MuLV and replication

Author Summary

Susceptibility to acutely leukemogenic Friend virus (FV) retrovirus infection varies among different mouse strains and is governed by several genetic factors, one of which is allelic variations at the mouse *Apobec3* locus. FV-resistant C57BL/6 (B6) mice express higher amounts of *Apobec3* transcripts than susceptible BALB/c mice. We previously showed that the differences in N-terminal amino acid sequences between B6 and BALB/c APOBEC3 proteins partly account for the distinct antiretroviral activities. In addition, B6 and BALB/c mice express major *Apobec3* transcripts of different sizes: the exon 5-lacking and the full-length transcripts, respectively. Here we asked if exon 5 has any role in the antiviral activity of mouse APOBEC3 and found that the presence of this exon resulted in a profound decrease in the efficiency of protein synthesis without affecting the mRNA expression levels. We also identified two genomic polymorphisms that control the inclusion of exon 5 into the *Apobec3* message: the number of TCCT repeats in intron 4 and a single nucleotide polymorphism within exon 5. The distribution of these functional polymorphisms among *Mus* species and wild mouse populations indicates that the exon 5 inclusion occurred recently in *Mus* evolution, and the full-length variant may have selective advantages in some mouse populations.

-defective spleen focus-forming virus (SFFV). Susceptibilities to FV-induced disease development differ among various inbred strains of mice, and these are controlled by several host factors that either directly affect FV replication or influence host immune responses to the viral antigens [31,32]. We and others have reported that the mouse *Apobec3* locus is polymorphic, and its genotypes are associated with the levels of viremia after F-MuLV or FV inoculation [26,27]. Mice of the prototypic FV-resistant strains C57BL/6 (B6) and C57BL/10 exhibit restricted replication of F-MuLV and earlier production of FV-neutralizing antibodies, while FV-susceptible BALB/c and A strains are less restrictive of F-MuLV replication and show delayed production of neutralizing antibodies, all of which are linked with *Apobec3* genotypes [26,27,33,34], although the production of neutralizing antibodies is also influenced by genotypes at the major histocompatibility complex and the *Tnfrsf13c* loci, the latter of which encode the receptor for B-cell activating factor belonging to the tumor necrosis factor family [31–33,35].

Mouse APOBEC3 and hA3G contain two cytidine deaminase domains (CDDs), each harboring the conserved zinc-coordinating motif; however, deaminase activity is exerted only by the N-terminal CDD of mA3 and the C-terminal CDD of hA3G [36]. We showed that the increased efficiency of B6 mA3 in inhibiting F-MuLV replication is associated with differences in the primary amino acid sequence within the active N-terminal CDD [27]; the functional importance of these residues was further implicated by the demonstration that they have been under positive selection in *Mus* [37]. In addition to the above sequence differences in the protein-coding regions, efficient virus restriction is also associated with higher levels of mA3 transcripts in FV-resistant B6 than in -susceptible BALB/c mice [27,29,38,39]. This enhanced transcription was linked with the presence of the long terminal repeat (LTR) of an endogenized xenotropic MuLV in the B6, but not in the BALB/c, *Apobec3* locus [37]. A third factor associated with virus resistance is the presence or absence of exon 5, which encodes a 33-amino acid segment separating the C-terminal and N-terminal CDDs [27]. The mA3 isoform of the FV-resistant

strains of mice lacks exon 5, while the predominant transcript in FV-susceptible mice is the exon 5-containing isoform [27,29,37]. Genetic sequences controlling mA3 splicing have not been identified, although it has been pointed out that polymorphisms between the B6 and BALB/c *Apobec3* alleles at the end of intron 4 include a putative splice acceptor site and possible mRNA branch selection site structures [38]. It has also not been shown whether the well-confirmed differences in transcript levels result in altered expression levels of mA3 protein in FV-resistant and -susceptible strains of mice.

In the present report, we show that mA3 protein is indeed more abundant in B6 than in BALB/c mice. This difference is due in part to more efficient translation of the exon 5-deficient message. We further show extensive functional evidence that two distinct polymorphisms within the *Apobec3* locus regulate exon 5 inclusion during its splicing: the previously predicted [38] TCCT repeat numbers in intron 4 and a newly identified single nucleotide polymorphism (SNP) within exon 5. We also describe the linkage between these splicing regulatory sequences in wild mouse species, their acquisition in *Mus* evolution, and their distribution in wild mouse populations.

Results

APOBEC3 expression is higher in B6 than in BALB/c spleens at both transcriptional and protein levels

It has been reported that mA3 mRNA expression is higher in B6 than in BALB/c mice, and the mA3 transcripts detected in B6 mice are predominantly the exon 5-lacking Δ5 isoform, while the majority of mA3 transcripts in BALB/c mice contain exon 5 (5+) [27,29,38]. To more accurately describe the above quantitative differences in mA3 transcripts, we performed two types of PCR analyses with two different sets of specific primers (Figure 1A). The primer set c–d, which is the same as the one previously used [27], detected both the 5+ and Δ5 transcripts in reverse-transcription PCR (RT-PCR) assays; however, mA3 transcripts were barely detectable after 30 cycles of amplification in BALB/c mice while the Δ5 transcript was readily detectable in B6 mice (Figure 1B). After 35 cycles of amplification, both the 5+ and Δ5 transcripts became detectable in BALB/c mice, although the 5+ mRNA was more abundant. Quantitative real-time PCR (qPCR) assays revealed mA3 expression levels that were approximately 18-times higher in B6 than in BALB/c spleens, although the difference in transcript lengths might have influenced the efficiencies of amplification in the real-time PCR reactions. To more precisely quantify the mA3 transcripts, we utilized the second primer set, a–b, which generates amplicons of the same size from both alleles (Figure 1A). The qPCR assay performed by using this latter primer set clearly demonstrated that B6 mice expressed >7-times higher amounts of mA3 transcript than BALB/c mice did (Figure 1C), consistent with the previous report [27]. The normalization of mA3 mRNA levels with TATA box-binding protein or GAPDH transcripts instead of β-actin gave similar results (data not shown).

Protein levels of mA3 in the above prototypic FV-resistant and -susceptible strains of mice were also compared. The spleen lysate from the mA3-knockout mice [40] was used as a negative control. In B6 spleens, an immunoreactive protein corresponding to Δ5 mA3 was detected as a prominent band, but the higher molecular weight 5+ mA3 was also faintly detectable (Figure 1D), even though the 5+ message was barely detectable by RT-PCR in the present study (Figure 1B) and in the previous reports [27,29]. On the other hand, only the 5+ mA3 protein was detected, with a much lower intensity, in BALB/c spleens. The immunoblotting assays thus demonstrated that the level of total mA3 protein
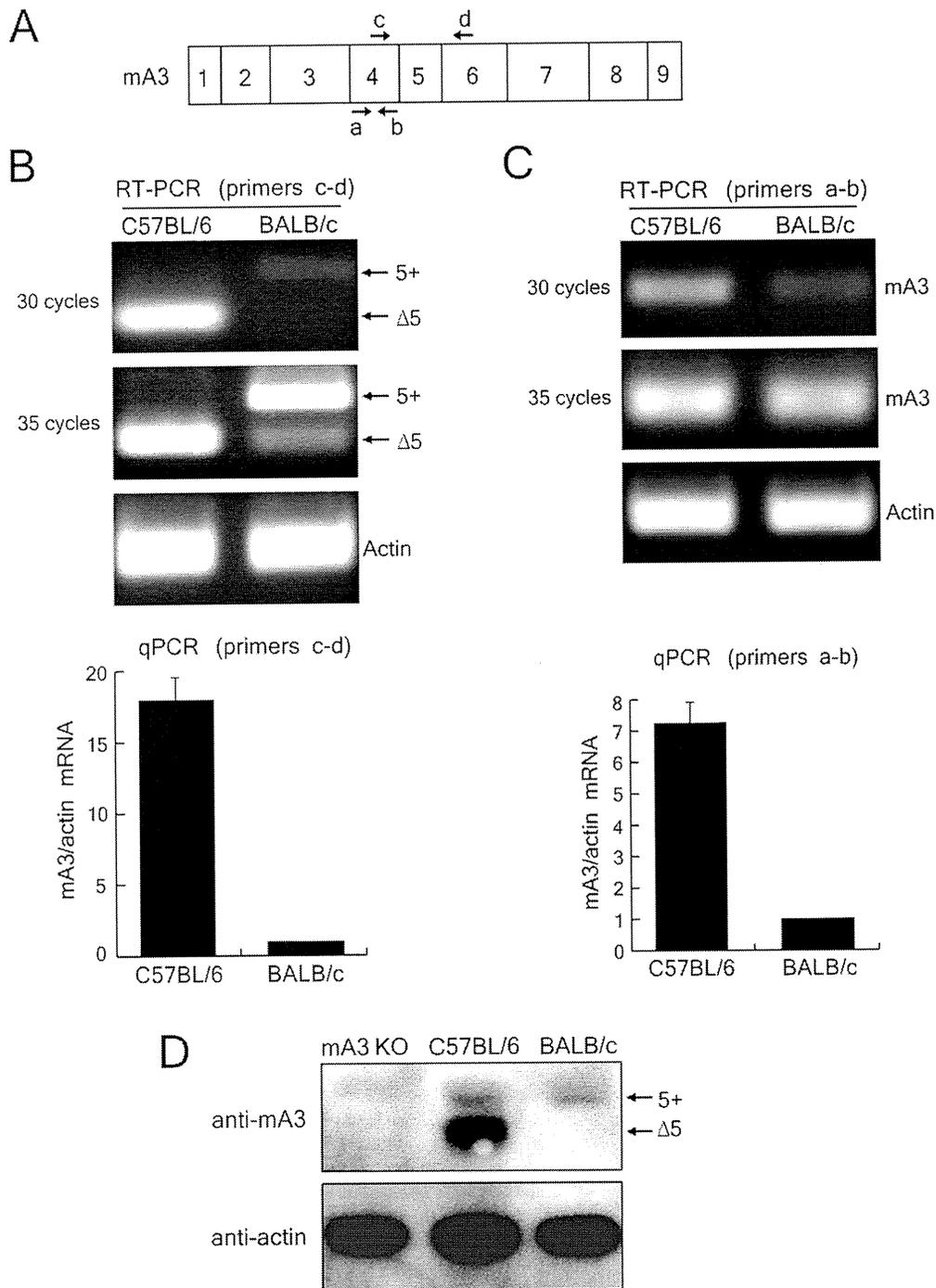
Figure 1. Expression of mA3 transcripts and protein in the spleens of B6 and BALB/c mice. (A) Schematic representation of the mouse APOBEC3 exon composition. The exons are indicated by numbers and their relative sizes are depicted. The arrows indicate the locations of the PCR primers used in the following experiments. (B and C) RT-RCR and quantitative real-time PCR (qPCR) analyses showing splicing patterns and quantities of mA3 mRNA in B6 and BALB/c spleen cells. The primers utilized were c-d for assays shown in (B) and a–b for those in (C). 5+, exon 5-containing mA3; Δ5, mA3 lacking exon 5. Actin was used as an internal control. Quantitative real-time PCR data show averages of three reaction wells and SD. (D) Immunoblot detection of mA3 protein expressed in mouse spleen cells. The spleen cell extracts were tested for their total protein content by Bradford assays and the same amount of protein was loaded in each lane. Anti-mouse APOBEC3 NT antibody reactive with an N-terminal epitope was first pre-absorbed with the spleen lysate prepared from an mA3-knockout (KO) mouse, and was used as the primary antibody. Mouse actin was used as a loading control. Three independent experiments were done for all assays, and the results were consistent with the representative ones shown here.
doi:10.1371/journal.ppat.1002478.g001

expression in the spleens of B6 mice is much higher than that in BALB/c mice.

