Journal of General Virology (1998), **79**, 1337–1345. Printed in Great Britain

REVIEW ARTICLE

Retrovirus mutation rates and their role in genetic variation

Louis M. Mansky

Department of Medical Microbiology and Immunology, Creighton University School of Medicine, 2500 California Plaza, Omaha, NE 68178-0213, USA

Introduction

Retrovirus variation has important implications not only on virus diversity and evolution, but also on virulence, pathogenesis and the ability to develop effective antiviral drugs and vaccines. This point is easily made when one considers human immunodeficiency virus type 1 (HIV-1) and the AIDS pandemic. The high variability of retroviruses (and of RNA viruses) has its basis in the lack of elaborate proofreading and repair mechanisms during some (or all) of the steps in replication of the virus genome. In this review, I will discuss some highlights in the literature that have helped shape our understanding of retrovirus variation.

Early studies on the variation of retrovirus populations

The earliest published report of retrovirus variation was by Rous & Murphy (1913), where they observed different varieties of tumours as the result of infection by a virus of chicken sarcoma no. 1 (now called Rous sarcoma virus, RSV). By passaging the virus in animals, variations in the tumours was observed, including those that were haemorrhagic, and those made up of either giant cells or spindle cells. In later studies, Duran-Reynals (1942) noted variations in Rous sarcomas when RSV was passaged in ducklings, then returned to chickens. These variations included rapidly growing tumours with non-infectious virus and late-developing tumours with infectious virus.

The development of the focus assay for transformation by RSV (Temin & Rubin, 1958) allowed for more quantitative measurements of retrovirus variation. Temin (1960) used this assay with different clonal stocks of RSV to show that genetic differences in RSV could lead to different morphological changes of infected cells. Two main phenotypes were *morph*^{r} (virus producing foci of round refractile cells) and *morph*^{r} (virus producing foci of long fusiform cells). Temin noted that virus mutations occurred during virus replication and not at the time of infection.

Author for correspondence: Louis M. Mansky. Present address: Department of Medical Microbiology & Immunology, Ohio State University, 2078 Graves Hall, 333 West 10th Ave, Columbus, OH 43210-1239, USA. Fax +1 614 292 9805.
e-mail mansky.3@osu.edu

Molecular approaches to analyse retrovirus variation and mutation

Technical advances in biochemistry and molecular biology allowed for more quantitative measurements of retrovirus variation. For example, when ten proviruses of spleen necrosis virus (SNV, an avian type C retrovirus) from a clonal stock were molecularly cloned, only six of the proviruses were found to be infectious (O'Rear *et al*., 1980). Each provirus could be distinguished by restriction enzyme digestion patterns and by biological properties. In addition, DNA sequencing analyses of Moloney murine leukaemia virus (M-MLV) indicated that many closely related strains existed (Weiss *et al*., 1985).

Further studies utilizing various molecular biological techniques reported the extent of retrovirus variation. For example, an avian sarcoma virus stock was analysed for variation after repeated passages in cell culture by Coffin *et al*. (1980). Variation was determined by fingerprinting of RNase T1 digests of radiolabelled viral RNA in two-dimensional polyacrylamide gels. Estimates of mutation frequency based on mutations at single nucleotide bases were made and were in the range of 10^{-3} to 10^{-4}. In another study, the extent of variation was compared between the v-*mos* oncogene of Moloney murine sarcoma virus (M-MSV) and the c-*mos* and *gag* genes of M-MSV and M-MLV. This was done by determining the extent of variation based upon the time of transduction of c-*mos* to v-*mos*. The extent of variation of nucleotide substitution for v-*mos*, the *gag* gene and c-*mos* was 1×10^{-3}, 6×10^{-4} and 2×10^{-9}, respectively. This work supported the notion that retroviruses, like other RNA viruses, evolve at a rate 1 million times that of eukaryotic DNA genomes.

