Table 2. CRF01_AE domestic micro-clades (MCs) detected in the study population.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	Median	95% HPD			Region	Age	Gender	Behavior	Nationality
Transmission Clusters									
MC01	Nov-'96	Sep-'93	–	Oct-'01	Region 2	35	M	Hetero	Japan
					Region 2	42	F	Unknown	Japan
					Region 2	31	M	Hetero	Japan
					Region 3	44	F	Hetero	Thailand
					Region 2	37	F	Hetero	Japan
					Region 2	40	M	Hetero	Japan
MC02	Mar-'00	Dec-'96	–	Aug-'02	Region 3	30	F	Hetero	Japan
					Region 2	29	M	Hetero	Vietnam
					Region 4	21	M	IVDU	Vietnam
					Reference	-	-	Unknown	Taiwan
					Reference	-	-	Unknown	Vietnam
MC03	Mar-'01	Apr-'98	–	Jun-'03	Region 2	26	M	IVDU	Indonesia
					Region 2	47	M	Hetero	Japan
					Region 4	47	M	Hetero	Japan
					Region 2	32	M	IVDU	Indonesia
					Region 4	23	F	Hetero	Indonesia
					Region 4	26	F	Unknown	Indonesia
					Reference	-	-	Unknown	Korea
MC04	Feb-'02	Sep-'99	–	Jan-'04	Region 2	21	F	IVDU	Japan
					Region 2	66	M	Unknown	Japan
					Region 2	45	F	Hetero	Japan
					Region 2	67	M	Hetero	Japan
					Region 2	60	M	MSM	Japan
					Region 2	63	M	Hetero	Japan
MC05	Jun-'02	Feb-'00	–	Apr-'04	Region 2	28	F	Hetero	Japan
					Region 2	56	M	Hetero	Indonesia
					Region 4	24	F	IVDU	Indonesia
					Region 4	34	M	IVDU	Japan
					Region 4	35	M	IVDU	Indonesia
Heterosexual Male-Female Pairs									
MC06	Feb-'03	Aug-'00	–	Nov-'04	Region 2	68	M	Hetero	Thailand
					Region 2	62	F	Hetero	Japan
MC07	May-'03	Oct-'01	–	Apr-'04	Region 4	50	M	Hetero	Unknown

Table 2. Cont.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	Median	95% HPD			Region	Age	Gender	Behavior	Nationality
					Region 4	50	F	Hetero	Japan
MC08	Nov-'03	Jan-'02	–	Dec-'04	Region 4	58	M	Hetero	Japan
					Region 4	51	F	Hetero	Japan
MC09	Jan-'06	Oct-'02	–	Jul-'08	Region 2	40	M	Hetero	Myanmar
					Region 2	33	F	Hetero	Myanmar
MC10	Sep-'06	Aug-'03	–	Oct-'08	Region 2	46	M	Hetero	Japan
					Region 2	29	F	Hetero	Japan
MC11	Apr-'07	Jan-'06	–	Dec-'07	Region 4	66	M	Hetero	Japan
					Region 4	65	F	Hetero	Japan
MSM pairs									
MC12	Nov-'01	Apr-'00	–	Nov-'02	Region 2	29	M	MSM	Japan
					Region 2	36	M	MSM	Japan
MC13	Aug-'03	Jan-'01	–	Jun-'05	Region 6	43	M	MSM	Japan
					Region 6	56	M	MSM	Japan
MC14	Jan-'07	Mar-'05	–	May-'08	Region 6	27	M	MSM	Japan
					Region 6	34	M	MSM	Japan
MC15	Feb-'07	Mar-'05	–	Apr-'08	Region 2	46	M	MSM	Japan
					Region 2	40	M	MSM	Japan
Discordant couples with possible missing cases									
MC16	May-'99	Jan-'99	–	May-'01	Region 2	28	M	Hetero	Japan
					Region 2	46	M	Hetero	Japan
MC17	Sep-'99	May-'96	–	Apr-'02	Region 2	49	M	Hetero	Japan
					Region 2	28	M	MSM	Japan
MC18	Sep-'00	Nov-'96	–	Feb-'03	Region 2	41	M	Hetero	Japan
					Region 2	52	M	Hetero	Japan
MC19	Oct-'00	Nov-'98	–	Jun-'02	Region 4	52	M	Hetero	Japan
					Region 4	65	M	Hetero	Japan
MC20	Dec-'00	May-'98	–	Mar-'03	Region 1	32	M	MSM	Japan
					Region 2	58	M	Hetero	Japan
MC21	Apr-'03	Oct-'99	–	Dec-'05	Region 2	25	F	Hetero	Japan
					Region 2	39	M	MSM	Japan
MC22	Apr-'05	Aug-'03	–	Jun-'06	Region 1	35	F	Hetero	Japan
					Region 2	59	F	Hetero	Japan
MC23	Mar-'07	Sep-'05	–	Jan-'08	Region 2	44	F	Hetero	Japan

Table 2. Cont.

	tMRCA				Characteristics of individuals in micro-clades				
MC-ID	Median	95% HPD			Region	Age	Gender	Behavior	Nationality
MC24	Sep-'07	Dec-'05	–	May-'08	Region 2	34	M	IVDU	non-Japanese
					Region 6	26	M	MSM	Japan
					Region 2	25	M	Hetero	Japan
Insufficient information								?	?
MC25	Aug-'04	Sep-'02	–	Oct-'05	Region 2	20s	M	Unknown	Unknown
					Region 2	28	M	MSM	Japan
MC26	Dec-'05	Nov-'02	–	Mar-'08	Region 2	23	M	Unknown	Unknown
					Region 2	31	F	Hetero	Myanmar
MC27	Mar-'07	Apr-'06	–	Jun-'07	Region 2	45	M	Hetero	Japan
					Region 4	44	M	Unknown	Japan
MC28	Apr-'07	Jun-'05	–	Apr-'08	Region 2	35	F	Unknown	Unknown
					Region 2	29	F	Hetero	Thailand
MC29	May-'07	Feb-'05	–	Dec-'08	Region 2	40s	M	Unknown	Unknown
					Region 2	73	M	Hetero	Japan
MC30	Jun-'07	Jun-'05	–	Oct-'08	Region 2	47	M	Unknown	Unknown
					Region 2	35	M	Unknown	Unknown
MC31	Jul-'08	Apr-'07	–	Jan-'09	Region 2	46	M	Hetero	Japan
					Region 4	39	M	Unknown	Japan
MC32	Oct-'08	Oct-'07	–	Jan-'09	Region 2	24	M	Unknown	Unknown
					Region 2	25	M	MSM	Japan
MC33	Feb-'09	Oct-'07	–	Oct-'09	Region 2	23	M	MSM	Japan
					Region 2	23	M	Unknown	Unknown

MC: micro-clade, HPD: highest posterior density, tMRCA: time of most recent common ancestor.
doi:10.1371/journal.pone.0102633.t002

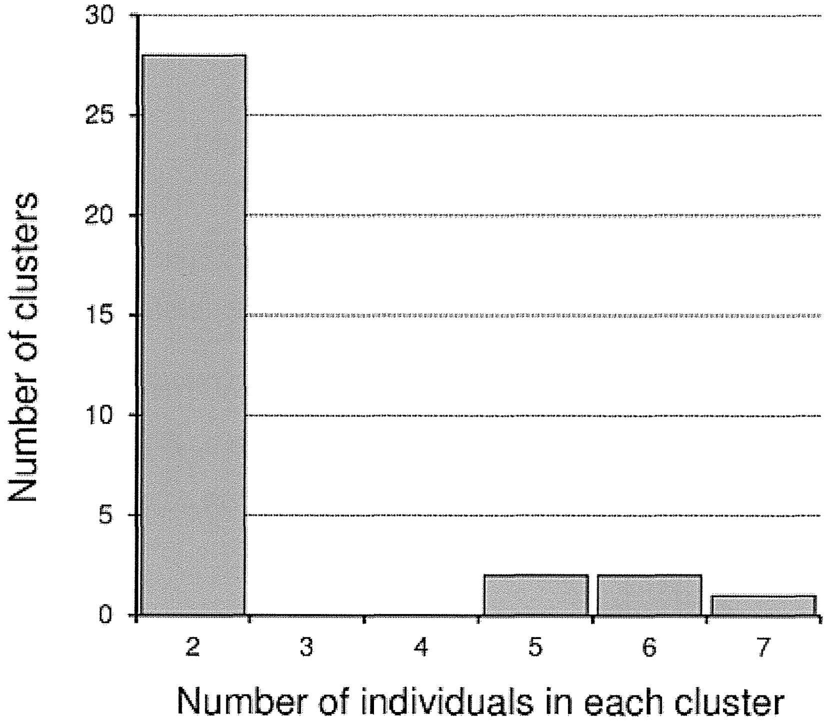

Figure 2. Histogram of CRF01_AE micro-clades by the number of individuals in each micro-clade. Twenty-eight micro-clades consisted of two patients and the remaining 5 micro-clades consisted of more than 4 individuals.
doi:10.1371/journal.pone.0102633.g002

Statistical Analysis

The distribution of IVDUs among micro-clades was tested by χ^2 goodness of fit test to the Poisson distribution. All statistical analyses were calculated using R version 2.10.0.

Results

Major Risk Behavior among CRF01_AE Cases is Heterosexual Contact

We analyzed 297 protease-RT sequences from 243 CRF01_AE-infected cases, which corresponds to 6.7% of the collected study population of 3618 newly diagnosed patients from 2003 to 2010. The highest transmission category in both male and female CRF01_AE populations was high-risk heterosexual contact (Table 1). Among CRF01_AE patients, Japanese males were the majority (52.7%), but this predominance was significantly lower (two sided χ^2 test, $p<0.001$) than in the overall population of HIV infected individuals in Japan (male: 93.2%, Japanese: 90.1%) [16]. The median ages for males, females and the total population were 45, 35 and 40 years, respectively. While patient nationality was not associated with heterosexual behavior (OR = 0.556; $p = 0.183$), MSMs and IVDUs were significantly more (OR = 15.444; $p<0.001$) and less frequent (OR = 0.086; $p = 0.001$) among Japanese patients, respectively (Table S3).

Phylogenetic Analysis Identifies 33 Micro-clades of CRF01_AE-Infected Patients in Japan

The estimated evolutionary diversity of the CRF01_AE protease-RT region is shown in Table S4. Coefficients of differentiation were low among the risk behaviors and collection areas. The mean evolutionary rate of the protease-RT region estimated by Bayesian MCMC inference was 1.07×10^{-3} substitutions per year (Table S5), consistent with previous estimates for the HIV genome [43–45]. The estimated mean coefficient of variation was 0.597, indicating substantial heterogeneity in the evolutionary rate in viral lineage (Table S5). The tree topology showed no association with any demographic parameters of patients. We identified 33 clusters that mainly spread in Japan, i.e., micro-clades MC01 to MC33 (Figure 1, Figure S2). Some of these micro-clades might have been due to intra-patient sequence variation, as suspected for MC27, 32 and 33. Other micro-clades seemed to come from inter-patient diversity because patients' characteristics were clearly distinct from each other (Table 2). These micro-clades consisted of 76 patients (Table 2), corresponding to 31% of all CRF01_AE patients collected in our study. The distribution of cluster sizes is shown in Figure 2. Most micro-clades (n = 28, 85%) consisted of two patients, with only five micro-clades containing more than a pair.