Effects of exon 5 inclusion on mA3 protein expression

The mA3 expression data shown in Figure 1 indicate that levels of mA3 protein expression roughly correlate with the corresponding transcript levels in mice with different *Apobec3* alleles. However, we previously observed that even when FLAG-tagged mA3 was expressed under the control of the cytomegalovirus (CMV) promoter, the levels of expression of the 5+ mA3 protein tended to be much lower than those of the Δ5 protein, despite similar levels of mRNA expression in transfected cells [27]. This implies that exon 5 might affect either the efficiency of mA3 translation or protein stability. To evaluate these possibilities, 5+ and Δ5 mA3 expression plasmids were constructed with either B6- or BALB/c-derived mA3 cDNA (Figure 2A). The mA3 cDNA were fused with a FLAG tag at their N-terminus and the transcription was driven by the CMV promoter. After transfection of 293T cells with either one of the FLAG-tagged expression constructs, we quantified mA3 transcripts by qPCR and mA3 protein by immunoblotting (Figure 2, B–D). Based on the co-expressed luciferase activities, similar transfection efficiencies for the 5+ and Δ5 plasmids were confirmed, and the data from the qPCR showed that the 5+ mA3 transcript was more highly expressed than the Δ5 mRNA for both alleles. Nevertheless, the expression levels of the 5+ mA3 protein were lower compared to those of the Δ5 counterpart regardless of the allelic differences, although B6 Δ5 protein was expressed more abundantly than BALB/c Δ5 (Figure 2, B and C). These results indicate that the inclusion of exon 5 does not compromise mA3 mRNA expression, but decreases the steady-state levels of mA3 proteins for both allelic variants.

Exon 5 inclusion does not affect mA3 protein stability but influences its translation

As higher mRNA levels of the 5+ mA3 resulted in lower protein levels in comparison with those for the Δ5 isoform in transfected cells, we next examined the steps at which the synthesis or degradation of mA3 protein was affected by the presence of exon 5. To this end, we transfected 293T cells with an expression plasmid harboring the 5+ or Δ5 mA3 cDNA, and cycloheximide was added to stop protein synthesis. The inhibitor of nuclear factor-κB, IκBα, was utilized as a control, as this protein is known to be in the process of constant degradation and regeneration under serum-containing culture conditions [41]. As expected, immunodetectable IκBα decreased upon cycloheximide treatment, while the solvent alone did not affect the protein content (Figure 3A). On the other hand, mA3 expressed in the same cells did not exhibit any reduction upon cycloheximide treatment regardless of the presence or absence of exon 5. Further, the Δ5 protein was again detected at higher levels than the 5+ protein. These results support the conclusion that the effect of exon 5 inclusion in reducing the amount of expressed mA3 protein is not due to accelerated protein degradation.

We next utilized an *in vitro* transcription and translation procedure to determine if exon 5 affects the translation of mA3 (Figure 3B). When the 5+ or Δ5 mA3 templates were added to the *in vitro* transcription and translation reaction, similar amounts of each transcript were detected by RT-PCR assays both at 30 min and 60 min after the beginning of incubation. However, the tempos of appearance and amounts of the 5+ mA3 protein were different from those of the Δ5 counterpart: a large amount of the Δ5 protein was detected at as early as 30 min after the beginning of incubation, while the 5+ mA3 was undetectable at the same time-point. The 5+ mA3 protein became detectable after 60 min

of incubation, but its protein level was still markedly lower than that of the Δ5 counterpart. These results collectively indicate that the inclusion of exon 5 modulates the translation efficiency of mA3 rather than its protein degradation.

The TCCT repeat number variation upstream of exon 5 is partially responsible for exon 5 inclusion into the mA3 transcript

As the inclusion of exon 5 is associated with a reduced level of mA3 protein, we next attempted to identify genetic polymorphisms that affect the splicing patterns of mA3 transcripts in terms of exon 5 inclusion. One possible allelic difference in the mouse *Apobec3* locus putatively associated with its splicing patterns is the possible pre-mRNA branch site polymorphisms found in the intron upstream of exon 5: a T/C SNP that lies within a preferred branch site sequence, TA(T/C)CAAC, and TCCT repeat numbers between this and the acceptor site [38]. As described previously [38], the intron 4 of the BALB/c allele contains a tandem repeat of the TCCT sequence near the intron 4/exon 5 boundary, while the B6 allele contains only a single TCCT copy, changing the length of the putative pyrimidine-rich lariat intermediate. The adjacent T/C SNP at position 741 from the first nucleotide of exon 4 is in linkage with the TCCT repeat number. To directly investigate the possible effect of these polymorphisms, we constructed splicing assay plasmids that harbor the *Apobec3* genomic fragment encompassing exons 4 through 7 from either the B6 or BALB/c allele. The resultant plasmids were designated B6 exon 4–7 and BALB exon 4–7 (Figure 4). Three mutants of each of these constructs were made with different combinations of the position 741 T/C SNP and the TCCT copy number variation as depicted in Figure 4B.

