



Strong Epistatic Selection on the RNA Secondary Structure of HIV

Raquel Assis*

Department of Biology, Huck Institutes of the Life Sciences, Center for Medical Genomics, Pennsylvania State University, University Park, Pennsylvania, United States of America

Abstract

A key question in evolutionary genomics is how populations navigate the adaptive landscape in the presence of epistasis, or interactions among loci. This problem can be directly addressed by studying the evolution of RNA secondary structures, for which there is constraint to maintain pairing between Watson-Crick (WC) sites. Replacement of a nucleotide at one site of a WC pair reduces fitness by disrupting binding, which can be restored via a compensatory replacement at the interacting site. Here, I present the first genome-scale analysis of epistasis on the RNA secondary structure of human immunodeficiency virus type 1 (HIV-1). Comparison of polymorphism frequencies at ancestrally conserved sites reveals that selection against replacements is ~ 2.7 times stronger at WC than at non-WC sites, such that nearly 50% of constraint can be attributed to epistasis. However, almost all epistatic constraint is due to selection against conversions of WC pairs to unpaired (UP) nucleotides, whereas conversions to GU wobbles are only slightly deleterious. This disparity is also evident in pairs with second-site compensatory replacements; conversions from UP nucleotides to WC pairs increase median fitness by $\sim 4.2\%$, whereas conversions from GU wobbles to WC pairs only increase median fitness by $\sim 0.3\%$. Moreover, second-site replacements that convert UP nucleotides to GU wobbles also increase median fitness by $\sim 4\%$, indicating that such replacements are nearly as compensatory as those that restore WC pairing. Thus, WC peaks of the HIV-1 epistatic adaptive landscape are connected by high GU ridges, enabling the viral population to rapidly explore distant peaks without traversing deep UP valleys.

Citation: Assis R (2014) Strong Epistatic Selection on the RNA Secondary Structure of HIV. *PLoS Pathog* 10(9): e1004363. doi:10.1371/journal.ppat.1004363

Editor: Adam Lauring, University of Michigan, United States of America

Received: March 6, 2014; **Accepted:** July 29, 2014; **Published:** September 11, 2014

Copyright: © 2014 Raquel Assis. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All data are included in the manuscript.

Funding: This work was funded by a NIH postdoctoral fellowship to RA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

* Email: rassis@psu.edu

Introduction

Epistasis is an evolutionary phenomenon whereby the fitness effect of a mutation is conditional on the genetic background in which it arises [1,2]. One of the simplest forms of epistasis occurs between WC sites of RNA secondary structures. Replacement of a nucleotide at one site of a WC pair is often deleterious because it disrupts binding, decreasing the stability of the RNA secondary structure. However, a second-site replacement in the modified genetic background may be beneficial, or compensatory, if it restores binding by creating a new WC pair [3]. Due to these complex interactions, RNA secondary structures evolve along rugged, or multi-peaked, adaptive landscapes [4], on which certain mutational trajectories may be inaccessible due to highly deleterious intermediate states. Thus, understanding how a population navigates from one peak to another on an epistatic adaptive landscape is a fundamental problem in evolutionary biology.

Selection to maintain WC pairing in RNA secondary structures leaves distinct evolutionary footprints. For one, theoretical work shows that evolutionary rates should be lower at WC than at unpaired sites [5,6], a pattern that has been observed in noncoding and synonymous regions of many viral RNA secondary structures, including those of influenza A [7], hepatitis C [8,9], and HIV-1

[10–12]. The negative correlation between extent of WC pairing and amino acid variability in HIV-1 suggests that selection to maintain WC pairing may also decrease evolutionary rates at nonsynonymous WC sites [13]. Moreover, because transitions occur more frequently and are thus compensated more rapidly than transversions, transition-to-transversion ratios are elevated at WC sites of RNA secondary structures [14]. Thus, epistatic selection can significantly alter the genomic landscape by modulating the numbers and types of mutations at WC sites of RNA secondary structures. These signatures of epistatic selection have been used both to predict and evaluate RNA secondary structures [14–16].

Epistatic interactions between WC sites have been studied in a variety of RNA molecules, including mRNAs [3,17], rRNAs [18–20], tRNAs [21,22], and RNA viruses [7,23–41]. These analyses have uncovered several key evolutionary principles. First, constraint to maintain WC pairing can result in strong long-term conservation of RNA secondary structures, yet weak conservation at the nucleotide level [18,24]. A striking example of this phenomenon involves the nearly identical secondary structures of the R regions of HIV-2 and simian immunodeficiency virus in mandrills, which have highly conserved WC pairing interactions despite a sequence homology of only 40% [24]. Second, introduction of a mutation at a WC site typically results in

Author Summary

Epistasis is an evolutionary process in which the effect of a nucleotide at one site in the genome is dependent on the presence or absence of particular nucleotides at other sites in the genome. One of the simplest types of epistasis occurs between Watson-Crick (WC) nucleotides in RNA secondary structures, which are under constraint to maintain base-pairing. In this study, I examine the effects of mutations at WC sites in the RNA secondary structure of HIV-1. I show that while epistasis plays a major role in the evolution of the HIV-1 secondary structure, different types of mutations have variable effects on fitness. Therefore, by favoring certain mutational trajectories, HIV-1 can evolve rapidly despite strong epistatic constraint on its RNA secondary structure.

impaired function, decreased thermodynamic stability, and lower fitness of a RNA secondary structure [3,5,7,17,20–23,25–41]. Third, compensatory replacements at WC sites often fully restore the function, thermodynamic stability, and fitness of a RNA secondary structure [3,17,21–23,27,28,30,32–35,38–41]. Fourth, second-site compensatory replacements may be preferred over back mutations [23], an intriguing finding that is also supported by studies of compensatory evolution in other interaction schemes [42–44]. Finally, compensatory evolution often proceeds through GU wobble intermediates [3,19,22,23], which are nearly as thermodynamically stable as WC pairs and are ubiquitous in RNAs from organisms in all three domains of life [23,45]. In some cases, GU wobbles may even confer higher fitness than WC pairs, resulting in their long-term retention [18,19].