Origin of the Transmission Clusters Introduced into Japan between 1996 and 2002

The median tMRCAs of micro-clades found in Japan and in foreign reference sequences are shown in Tables 2 and 3, respectively. The median tMRCA of CRF01_AE viruses in Japan was estimated using Bayesian MCMC inference as 1968 (95% HPD: 1975–1956), identical to that of the all CRF01_AE viruses measured in this study, including reference sequences. The earliest transmission cluster originating in Japan was MC01, with tMRCA of 1996. This cluster consisted of 6 individuals (3 males and 3 females), of whom 5 reported their risk behavior as heterosexual contact. Five individuals in MC01 were from Region 2, and one Thai female was detected in Region 3 (Figure 3). Four other transmission clusters (MC02-05) showed median tMRCAs between 2000 and 2002. As shown in Figure 3, these clusters comprised individuals from geographically wide areas of Japan, from eastern to central Japan (Regions 2, 3, 4 and 5; see also

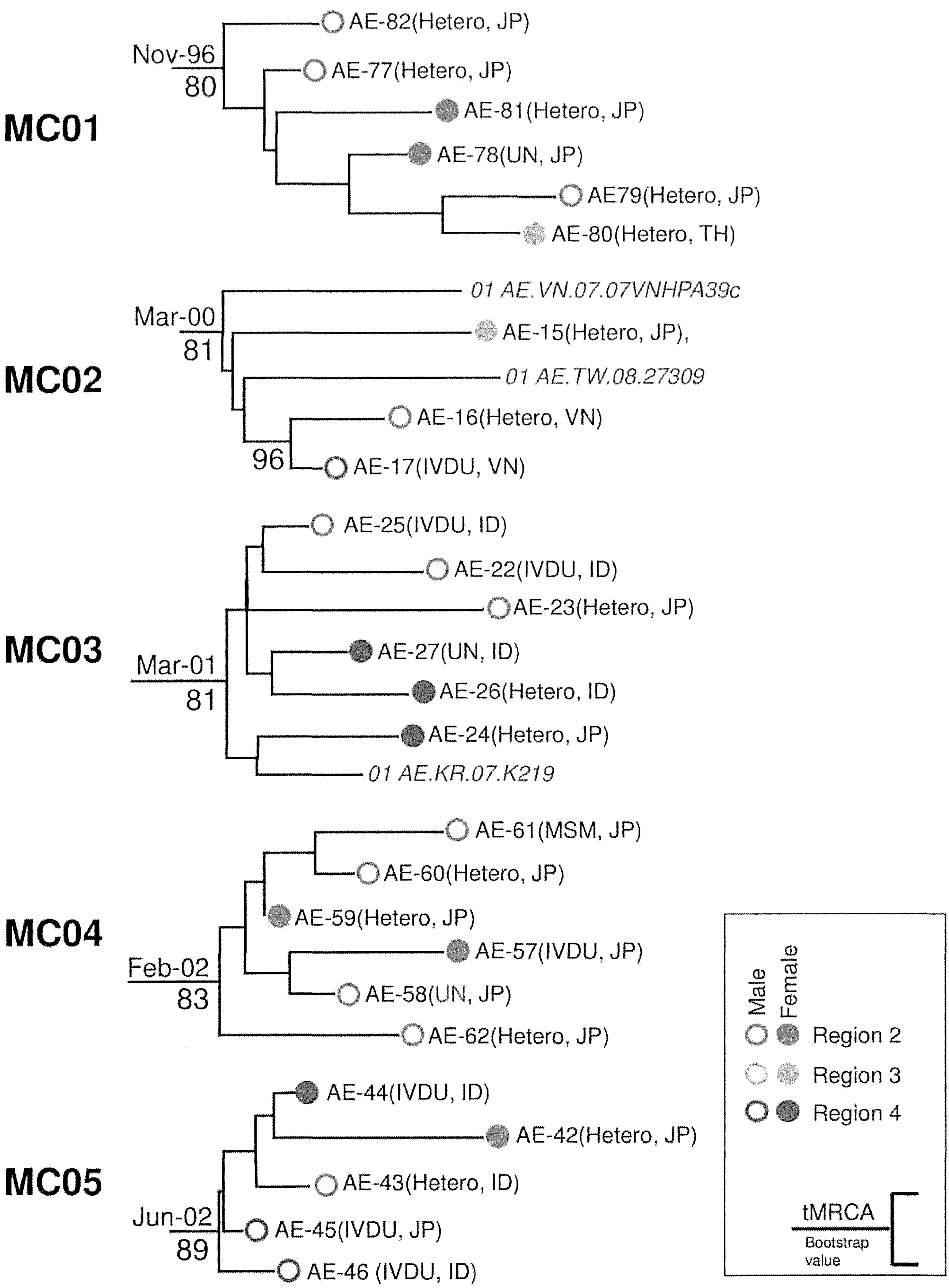
MC01
Nov-96
80
AE-82(Hetero, JP)
AE-77(Hetero, JP)
AE-81(Hetero, JP)
AE-78(UN, JP)
AE79(Hetero, JP)
AE-80(Hetero, TH)
MC02
Mar-00
81
01 AE.VN.07.07VNHPA39c
AE-15(Hetero, JP),
01 AE.TW.08.27309
AE-16(Hetero, VN)
96
AE-17(IVDU, VN)
MC03
Mar-01
81
AE-25(IVDU, ID)
AE-22(IVDU, ID)
AE-23(Hetero, JP)
AE-27(UN, ID)
AE-26(Hetero, ID)
AE-24(Hetero, JP)
01 AE.KR.07.K219
MC04
Feb-02
83
AE-61(MSM, JP)
AE-60(Hetero, JP)
AE-59(Hetero, JP)
AE-57(IVDU, JP)
AE-58(UN, JP)
AE-62(Hetero, JP)
MC05
Jun-02
89
AE-44(IVDU, ID)
AE-42(Hetero, JP)
AE-43(Hetero, ID)
AE-45(IVDU, JP)
AE-46 (IVDU, ID)
Male
Female
Region 2
Region 3
Region 4
tMRCA
Bootstrap value

Figure 3. Detailed phylogenetic structure of five large micro-clades. Open and solid circles indicate male and female cases, respectively. Circle color indicates geographic origin of the samples, followed by risk behavior, and nationality, if known. Dates at the sub-tree's root show the tMRCA of each micro-clade. Numbers below each node show bootstrap values above 80%. JP = Japanese, KR = Korean, VN = Vietnamese, ID = Indonesian, TH = Thai, TW = Taiwanese, UN = unknown, IDVU = intravenous drug user.
doi:10.1371/journal.pone.0102633.g003

Figure S1). These clusters also included both genders and individuals of different nationalities, such as Japanese, Indonesian, Vietnamese, Taiwanese and South Korean, suggesting complex transmission networks in CRF01_AE. The most striking finding was that 7 of 9 IVDUs were significantly concentrated in these four large clusters ($\chi^2 = 528.5$; $p < 2.2 \times 10^{-16}$).

Small CRF01_AE clusters frequently included MSM

In contrast to the large CRF01_AE clusters, which contained only 1 MSM/29 total individuals in clusters, the small CRF01_AE clusters frequently included individuals reporting MSM risk factors. The median tMRCA of small micro-clades ranged between 1999 and 2009 (Table 2). Among these micro-clades, 23 consisted of samples collected in the same region, whereas individuals in the remaining 5 micro-clades were widely distributed over mainland Japan (MC20, 22, 24, 27, and 31). Five micro-clades (MC06, 09, 23, 26, and 28) included at least one non-Japanese. The proportion of females in these small micro-clades was 23%, lower than that of the entire study population (30.5%, Table 1). Six micro-clades (MC06-11) consisted of heterosexual pairs that were unlikely to expand their transmission networks. The median tMRCAs of such closed micro-clades ranged from 2003 to 2007, and those of 4 micro-clades (MC12-15) consisting of MSM pairs ranged from 2001 to 2007. Other 26 small micro-clades had diverse demographic, and risk factor characteristics (Table 2). Taken together with the findings of the large clusters, these data highlight the complexity of CRF01_AE transmission in Japan.

CRF01_AE Epidemic in Japan Occurred in at least Two Waves

The distributions of numbers of individuals in micro-clades and their demographic parameters are graphed versus micro-clade tMRCA in Figure 4. Analysis of tMRCA revealed two distinct groups of CRF01_AE infected individuals: the first wave coming in the early 2000s and the second wave in 2007 to 2008 (Figure 4A). Before the first wave, the major risk factor in the clusters was heterosexual behavior, with a remarkable number of IVDUs around the first wave (Figure 4B). After the first wave, MSM increases gradually and reaches a peak at the second wave (Figure 4B). The clusters in the first wave included many Indonesians (Figure 4C). Thailand, where the CRF01_AE outbreak started, had few contributions to cluster formation in any years, and Japanese mainly contributed in the recent 4 years from 2006 to 2010 (Figure 4C). Individuals from Region 6 were found in the clusters only after 2003, showing that the infection seems to have spread from eastern Japan to the rest of the mainland around this year (Figure 4D). These patterns were not found in analyses of the relationship between demographics and collection year (Figure S3). These data suggest CRF01_AE was introduced at least twice, with one early wave occurring in the early 2000's largely transmitted through heterosexual contact, and a second distinct wave of transmission occurring later that has been sustained by transmission through multiple risk factors, that includes a substantial contribution of MSM transmission.

Some Individuals Infected with Viruses from the Transmission Clusters Spread in East and Southeast Asian Countries

Besides domestic micro-clades, 6 international micro-clades (IMC-1-6) were determined in reference CRF01_AE sequences from the Los Alamos HIV database (Figure 1, Table S1). Among these international micro-clades, 4 (IMCs 1-3 and 5) included sequences from our study (Table 3, Figures S4 and S5). Seven individuals from a wide range of regions in Japan were grouped into IMC-1, the largest international micro-clade with reference sequences collected in Vietnam and China. IMC-2, the second largest cluster consisting of 27 sequences from China, contained one Chinese female detected in Region 4. IMC-3 consisted of 12 sequences from China and included MC33, which included 1 or 2 Japanese males from Region 2. IMC-5, which contained reference sequences from the Philippines, included one MSM Japanese male from Region 2. Thus, our data clearly demonstrate ongoing international transmission of CRF01_AE, especially between China and Japan.

Discussion

The results of this study reveal a process of multiple transmissions and a subsequent spreading pattern of CRF01_AE infection into Japan. Our sample represents 18.7% of the officially reported HIV-1-infected cases in Japan [16]. Although our results were obtained from the relatively short and genetically conserved *pol* sequence region, we collected the largest number of sequences for this region and we considered that this greater number would be more advantageous for our study. The age distribution of CRF01_AE is shifted more toward older populations compared to subtype B cases. However, as non-B subtype HIV-1-infected individuals tend to visit a clinic when they have not recently seroconverted (>155 days) [16], the estimated age at infection may be similar to that for subtype B HIV-infected cases. Unlike subtype B-infected cases, among which MSM is the major risk factor, more than half of the CRF01_AE-infections in Japan were estimated to be transmitted through heterosexual contact.