Because of the possible presence of species-specific regulatory factors, the resultant genomic constructs were transfected into BALB/3T3 instead of 293T cells along with the luciferase expression plasmid as a control for transfection efficiency. To avoid the amplification of the endogenous mA3 message expressed in BALB/3T3 cells [27], we utilized the primers g and h (Figure 4B) for RT-PCR assays, which were designed to hybridize to the T7 promoter and V5 tag regions of the expression vector. Transfection with the B6 exon 4–7 plasmid resulted in the expression of only the Δ5 transcript, while the BALB exon 4–7 generated both the 5+ and Δ5 transcripts with much higher intensity of the 5+ one (Figure 4C), reproducing the splicing patterns observed in B6 and BALB/c spleens, respectively (Figure 1B). The ratio between the 5+ and Δ5 transcripts was reduced in the samples transfected with BALB ΔTCCT or BALB C741T ΔTCCT plasmid harboring only a single copy of TCCT, which was further confirmed by utilizing the primer i hybridizing to the sequence within exon 5 along with primer h (Figure 4C). Quantitative real-time PCR analyses were done by utilizing another primer (primer j), designed to hybridize with the sequence within exon 5, and further confirmed reduced levels of the exon 5-containing message expressed from the modified BALB exon 4–7 constructs lacking the TCCT repeat. These results imply that the TCCT repeat observed in the BALB/c allele, but not the T/C substitution at position 741, at least partly facilitates the exon 5 inclusion. The lack of an effect of the T/C SNP within the putative branch site sequence was further confirmed by the abundant expression of the 5+ message from the BALB C741T construct. However, both the 5+ and Δ5 transcripts were still produced from the BALB ΔTCCT plasmid, suggesting that the TCCT repeat number is not the only determinant controlling the exon 5 inclusion. In fact, despite the presence of the repeated TCCT, B6 +TCCT and B6 T741C +TCCT did not generate the 5+ message,
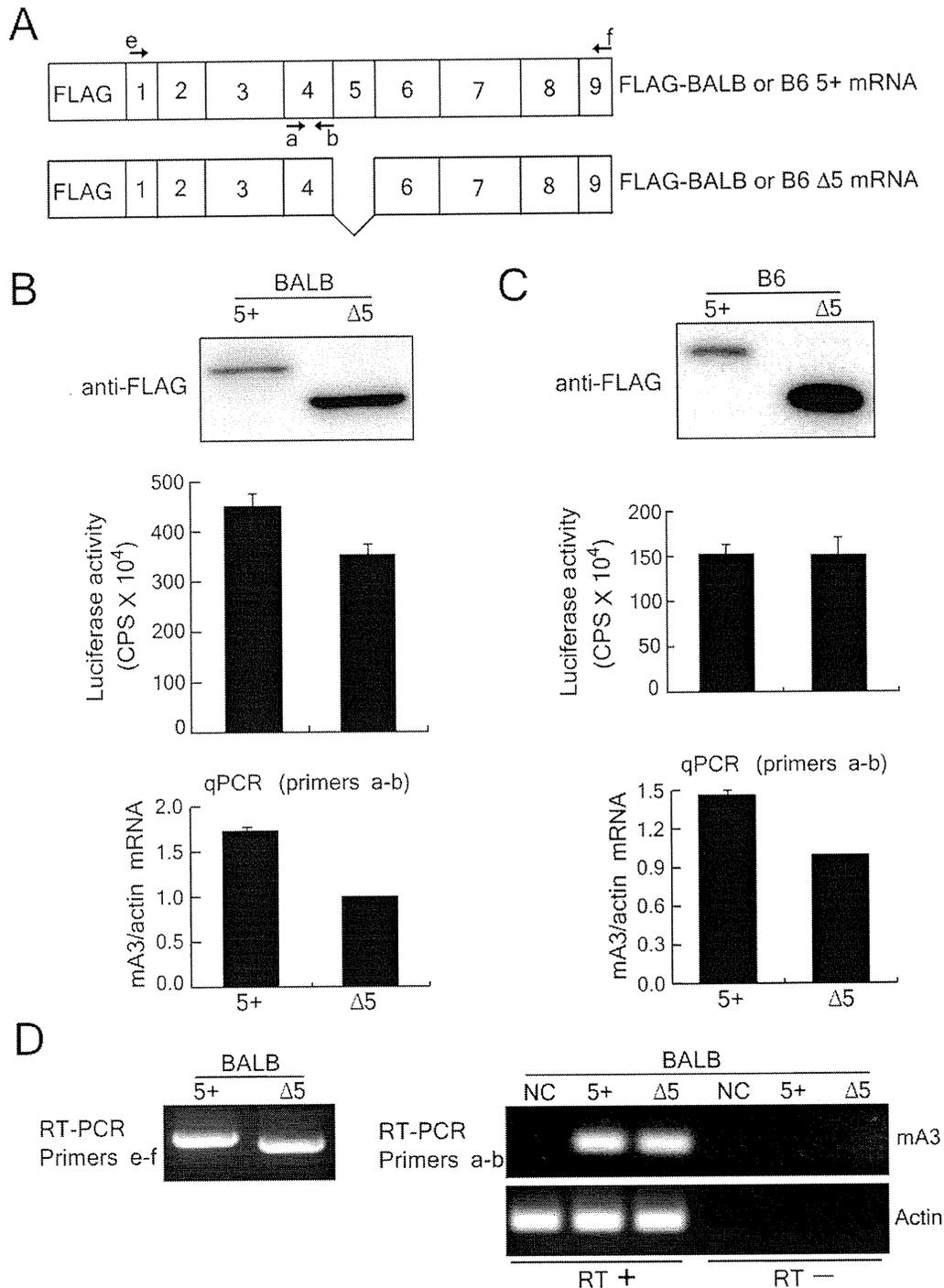
Figure 2. Protein expression of the full-length and Δ5 mA3 in transiently transfected 293T cells. (A) Exon 5-containing (5+) and Δ5 mA3 cDNA were tagged with the FLAG epitope and inserted into the expression vector. The arrows indicate the positions of the PCR primers. The primer set a-b is the same as shown in Figure 1. (B and C) 293T cells were transfected with pFLAG-CMV2-$mA3^d$ or pFLAG-CMV2-$mA3^d$Δ5 in B or with pFLAG-CMV2-$mA3^b$ or pFLAG-CMV2-$mA3^b$Δ5 in C, which express either the 5+ or Δ5 mA3 cDNA cloned from BALB/c or B6 mice, respectively [27]. A luciferase-expressing plasmid, pluc, was co-transfected to standardize transfection efficiencies. At 24 hours after transfection, each one-third of the transfected cells was used for immunoblotting, RNA extraction, and luciferase assays. For immunoblotting, the full-length and Δ5 mA3 proteins were detected with the anti-FLAG antibody. Quantitative real-time PCR reactions were carried out with primer set a-b and were normalized with the levels of actin transcripts expressed in 293T cells. Data shown are averages of three reaction wells and SD. (D) RT-PCR assays were performed with (RT+) or without (RT−) reverse transcription to demonstrate specific detection of expected mRNA. Both primer sets e-f and a-b were used to detect mA3 mRNA. NC, negative control transfected with the empty vector, pFLAG-CMV2. RT− samples were included to evaluate the possible contamination of transfected DNA. Specific amplification of cDNA generated in the presence of RT was observed in the cells transfected with a plasmid expressing the 5+ or Δ5 mA3 cDNA. Similar results were obtained for the corresponding B6-derived cDNA clones.
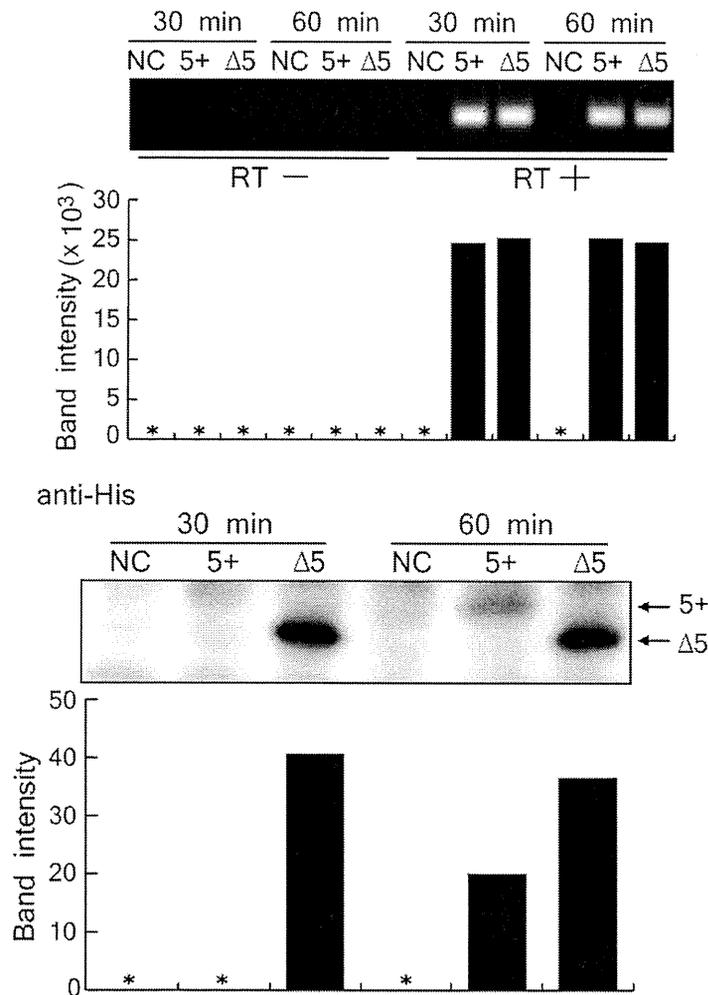doi:10.1371/journal.ppat.1002478.g002
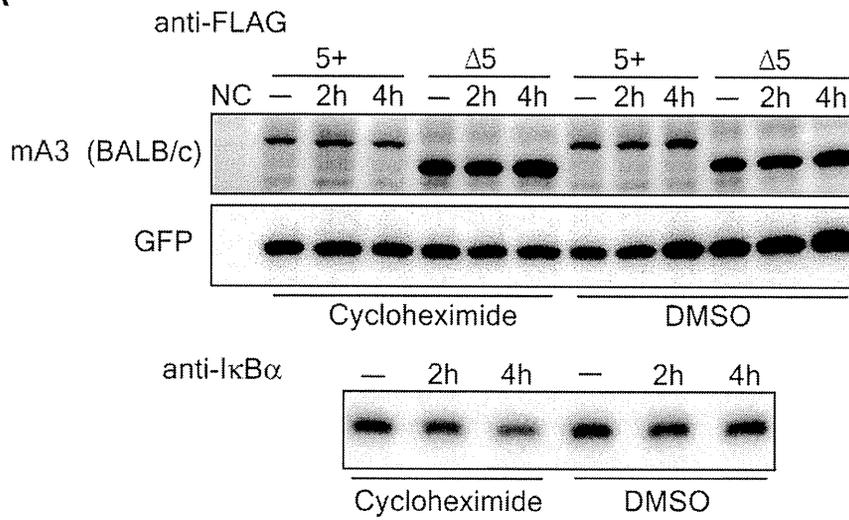
Figure 3. The presence of exon 5 does not affect protein degradation but exhibits a profound impact on mA3 protein synthesis.
(A) 293T cells were transfected with pFLAG-CMV2-$mA3^d$ or pFLAG-CMV2-$mA3^d\Delta5$ along with pFLAG-CMV2-GFP, which expresses green fluorescent protein (GFP) as a loading control. The cells were treated with cycloheximide for the indicated duration to stop protein synthesis and then harvested. DMSO was used as a solvent control. The FLAG-tagged mA3 and GFP were detected with the anti-FLAG antibody in immunoblotting. Endogenous IκBα expression was used as a positive control to confirm the effect of cycloheximide. (B) *In vitro* transcription and translation assays. DNA templates containing the entire coding region of the BALB/c full-length or Δ5 mA3 cDNA with the T7 promoter and His-tag sequence were subjected to an *in vitro* transcription/translation reaction by incubating for 30 or 60 min. The levels of mA3 protein synthesis and mRNA expression were evaluated by immunoblotting using the anti-His antibody and RT-PCR using the primer set a-b, respectively. Intensities of protein bands on the immunoblot membrane and DNA bands after the RT-PCR reaction and electrophoresis were measured by densitometry and are shown below each corresponding band. *, signals below detection limits.
doi:10.1371/journal.ppat.1002478.g003

while transfection with the corresponding BALB C741T and the wild-type BALB exon 4–7 clearly resulted in the generation of the 5+ mRNA. These results indicate that polymorphisms other than those in intron 4 might play more important roles in determining exon 5 inclusion into mA3 messages.

A region around the exon 5/intron 5 boundary contains a major determinant for exon 5 inclusion

As abundant expression of the exon 5-containing message was found with the BALB C741T but not with the B6 +TCCT construct, we were prompted to examine other polymorphisms

downstream of intron 4 that might be involved in the splicing of exon 5. To explore other possible sequence variations required for exon 5 inclusion, we focused on intron 5, since this region contains numerous polymorphisms between reported genomic sequences of different mouse strains (Figure 5A). Splicing assay vectors harboring the genomic DNA fragment containing exons 5 and 6 along with the entire intron 5, from either the B6 or the BALB/c allele, were constructed and designated B6 exon 5–6 or BALB exon 5–6 (Figure 5B). BALB/3T3 cells transfected with the BALB exon 5–6 plasmid generated the properly spliced product of the expected size (Figure 5C), indicating that the splice donor site
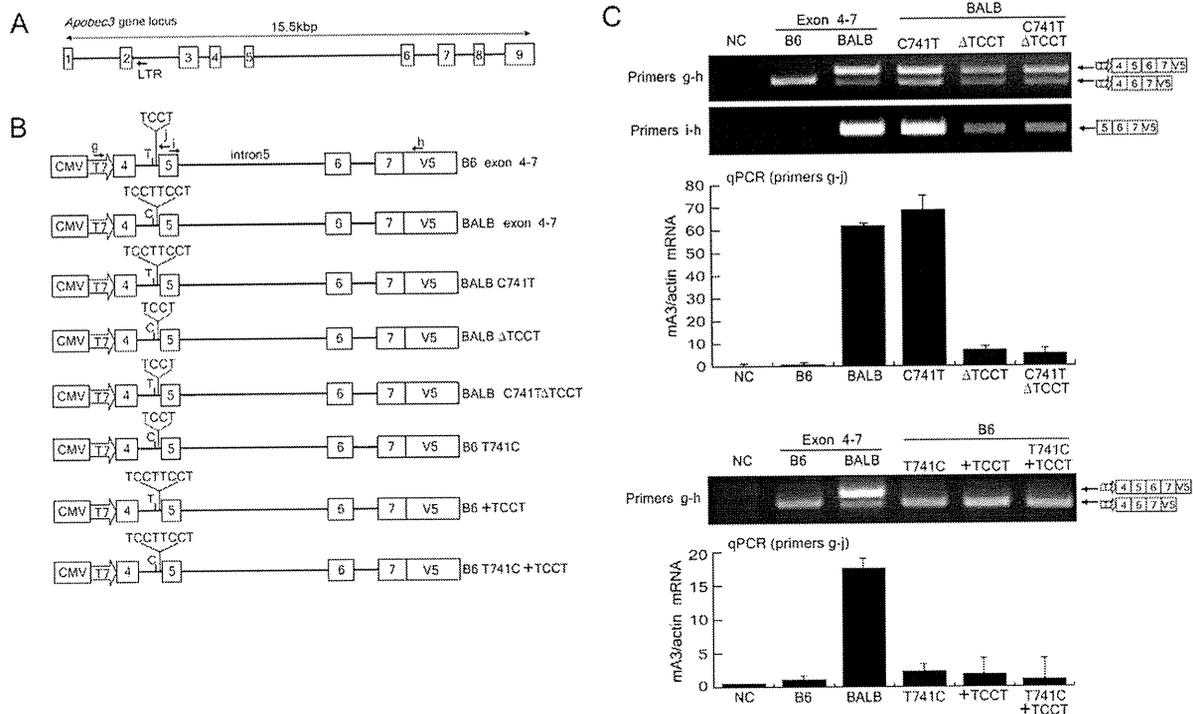