While the dynamics of WC pairing have been extensively studied in HIV-1 [24–41], previous analyses primarily focused on secondary structures located in the 5′LTR, which regulates the transcription of viral genes. Little is known about the evolution of secondary structures across the HIV-1 genome. Recently, the RNA secondary structure of the entire HIV-1 subtype B NL4-3 genome was experimentally derived with high confidence via high-throughput selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) reactivity [46]. The availability of this structure provides a novel opportunity to study the evolution of WC pairing in HIV-1 on a genome-wide scale.

HIV-1 is an ideal model system in which to study epistasis at WC sites for a number of reasons. First, there is an abundance of publicly available sequence data for HIV-1. Second, HIV-1 has one of the highest observed spontaneous mutation rates and a relatively small genome and, thus, the waiting time for new mutations is short [47]. Third, experimental analyses of HIV-1 have demonstrated the importance of its RNA secondary structure at all stages of the viral life cycle, including reverse transcription [33,35], frameshifting [36–38], mRNA splicing [39,41], and viral packaging and transport [26,31,33]. Mutations that disrupt WC pairing in important domains often have severe phenotypic consequences [25–41], and site-directed mutagenesis studies have shown that compensatory mutations that re-establish WC pairing can restore wild type functions [27,28,30,32–35,38–41]. Finally, HIV-1 is a virus of great clinical significance, and knowledge about its evolutionary dynamics at the structural level may inform public health studies.

Results and Discussion

To investigate epistatic interactions between WC sites of HIV-1, I utilized the subtype B NL4-3 genomic sequence and RNA secondary structure [38] as a reference, 197,863 subtype B sequences (1,867 genomic) for intra-population comparison, and 66 subtype D (closest relative to subtype B) genomic sequences as an outgroup. Pairing probabilities associated with the RNA secondary structure [38] were not used because they were computed via phylogenetic analyses of covariation between sites. I restricted my analysis to sites at which reference nucleotides are ancestral, *i.e.*, conserved in all 66 outgroup sequences, both enabling polarization of mutations in the subtype B population and ensuring that reference nucleotides have been under long-term selective constraint and are therefore likely important to the RNA secondary structure. Additionally, I only considered noncoding and synonymous sites so as not to confound selection on pairing with selection on amino acid composition, though nonsynonymous sites were also analyzed separately (see Materials and Methods). Using these criteria, I identified 562 WC sites (281 pairs) and 2,868 non-WC (nWC) sites in the secondary structure of the HIV-1 reference genome.

Stems utilized in this analysis have a median length of 4 bp and are distributed across the HIV-1 genome (Table 1). Most stems are located in *Pol*, which is the gene responsible for transcribing

Table 1. Number of stems, median stem size, and number of ancestral WC sites in different regions of the HIV-1 genome.

	Number of stems	Median stem size (bp)	Number of ancestral WC sites
5′LTR	13	4	101
Gag	78	4	66
Pol	133	4	119
Vif	24	4.5	25
Vpr	19	4	13
Tat	24	3	16
Rev	24	3	21
Vpu	15	5	17
Env	125	4	87
Nef	37	4	24
3′LTR	34	4	104

Note that many genomic regions of HIV-1 overlap.
doi:10.1371/journal.ppat.1004363.t001

viral RNA into double-stranded DNA, and which also contains the greatest number of ancestral WC sites. Surprisingly, while most studies of WC pairing interactions in HIV-1 have focused on the 5'LTR, this region contains the fewest number of stems, indicating that much information may be gained from studying the evolution of WC pairing across the entire HIV-1 genome.

Comparisons of subtype B sequences to the reference sequence yielded 1,105 polymorphisms at WC sites and 7,723 polymorphisms at nWC sites. These counts indicate that there are, on average, 1.97 polymorphisms per WC site and 2.69 polymorphisms per nWC site. Thus, because there are three possible replacement nucleotides per site, WC sites are ~66% saturated and nWC sites are ~90% saturated, and this difference is highly significant ($p < 2.20 \times 10^{-16}$, Binomial test; see Materials and Methods for details). Additionally, of the 1,105 WC polymorphisms, only 306 occur as single-site replacements; the remaining 799 polymorphisms correspond to 669 double-site replacements (see Materials and Methods). Together, the lower mutational saturation at WC sites and tendency for polymorphisms at

interacting WC sites to co-segregate highlight the importance of epistasis in the evolution of the HIV-1 secondary structure.

If conservation of WC interactions in the HIV-1 secondary structure is important, destruction of WC pairing should result in a significant fitness loss. Comparison of the intra-population frequencies of single-site WC polymorphisms to those of nWC polymorphisms revealed that this is indeed the case (Figure 1). WC polymorphisms segregate at frequencies that are ~82% lower than those of nWC polymorphisms, and this difference is highly significant ($p = 1.07 \times 10^{-8}$, Mann-Whitney U test), indicating that there is strong selection against destruction of WC pairing in the HIV-1 secondary structure. To estimate the strength of this constraint, I computed selection coefficients against WC and nWC polymorphisms by $s = \mu/p_{med}$ [48], where μ is mutation rate, which has been estimated as 3.0×10^{-5} replacements/site/replication cycle [47], and p_{med} is the median frequency of the segregating polymorphism (Table 2). Differences between selection coefficients against WC and nWC polymorphisms indicate that ~63% of constraint at WC sites, and ~46% of genome-wide