Using composite phylogenetic analysis, we determined at least 30 clusters of CRF01_AE-infected individuals in Japan; we noted 6 groups had international members, as they contained both our surveillance data and reference sequences from other Asian countries. Among 243 patient samples collected in the study, 76 were from patients involved in these micro-clades, demonstrating transmission networks of the CRF01_AE infection in Japan. The remaining 167 samples failed to be identified in transmission clusters; the CRF01_AE epidemic is quite large, and it is possible that sequences without identifiable partners is due to the low coverage rate of our surveillance or that they were newly imported from neighboring countries.

An original outbreak of CRF01_AE among the high-risk heterosexual population was first reported in Thailand in the late 1980s [4–6], then disseminated to neighboring countries [8–12]. Soon after the outbreak in Thailand, the first extensive colonization of the virus was estimated in Japan. Indeed, our phylodynamic analysis suggests that the primary domestic transmission cluster of CRF01_AE was formed in the 1990s. This possibility was confirmed from IMC-1, which includes 7 domestic sequences

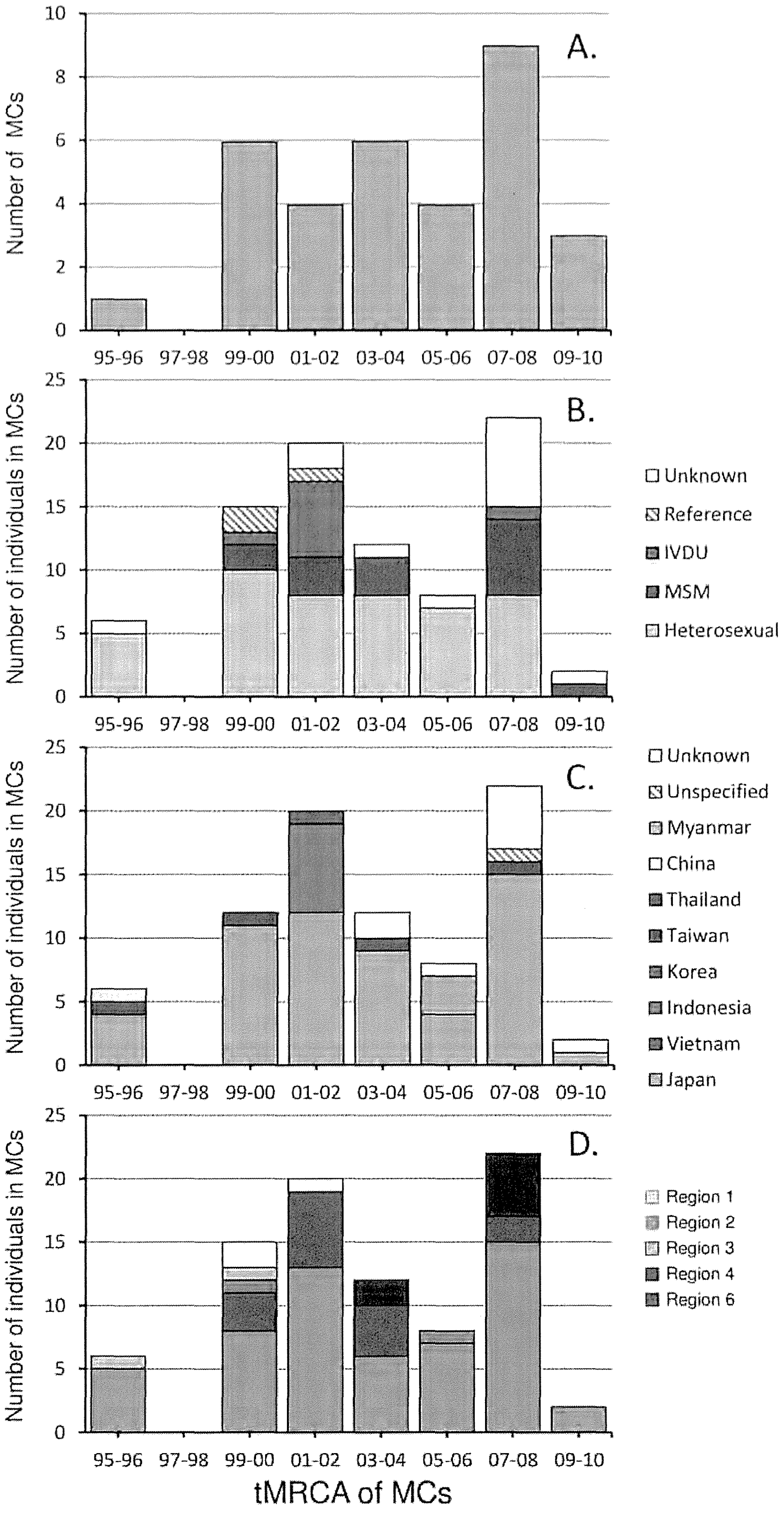

Figure 4. Dynamics of transmission clusters in Japan according to member individuals' demographic characteristics. A) Number of individuals in micro-clades is shown versus micro-clade tMRCA (time of the most recent common ancestor). Bars in each panel are colored by B) risk behavior, C) nationality, and D) geographical region. MC = micro-clade.
doi:10.1371/journal.pone.0102633.g004

Table 3. International micro-clades (IMCs) detected in the study population and outlier sequences.

	tMRCA				Characteristics of individuals in micro-clades				
IMC-ID	Median	95% HPD			Region	Age	Gender	Behavior	Nationality
IMC-1	Dec-'94	Mar-'93	–	Mar-'96	Region 7	26	F	Hetero	Japan
					Region 4	24	M	Hetero	Japan
					Region 4	62	M	Hetero	Japan
					Region 2	35	M	MSM	Japan
					Region 2	54	M	Hetero	Japan
					Region 4	43	M	Unknown	Japan
					Region 2	66	M	Hetero	Japan
					Vietnam	35 ref seq (see Table S1)			
					China	7 ref seq (see Table S1)			
					France	2 ref seq (see Table S1)			
					Czech	2 ref seq (see Table S1)			
IMC-2	Oct-'99	Dec-'96	–	Dec-'01	Region 4	32	F	Unknown	China
					China	27 ref seq (see Table S1)			
IMC-3	Feb-'00	Aug-'99	–	Dec-'01	MC33	2 individuals in Japan (see Table 2)			
					China	12 ref seq (see Table S1)			
IMC-4	ND	ND		ND	China	4 ref seq (see Table S1)			
IMC-5	Aug-'02	Jun-'99	–	Feb-'05	Region 2	30	M	MSM	Japan
					Philippines	3 ref seq (see Table S1)			
IMC-6	ND	ND		ND	China	23 ref seq (see Table S1)			
Sequence groups in CRF01_AE									
Japan AE	Nov-'68	Sep-'56	–	Apr-'75	-	-			
Whole AE	Nov-'68	Sep-'56	–	Apr-'75	-	-			

HPD: highest posterior density, tMRCA: time of most recent common ancestor, ND: not determined.
doi:10.1371/journal.pone.0102633.t003

and "cluster 3″ containing sequences from North Vietnam and a neighboring Chinese province described by Liao et al. [46], suggesting simultaneous disseminations to these countries, including Japan (Figure S5). Our tMRCA analysis demonstrated that some transmission clusters independently spread in East Asia from around 2000 (Table 3). Among these, IMC-1 was the same as the cluster found in IVDU populations in Northern Vietnam and Southeastern China [46,47]. The estimated median tMRCA of IMC-1 was similar to that of previous studies [46,47], and our subjects from regions 2 and 4 were scattered within the cluster, suggesting repeated transmission events of CRF01_AE from these regions to Japan around 2000. In turn, IMC-3 was the same as a cluster recently found in a northeastern Chinese MSM population (named "cluster 1″ [48–50] or CN.MSM-01-01 [51]) and included the domestic micro-clade, MC33.e constructed a maximum likelihood tree using our subjects with 4 CN.MSM-01-01 sequences derived from Japan with the same substitution model and confirmed that members of MC33 were involved in a Japanese sub-cluster of CN.MSM-01-01 (Figure S6). Therefore, one can observe that MC33-related viruses may have contributed an outbreak in an MSM community in metropolitan areas of Japan. Additionally, domestic clusters MC02 and MC03 included reference sequences originating from Taiwan and Vietnam, and South Korea, respectively (Figure 3). Thus our data clearly suggest that transmission networks of CRF01_AE have developed between Japan and other Asian countries from the first colonization wave. The CRF01_AE epidemic scenario would be that CRF01_AE was initially imported to Japan in the 1990s from neighboring Asian countries especially by IVDU behavior. Then, Japanese variants may have influenced epidemics among MSM in East Asian countries, as suggested by Kondo et al [51], by exporting cases in the middle of the 2000s. Among such "connected" countries in Asia, generation of recombinant forms between CRF01_AE and other subtypes has been a recent concern [7,51].

After the primary dissemination to Japan, our tMRCA distribution data strongly suggest that CRF01_AE invaded Japan in two substantial waves, one in 2000 and the other in 2007. Our finding is based on the ability to detect common ancestors among circulating viral variants. We could not assign common ancestors to all of our sequences, and the CRF01_AE epidemic is substantial and genetically diverse. As a result, our sampling, as well as sampling in other countries, may not be sufficient to detect additional, ongoing introductions of CRF01_AE. A serious concern drawn from our findings is the role of IVDU in the CRF01_AE transmission network. The concentration of IVDUs observed in large micro-clades (e.g., MC05) indicates a suspected linkage of high-risk sexual communities and drug addicts, including international partners (MC05) As stigma often limits standard risk assessment for HIV, the phylodynamic approach described here offers an effective method to track ongoing epidemics within suspected IVDU communities, with the results of our analysis indeed clarifying critical contributions of IVDU to the CRF01_AE outbreak in Japan. These results aid in calling attention to the need to focus resources on interventions designed to specifically limit spread among specific risk groups to curtail CRF01_AE transmissions through the IVDU route.