Figure 4. The impact of TCCT repeat numbers on exon 5 inclusion in mA3 mRNA splicing. (A) Schematic representation of the genomic DNA structure of the *Apobec3* gene locus. Boxes indicate exons. LTR, the endogenous retroviral LTR inserted into intron 2 in some strains of mice [37]. (B) Insert structures of the plasmids used for the splicing assays. BALB/c exon 4–7 and BALB/c ΔTCCT plasmids harbored the same DNA fragment amplified from BALB/c genomic DNA encompassing exons 4 and 7, except that the BALB/c ΔTCCT contains only a single TCCT quadruplet. B6 exon 4–7 plasmid harbored the B6 genomic DNA fragment encompassing exons 4 and 7. C741T and T741C indicate a C to T or reciprocal nucleotide substitution, respectively, within intron 4 at 741-bp downstream from the first nucleotide of exon 4 on the backbone of the BALB/c or B6 exon 4–7 insert. An additional repeat of the TCCT quadruplet was added to the B6 exon 4–6 construct to generate B6 +TCCT. The positions of primers g, h, i, and j used for RT-PCR assays are indicated with the arrows. (C) The plasmid harboring each insert depicted in (B) was transfected into BALB/3T3 cells, total RNA was extracted, and RT-PCR reactions were performed with primer pairs g-h and i-h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). The predicted splicing products are schematically indicated on the right side of the panel. Quantitative real-time PCR assays were performed with primers g and j, and data are shown here by averages of three reaction wells and SD.
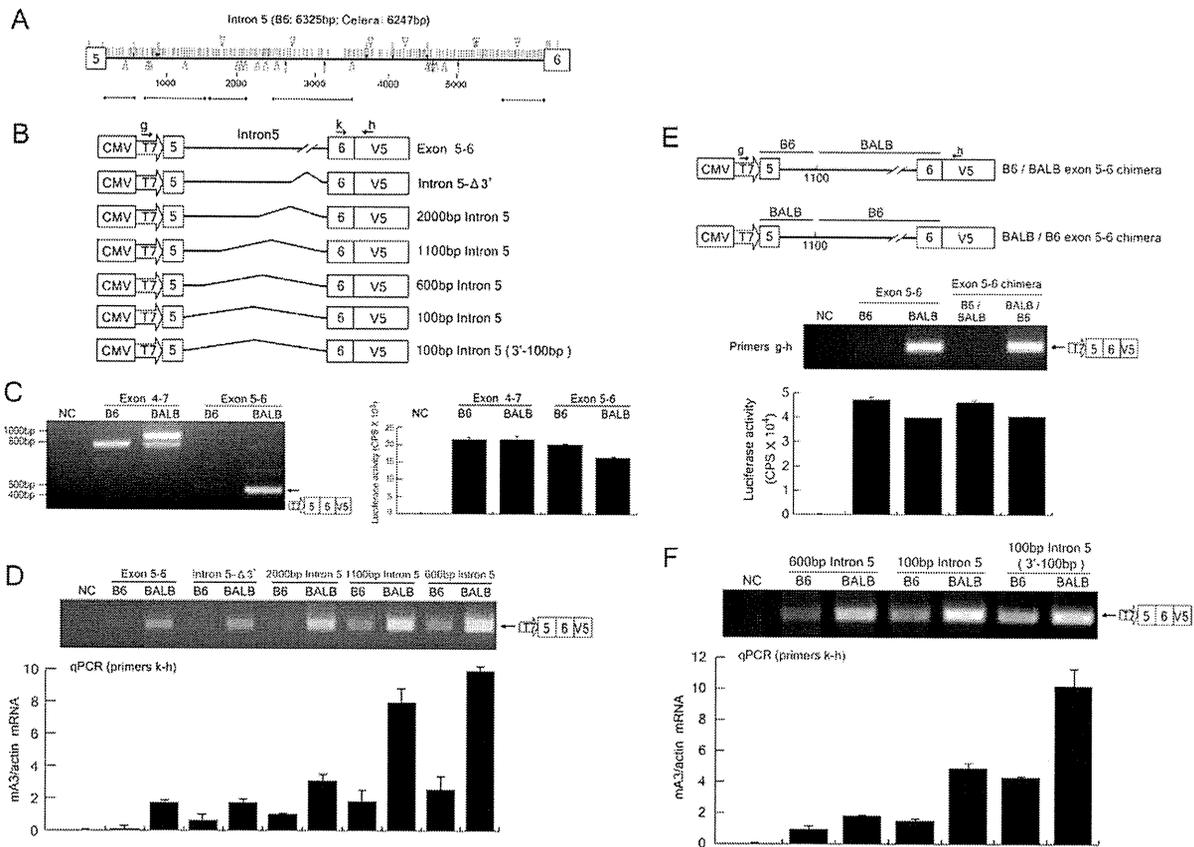doi:10.1371/journal.ppat.1002478.g004

Figure 5. The effect of exon 5 and its downstream sequences on mA3 intron 5 splicing. (A) Distributions of sequence polymorphisms in the intron 5 of the *Apobec3* gene locus between the B6 allele [NT_039621] and the Celera database sequence of mixed mouse DNA [NW_001030577.1]. Shorter horizontal lines below the thick one representing intron 5 indicate regions of sequenced BALB/c genome [DDBJ accession No. AB646261-AB646265], which show nucleotide sequences identical to corresponding Celera database sequences. Spans of the analyzed regions are 1–620, 661–1597, 1643–2058, 2759–3443, and 5558–6247 in base numbers starting from the first nucleotide of intron 5. SNP are shown with vertical lines, single-base indels with arrows, and deletions of ≥2 bases with triangles. Indels and deletions above the thick horizontal line that represents intron 5 are deletions in the B6 allele relative to the Celera sequence, while those underneath the horizontal line are deletions in the Celera sequence relative to the B6 allele sequence. (B) Plasmid constructions for the splicing assays. The exon 5–6 plasmids harbored either the B6 or BALB/c genomic fragment encompassing exons 5 and 6, including the entire intron 5 of the corresponding allele. Each of the remaining plasmids possessed a sequentially reduced length of the intron 5 as indicated. The precise size of each PCR-generated 5′ fragment included was as follows: 3177bp and 3185bp for B6 and BALB/c intron 5-Δ3′; 2106bp and 2086bp for B6 and BALB/c 200bp intron 5; 1120bp and 1110bp for B6 and BALB/c 1100bp intron 5; and 634bp and 620bp for B6 and BALB/c 600bp intron 5, respectively. The primers g-h are the same as shown in Figure 4. (C) RT-PCR detection of spliced messages expressed from the B6 and BALB/c exon 5–6 plasmids along with those from the exon 4–7 plasmids as controls. Primers g and h were used. A lack of expression of the spliced message in cells transfected with the B6 exon 5–6 plasmid was evident. (D and F) Splicing assays using intron 5 deletion plasmids. The intron 5 fragment included in each plasmid is shown in (B). RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples as shown in (C). Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). Quantitative real-time PCR data show averages of three reaction wells and SD. (E) Reciprocal chimeras were produced between B6 and BALB/c exon 5–6 by exchanging the cloned genomic DNA fragment at position 1100 within intron 5. The exact location of the above position 1100 for B6 and BALB/c intron 5 is described in the legend for (B). RT-PCR detection of spliced messages was performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies.
doi:10.1371/journal.ppat.1002478.g005

downstream of exon 5 and the acceptor site upstream of exon 6 in the BALB/c allele are both functional. In contrast, no spliced product was detectable in cells transfected with the B6 exon 5–6 plasmid despite a higher transfection efficiency. These results indicate that the B6 allele may not carry a functional splice donor site at the exon 5/intron 5 boundary, as the splice acceptor site upstream of exon 6 seems intact and thus can generate a message corresponding to the mA3 Δ5 when exon 4 and intron 4 are included (Figures 4 and 5C).

To narrow down the region affecting the intron 5 splicing, serial deletion constructs were produced and subjected to the splicing

assay (Figure 5B). The intron 5-Δ3′ is a deletion construct which lacks the 3′ half of intron 5 but retains the 261-bp sequence adjacent to exon 6 to include the putative acceptor site. The cells transfected with the BALB intron 5-Δ3′ plasmid produced the spliced message as efficiently as those transfected with the parental BALB exon 5–6 construct; however, the B6 intron 5-Δ3′ plasmid generated a barely detectable band representing spliced message (Figure 5D). Quantitative real-time PCR assays revealed markedly lower levels of properly spliced message generated from the B6 intron 5-Δ3′ plasmid compared to that generated from the BALB/c counterpart. These data show that the 5′ half of intron 5 is

responsible for the observed differences in splice site functions between the B6 and BALB/c alleles. Thus, further deletions progressively closer to exon 5 were introduced into the intron 5 fragment. As expected, the shorter the included intron 5 fragment, the more abundant the spliced product generated was, regardless of the expressed *Apobec3* alleles, indicating that the intron length of the primary transcript influences the splicing efficiency (Figure 5D). Nevertheless, the spliced product was generated much more efficiently from the BALB/c allele than from the B6 allele with all examined construct pairs, despite comparable levels of transfection efficiency, indicating that each shorter intron fragment still retained the polymorphism responsible for the inefficient splicing of the B6 mRNA.