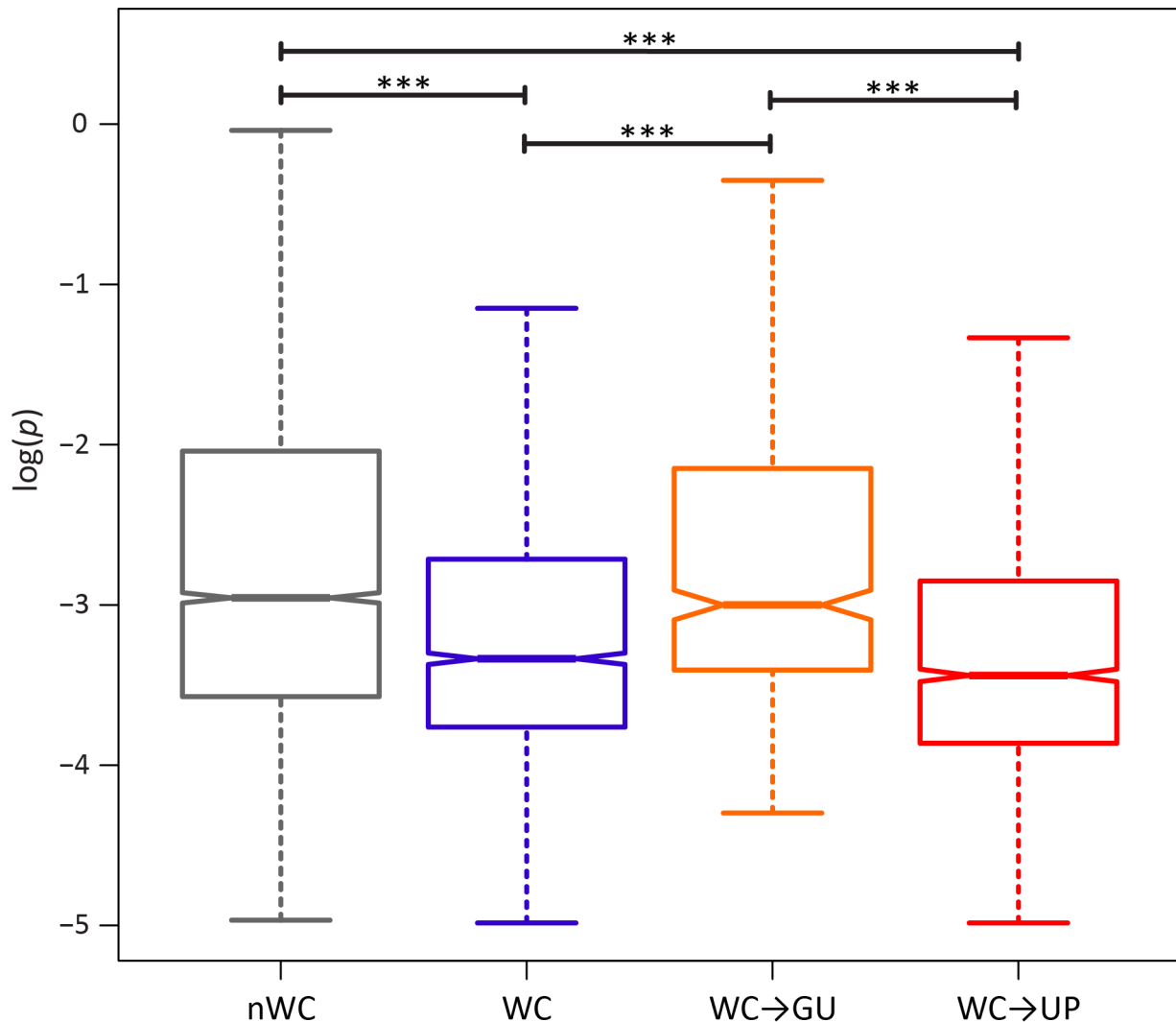


Figure 1. Intra-population frequencies of nWC and single-site WC replacement polymorphisms in the HIV-1 genome. Frequencies are normalized to enable comparisons among classes (see Materials and Methods for details) and plotted on a \log_{10} -scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***)

doi:10.1371/journal.ppat.1004363.g001

Table 2. Median frequencies, selection coefficients, and relative fitnesses of nWC and single-site WC replacement polymorphisms in the HIV-1 genome.

	P_{med}	s	W_{rel}
nWC sites	2.37×10^{-3}	1.26×10^{-2}	0.987
WC sites	8.76×10^{-4}	3.43×10^{-2}	0.966
WC→GU	2.04×10^{-3}	1.47×10^{-2}	0.985
WC→UP	6.47×10^{-4}	4.64×10^{-2}	0.954

doi:10.1371/journal.ppat.1004363.t002

constraint, is due to epistatic interactions between nucleotides in the RNA secondary structure of HIV-1 (see Materials and Methods for details).

A single replacement at a WC site can either result in two unpaired nucleotides (WC→UP) or in a GU pair (WC→GU). Because previous studies have demonstrated that WC→GU replacements are typically slightly deleterious, and can sometimes even be beneficial, selection against WC→GU replacements should be weaker than selection against WC→UP replacements [3,19,22,23,45]. Consistent with this expectation, WC→GU polymorphisms segregate at significantly higher frequencies than WC→UP polymorphisms ($p = 3.75 \times 10^{-6}$, Mann-Whitney U test; Figure 1), such that selection against WC→UP replacements is approximately three times stronger than selection against WC→GU replacements (Table 2). Moreover, frequencies of WC→GU polymorphisms are comparable to those of nWC polymorphisms ($p = 0.86$, Mann-Whitney U test), and selection against WC→GU replacements is marginally stronger than selection against replacements at nWC sites (see Table 2), indicating that WC→GU replacements are slightly deleterious in the HIV-1 secondary structure.

A potential factor in the effect of a mutation at a WC site is location. In particular, location-specific effects of replacements

may be due to position within a stem or within the HIV-1 genome. Surprisingly, frequencies of both WC→GU and WC→UP polymorphisms are similar for interior and exterior stem sites ($p = 0.65$, $p = 0.97$, Mann-Whitney U tests; Figure 2). However, frequencies of both replacement classes differ among sites located in 5'LTR, protein-coding, and 3'LTR regions of the HIV-1 genome (Figure 3). In particular, WC→GU and WC→UP polymorphisms segregate at significantly higher frequencies in 5'LTR than in either protein-coding ($p = 8.01 \times 10^{-3}$, $p = 5.49 \times 10^{-7}$, Mann-Whitney U tests) or 3'LTR ($p = 5.20 \times 10^{-3}$, $p = 7.05 \times 10^{-8}$, Mann-Whitney U tests) regions, which contain similar distributions of polymorphism frequencies ($p = 0.94$, $p = 0.33$, Mann-Whitney U tests). Thus, WC sites in the 5'LTR are less constrained than those in any other genomic region, underlining the importance of regulatory changes in the evolution of HIV-1.