Supporting Information

Figure S1 Geographic location of HIV-1 sample collection regions in Japan. Regions of sample collection are designated by the same colors used to indicate sample origin in other figures.
(PDF)

Figure S2 Distance-based neighbor joining phylogeny of the protease-RT region of CRF01_AE HIV-1 in Japan. Numbers on each branch show the results of interior branch testing, where probabilities >95%. The sequences obtained in our surveillance network are designated by circles in different colors according to the region of sample collection. Reference sequences from the Los Alamos database are designated by black triangles. Micro-clades and significant clusters are annotated by red branches with brackets on the right of the tree. Scale bar at the bottom shows the number of nucleotide substitutions per site.
(PDF)

Figure S3 Distribution of sample collection time of CRF01_AE HIV-1-infected individuals in Japan. The cumulative numbers of CRF01_AE HIV-1-infected individuals are shown by year of sample collection. Bars in each panel are colored by individuals' A) gender and risk behavior, B) nationality, and C) geographical region.
(PDF)

Figure S4 Partial chronological phylogenetic tree of IMC-1. An international micro-clade including 7 sequences from our study population and a cluster that spread mainly in Vietnam [46] extracted from the Bayesian MCMC phylogeny is shown. The 7 sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Figure S5 Partial chronological phylogenetic tree of IMC-3. An international micro-clade composed of CRF01_AE sequences found in China extracted from the Bayesian MCMC phylogeny is shown. This cluster included MC15. Sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Figure S6 Maximum likelihood phylogenetic tree of the large Chinese cluster CN.MSM.01-01 with MC33. Protease-RT sequences belonging to CN.MSM.01-01 [51] were selected from the Los Alamos database and aligned with our study subjects and outlier sequences. Maximum likelihood phylogeny was inferred from the alignment as described in Materials and Methods. A partial tree including CN.MSM.01-01 is represented. Numbers below branches indicate bootstrap probability. Japanese sequences in CN.MSM.01-01 are underlined. Our sequences are designated by symbols according to their gender and the region of sample collection. JP = Japanese; UN = unknown.
(PDF)

Table S1 CRF01_AE outlier sequences from the Los Alamos HIV database.
(PDF)

Table S2 Bayesian factor analysis of molecular clock models compared for constant demographic size.
(PDF)

Table S3 Independence test of risk behaviors and nationalities in CRF01_AE-infected patients in Japan.
(PDF)

Table S4 Estimates of the mean evolutionary diversity for categories of CRF01_AE sequences.
(PDF)

Table S5 Evolutionary parameters obtained in Bayesian MCMC inference with constant size and lognormal relaxed.
(PDF)

Acknowledgments

We are grateful to all the patients who participated in our surveillance study. We appreciate the support of and helpful discussions with members of the Japanese Drug Resistance HIV-1 Surveillance Network: Atsushi Ajisawa, Hitoshi Chiba, Teruhisa Fujii, Yuko Fujikawa, Akira Fujita, Katsuyuki Fukutake, Hiroyuki Gatanaga, Tetsushi Goto, Shuji Hatakeyama, Igen Hongo, Masahide Horiba, Makiko Kondo, Makiko Kondo, Mitsuru Konishi, Shuzo Matsushita, Motoo Matsuura, Rumi Minami, Naoko Miyazaki, Haruyo Mori, Masaaki Noda, Chiho Otani, Takeyuki Sato, Satoshi Shirahata, Masashi Taki, Sadahiro Tamashima, Masap Tateyma, Masanori Tei, Kazue Uchida, Dai Watanabe, Yasuyuki Yamamoto, Kunio Yano, Mihoko Yotsumoto, Shigeru Yoshida.

We thank Dr. Frank Maldarelli for reading the manuscript critically. We also thank Claire Baldwin for her help in preparing the manuscript.

Author Contributions

Conceived and designed the experiments: TS JH WS. Performed the experiments: JH TS YI. Analyzed the data: TS JH. Contributed reagents/materials/analysis tools: YY YI WS. Wrote the paper: TS JH WS.

References

1. Infectious Diseases Surveillance Center (2011) HIV/AIDS in Japan, 2010. IASR 32: 282–283.
2. Yamanaka T, Fujimura Y, Ishimoto S, Yoshioka A, Konishi M, et al. (1997) Correlation of titer of antibody to principal neutralizing domain of HIVMN strain with disease progression in Japanese hemophiliacs seropositive for HIV type 1. AIDS Res Hum Retroviruses 13: 317–326.
3. Kato S, Saito R, Hiraishi Y, Kitamura N, Matsumoto T, et al. (2003) Differential prevalence of HIV type 1 subtype B and CRF01_AE among different sexual transmission groups in Tokyo, Japan, as revealed by subtype-specific PCR. AIDS Res Hum Retroviruses 19: 1057–1063.
4. Weniger BG, Takebe Y, Ou CY, Yamazaki S (1994) The molecular epidemiology of HIV in Asia. AIDS 8 Suppl 2: S13–S28.
5. Ou CY, Takebe Y, Luo CC, Kalish M, Auwanit W, et al. (1992) Wide distribution of two subtypes of HIV-1 in Thailand. AIDS Res Hum Retroviruses 8: 1471–1472.
6. Ou CY, Takebe Y, Weniger BG, Luo CC, Kalish ML, et al. (1993) Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet 341: 1171–1174.
7. Lau KA, Wang B, Saksena NK (2007) Emerging trends of HIV epidemiology in Asia. AIDS Rev 9: 218–229.
8. Beyrer C, Vancott TC, Peng NK, Artenstein A, Duriasamy G, et al. (1998) HIV type 1 subtypes in Malaysia, determined with serologic assays: 1992–1996. AIDS Res Hum Retroviruses 14: 1687–1691.
9. Kusagawa S, Sato H, Kato K, Nohtomi K, Shiino T, et al. (1999) HIV type 1 env subtype E in Cambodia. AIDS Res Hum Retroviruses 15: 91–94.
10. Menu E, Truong TX, Lafon ME, Nguyen TH, Müller-Trutwin MC, et al. (1996) HIV type 1 Thai subtype E is predominant in South Vietnam. AIDS Res Hum Retroviruses 12: 629–633.
11. Nerurkar VR, Nguyen HT, Woodward CL, Hoffmann PR, Dashwood WM, et al. (1997) Sequence and phylogenetic analyses of HIV-1 infection in Vietnam: subtype E in commercial sex workers (CSW) and injection drug users (IDU). Cell Mol Biol (Noisy-le-grand) 43: 959–968.
12. Porter KR, Mascola JR, Hupudio H, Ewing D, VanCott TC, et al. (1997) Genetic, antigenic and serologic characterization of human immunodeficiency virus type 1 from Indonesia. J Acquir Immune Defic Syndr Hum Retrovirol 14: 1–6.
13. Wang W, Jiang S, Li S, Yang K, Ma L, et al. (2008) Identification of subtype B, multiple circulating recombinant forms and unique recombinants of HIV type 1 in an MSM cohort in China. AIDS Res Hum Retroviruses 24: 1245–1254.
14. Wang W, Xu J, Jiang S, Yang K, Meng Z, et al. (2011) The dynamic face of HIV-1 subtypes among men who have sex with men in Beijing, China. Curr HIV Res 9: 136–139.
15. Zhang X, Li S, Li X, Li X, Xu J, et al. (2007) Characterization of HIV-1 subtypes and viral antiretroviral drug resistance in men who have sex with men in Beijing, China. AIDS 21 Suppl 8: S59–S65.
16. Hattori J, Shiino T, Gatanaga H, Yoshida S, Watanabe D, et al. (2010) Trends in transmitted drug-resistant HIV-1 and demographic characteristics of newly diagnosed patients: nationwide surveillance from 2003 to 2008 in Japan. Antiviral Res 88: 72–79.
17. Kihara M, Ichikawa S, Kihara M, Yamasaki S (1997) Descriptive epidemiology of HIV/AIDS in Japan, 1985–1994. J Acquir Immune Defic Syndr Hum Retrovirol 14 Suppl 2:S3–12.
18. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303: 327–332.
19. Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. Nat Rev Genet 8: 196–205.
20. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC (2007) Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. PLoS Pathog 3: 1220–1228.
21. Shiino T, Okabe N, Yasui Y, Sunagawa T, Ujike M, et al. (2010) Molecular evolutionary analysis of the influenza A(H1N1)pdm, May-September, 2009: temporal and spatial spreading profile of the viruses in Japan. PLoS One 5: e11057.
22. Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, et al. (2009) The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Med 6: e1000198.
23. Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A (2003) The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol 20: 381–387.
24. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, et al. (2009) Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS Pathog 5: e1000590.
25. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, et al. (2011) Transmission Network Parameters Estimated From HIV Sequences for a Nationwide Epidemic. J Infect Dis 204: 1463–1469.
26. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ (2008) Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med 5: e50.
27. Gatanaga H, Ibe S, Matsuda M, Yoshida S, Asagi T, et al. (2007) Drug-resistant HIV-1 prevalence in patients newly diagnosed with HIV/AIDS in Japan. Antiviral Res 75: 75–82.
28. UNGASS (2012) National Commitments and Policies Instrument (NCPI). Available: http://www.unaids.org/en/dataanalysis/knowyourresponse/ncpi/2012countries/Japan NCPI 2012.pdf. Accessed 20 June 2012.
29. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
30. Lloyd AL, May RM (2001) Epidemiology. How viruses spread among computers and people. Science 292: 1316–1317.
31. Romano CM, de Carvalho-Mello IM, Jamal LF, de Melo FL, Iamarino A, et al. (2010) Social networks shape the transmission dynamics of hepatitis C virus. PLoS One 5: e11170.
32. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.
33. R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available: http://www.r-project.org/.
34. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101: 11030–11035.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
36. Swofford DL (2003) PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
37. Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Uppsala University: Evolutionary Biology Centre.
38. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
39. Suchard MA (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013.
40. Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P (2013) Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol 30: 239–243.
41. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst Biol 60: 150–160.
42. Prosperi MC, Ciccozzi M, Fanti I, Saladini F, Pecorari M, et al. (2011) A novel methodology for large-scale phylogeny partition. Nat Commun 2: 321.
43. Drummond A, Forsberg R, Rodrigo AG (2001) The inference of stepwise changes in substitution rates using serial sequence samples. Mol Biol Evol 18: 1365–1371.
44. Drummond A, Rodrigo AG (2000) Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. Mol Biol Evol 17: 1807–1815.
45. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 73: 10489–10502.

46. Liao H, Tee KK, Hase S, Uenishi R, Li XJ, et al. (2009) Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. Virology 391: 51–56.

47. Li L, Liang S, Chen L, Liu W, Li H, et al. (2010) Genetic characterization of 13 subtype CRF01_AE near full-length genomes in Guangxi, China. AIDS Res Hum Retroviruses 26: 699–704.

48. An M, Han X, Xu J, Chu Z, Jia M, et al. (2012) Reconstituting the epidemic history of HIV strain CRF01_AE among men who have sex with men (MSM) in Liaoning, northeastern China: implications for the expanding epidemic among MSM in China. J Virol 86: 12402–12406.

49. Feng Y, He X, Hsi JH, Li F, Li X, et al. (2013) The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. AIDS 27: 1793–1802.

50. Ye J, Xin R, Yu S, Bai L, Wang W, et al. (2013) Phylogenetic and temporal dynamics of human immunodeficiency virus type 1 CRF01_AE in China. PLoS One 8: e54238.

51. Kondo M, Lemey P, Sano T, Itoda I, Yoshimura Y, et al. (2013) Emergence in Japan of an HIV-1 variant associated with transmission among men who have sex with men (MSM) in China: first indication of the International Dissemination of the Chinese MSM lineage. J Virol 87: 5351–5361.

Mitra *et al. Retrovirology* (2015) 12:3
DOI 10.1186/s12977-014-0130-8

RETROVIROLOGY

RESEARCH Open Access

Sequence and structural determinants of human APOBEC3H deaminase and anti-HIV-1 activities

Mithun Mitra[1,6], Dustin Singer[1], Yu Mano[2], Jozef Hritz[3,4,7], Gabriel Nam[1], Robert J Gorelick[5], In-Ja L Byeon[3,4], Angela M Gronenborn[3,4], Yasumasa Iwatani[2] and Judith G Levin[1*]

Abstract

Background: Human APOBEC3H (A3H) belongs to the A3 family of host restriction factors, which are cytidine deaminases that catalyze conversion of deoxycytidine to deoxyuridine in single-stranded DNA. A3 proteins contain either one (A3A, A3C, A3H) or two (A3B, A3D, A3F, A3G) Zn-binding domains. A3H has seven haplotypes (I-VII) that exhibit diverse biological phenotypes and geographical distribution in the human population. Its single Zn-coordinating deaminase domain belongs to a phylogenetic cluster (Z3) that is different from the Z1- and Z2-type domains in other human A3 proteins. A3H HapII, unlike A3A or A3C, has potent activity against HIV-1. Here, we sought to identify the determinants of A3H HapII deaminase and antiviral activities, using site-directed sequence- and structure-guided mutagenesis together with cell-based, biochemical, and HIV-1 infectivity assays.