To exclude the possibility that the 3′ region of intron 5 and/or exon 6 from the B6 allele might harbor inhibitory sequences that interfere with splicing, we constructed chimeras between the BALB and B6 exon 5–6 plasmids by exchanging the corresponding genomic DNA fragments at approximately position 1100 within intron 5 (Figure 5E). BALB/B6 chimera harboring the 5′ donor site sequence from the BALB/c allele and 3′ acceptor site from the B6 allele generated properly spliced mRNA as efficiently as the BALB exon 5–6 plasmid did, while the reciprocal construct harboring the B6 donor and BALB/c acceptor sequences did not (Figure 5E), indicating that there are no inhibitory elements in the fragment harboring the 3′ intron 5 and exon 6 from the B6 allele. Thus, we continued to narrow down the sequence affecting the intron 5 splicing toward the 5′ end of the genomic constructs.

Even the shortest construct pair, B6 100bp intron 5 and BALB 100bp intron 5, which harbor the 5′ intron 5 fragment a mere 100-bp from the exon 5 boundary, still exhibited a readily discernible difference in the amounts of the spliced messages, and this was also true for another pair of deletion constructs that harbor a 100-bp acceptor region fragment of 3′ intron 5 (Figure 5F). These results clearly indicate that the region including exon 5 and the 5′ 100-bp of intron 5 carries the primary determinants responsible for the different efficiencies in splicing of intron 5 shown by the B6 and BALB/c alleles.

A single nucleotide polymorphism within exon 5 is primarily responsible for exon 5 inclusion

Within the above narrowed-down region of exon 5 and the 5′ 100 bp of intron 5, there are only 4 SNPs between the B6 and BALB/c alleles, as previously described [38]: the T/C SNP at 14-bp downstream from the first nucleotide of exon 5, G/C at 88-bp downstream from the same first nucleotide, G/C at 153-bp downstream within intron 5, and A/C at 163-bp downstream within intron 5 (Figure 6A). We therefore produced a series of point mutants based on the B6 intron 5–100 (3′-100bp) and BALB/c intron 5–100 (3′-100bp) constructs to precisely identify the critical SNP responsible for the functionalities of the splice donor site. In the first set of experiments, shown in Figure 6B, each indicated nucleotide within the BALB/c genomic sequence was substituted with the corresponding nucleotide found in the B6 allele. The substitution of the C at position 14 to the B6-type T slightly reduced the amount of the spliced message; more importantly, however, the similar substitution of C at position 88 to the B6-type G totally abrogated the intron 5 splicing. Simultaneous substitutions of C to G at position 153 and G to A at 163 did not affect the generation of the spliced message. Transfection efficiencies were essentially equivalent for all samples and even higher than those for the C88G construct as confirmed by cotransfecting the luciferase expression plasmid (data not shown). These results indicate that the G/C SNP at position 88 in exon 5 is most critical for the splice donor function.

In a reciprocal set of experiments, we attempted to alter splicing of the transcript of the B6 allele by replacing the polymorphic nucleotides with the BALB/c-types (Figure 6C). When all 4 SNPs in the B6 intron 5–100 fragment were replaced with those corresponding to the BALB/c allele, abundant generation of the spliced message was observed (Figure 6C). A combined substitution of T to C and G to C at positions 14 and 88, respectively, also resulted in efficient splicing of intron 5, and a single substitution of T to C at position 14 alone slightly increased the generation of the spliced message from the plasmid harboring the truncated intron. Importantly, a single substitution of G to C at position 88 resulted in as abundant expression of the spliced message as that observed with the BALB/c allele. Thus, these results collectively indicate that the G at position 88 is critical for the exclusion of exon 5, although the C at position 14 may also play a minor role in the inclusion of this exon.

Site-directed mutagenesis of the plasmids harboring the entire intron 5

To determine the association of the above 4 SNPs with splicing efficiency under more physiological conditions, we further introduced point mutations into the B6 exon 5–6 plasmid harboring the entire intron 5 of 6kb in length, and the resultant plasmids were subjected to the splicing assays (Figure 7A). In contrast to the result with the B6 100bp intron 5 T14C mutant (Figure 6C), we could not detect the generation of spliced mRNA after transfection of the B6 exon 5–6 T14C mutant (Figure 7B). However, the single substitution at position 88 from the B6-type G to BALB/c-type C led to readily detectable production of the spliced mRNA, despite the general inefficiency in splicing of the full-length intron 5 depicted in Figure 5. Quantitative real-time PCR assays revealed levels of expression of the exon 5-containing spliced message from the B6 exon 5–6 G88C mutant that were about 70% of those expressed from the BALB exon 5–6 plasmid. Single nucleotide substitutions at two other positions did not result in a detectable generation of the spliced mRNA from the B6 allele (Figure 7B). Combined mutations at positions 14 and 88 within exon 5 on the B6 exon 5–6 construct (B6 exon 5–6 T14C G88C) and further addition of the nucleotide replacements at positions 153 and 163 also resulted in detectable generation of the spliced message, but did not noticeably enhance the splicing over what resulted from the G88C substitution alone (Figure 7C).

Combined effects of the intron 4 TCCT repeat number and exon 5 G/C SNP on exon 5 inclusion into mA3 mRNA

To further compare the effects of the TCCT repeat number and exon 5 G/C SNP at position 88, we constructed B6 and BALB exon 4–7 plasmids that harbored reciprocal substitutions at these two polymorphic sites (Figure 8). Introduction of an additional copy of TCCT into the B6 exon 4–7 construct did not result in the generation of 5+ message, while the combination of the additional TCCT and G88C substitution resulted in the generation of exon 5-containing message at levels comparable to those expressed from the BALB exon 4–7 plasmid. On the other hand, a deletion of a TCCT copy from the BALB exon 4–7 construct resulted in much reduced expression of the 5+ message, and the combination of ΔTCCT and C88G substitution totally abrogated the generation of exon 5-containing message. Thus, these results clearly demonstrate that the most critical polymorphism for exon 5 inclusion into the mA3 mRNA is the G/C SNP at position 88 within exon 5, but for the full-level expression of the 5+ mRNA as observed in BALB/c mice the intron 4 TCCT repeat is also required.

Figure 6. Identification of critical residues for exon 5 inclusion into mA3 mRNA. (A) Polymorphic nucleotides indicated with the shades were exchanged at the boxed sites between the B6 and BALB/c alleles. The numbers 14, 88, 153, and 163 above the shown nucleotide sequence are base numbers from the first nucleotide of exon 5. (B and C) The results of splicing assays performed with the plasmids depicted in (A). The predicted splicing product is shown on the right side of each panel. Band intensities are averages of 3 independent RT-PCR reactions with SD.
doi:10.1371/journal.ppat.1002478.g006

A

B

qPCR (primers k-h)

C

qPCR (primers k-h)