To further elucidate the nature of epistatic interactions at WC sites of the HIV-1 secondary structure, I investigated the fitness effects of second-site WC replacements. A second-site replacement after an initial WC→GU replacement can result in a WC pair (GU→WC) or unpaired nucleotides (GU→UP), while a second-site replacement after an initial WC→UP replacement can result in a WC pair (UP→WC), a GU wobble (UP→GU), or unpaired

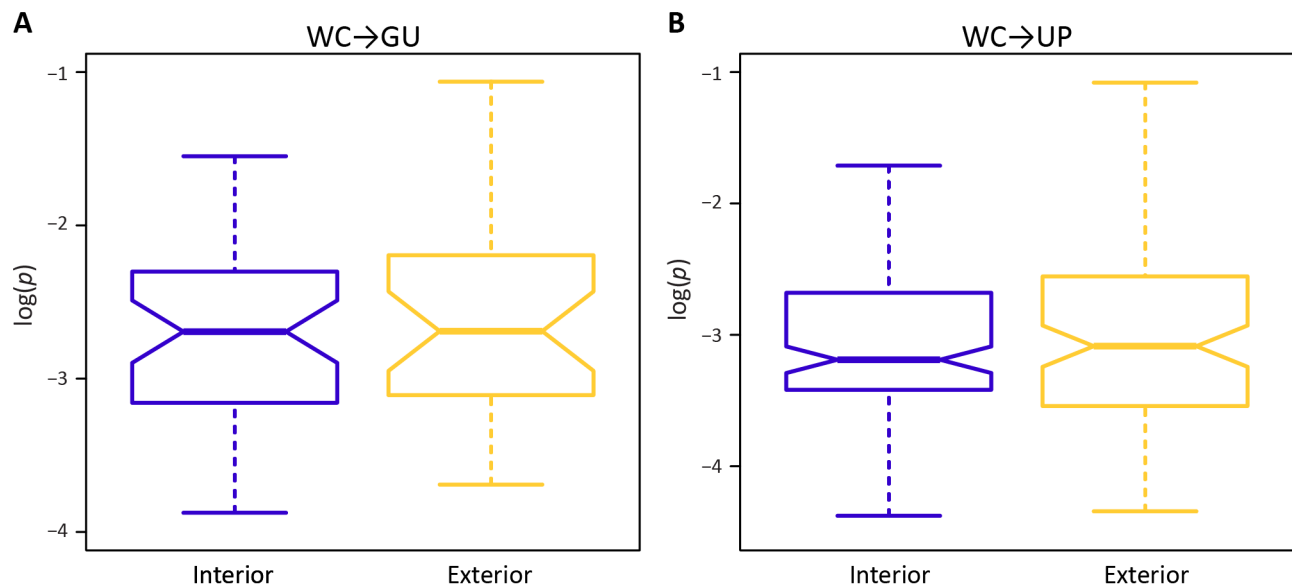


Figure 2. Intra-population frequencies of single-site WC replacement polymorphisms at interior and exterior stem positions. A) Frequencies of WC→GU polymorphisms. B) Frequencies of WC→UP polymorphisms. Frequencies are normalized to enable comparisons among classes (see Materials and Methods for details) and plotted on a \log_{10} -scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). doi:10.1371/journal.ppat.1004363.g002

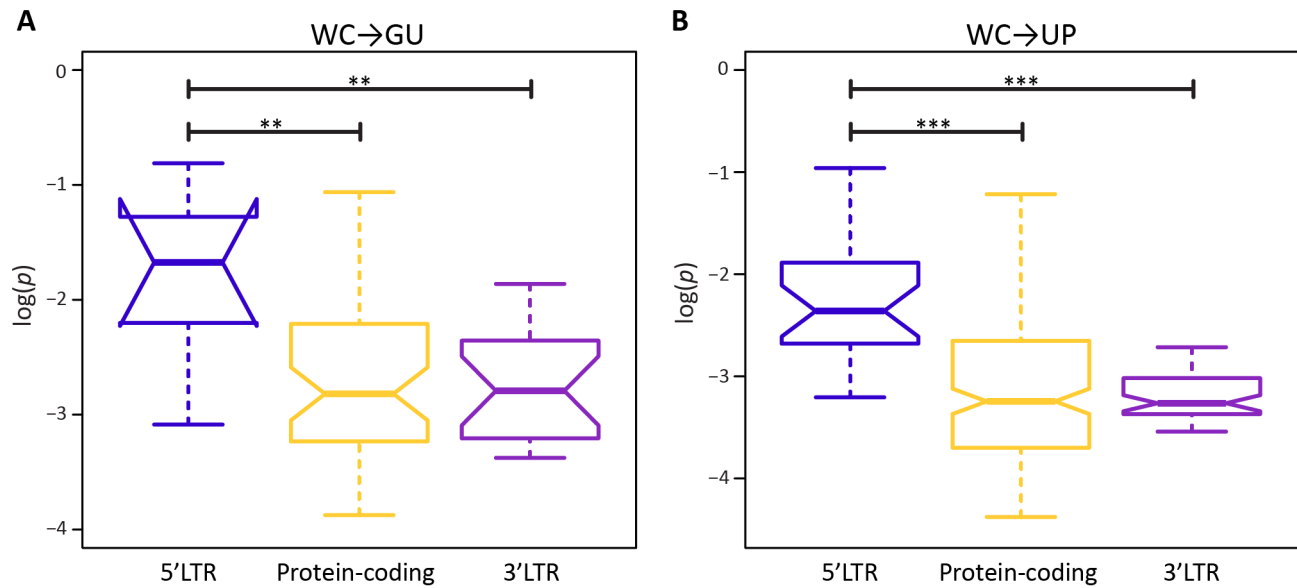


Figure 3. Intra-population frequencies of single-site WC replacement polymorphisms in 5'LTR, protein-coding, and 3'LTR regions of the HIV-1 genome. A) Frequencies of WC→GU polymorphisms. B) Frequencies of WC→UP polymorphisms. Frequencies are normalized to enable comparisons among classes (see Materials and Methods for details) and plotted on a \log_{10} -scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