Results: We have constructed a homology model of A3H HapII, which is similar to the known structures of other A3 proteins. The model revealed a large cluster of basic residues (not present in A3A or A3C) that are likely to be involved in nucleic acid binding. Indeed, RNase A pretreatment of 293T cell lysates expressing A3H was shown to be required for detection of deaminase activity, indicating that interaction with cellular RNAs inhibits A3H catalytic function. Similar observations have been made with A3G. Analysis of A3H deaminase substrate specificity demonstrated that a 5′ T adjacent to the catalytic C is preferred. Changing the putative nucleic acid binding residues identified by the model resulted in reduction or abrogation of enzymatic activity, while substituting Z3-specific residues in A3H to the corresponding residues in other A3 proteins did not affect enzyme function. As shown for A3G and A3F, some A3H mutants were defective in catalysis, but retained antiviral activity against HIV-1*vif* (–) virions. Furthermore, endogenous reverse transcription assays demonstrated that the E56A catalytic mutant inhibits HIV-1 DNA synthesis, although not as efficiently as wild type.

Conclusions: The molecular and biological activities of A3H are more similar to those of the double-domain A3 proteins than to those of A3A or A3C. Importantly, A3H appears to use both deaminase-dependent and -independent mechanisms to target reverse transcription and restrict HIV-1 replication.

Keywords: HIV-1, APOBEC3H, Homology model, Deaminase activity, Antiviral activity, Deaminase-independent restriction, Reverse transcription

Background

The human APOBEC3 (A3) family consists of seven cytidine deaminases that catalyze the conversion of deoxycytidine (dC) to deoxyuridine (dU) in single-stranded (ss) DNA, thereby inducing G-to-A hypermutation in double-stranded DNA [1-5]. A3 proteins play an important role in the innate immune defense system by inhibiting a broad range of exogenous viruses such as human immunodeficiency virus type 1 (HIV-1) (reviewed in refs. [6-13]), human T-lymphotropic virus type 1 (HTLV-1) [14,15], and hepatitis B virus (HBV) [16,17] as well as endogenous retrotransposons such as LINE-1 and Alu elements (reviewed in refs. [7,18]). These proteins contain either one (A3A, A3C, and A3H) or two (A3B, A3D (formerly known as A3D/E), A3F, and A3G) Zn-binding domains with the conserved motif $HX_1EX_{23-24}CX_{2-4}C$ (X is any amino acid) [19] (reviewed in refs.

* Correspondence: levinju@mail.nih.gov
[1]Section on Viral Gene Regulation, Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892-2780, USA
Full list of author information is available at the end of the article

[20,21]). The histidine and two cysteines coordinate a Zn ion, while the glutamic acid residue is thought to act as a proton shuttle during catalysis [6,22]. Based on phylogenetic analysis, the Zn-binding domains were further classified into the following groups: Z1 (A3A and C-terminal domains (CTD) of A3B and A3G), Z2 (A3C, N-terminal domains (NTD) of A3B and A3G, and both NTD and CTD of A3D and A3F), and Z3 (A3H) [23,24].

A3H is the most divergent member of the A3 family and has a single Zn-binding domain that belongs to the unique Z3 group [24,25]. The A3H message undergoes alternative splicing to generate variants containing distinct C-terminal regions [26,27]. Furthermore, unlike other *A3* genes, *A3H* is present in the human population as different haplotypes containing functional polymorphisms. At present, seven haplotypes of A3H (Hap I-VII) have been identified that differ in their antiviral activities: only Hap II, Hap V, and Hap VII are stably expressed and are able to restrict Vif-deficient HIV-1 [26-29]. Interestingly, the distribution of A3H haplotypes in the human population is correlated with geographical location [26,29]. For example, a higher frequency of HapII is present in Africa, compared to Europe and Asia, possibly due to a greater selection pressure against pathogens endogenous to that region [26].

HIV-1 Vif, which counteracts antiviral activity by promoting proteasomal degradation of A3C, A3D, A3F, and A3G, exhibits different degrees of potency against the individual A3H haplotypes [12,29-32]. In cell-based assays, the sensitivity of antiviral A3H HapII towards Vif was shown to be dependent upon the Vif subtype [33,34] and a remarkable study involving recently infected HIV-1 patients revealed adaptive changes in viral Vif sequences that were attributed to the presence of the different antiviral A3H haplotypes [32]. These observations provide strong evidence for a significant role of A3H as an antiviral defense protein.

A3 proteins deaminate dC residues in a sequence-specific manner. For example, A3A exhibits a greater preference for the dC in the center of a T**C**A target [35-42], while A3G specifically deaminates the dC in a C**C**C motif [21,43-45]. Evaluation of the structural basis of the sequence specificity suggested that it is determined by the architecture of the active site and surrounding amino acids, in particular, residues in loop 7 [41-46] (reviewed in refs. [21,47]). Although, A3H is known to deaminate dC in T**C** motifs of HIV-1 minus-strand DNA [27,48], a detailed investigation of the nucleotide context immediately 5′ and 3′ of the dC on sequence-specific deamination has not been reported. In addition, the amino acid residues that are important for A3H deaminase activity and the role of structure in dictating biological function have not been investigated.

In the present study, we focus on the biochemical and structural determinants of A3H HapII (to be referred to as "A3H") deaminase and antiviral activities, using site-directed and structure-guided mutagenesis. We have constructed a homology model of A3H and find that the A3H structure, as expected, is similar to the known structures of A3A [41], A3C [49], A3G-CTD [43,50-53], and A3F-CTD [54,55], with differences mainly in flexible loop regions. Our model resembles the ones generated by (i) MODELLER [56], based on the A2 and A3G-CTD structures [57], and (ii) the automated structure-homology-modeling server, SWISS-MODEL, using the A2 structure [13], although details may be different. Interestingly, our model also reveals a large cluster of basic residues, which is not present in other A3 deaminase-active domains, and is consistent with the observation that deaminase activity in cell-free extracts is detected only after removal of RNA by treatment with RNase A. In addition, we have evaluated the deaminase and antiviral activities of a series of A3H mutants. Although these activities can be correlated in most cases, a significant number of mutants lacking enzymatic activity are still able to inhibit HIV-1 replication, albeit at a lower efficiency than wild type (WT). This result raises the possibility that A3H restricts HIV-1 by catalytic-dependent and -independent mechanisms. Indeed, assays of endogenous reverse transcription (ERT) support this hypothesis. Taken together, our findings provide new insights into the role of A3H as a naturally occurring human restriction factor and should contribute to continuing efforts to combat HIV infection in the African human population.

Results

Sequence- and structure-based design of A3H mutants

In this work, we set out to investigate the determinants of A3H cytidine deaminase and antiviral activities, using a mutagenic approach. Given A3H's unique Z3-type Zn-binding domain [24], we initially carried out a sequence comparison of the Z3 domain of A3H with the Z1 and Z2 domains of other A3 proteins to identify conserved and distinct regions in A3H (Additional file 1: Figure S1). Sequence identities range from 28-43% and the Z3 domain shares the greatest identity with the Z2 domains of A3C and A3F-CTD and the least with A3D-NTD. The sequence alignment also identified four residues unique to the Z3 domain: T81, L102, S109, and V135, which are replaced by S, V, A, and I in other Z domains.

A more extensive sequence alignment was performed by comparing the residues in the complete A3H protein with the sequences of A3 proteins whose three-dimensional structures have been solved at high resolution, i.e., A3A [41], A3C [49], A3F-CTD [54,55] and A3G-CTD [43,50-53] (Figure 1). The overall sequence identity between A3H and each of these proteins is very similar and

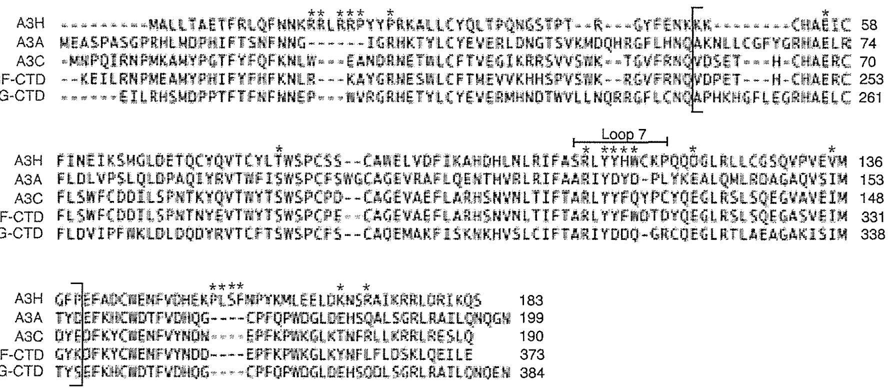

Figure 1 Sequence alignment of A3H HapII (residues 1 to 183) and four A3 deaminase-active proteins whose structures have been solved: A3A (1 to 199) [41], A3C (1 to 190) [49], A3F-CTD (185–373) [54,55], and A3G-CTD (191–384) [53] (also see refs. [43,50-52]). The A3H residues mutated in this study are highlighted with green asterisks and the active site residues are highlighted in blue. The region inside the square brackets represents the Z domain sequences for all of the proteins. The amino acids that comprise A3H loop 7 are also shown. The sequence alignment was generated using Lasergene software (DNASTAR, Inc., Madison, WI, USA).

ranges from 35 to 38% (Table 1). The Z domains (bracketed) include the loop 7 residues, a region that is involved in deaminase substrate specificity [41-46] (reviewed in [21,47]). Interestingly, loop 7 sequences for different A3 proteins display alternative arrangements of polar and non-polar residues. For example, A3H contains a stretch of aromatic residues (YYHW, 112–115), while in A3G-CTD, the corresponding residues are polar (YDDQ, 315–318). In addition, a unique stretch of residues was noted outside the Z domains, namely 154–157 (PLSF), which is absent in the other A3 proteins.

To probe the role of A3H residues in enzymatic activity in relation to their location in the structure, we constructed a homology model based on the X-ray structure of A3G-CTD [53] (Figure 2A). Comparison of our A3H model with A3G-CTD (PDB: 3IR2) [53], A3A (PDB: 2 M65) [41], A3C (PDB: 3 VOW) [49], and A3F-CTD (PDB: 4IOU) [54] showed that the A3H model is similar to the other A3 structures, which have r.m.s.d. values ranging from 3.1 to 3.5 Å. The major differences between the various A3 structures occur in the loops, as suggested previously [41] (Additional file 2: Figure S2).

Table 1 Sequence comparison of A3H with other deaminase-active A3 proteins and theoretical isoelectric points (pI)

	Zinc-binding domain (Z) type	% Sequence identity to A3H[a]	Theoretical pI[b]
A3A (1–199)	Z1	36	6.3
A3C (1–190)	Z2	38	7.5
A3F-CD2 (185–373)	Z2	36	5.0
A3G-CD2 (191–384)	Z1	35	6.2
A3H HapII (1–183)	Z3	100	8.9

[a] Sequence identity analysis was performed using Lasergene software (DNASTAR, Inc.).
[b] The theoretical pI values were calculated using the Protparam online web-based program (http://web.expasy.org/compute_pi/).