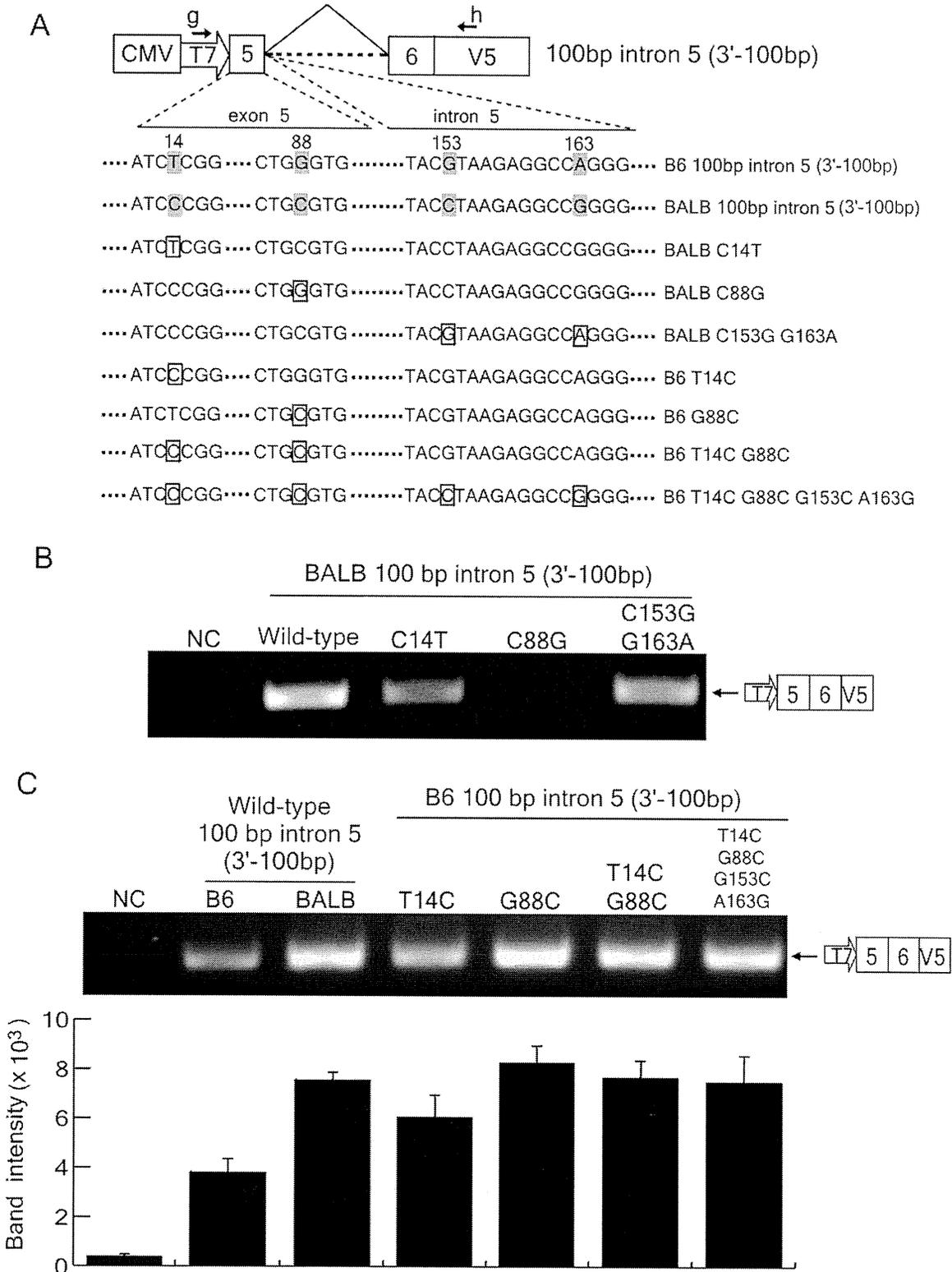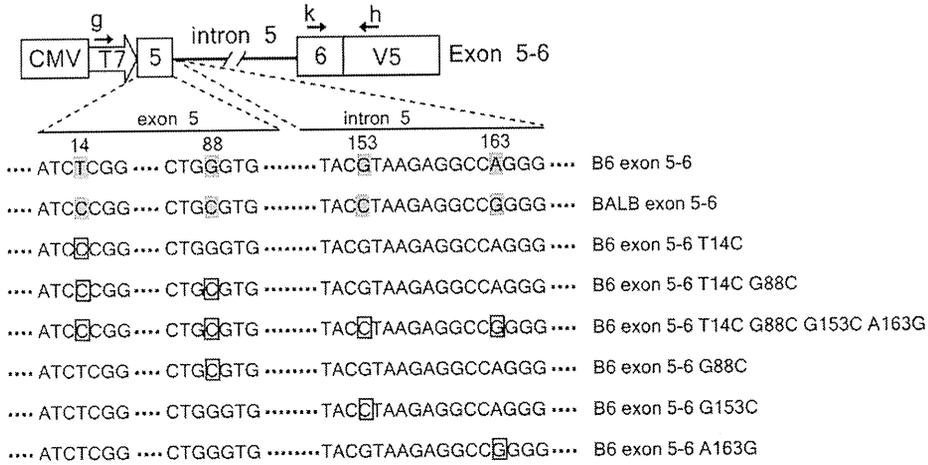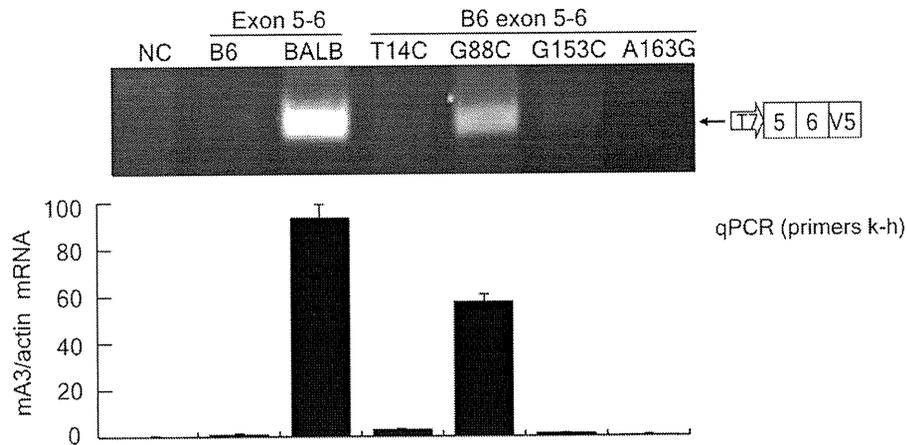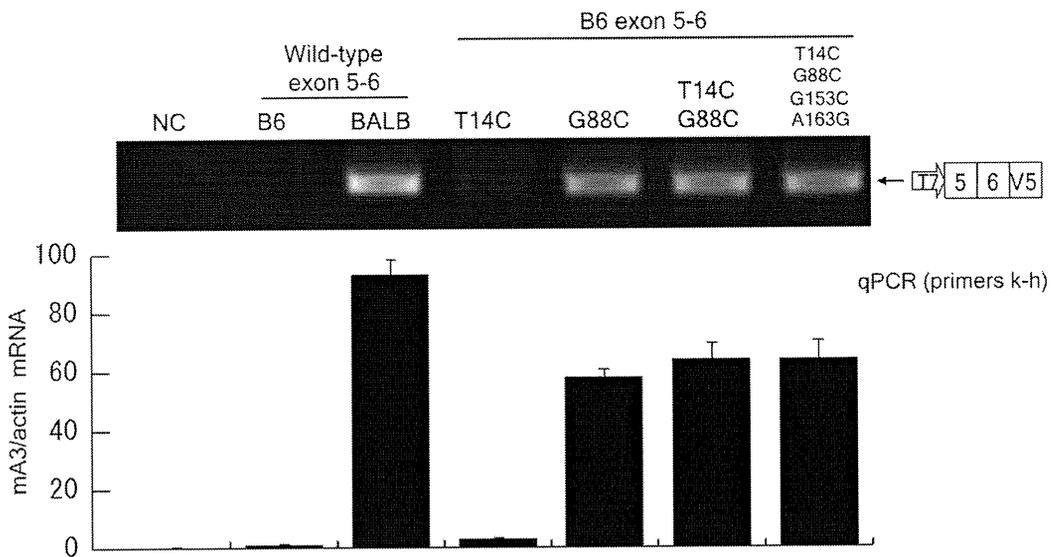**Figure 7. The G88C nucleotide substitution in exon 5 is most critical for exon 5 inclusion into mA3 mRNA.** (A) Polymorphic nucleotides indicated with the shades were exchanged between the B6 and BALB/c alleles at the sites indicated with the boxes. Plasmids used here carry the entire intron 5 instead of the deleted one used in the experiments shown in Figure 6. (B and C) The results of splicing assays using the plasmids depicted in panel (A). The predicted splicing product is shown on the right side of each panel. RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). Quantitative real-time PCR data show averages of three reaction wells and SD.
doi:10.1371/journal.ppat.1002478.g007

Distribution of the TCCT repeat and G/C88 polymorphism among *Mus* species

Because the inclusion of exon 5 into mA3 message depends on the intron 4 TCCT repeat and the exon 5 G/C polymorphism at position 88 (G/C88), we screened additional inbred strains and wild mouse species for these sequence variations and for the presence or absence of exon 5 in the mA3 message. We sequenced exon 5 and segments of the flanking introns from 39 mice that represent different taxa or members of the same species trapped in different geographic locations (Table S1), as well as those from the inbred laboratory strains B10.A and A/WySn, prototypic strains with FV-restrictive and -permissive phenotypes [26,27,31,32]. The complete dataset is shown in Figure S1. Genomic sequencing showed that A/WySnJ mice share the TCCT duplication and exon 5 C88 SNP with BALB/c, while B10.A mice are identical to B6 mice at both sites (Figure 9A); this is consistent with previous observations on mA3 expression levels and inclusion of exon 5 in these strains of mice [27]. Among the 39 wild mice, the TCCT duplication was found in 10 mice that possessed the exon 5 C88 SNP, and all of the 29 other mice with G88 had a single TCCT copy as representatively shown in Figure 9A. The observed linkage between the repeated TCCT and C88 is reasonable, as both are required for efficient expression of 5+ message from the BALB/c allele (Figures 4, 6, and 8). RNA samples from 23 of the above mice were typed for the splicing phenotype. The presence or absence of exon 5 in mA3 mRNA correlated with the above identified sequence polymorphisms: exon 5 was present only in 7 mice, all of which had the linked TCCT duplication and C88, consistent with our functional assays. On the other hand, there were at least 7 discrepancies with respect to exon 5 inclusion and the C14 SNP confirming that this SNP has no major role in exon 5 splicing.

The coding sequence for exon 5 was present in all 39 *Mus* DNA specimens, as well as in rat DNA specimens (GenBank EDM15775) as described previously [29]. There were frameshift mutations within exon 5 in *M. setulosus* and both *M. cervicolor* subspecies, all of which lack exon 5 in mA3 mRNA, and the exon 5-containing splice variant was found only in mA3 mRNA of *M. spretus* and 3 of the house mouse species (*M. domesticus*, *M. musculus*, *M. castaneus*) (Figure 9B). The wild mice were also tested for the presence or absence of another genetic variation associated with variable expression levels of mA3, the exon 2-associated xenotropic MuLV LTR insertion [37]. Results indicated that the presence of this LTR (LTR+), like the inclusion of exon 5 due to the above allelic variation (exon 5+), was introduced into the *Apobec3* locus at about the time of the house mouse radiation (Figure 9B). These two features are not found together in any mouse, and the LTR+ and exon 5+ variants are both found in mice classed as *M. musculus* or *M. domesticus*, while *M. castaneus* mice have both variants as well as the ancestral Δ5 LTR− mA3 type.

To better understand this species distribution, we typed additional samples of European *M. domesticus* and *M. musculus* for the presence of the LTR and exon 5 (Figure 9C). The geographic distribution of the trapping sites for these 19 house mice and for the exon 5+ *M. spretus* indicated that the exon 5+ mA3 was

characteristic of *M. domesticus* of north Africa and western Europe, whereas LTR+ mA3 was found in eastern Europe (*P* = 0.049 by Fisher's exact test). Four of the 9 eastern European *M. musculus* mice carried the *M. domesticus* variant, which is consistent with previous reports that gene flow across the hybrid zone is biased in the direction from *M. domesticus* to *M. musculus* ([42], for example). Additional samples identified as *M. domesticus* were trapped in the Americas (HAF, CL, WSA, PGN2, SC1, BQC, SAF, JJD, PERA; see Table S1), and these carried either the exon 5+ mA3 (like *M. domesticus*) or the LTR+ mA3 (like *M. musculus*); this is likely due to the fact that *Mus* are non-native species that were introduced into the Americas by passive transport, and although generally classed as *M. domesticus*, these mice show evidence of hybridization with other introduced house mouse species [43,44].

Discussion

Polymorphisms in mA3 are responsible for the fact that the B6 and C57BL/10 mouse strains are more restrictive to the replication of both beta- and gammaretroviruses than the BALB/c and A strains [25–29]. These differences in mA3 antiviral activities have been associated with sequence differences in the N-terminal region of mA3 [27], different levels of its mRNA expression [27,29,38,39], and the presence or absence of exon 5 in mA3 mRNA and protein [27], although the relative importance of these three factors for antiviral activity has not been established. In the present study, we focused on the role of exon 5 and have shown that the Δ5 mRNA directs more efficient mA3 protein synthesis than the exon 5-containing message. We have also identified 2 genetic determinants responsible for the inclusion of this exon into mA3, a TCCT repeat in intron 4 and a C88 substitution in exon 5 relative to the B6 allele. We further showed that these two determinants were coordinately acquired within the last 0.5 million years by house mouse species of *Mus*.