doi:10.1371/journal.ppat.1004363.g003

nucleotides (UP→UP). Consistent with the prediction that restoration of WC pairing is compensatory, GU→WC polymorphisms segregate at significantly higher frequencies than GU→UP polymorphisms ($p = 2.44 \times 10^{-5}$, Mann-Whitney U test), and UP→WC polymorphisms segregate at significantly higher frequencies than either UP→GU ($p = 0.04$, Mann-Whitney U test) or UP→UP ($p = 8.21 \times 10^{-7}$, Mann-Whitney U test) polymorphisms (Figure 4). However, frequencies of UP→WC polymorphisms are also significantly greater than those of GU→WC polymorphisms ($p = 4.07 \times 10^{-3}$, Mann-Whitney U test; Figure 4). Moreover, while UP→WC replacements increase median fitness by $\sim 4.2\%$ relative to initial WC→UP replacements, GU→WC replacements only increase median fitness by $\sim 0.3\%$ relative to initial WC→GU replacements (see Materials and Methods for details). This difference indicates that conversion of a GU wobble back to a WC pair results in a small fitness gain that is comparable to the small fitness loss from an initial WC→GU replacement.

Because GU wobbles confer higher fitness than UP nucleotides in the HIV-1 secondary structure, the difference between frequencies of second-site GU→WC and GU→UP polymorphisms may be due not only to the compensatory nature of second-site GU→WC replacements, but also to fitness losses from second-site GU→UP replacements. To test this hypothesis, I compared the frequencies of second-site GU→UP and UP→UP polymorphisms, since the latter second-site replacement should not result in a fitness loss relative to the initial WC→UP replacement. Indeed, GU→UP polymorphisms segregate at significantly lower frequencies than UP→UP polymorphisms ($p = 2.42 \times 10^{-5}$, Mann-Whitney U test), such that selection against GU→UP replacements is ~ 2.6 times stronger than selection against UP→UP replacements (see Table 3). Moreover, second-site GU→UP replacements result in a median fitness loss of $\sim 3.5\%$ relative to initial WC→GU replacements (see Materials and Methods for details), illustrating the highly deleterious effect of losing all complementarity at ancestral WC sites in the HIV-1 secondary structure.

While restoration of WC pairing is thought to be the only mechanism for compensatory evolution, the findings from this study prompt a key question: Are second-site UP→GU replacements also compensatory in the HIV-1 secondary structure? Examination of frequencies of UP→GU polymorphisms shows that they are intermediate to those of UP→UP and UP→WC polymorphisms (Figure 4). Thus, second-site UP→GU replacements are indeed compensatory, though not to the degree of UP→WC replacements. In particular, UP→GU replacements result in a $\sim 4\%$ fitness gain relative to initial WC→UP replacements, in contrast to the $\sim 4.2\%$ fitness gain conferred by UP→WC replacements (see Materials and Methods for details). Hence, as expected given the small fitness losses from initial WC→GU replacements and large fitness losses from second-site GU→UP replacements, second-site UP→GU replacements are nearly as compensatory as GU→WC replacements at ancestral WC sites in the HIV-1 secondary structure.

This analysis highlights the complexities of epistatic interactions between WC sites in the HIV-1 secondary structure. In particular, although epistatic selection is strong and accounts for nearly half of all constraint on the HIV-1 secondary structure, it primarily targets replacements that completely abolish pairing interactions. In contrast, GU wobbles are typically slightly deleterious and can even compensate for the loss of fitness from initial WC→UP replacements. While the GU wobble as an intermediate is not a novel theme in the evolution of RNA secondary structures, these findings suggest that the GU wobble may play a more central role in compensatory evolution via its ability to “rescue” a RNA secondary structure after an initially deleterious WC→UP replacement. Thus, GU wobbles act not just as intermediates, but also as compensators. Moreover, this study provides the first numerical analysis of the fitness effects of various initial and second-site replacements, including those involving GU wobbles, at WC sites in the RNA secondary structure of HIV-1.