Comparisons of the electrostatic surface features of A3H (Figure 2B, Additional file 3: Figure S3) and other A3 proteins (A3G-CTD, A3A, A3C, and A3F-CTD) (Additional file 3: Figure S3) reveal a striking difference in the clustering of basic residues (blue regions), consistent with the fact that A3H is highly basic (theoretical pI, 8.9) compared to the other single-domain A3 proteins A3A (pI, 6.3) and A3C (pI, 7.5) (Table 1). Indeed, the basic character of A3H is more similar to the positively charged N-terminal domains (NTDs) of the double-domain proteins, A3F (residues 1–180; pI, 8.6) and A3G (residues 1–185; pI, 9.4). The NTDs of A3F and A3G play an important role in binding viral RNA that is packaged; however, they are enzymatically inactive [58-61]. Since A3H is a single-domain protein, it is likely that the basic residues serve a dual function, i.e., binding both viral and/or cellular RNA as well as interacting with the ssDNA substrate for deamination.

RNA inhibits the deaminase activity of A3H

Deaminase activity of A3H variants was tested in 293T cell lysates using a TTCA-containing 40-nt oligonucleotide as described in Methods; protein expression was monitored by Western blot analysis, using an A3H-specific antibody probe (Figure 3A). We selected the TTCA motif, since analysis of HIV-1 viral DNA showed that the dC residue in TC motifs was preferentially deaminated by A3H [27,32,48,62]. Surprisingly, we observed only trace amounts of deaminase activity in (WT) A3H cell lysates (Figure 3B and C). However, activity was clearly detectable after treating these lysates with RNase A, while no effect was seen with cells transfected with the empty vector control. These results suggest that A3H is

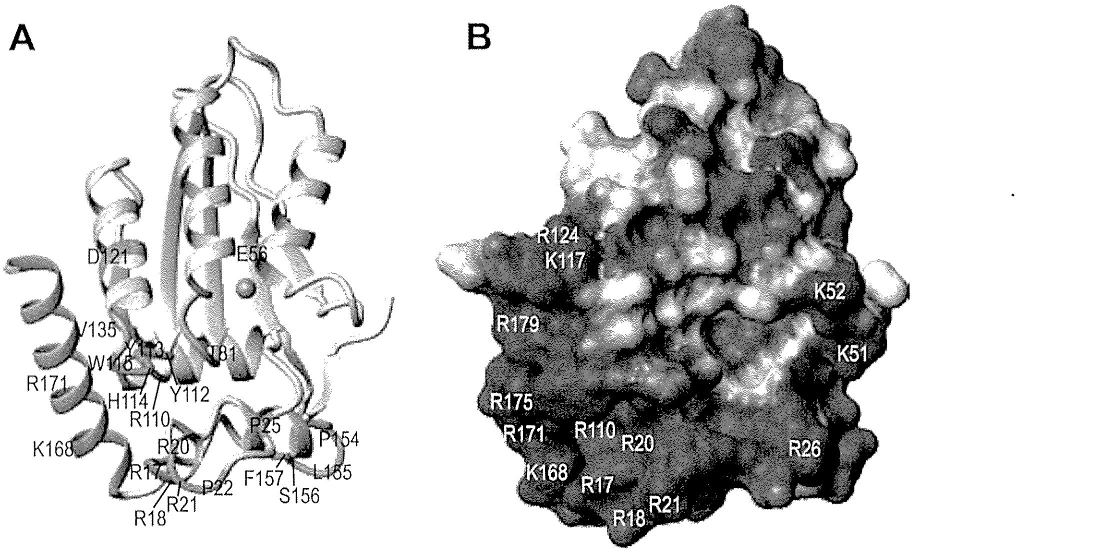

Figure 2 A3H model structure. (A) Ribbon representation of the A3H model showing the positions of residues mutated in this study in green. The Zn ion is colored in brown. **(B)** Electrostatic map of the A3H model depicting regions with positive (blue) and negative (red) electrostatic potentials. Basic residues are shown in white.

bound to RNA present inside the cell and that this interaction with RNA inhibits A3H deaminase activity, presumably in a competitive manner. Interestingly, a similar inhibitory effect on enzyme function was also observed with A3G [7,63].

Deaminase target specificity of A3H

Until now, detailed analysis of A3H deaminase specificity has not been reported. We therefore performed deaminase assays using oligonucleotides containing different motifs: TTCA, TTCT, TTCG, TGCA, and ACCCA, where either a purine (G) or a pyrimidine (T or C) is present at the position immediately 5' of the target dC residue. As shown in Figure 3D, the presence of a 5′ T (e.g., TTCA, TTCT, and TTCG) yielded the highest amounts of deamination product (~90% substrate conversion with 5 μg of total protein after incubation for 1 h at 37°C), while a 5′ G, as in TGCA, was poorly tolerated (~17%). The rank order of deamination efficiency of the substrates is: TTCA ~ TTCT ~ TTCG > ACCCA > TGCA. These data suggest a less stringent requirement for the 3′ position, where either a purine or a pyrimidine is tolerated, since similar levels of deamination product were formed with substrates containing TTCA, TTCT, and TTCG motifs. (Note that some difference in the deamination efficiency of substrates with TTCA, TTCT, and TTCG motifs might occur if less than 1 μg of lysate were added to the reaction.)

Identification of A3H residues important for deaminase activity

The sequence-structure analysis (Figures 1 and 2) provided a rationale for mutagenesis of A3H residues that could potentially play a role in catalytic activity. Deaminase and HIV-1 infectivity assays of WT and mutant constructs were performed in parallel to evaluate enzymatic (Figure 4) and antiviral (Figure 5) activities. The data in Figure 4A and D are arranged according to residue position along the polypeptide chain from the N- to C-terminus. As expected, the active site mutant E56A (negative control) did not display any deaminase activity (Figure 4B).

In our initial screen, we focused on residues that are Z3-specific (Additional file 1: Figure S1), in order to probe whether these residues were selected throughout evolution to ensure A3H enzymatic activity. We chose two examples from this group, T81 and V135, which correspond to S and I, respectively, in Z1 and Z2 Zn-binding domains (Figure 1). Mutants with these Z1-Z2 substitutions were expressed at levels comparable to or higher than WT (Figure 4A and D). Interestingly, changing T81 to S did not reduce deaminase activity, whereas the T81 to A mutation completely abolished activity (Figure 4B). Note that the mutant protein T81A was expressed at levels comparable to WT (Figure 4A). Since both threonine and serine possess a hydroxyl group, while alanine does not, this may suggest that the side chain of T81 is involved in a polar interaction. Changing the analogous residue S97 in A3A (Z1 type) to T did not affect deaminase activity (Figure 4B), consistent with retention of the important hydroxyl group. Similarly, mutating V135 to I also represents a conservative change and not surprisingly, there was no effect on enzymatic activity (Figure 4E).

In contrast, deletion of the unique stretch of amino acids, PLSF (aa 154–157), which contains three bulky hydrophobic residues, proline, leucine, and phenylalanine,

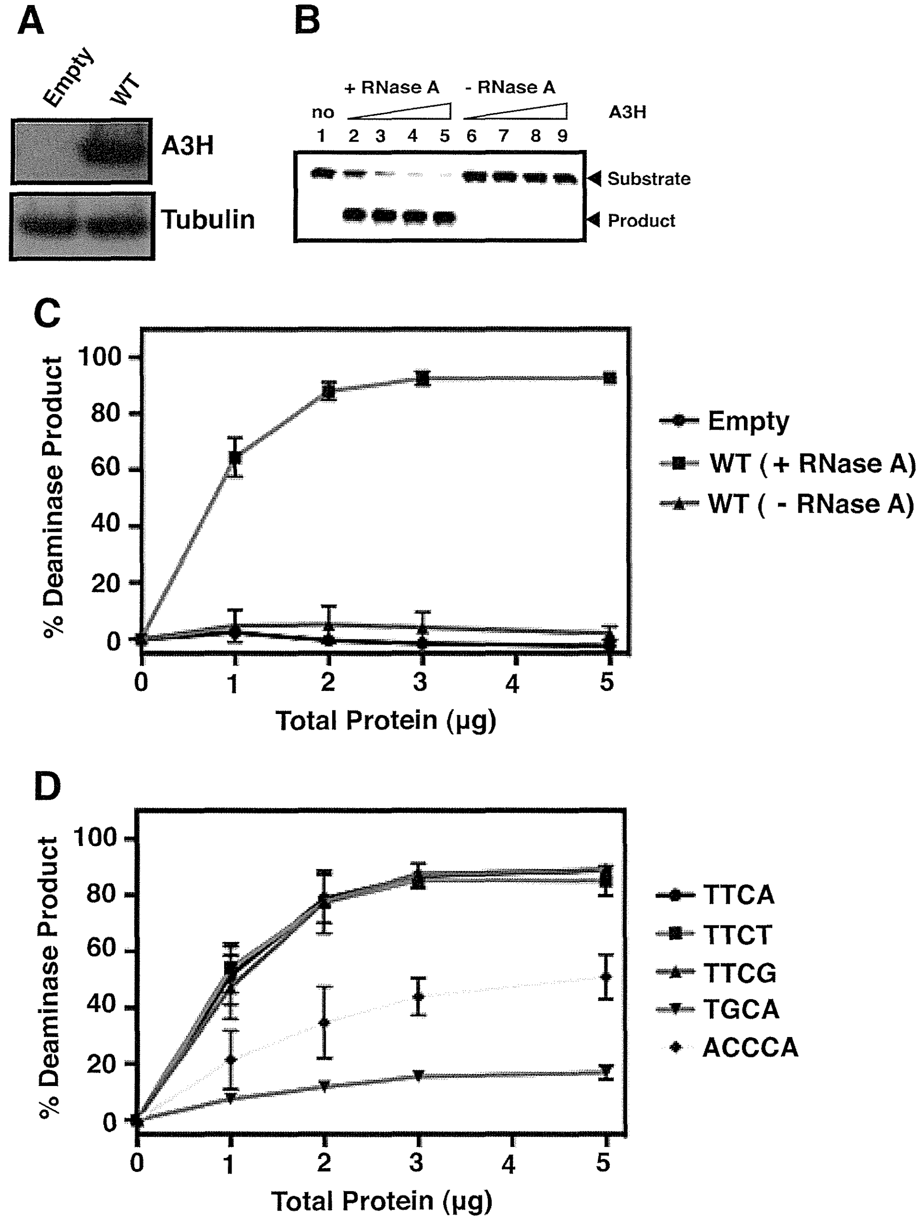

Figure 3 A3H deaminase activity dependence on RNase A treatment of extracts and substrate specificity. (A) Western blot analysis showing the WT A3H protein levels in 293T cells. Transfection of the empty vector (no A3H) served as a negative control and showed that 293T cells do not contain detectable levels of endogenous A3H. The tubulin loading control is also shown. **(B)** Representative gel illustrating assay of WT A3H deaminase activity in a cell extract using a 40-nt TTCA-containing oligonucleotide substrate. The oligonucleotide was incubated with increasing amounts of A3H extract in the presence and absence of RNase A. The positions of the substrate (40 nt) and the deamination product are indicated by arrows to the right of the gel. Lane 1, empty vector control; lanes 2 and 6, lanes 3 and 7, lanes 4 and 8, and lanes 5 and 9 represent reactions containing 1 μg, 2 μg, 3 μg, and 5 μg of total protein, respectively. **(C)** The percent (%) deamination product was calculated as described in Methods and was plotted against the amounts of total protein. **(D)** Deaminase assay using WT A3H extract and 40-nt oligonucelotides containing the following deaminase motifs: TTCA, TTCT, TTCG, TGCA, and ACCCA. The data were analyzed and plotted as described in **(C)**.

reduced enzymatic activity to background level (Figure 4E). This suggests that a major conformational change was introduced by the deletion, causing an overall structural defect. In turn, this would lead to destabilization of the protein, which could explain the low protein expression level of this mutant (Figure 4D) and inability to be packaged efficiently (see below) (Figure 5A). Changing another unique A3H residue, P25 to R, abrogated A3H deaminase activity (Figure 4B), although its expression level was normal (Figure 4A).