Previous studies have recognized differences in mRNA expression levels of mA3 between B6 and BALB/c mice [27,29,38,39]. Higher constitutive expression of mA3 mRNA is thought to contribute to better antiviral activity as suggested by the facts that higher levels of protein production result in more efficient incorporation of mA3 into viral particles [45] and that lipopolysaccharide (LPS)-induced enhancement of mA3 expression results in better restriction of MMTV *in vivo* [46]. It has also been shown that increased APOBEC3G and APOBEC3F expression is associated with lower viral load in rhesus macaques infected with simian immunodeficiency virus [47]. Higher protein levels of hA3G in CD14+ monocytes are also associated with HIV-1-exposed but uninfected status in humans [48]. The increased level of *Apobec3* gene transcription has been associated with the presence in the B6 allele of an intact MuLV LTR, a sequence capable of driving enhanced transcription [37]. While increased levels of the mA3 transcript could be solely responsible for the presently observed increase in mA3 protein levels in B6 mice (Figure 1D), it was also reported that, when driven by the same strong promoter, the B6 mA3 cDNA lacking exon 5 produced more protein product than the exon 5-containing BALB/c cDNA and resulted in more efficient incorporation into viral particles
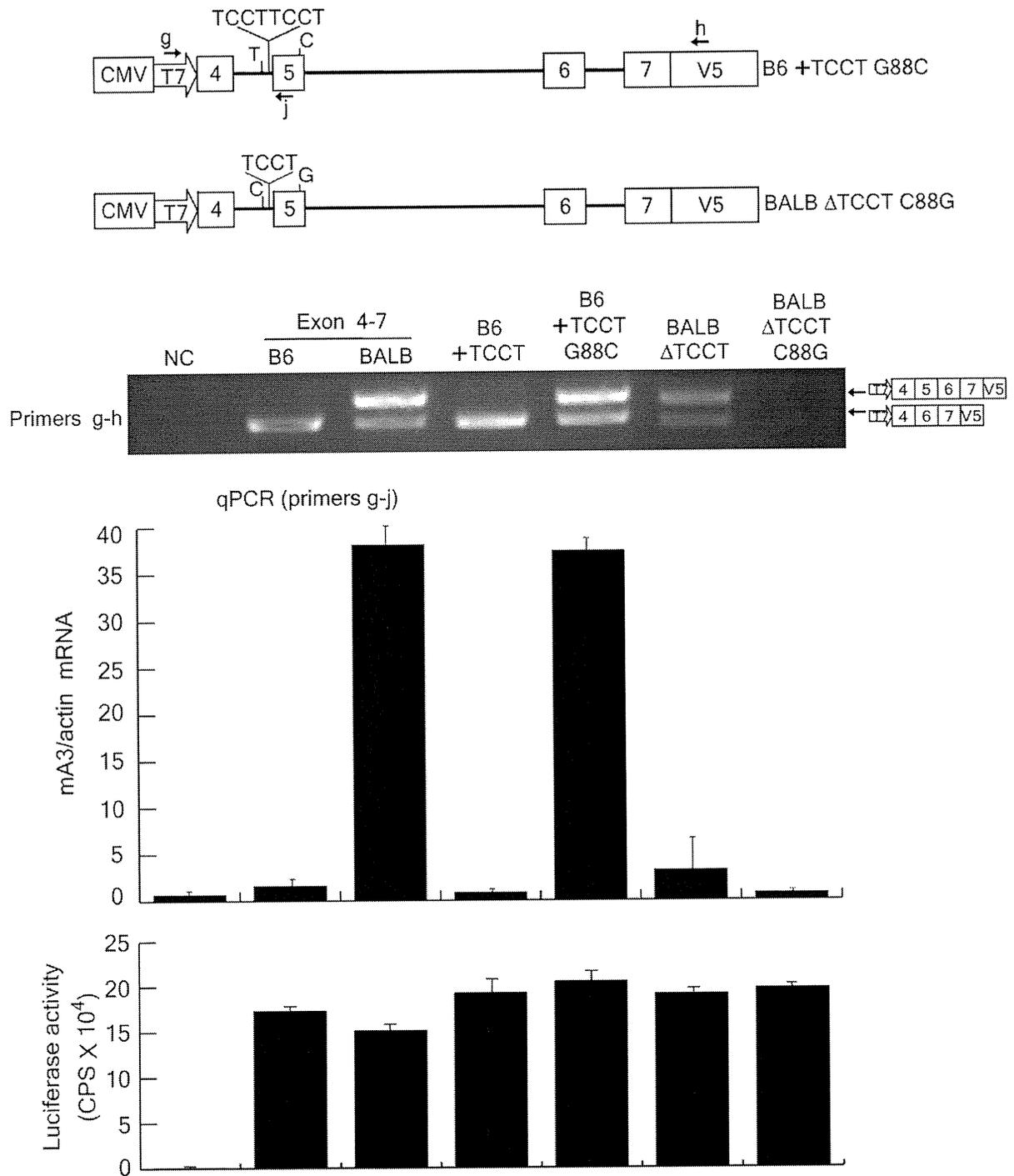
Figure 8. Effect of the combination of intron 4 TCCT repeat and exon 5 G/C SNP on exon 5 inclusion into the mA3 transcripts. B6 exon 4–7 + TCCT and BALB/c exon 4–7 ΔTCCT plasmids, as described in Figure 4, were further modified so that the exon 5 G/C SNP nucleotide at position 88 was reciprocally exchanged. The resultant plasmids were used for *in vitro* transcription assays as described for Figures 4–7. RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Quantitative real-time PCR data show averages of three reaction wells and SD.
doi:10.1371/journal.ppat.1002478.g008

[27,45]. Thus, it was unclear whether the experimentally observed increased translation was due to the lack of exon 5 in the B6 cDNA or to the other sequence differences that distinguish the two

Apobec3 alleles [27,37,38]. Here we have clearly shown that the Δ5 mRNA directs much more efficient protein synthesis than the 5+ mA3 mRNA regardless of other allelic differences (Figures 2 and

A

intron 4   intron 5

4   5   6

intron 4   exon 5   intron 5   Ex5   LTR

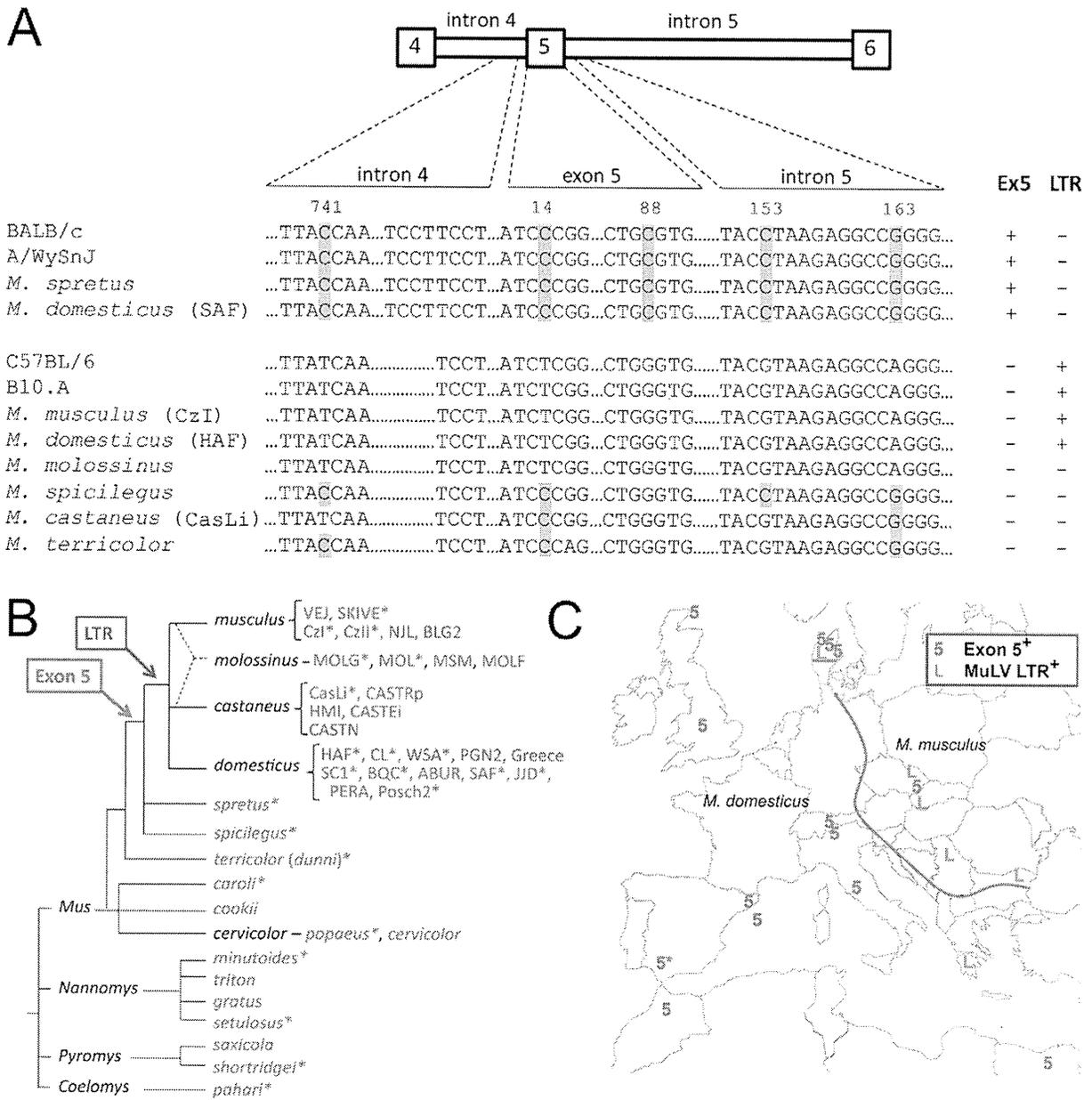|	741		14	88	153	163	Ex5	LTR
BALB/c	...TTACCAA...	TCCTTCCT...	ATCCCGG...	CTGCCGTG......	TACCTAAGAGGCCGGGG...		+	–
A/WySnJ	...TTACCAA...	TCCTTCCT...	ATCCCGG...	CTGCCGTG......	TACCTAAGAGGCCGGGG...		+	–
M. spretus	...TTACCAA...	TCCTTCCT...	ATCCCGG...	CTGCCGTG......	TACCTAAGAGGCCGGGG...		+	–
M. domesticus (SAF)	...TTACCAA...	TCCTTCCT...	ATCCCGG...	CTGCCGTG......	TACCTAAGAGGCCGGGG...		+	–
C57BL/6	...TTATCAA............	TCCT...	ATCTCGG...	CTGGGTG......	TACGTAAGAGGCCAGGG...		–	+
B10.A	...TTATCAA............	TCCT...	ATCTCGG...	CTGGGTG......	TACGTAAGAGGCCAGGG...		–	+
M. musculus (CzI)	...TTATCAA............	TCCT...	ATCTCGG...	CTGGGTG......	TACGTAAGAGGCCAGGG...		–	+
M. domesticus (HAF)	...TTATCAA............	TCCT...	ATCTCGG...	CTGGGTG......	TACGTAAGAGGCCAGGG...		–	+
M. molossinus	...TTATCAA............	TCCT...	ATCTCGG...	CTGGGTG......	TACGTAAGAGGCCAGGG...		–	–
M. spicilegus	...TTACCAA............	TCCT...	ATCCCGG...	CTGGGTG......	TACCTAAGAGGCCGGGG...		–	–
M. castaneus (CasLi)	...TTATCAA............	TCCT...	ATCCCGG...	CTGGGTG......	TACGTAAGAGGCCGGGG...		–	–
M. terricolor	...TTACCAA............	TCCT...	ATCCCAG...	CTGGGTG......	TACGTAAGAGGCCGGGG...		–	–