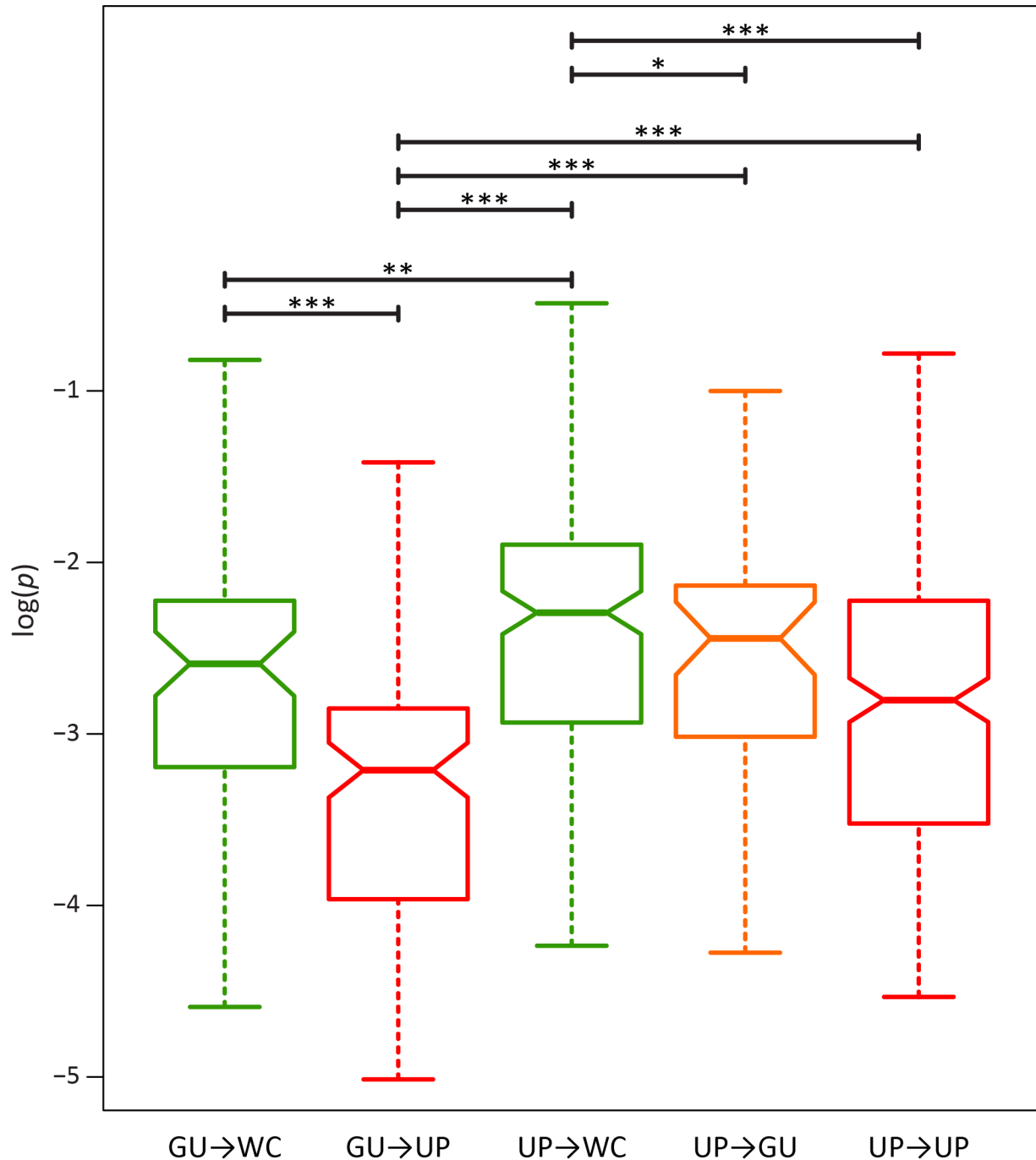


Figure 4. Intra-population frequencies of second-site WC replacement polymorphisms in the HIV-1 genome. Frequencies are normalized to enable comparisons among classes (see Materials and Methods for details) and plotted on a \log_{10} -scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).
doi:10.1371/journal.ppat.1004363.g004

Together, these findings suggest that epistatic selection on the RNA secondary structure of HIV-1 operates under a fitness hierarchy in which $w_{rel}^{WC} > w_{rel}^{GU} \gg w_{rel}^{UP}$, and the ability of a new state to increase in frequency is based on its position in the hierarchy relative to that of the previous state. Because the fitness of a GU wobble is nearly equivalent to that of a WC pair, a GU wobble can be maintained stably at a relatively high frequency in the population, likely until fitness is completely restored by a

replacement that re-establishes WC pairing. Thus, GU wobbles compose ridges that connect WC peaks in the epistatic adaptive landscape of the HIV-1 secondary structure, forming relatively flat high-fitness mutational paths to distant peaks. Moreover, while the HIV-1 population will inevitably fall into UP valleys as it traverses the adaptive landscape, it can be rescued from such a valley by a mutation that lifts it to either a WC peak or a GU ridge. Because of the high mutation rate and small genome of HIV-1, such a

Table 3. Median frequencies, selection coefficients, and relative fitnesses of second-site WC replacement polymorphisms in the HIV-1 genome.

	P_{med}	s	W_{rel}
GU→WC	2.56×10^{-3}	1.17×10^{-2}	0.988
GU→UP	6.15×10^{-4}	4.88×10^{-2}	0.951
UP→WC	5.09×10^{-3}	5.90×10^{-3}	0.994
UP→GU	3.60×10^{-3}	8.32×10^{-3}	0.992
UP→UP	1.57×10^{-3}	1.91×10^{-2}	0.981

doi:10.1371/journal.ppat.1004363.t003

mutation will arise quickly, preventing the population from becoming trapped in a UP valley and enabling its rapid evolution along the epistatic adaptive landscape.

Materials and Methods

Sequence retrieval and analysis

HIV-1 sequences were downloaded from the HIV Sequence Database at <http://www.hiv.lanl.gov/> and aligned by HMMER [49] using the HIVAlign [50] tool. The subtype B NL4-3 genome sequence (accession M19921) and corresponding positions of WC pairs in the RNA secondary structure [46] were used as a reference set for all analyses. Protein-coding nonsynonymous sites were removed from analyses to minimize confounding effects of selection on amino acids. However, as expected, findings for a separate analysis of nonsynonymous sites (Figures S1 and S2) are generally consistent with those obtained with their exclusion (Figures 1 and 4). Also, it is important to note that splice sites, which may be under additional selective constraint, were not removed from analyses, although these should not affect overall patterns observed. A site in the reference genome was considered ancestral if it is conserved in all 66 subtype D genomic sequences, and a replacement mutation was inferred when an ancestral site is polymorphic in the subtype B population. WC sites were considered to have undergone a single-site replacement when a polymorphism at one site segregates with the ancestral nucleotide at the interacting site, and a double-site replacement when polymorphisms at both sites segregate together in the population. In cases of double-site replacements, polymorphism frequencies were used to distinguish between initial and second-site replacements. In particular, the polymorphism segregating at a higher frequency (with the ancestral nucleotide at the interacting site) was designated as the initial replacement.

Normalization of polymorphism frequencies

Selection coefficients were inferred from polymorphism frequencies and the average spontaneous mutation rate of HIV-1. However, mutation rates and effects of selection may vary among different classes of nucleotide replacements. Thus, to enable comparisons of polymorphism frequencies among different classes of replacements, I normalized polymorphism frequencies by multiplying the frequency of each nucleotide replacement by its observed/expected ratio. The expected number of a particular nucleotide replacement (e.g., A→U) was computed by multiplying the number of sites with the ancestral state (e.g., A) by the corresponding nucleotide replacement rate (e.g., A→U), which was estimated from replacements at nWC sites. For example, the A→U replacement rate was computed by dividing the total number of A→U replacements at nWC sites by the total number

of nWC sites with replacements of an A (A→U+A→G+A→C). The rates for all replacement types are given in Table S1. As expected, transitions are more common than transversions at nWC sites.