Basic residues in A3 proteins are often involved in specific and non-specific nucleic acid interactions [21,47,64].

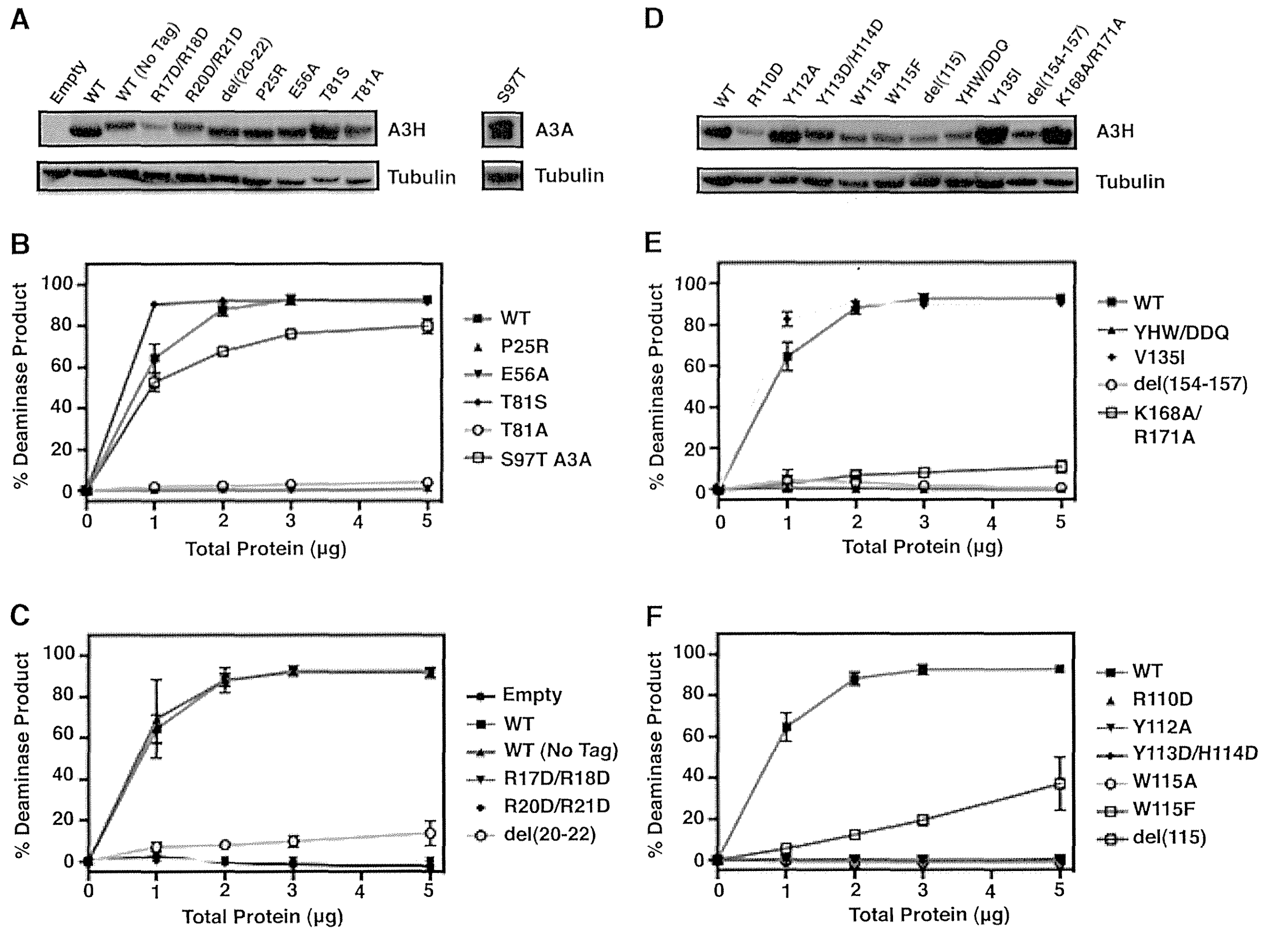

Figure 4 Deaminase activities of A3H WT and mutants. (A) and (D) Western blot analysis showing the amounts of A3H WT and mutants and A3A S97T in 293T cells. **(B, C, E, F)** Deaminase assays using 293T cell extracts expressing A3H WT and mutants and A3A S97T. The deaminase assays were performed as described in Methods.

To examine the role of these residues in A3H deaminase activity (Figure 4C and F) and to determine whether the location of these residues in the A3H model structure is related to their function, we focused on residues in the "basic patch" (Figure 2B): R17, R18, R20, R21, R110, K168, and R171. Note that R18, R20, and R21 are not present in other A3 proteins (Figure 1). We constructed a single mutant with an R → D change (R110D) as well as double mutants R17D/R18D, R20D/R21D, and K168A/R171A. With the exception of R17D/R18D and R110D, the mutant proteins were efficiently expressed (Figure 4A and D), but almost all lacked deaminase activity (Figure 4C, 4E, and F), even at high amounts of total protein. Two mutants, del (20–22) (Figure 4C) and K168A/R171A (Figure 4E), displayed greatly reduced, but measurable activity (~15% and ~11% product, respectively, at 5 µg total protein). These results suggest that positive charges are necessary for binding of A3H to the ssDNA substrates and that the introduction of a single negative charge in the basic patch disrupts the favorable charge-charge interaction. The change to the non-polar alanine or deletion of residues 20–22, which are present only in A3H as a unique insertion, is less detrimental. Taking all of the above data together, it appears likely that these basic residues form part of the A3H nucleic acid binding interface.

Another region of interest, loop 7 (Figure 1), which in other A3 proteins has been shown to be important for substrate binding and recognition [41-46] (reviewed in refs. [21,47]) was also subjected to mutagenesis. Several of the mutant proteins e.g., W115A, W115F, del(115), and Y113D/H114D/W115Q (YHW/DDQ) were expressed at low levels compared to WT A3H (Figure 4D), possibly resulting from reduced protein stability due to removal of the large tryptophan side chain. Changing residues Y112, Y113, and H114, e.g., Y112A, Y113D/H114D (Figure 4F), and YHW/DDQ, which introduces polar residues from the A3G-CTD loop 7 (Figure 4E), led to the complete loss of deaminase activity. A similar result was obtained when W115 was deleted (del115) or changed to

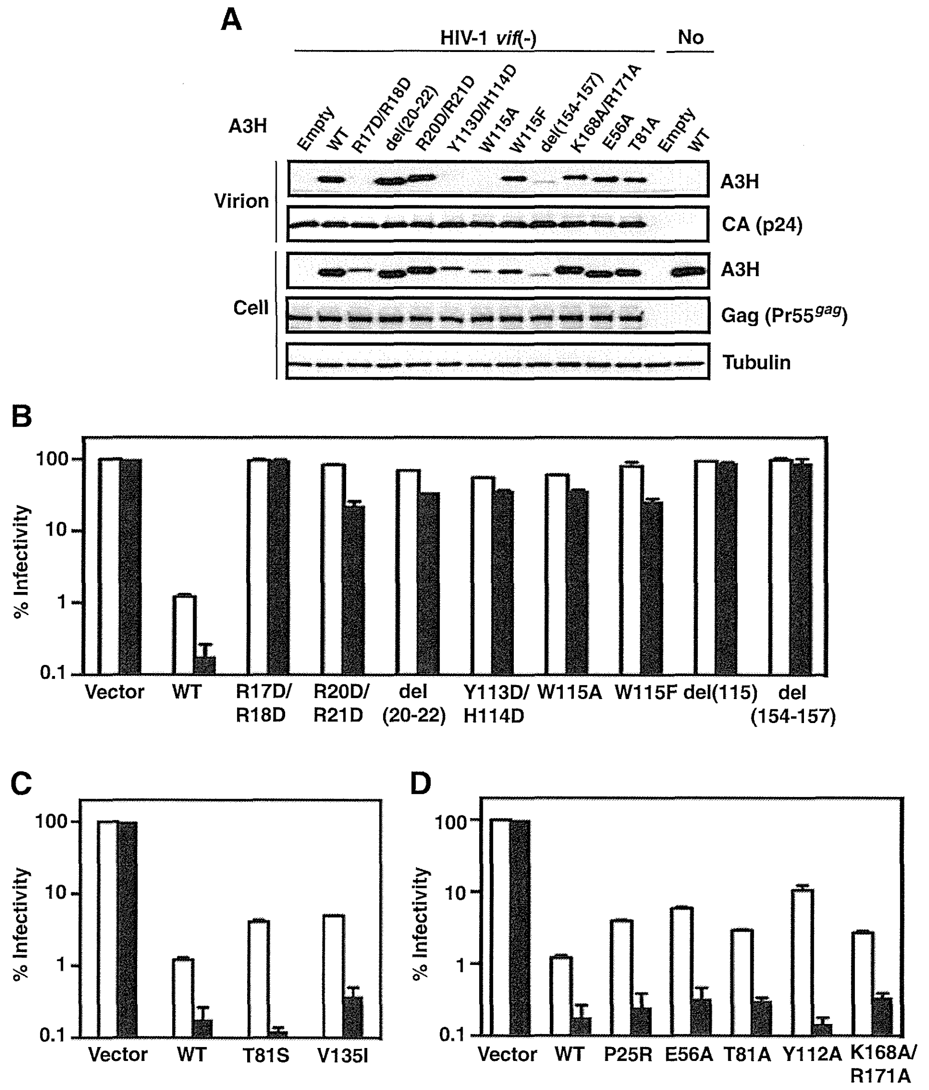

Figure 5 Antiviral activities of A3H WT and mutants. (A) Western blot analysis showing the amounts of A3H WT and mutants in HIV-1*vif* (–) virions and 293T cells. 293T cells were transfected with HIV-1*vif* (–) and A3H WT or mutant plasmids at a 1:1 ratio (1 μg each for determination of expression in cell extracts and 8 μg each for determination of A3H in viral lysates). Cell extracts (10 μg of total protein) as well as viral lysates (8 μl of viral pellet resuspended in 200 μl of loading buffer) were subjected to Western blot analysis. Viral lysates were probed with antibodies to the N-terminal FLAG tag of A3H and HIV-1 CA; cell extracts were probed with antibodies to the N-terminal FLAG tag of A3H, HIV-1 Gag (Pr55^{gag}), and tubulin. Controls: Left side, EMPTY refers to HIV-1*vif* (–) and empty vector (pTR600); Right side, EMPTY refers to empty vector alone; WT, refers to A3H plasmid DNA alone. **(B-D)** Antiviral activity was determined as described in Methods. The gray and black bars respresent transfection with 0.1 μg or 1 μg of the indicated A3H plasmid, respectively.

alanine (W115A) (Figure 4F). However, changing tryptophan to another aromatic residue, phenylalanine (W115F) (Figure 4F), led to only partial loss of deaminase activity, suggesting that these aromatic residues could be involved in base stacking interactions with the nucleic acids. Finally, we tested the deaminase activity of W115A, del115, Y113D/H114D, YHW/DDQ, and R110D at 10 μg total protein, but again, no activity was observed (data not shown). Collectively, these results suggest that the residues in A3H loop 7 are also likely to participate in specific interactions with nucleic acids.