B

LTR

Exon 5

musculus — VEJ, SKIVE*, CzI*, CzII*, NJL, BLG2

molossinus – MOLG*, MOL*, MSM, MOLF

castaneus — CasLi*, CASTRp, HMI, CASTEi, CASTN

domesticus — HAF*, CL*, WSA*, PGN2, Greece, SC1*, BQC*, ABUR, SAF*, JJD*, PERA, Posch2*

spretus*

spicilegus*

terricolor (dunni)*

caroli*

cookii

cervicolor – popaeus*, cervicolor

Mus

Nannomys — minutoides*, triton, gratus, setulosus*

Pyromys — saxicola, shortridgei*

Coelomys — pahari*

C

5   Exon 5+
L   MuLV LTR+

M. musculus

M. domesticus

Figure 9. Distribution of sequence variations in and around the *Apobec3* **exon 5 in** *Mus* **species and strains.** (A) The *Apobec3* genomic sequences of the indicated strains and representative species are shown. The five SNP tested for their effects on exon 5 splicing are indicated with shaded bars. Also shown are the number of TCCT repeats and typing results indicating the presence or absence of exon 5 (Ex5) in spliced messages and the presence or absence of an MuLV LTR in intron 2. (B) The phylogenetic tree is a synthetic tree adapted from other sources [66–68]. Species are grouped into the 4 *Mus* subgenera and are listed to the right of each branch. Subspecies and stock or inbred line designations are given where multiple samples of a given species were tested. Red, exon 5+/MuLV LTR−; blue, Δ5/MuLV LTR+; green, Δ5/MuLV LTR−. *, mice tested for exon 5 splicing pattern. All were sequenced across exon 5 (Figure S1); PERA and MOLF sequence information was taken from Okeoma et al. [38]. (C) Distribution of expression variants of mA3 in mice trapped in Europe. The red line represents the 20 km-wide hybrid zone separating the ranges of *M. domesticus* and *M. musculus*. Symbols representing mA3 allelic variants are placed at trapping sites.
doi:10.1371/journal.ppat.1002478.g009

3), and the combination of the higher mRNA expression levels with preferential generation of the Δ5 isoform in mRNA splicing results in much higher mA3 protein levels in B6 in comparison with BALB/c mice (Figure 1).

The mechanisms by which the presence of exon 5 in the mRNA interferes with protein synthesis are currently unclear: however, predictions of the secondary structure for the mA3 mRNA suggest that the portion encoding exon 5 may form stable stem-loop structures (Figure S2), which might interfere with efficient translation. The predicted stem-loop structure is thermodynamically more stable for B6 than for BALB/c exon 5, and this portion is actually spliced out from the major transcript in B6 mice. It is

possible, therefore, that the elimination of this exon from mA3 mRNA confers a functional advantage to B6 mice through more efficient translation of the Δ5 mRNA and resultant higher levels of mA3 protein expression. It should be noted, however, that introduction of this exon between the two CDDs does not physically disrupt the functional domains, as the 5+ product of the BALB/c allele still exerts deaminase activity and restricts AKV replication *in vitro* [29]. The possible effect of this 33-amino acid exon on interactions between the N-terminal and C-terminal CDDs and on mA3 oligomerization remains to be elucidated.

Our results also show that the ancestral *Mus Apobec3* gene locus lacked both the MuLV LTR and the functional polymorphisms that determine exon 5 inclusion (Figure 9). These two genetic features were acquired independently at about the time of the house mouse radiation 0.5–1.0 million years ago, and are found in distinct lineages of wild mice and different inbred strains of laboratory mice. It should be noted that all 4 house mouse species originated and diverged from an ancestral population on the Indian subcontinent that carries many of the alleles found in peripheral Eurasian populations [49,50]. The observed distribution of the LTR$^+$ and exon 5$^+$ mA3 variants among the house mouse lineages is consistent with the appearance of these variants in this ancestral population. It is thus likely that *M. domesticus* progenitors carrying the exon 5$^+$ *Apobec3* allele moved west to north Africa and western Europe, and *M. musculus* (LTR$^+$) moved north to Russia, eastern Europe and north China. *M. castaneus*, which carries these two variants as well as the ancestral Δ5 LTR$^-$ mA3, migrated east to Thailand and China. Only the ancestral mA3 type is found in *M. molossinus*, which is a natural hybrid of *M. musculus* and *M. castaneus* [44], and these Japanese mice presumably inherited their *Apobec3* gene from their *M. castaneus* progenitors.

Numerous studies indicate that host antiviral factors co-evolve with the pathogens they restrict, and this "arms race" is responsible for mutational changes in APOBEC3 in primates and mice as evidenced by detectable positive selection [37,51]. The acquisition of the MuLV LTR that is associated with enhanced mA3 transcription in virus-infected mice also makes sense as an evolutionary adaptation to pathogen infection [37]. Thus, the B6 *Apobec3* allele has acquired two advantageous features that contribute to enhanced retrovirus restriction: high mRNA expression directed by the intron 2 LTR that is in linkage with the positively selected, more functional amino acid sequence in the N-terminal deaminase domain [27,37]. The BALB/c allele, in contrast, acquired the polymorphisms that direct exon 5 inclusion as shown in the present study, along with coding sequence polymorphisms associated with lower antiviral activity [27]. The far-flung distribution of exon 5$^+$ mA3 throughout western Europe (Figure 9C) is surprising, because this allele, having a reduced level of mA3 protein expression with inefficient restriction of MuLV replication at least *in vitro* [27], would seem to be evolutionarily deleterious. The observed distribution of exon 5$^+$ mA3 suggests either that the exon 5$^+$ variant provides sufficiently protective levels of antiviral activity in virus-infected mice and is therefore not subject to purifying selection, that these particular mice have not been exposed to significant challenge by mA3-sensitive pathogens, that this mA3 variant provides some other, unrecognized selective advantage, or that this allele is tightly linked to another advantageous polymorphism.

All 4 house mouse species possessing the exon 5$^+$ genetic variation carry beta- and gammaretroviruses along with other retroelements that are subject to APOBEC3 restriction [25–30,52]. In this regard, although the BALB/c mA3 is less antiviral than the product of the B6 allele, it does have measurable antiviral activity [29]. Further, analysis of 54 germline MuLV proviruses of

three envelope types in the sequenced B6 genome demonstrated evidence of APOBEC3-mediated editing of the polytropic and modified polytropic, but not the xenotropic, proviruses [30]. The mA3-edited polytropic and modified polytropic viruses originated in *M. domesticus* [52,53], mice that carry the exon 5$^+$ mA3, suggesting that this mA3 variant may be effectively antiviral *in vivo*. Finally, in mA3-deficient B6 mice, the absence of this protein differentially affects FV replication depending on target cell types [54]. As cell type-dependent and pathogen-induced changes in mA3 expression may differ between mice of the ancestral Δ5 LTR$^-$, exon 5$^+$, and LTR$^+$ variants, mice with the exon 5$^+$ *Apobec3* allele may still restrict mA3-sensitive pathogens in critical target cells.

It is also possible that some mA3 polymorphisms have been selected for reasons other than their antiviral functions, and these functions may take precedence in mice not threatened with retrovirus assault. It is unlikely that differences in mA3 expression levels and/or its amino acid sequence provide a significant selective advantage in terms of normal development, survival, or fertility, as *Apobec3*-knockout mice show no increased propensity for tumor development or disease, and both male and female mA3-deficient mice were fertile [40]. On the other hand, the presence of active cytidine deaminases can have costs that may become more significant when endogenous and exogenous retroviral activity is low. Transgenic mice overexpressing mouse activation-induced deaminase (AID) or rabbit apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1 (APOBEC1), two other members of the AID/APOBEC family of cytidine deaminases, developed neoplastic diseases and showed evidence of significant editing of various expressed genes [55,56]. This suggests that when the retroviral threat is low, the consequences of possessing highly active mutagenic enzymes, like the Δ5 mA3 expressed in B6 mice, can outweigh their advantages as antiviral factors. In this regard, it should be noted that infectious MuLV has not been isolated from European *M. domesticus*, which carry the less highly expressed 5+ mA3, whereas the more active *Apobec3* alleles are found in all 3 of the other species of *Mus*, all of which have been found to carry infectious MuLVs [57].

Finally, it has also been observed that there is a similarly widespread distribution in humans of a deletion of a genomic segment harboring the *Apobec3b* gene locus [58]. At least one copy of this deletion is present in >40% of humans. The selective advantage of this genetic variant has not been determined, but its retention may suggest a possible advantage in reducing the genotoxic activity of this cytidine deaminase, or alternatively, it may be linked to positively selected variants in one or more of the 6 linked human APOBEC3 paralogues [59,60]. Further comparative analyses of the geographic and species distributions of natural mouse pathogens and the various mA3 variants may help define previously unrecognized targets of APOBEC3-mediated restriction and provide further insight into the coevolution of pathogens and this host restriction factor.

Materials and Methods

Ethics statement

The studies utilizing laboratory animals were carried out in strict accordance with the Act on Welfare and Management of Animals of the government of Japan and the Regulations for the Care and Use of Laboratory Animals of Kinki University. The protocol was approved by the Institutional Animal Experimentation Committee of Kinki University School of Medicine (Permit Number: KAME-19-029). All surgery was performed under sodium pentobarbital anesthesia, and all efforts were made to