Also shown in Table S1 are replacement rates computed for experimentally derived mutation data from Mansky and Temin (1995) [51]. These rates were not appropriate for the current analysis for two reasons. First, Mansky and Temin did not observe any transversions at three of the four ancestral nucleotides (G, C, and A; see Table S1), which may have been due to their small sample size (42 replacements), and is an unrealistic expectation for the current dataset (7,723 replacements at nWC sites). Second, and more importantly, replacements observed by Mansky and Temin reflect mutation rates, while nWC replacements were likely affected by non-epistatic selection. Thus, I was able to use replacement rates at nWC sites to compare and quantify epistatic and non-epistatic components of selection on the HIV-1 secondary structure, which were major objectives of the current study.

Additionally, because WC→GU polymorphisms segregate at much higher frequencies than WC→UP polymorphisms, and the probability (and frequency) of a particular second-site replacement is proportional to the probability (and frequency) of the initial replacement polymorphism, I normalized frequencies of second-site replacements by median frequencies of single-site replacements.

Estimation of the proportion of constraint due to epistasis

Selective constraint against replacements at nWC and WC sites are given in Table 1 as $s_{nwc} = 1.26 \times 10^{-2}$ and $s_{wc} = 3.43 \times 10^{-2}$. Because s_{wc} cannot solely be attributed to epistasis, I estimated the epistatic component of s_{wc} by $e(s_{wc}) = s_{wc} - s_{nwc}$. Then, the proportion of s_{wc} that is due to epistasis can be estimated by $e(s_{wc})/s_{wc}$, and the proportion of constraint at all HIV-1 sites that is due to epistasis can be estimated by $e(s_{wc})/(s_{nwc} + s_{wc})$, where $(s_{nwc} + s_{wc})$ represents total constraint.

Estimation of fitness changes

Assuming that the optimal genotype has a fitness of 1, the relative fitness (w_{rel}) of each replacement class can be estimated by $1 - s$, where s is the estimated selection coefficient of the respective replacement class that was derived from normalized polymorphism frequencies (see above). Thus, w_{rel} is the relative median fitness of a particular replacement class based on normalized nucleotide polymorphism frequencies, rather than an experimentally derived fitness value based on viral replication capacity. This estimation enables the comparison of median, but not absolute, fitness effects between replacement classes. The change in relative

fitness due to conversion from state 1 to state 2, $\Delta w_{rel}^{1,2}$, was computed as the difference between the relative fitnesses of the two states ($w_{rel}^2 - w_{rel}^1$), divided by w_{rel}^1 . For example, if state 1 is an initial WC→GU replacement and state 2 is a second-site GU→WC replacement, the change in relative fitness due to the second-site replacement is given by

$$\Delta w_{rel}^{WC \rightarrow GU, GU \rightarrow WC} = \frac{w_{rel}^{GU \rightarrow WC} - w_{rel}^{WC \rightarrow GU}}{w_{rel}^{WC \rightarrow GU}}.$$

Statistical analyses

All statistical analyses were performed in the R software environment [52]. An exact binomial test was used to compare mutation saturation levels at WC and non-WC sites by setting x to the number of observed polymorphisms at WC sites (1,105), n to the total number of possible polymorphisms at WC sites (1,686), and p to the proportion of non-WC sites that are saturated by mutation (0.898). Mann-Whitney U tests were used to compare all pairs of frequency distributions.

Supporting Information

Figure S1 Intra-population frequencies of nonsynonymous nWC and single-site WC replacement polymorphisms in the HIV-1 genome. Frequencies are normalized to enable compar-

isons among classes (see Materials and Methods for details) and plotted on a log₁₀-scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***)

(PDF)

Figure S2 Intra-population frequencies of nonsynonymous second-site WC replacement polymorphisms in the HIV-1 genome. Frequencies are normalized to enable comparisons among classes (see Materials and Methods for details) and plotted on a log₁₀-scale. Asterisks indicate $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***)

(PDF)

Table S1 Nucleotide replacement rates computed from replacements at nWC sites and from experimental mutation data of Mansky and Temin (1995).

(PDF)

Acknowledgments

The author thanks Drs. Doris Bachtrog, Yegor Bazykin, Michael DeGiorgio, and four anonymous reviewers for their helpful comments and suggestions.

Author Contributions

Conceived and designed the experiments: RA. Performed the experiments: RA. Analyzed the data: RA. Contributed reagents/materials/analysis tools: RA. Wrote the paper: RA.