Role of A3H residues in A3H antiviral activity

Inhibition of HIV-1*vif* (–) replication by A3F and A3G is mediated by both deaminase-dependent and -independent mechanisms [20,61,65-79]. In an early study, it was concluded that A3H antiviral activity is deaminase-independent [80], but other reports indicated that this activity is dependent on catalysis [25,27]. It was therefore of interest to evaluate whether the presence or absence of deaminase activity (Figure 4) could be correlated with A3H antiviral activity (Figure 5) in our system.

To determine whether the inability of certain mutants to restrict HIV-1 replication was due to a defect in A3H packaging, Western blot analysis was performed (Figure 5A). Virions were probed for capsid protein (CA) as well as for A3H. For comparison, expression levels of A3H and tubulin in cells were also measured and the data were similar to the results in Figure 4A and D. Interestingly, several mutants that were expressed poorly in cells, packaged little or no A3H in virions (Figure 5A). These mutants include: R17D/R18D, Y113D/H114D, W115A, and del(154–157). Although W115F exhibited lower levels of protein than WT (Figures 4D and Figure 5A), a significant amount of A3H was encapsulated (Figure 5A).

Single-cycle infectivity assays of virions produced in cells expressing WT and mutant A3H proteins were performed using two different amounts of A3H plasmid (0.1 μg and 1 μg). Under conditions where the 0.1 μg dose was used, the antiviral activities of the mutants could be divided into three groups: (1) mutants that showed little or no deaminase or antiviral activities (i.e., having values similar to the empty vector control), such as R17D/R18D, R20D/R21D, del(20–22) (low level of deaminase activity), Y113D/H114D, W115A, W115F (reduced level of deaminase activity), del(115), and del (154–157) (Figure 5B); (2) mutants with WT levels of deaminase activity and appreciable antiviral activity, albeit lower than that of WT, such as T81S and V135I (Figure 5C); and (3) mutants completely lacking (P25R, E56A, T81A, Y112A) or having reduced levels (K168A/R171A) of deaminase activity that retain antiviral activity (Figure 5D).

The results obtained with mutants in groups 1 and 2 are consistent with deaminase-dependent antiviral activity, since group 1 mutants have neither activity and group 2 mutants have both. With the exception of R20D/R21D, del (20–22), and W115F, all of the group 1 mutants that were analyzed by Western blot exhibited packaging defects (see above), which would account for their lack of virion-associated deaminase and anti-HIV activities. Interestingly, the results with group 3 mutants were discordant and suggest that A3H may also utilize a deaminase-independent mechanism for HIV-1 restriction. Note that the levels of antiviral activity for these deaminase-negative mutants were still lower than WT values (0.1 μg condition), suggesting that deaminase activity is indeed required for maximal activity.

The antiviral activities of WT and a majority of the mutants increased upon increasing the transfected plasmid amount to 1 μg. Surprisingly, the antiviral activities of T81S and Y112A were similar to WT levels under this condition. The behavior of Y112A in our study differed from that of A3H Hap VII Y112A, which although expressed efficiently in cells, was poorly packaged and exhibited a very low level of anti-HIV-1 activity [29]. The explanation for this difference is not clear. We also performed a side-by-side comparison of the activities of the deaminase-negative catalytic mutants of A3G (E259Q) and A3H (E56A) as well as the respective WTs (Additional file 4: Figure S4) and found that both A3H and A3G WT and mutant samples displayed dose-dependent inhibition of HIV-1 infectivity. This suggests that A3G and A3H utilize a common mechanism for antiviral activity.

Mechanism of A3H antiviral activity

A3G and A3F deaminase-independent anti-viral activity targets nascent DNA synthesis during reverse transcription [20,68-79]. To determine whether A3H deaminase-independent inhibition of HIV-1 infectivity is also associated with a reduction in viral DNA synthesis, we performed ERT assays using WT A3H and the active site mutant E56A.

In our assays, HIV-1*vif* (–) and A3H plasmids were transfected at two different ratios: 10:1 or 3:1, respectively (see Methods). Synthesis of R-U5 DNA (minus-strand strong-stop DNA) and R-5'UTR DNA (plus-strand DNA synthesized after plus-strand transfer) was measured over a 4-h time interval (Figure 6A and B). With WT A3H using the 10:1 condition, the levels of R-U5 were decreased to about 40% of the minus A3H control (100%) at 2 h (Figure 6A). The levels were drastically reduced for WT (3:1 condition) (~15% at 2 h) and the time course showed no appreciable change in level over the 4-h window. The E56A mutant, which lacks deaminase activity, was also capable of reducing the R-U5 levels to about 60% at 2 h (10:1 condition). At the higher dose (3:1 condition), R-U5 levels were further reduced relative to the control (~25% at 2 h), but the inhibition did not saturate even after 4 h, indicating partial inhibition. Similar trends for the WT and E56A mutant were also observed when synthesis of R-5'UTR (plus-strand synthesis) was monitored (Figure 6B). These results demonstrate that A3H deaminase-independent HIV-1 restriction involves inhibition of viral DNA synthesis.

Discussion

In this work, we use sequence- and structure-guided mutagenesis to provide a detailed analysis of A3H deaminase activity and to correlate enzymatic function with antiviral activity. In addition, we identify the A3H structural elements associated with these activities. We also show that A3H deaminase activity in cell extracts is suppressed by binding to cellular RNAs (Figure 3), consistent with an earlier report indicating that A3H interacts strongly with 7SL, Y1, Y3, and Y4 RNAs in 293T cells [57]. Interestingly, extensive mutagenesis studies demonstrate that the determinants of deaminase and antiviral activities are not necessarily the same, as a number of deaminase-negative mutants retain antiviral activity (Figures 4 and 5). Thus, A3H appears to inhibit HIV-1

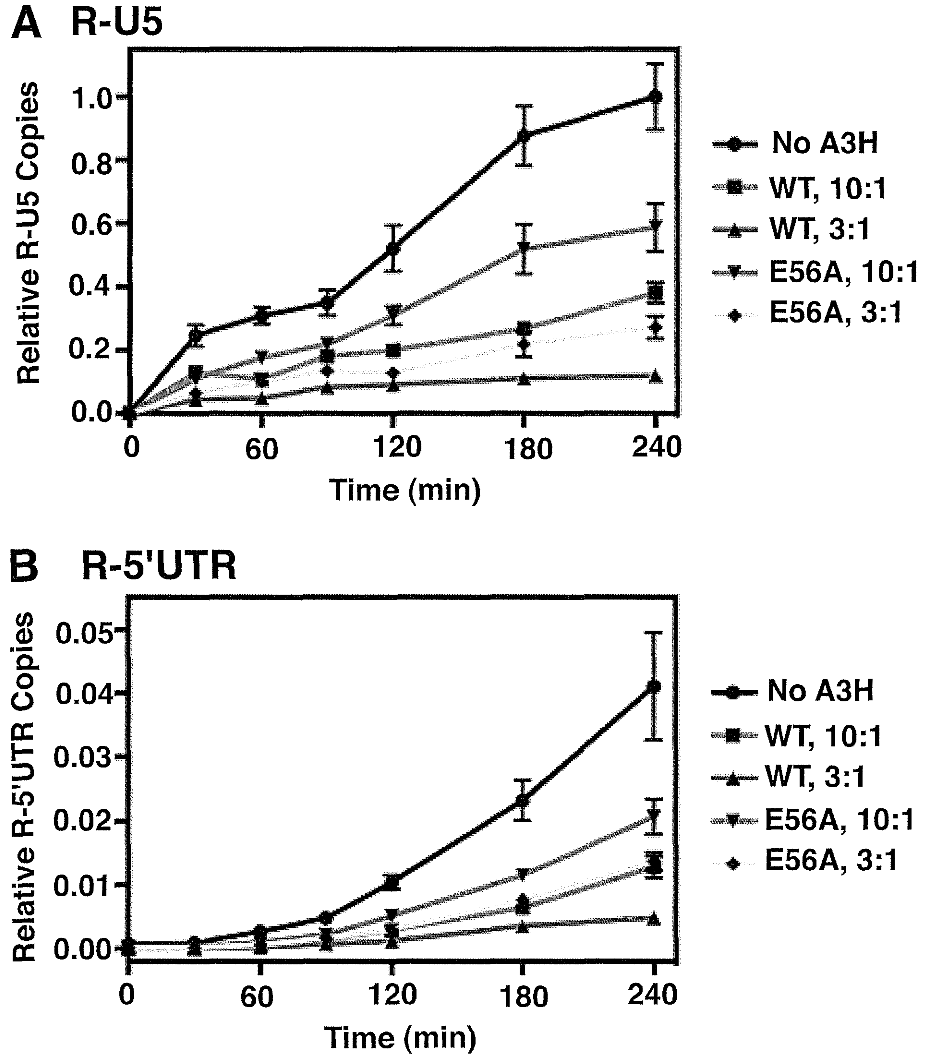

Figure 6 ERT assays of virions produced following transfection of 293T cells with HIV-1*vif* (–) and WT A3H or the E56A active site mutant. (A and B) Kinetics of DNA synthesis in ERT assays measuring the levels of R-U5 **(A)** and R-5′UTR **(B)**. The assays were performed as described in Methods. Note that synthesis of R-U5 DNA was more efficient than synthesis of R-5′UTR DNA in the presence of WT A3H under both the 10:1 and 3:1 conditions: the time required for 50% inhibition was 2 h for R-U5 and 3 h for R-5′UTR over the time course of the analysis.

infectivity via deaminase-dependent and -independent mechanisms.

To assess the molecular and structural properties of A3H Hap II, we generated a homology model based on the crystal structure of A3G-CTD (PDB: 3IR2) [53]. The electrostatic surface potential of the A3H model identified a "large basic patch" (Figure 2B, blue region), containing a cluster of basic residues that could be potentially involved in specific binding to ssDNA substrates as well as binding to cellular RNAs. While this manuscript was under review, a paper by Shandilya et al. [81] appeared, presenting homology models of the individual domains of several A3 proteins, including A3H. An electrostatic map of this A3H model also revealed a large basic patch. In fact, comparison of the two structures did not show any significant structural differences.

The basic nature of A3H is important and impacts catalytic function. Thus, deaminase activity is strongly inhibited upon mutating a subset of the basic residues (R17, R18, R20, R21, Figure 4C; R110, Figure 4F; K168, R171, Figure 4E), suggesting reduced binding to the 40-nt nucleic acid substrate. Arginine mutants that were tested also lacked antiviral activity (Figure 5B). However, the basic residues may not all function in the same manner, since the double mutant K168A/R171A retained some deaminase (Figure 4E) as well as restriction activity (Figure 5D). Interestingly, the region corresponding to R17 to P22 in A3H (containing four basic residues) is