References

- Kimura M (1985) The role of compensatory neutral mutations in molecular evolution. *J Genet* 64: 7–19.
- Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855–867.
- Kirby DA, Muse SV, Stephan W (1995) Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc Natl Acad Sci USA* 92: 9047–9051.
- Huynen MA, Hogeweg P (1994) Pattern generation in molecular evolution: exploitation of the variation in RNA landscapes. *J Mol Evol* 39: 71–79.
- Stephan W (1996) The rate of compensatory evolution. *Genetics* 144: 419–426.
- Innan H, Stephan W (2001) Selection intensity against deleterious mutations in RNA secondary structures and the rate of compensatory nucleotide substitutions. *Genetics* 159: 389–399.
- García M, Crawford JM, Latimer JW, Rivera-Cruz E, Perdue ML (1996) Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico. *J Gen Virol* 77: 1493–1504.
- Contreras AM, Hiasa Y, He W, Terella A, Schmidt EV, et al. (2002) Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system. *J Virol* 76: 8505–8517.
- Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P (2002) Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8: 824–841.
- Le SY, Chen JH, Braun MJ, Gonda MA, Maizel JV (1988) Stability of RNA stem-loop structure and distribution of non-random structure in the human immunodeficiency virus (HIV-1). *Nucleic Acids Res* 16: 5153–5168.
- Le SY, Chen JH, Chatterjee D, Maizel JV (1989) Sequence divergence and open regions of RNA secondary structures in the envelope regions of the 17 human immunodeficiency isolates. *Nucleic acids Res* 17: 3275–3288.
- Yoshida K, Nakamura M, Ohno T (1997) Mutations of the HIV type 1 V3 loop under selection pressure with neutralizing monoclonal antibody NM-01. *AIDS Res Hum Retroviruses* 13: 1283–1290.
- Sanjuán R, Bordería AV (2011) Interplay between RNA structure and protein evolution in HIV-1. *Mol Biol Evol* 28: 1333–1338.
- Knies JL, Dang KK, Vision TJ, Hoffman NG, Swanström R, et al. (2008) Compensatory evolution in RNA secondary structures increases substitution rate variation among sites. *Mol Biol Evol* 25: 1778–1787.
- Muse SV (1995) Evolutionary analyses of DNA sequences subject to constraints on secondary structure. *Genetics* 139: 1429–1439.
- Pederson J, Meyer I, Forsberg R, Simmonds P, Hein J (2004) A comparative method for finding and folding RNA secondary structures within protein-coding regions. *Nucleic Acids Res* 32: 4925–4936.
- Chen Y, Stephan W (2003) Compensatory evolution of a precursor messenger RNA secondary structure in the *Drosophila melanogaster Adh* gene. *Proc Natl Acad Sci USA* 100: 11499–11504.
- Woese CR, Gutell R, Gupta R, Noller HF (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* 47: 621.
- Rousset F, Pelandakis M, Solignac M (1991) Evolution of compensatory substitutions through G.U intermediate states in *Drosophila* rRNA. *Proc Natl Acad Sci USA* 88: 10032–10036.
- Dutheil JY, Jossinet F, Westof E (2010) Base pairing constraints drive structural epistasis in ribosomal RNA sequences. *Mol Biol Evol* 27: 1868–1876.
- Kern AD, Kondrashov FA (2004) Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nature Genet* 36: 1207–1212.
- Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA (2010) Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature* 464: 279–282.
- Olsthoorn RCL, Lics N, van Duin J (1994) Leeway and constraints in the forced evolution of a regulatory RNA helix. *EMBO J* 13: 2660–2668.
- Berkhout B (1991) Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic approach. *Nucleic Acids Res* 20: 27–31.
- Berkhout B, Jeang KT (1991) Detailed mutational analysis of TAR RNA: critical spacing between the bulge and loop recognition domains. *Nucleic Acids Res* 19: 6169–6176.
- Harrison GP, Lever AM (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol* 66: 4144–4153.
- Klaver B, Berkhout B (1994) Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *EMBO J* 13: 2650–2659.
- Emiliani S, Van Lint C, Fischle W, Paras P Jr, Ott M, et al. (1996) A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency. *Proc Natl Acad Sci USA* 93: 6377–6381.
- McBride MS, Panganiban AT (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* 70: 2963–2973.
- Berkhout B, Klaver B, Das AT (1997) Forced evolution of a regulatory RNA helix in the HIV-1 genome. *Nucleic Acids Res* 25: 94–947.
- Clever JL, Parslow TG (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* 71: 3407–3414.
- Das AT, Klaver B, Berkhout B (1999) A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J Virol* 73: 81–91.
- Harrich D, Hooker CW, Parry E (2000) The human immunodeficiency virus type 1 TAR RNA upper stem-loop plays distinct roles in reverse transcription and RNA packaging. *J Virol* 74: 5639–5646.

34. Kulinski T, Olejniczak M, Huthoff H, Bielecki L, Pachulska-Wieczorek K, et al. (2003) The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. *J Biol Chem* 278: 38892–38901.
35. Cobrinik D, Soskey L, Leis JA (1988) Retroviral RNA secondary structure required for efficient initiation of reverse transcription. *J Virol* 62: 3622–3630.
36. Wilson W, Braddock M, Adams SE, Rathjen PD, Kingsman SM, et al. (1988) HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* 55: 1159–1169.
37. Cassan M, Berteaux V, Angrand PO, Rousset JP (1990) Expression vectors for quantifying in vivo translational ambiguity: their potential use to analyze frameshifting at the HIV gag-pol junction. *Res Virol* 141: 597–610.
38. Parkin NT, Chamorro M, Varmus HE (1992) Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol* 66: 5147–5151.
39. Jacquenet S, Ropers D, Bilodeau PS, Damier L, Mougin A, et al. (2001) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res* 29: 464–478.
40. Garcia JA, Harrich D, Soutanakis E, Wu F, Mitsuyasu R, et al. (1989) Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. *EMBO J* 8: 765–778.
41. Abbink TEM, Berkhout B (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type I major splice donor. *J Virol* 82: 3090–3098.
42. Schrag SJ, Perrot V, Levin BR (1997) Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. *Proc R Soc Lond B Biol Sci* 264: 1287–1291.
43. Maisnier-Patin S, Berg OG, Liljas L, Anderson DI (2002) Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol Microbiol* 46: 355–366.
44. Hoffman NG, Schiffer CA, Swanstrom R (2005) Covariation of amino acid positions in HIV-1 protease. *Virology* 331: 206–207.
45. Varani G, McClain WH (2000) The GU wobble base pair. *EMBO Rep* 1: 18–23.
46. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess CW Jr, et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460: 711–719.
47. Perelson AS, Ribeiro RM (2008) Estimating drug efficacy and viral dynamic parameters: HIV and HCV. *Statist Med* 27: 4647–4657.
48. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, et al. (2001) Prediction of deleterious human alleles. *Hum Mol Genet* 10: 591–597.
49. Eddy SR (1995) Multiple alignment using hidden Markov models. *ISMB* 3: 114–120.
50. Gaschen B, Kuiken C, Korber B, Foley B (2001) Retrieval and on-the-fly alignment of sequence fragments from the HIV database. *Bioinformatics* 17: 415–418.
51. Mansky LM, Temin HM (2005) Lower in vivo mutation rate of human immunodeficiency virus type 1 than predicted from the fidelity of purified reverse transcriptase. *J Virol* 69: 5087.
52. R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.