Many estimates of HIV and simian immunodeficiency virus (SIV) variation have been made. In several cases, the PCR was used, allowing easier isolation of large numbers of genome variants for analysis by DNA sequencing. Such studies have indicated that retroviruses can exist as a complex viral quasispecies (Domingo *et al*., 1985) or swarm (Temin, 1989). For example, Hahn *et al*. (1986) recovered sequential isolates

0001-4787

from HIV-1 infected individuals (four to six isolates recovered from three individuals over a 3 year period). The extent of variation for nucleotide substitutions was estimated to be at least 10^{-3} for the *env* gene and 10^{-4} for the *gag* gene. Three reports used similar approaches to analyse SIV variation by inoculating a molecular clone into either rhesus monkeys or macaques. Burns & Desrosiers (1991) analysed the *env* gene sequence of 27 clones from rhesus monkeys and estimated the variation to be 9×10^{-3} in the SU region of *env*, with most mutations (81%) resulting in amino acid changes. Johnson *et al.* (1991) examined SIV *env* variation of clones from macaques and found that, on average, nucleotide substitutions occurred at a level of 10^{-2}, and observed many clones with a large degree of G-to-A transition mutations. Finally, Overbaugh *et al.* (1991) found that virtually all of the nucleotide changes in SIV *env* from macaques encoded amino acid changes (98%).

Willems *et al.* (1993) injected cloned bovine leukaemia virus (BLV) proviral DNA into sheep to analyse variation by PCR. Variation was estimated to be 9×10^{-5} in *env* and 3×10^{-4} in the long terminal repeat. This indicates that BLV diversity is considerably different from that of HIV or SIV. The relatively lower diversity of BLV and that of human T-cell leukaemia virus (HTLV), the type members of the BLV/HTLV genus of the family *Retroviridae*, may reflect the fact that replication of these viruses is primarily by clonal expansion rather than by reverse transcription (Wain-Hobson, 1996).

Determining rates of retrovirus mutation

In the previous studies, it is difficult to determine the mutation rate of the virus because of the many rounds of virus replication that have occurred. To put this into some perspective, it is worth considering the variables that define the rate of genetic variation of retroviruses. Retrovirus genetic variation can be viewed as the composite of the mutation rate per replication cycle, the number of replication cycles, the fixation rate of mutations and the rate of recombination (Coffin, 1996; Hu & Temin, 1990*b*; Zhang & Temin, 1993). Given the apparent high variability of retroviruses, it was thought that the mutation rate per replication cycle may be the main variable responsible for this. Such views were supported by the numerous cell-free studies that indicated the high error rates of reverse transcriptase purified from various retroviruses (Bebenek & Kunkel, 1993). To test this, approaches were used to determine the retrovirus mutation rate per replication cycle.

Leider *et al.* (1988) measured the mutation rate of RSV by harvesting virus from a transformed cell clone and infecting fresh chick embryo fibroblasts. Colonies of transformed cells were grown, virus harvested, and viral RNA analysed in seven regions by heteroduplex analysis using denaturing gradient gel analysis. Mutations were detected by altered gel migration. The overall mutation rate for RSV was calculated to be 1×10^{-4} mutations per nucleotide per replication cycle.

Monk *et al.* (1992) determined the point mutation rate of the murine leukaemia virus AKV strain. A clonal cell line was infected at a low m.o.i. and virus stocks from these cells were used to infect fresh cells to produce secondary clonal cell lines. RNase T1-oligonucleotide fingerprinting analyses of viral RNAs was used to detect base changes along with direct sequencing of the viral RNAs. These combined methods gave a mutation rate of 2×10^{-5} mutations per nucleotide per replication cycle.

Dougherty & Temin (1988) used an SNV-based vector containing two selectable markers (*neo* and *hygro*) to develop a protocol in which a retroviral vector could be replicated and would complete only one round of replication (see Fig. 1). The protocol was based on the use of retrovirus helper (or packaging) cell lines, which are specially engineered cell lines that produce viral proteins. Introduction of a retroviral vector (which is missing some or all viral genes) can be complemented in *trans* in a helper cell such that virus particles are produced in these cells. Vector virus produced from these cells can then be used to infect permissive target cells. By introducing the SNV vector into a characterized SNV helper cell clone, vector virus was produced that could be used to infect permissive target cells, which were subsequently placed under drug selection until individual colonies formed. The steps going from the

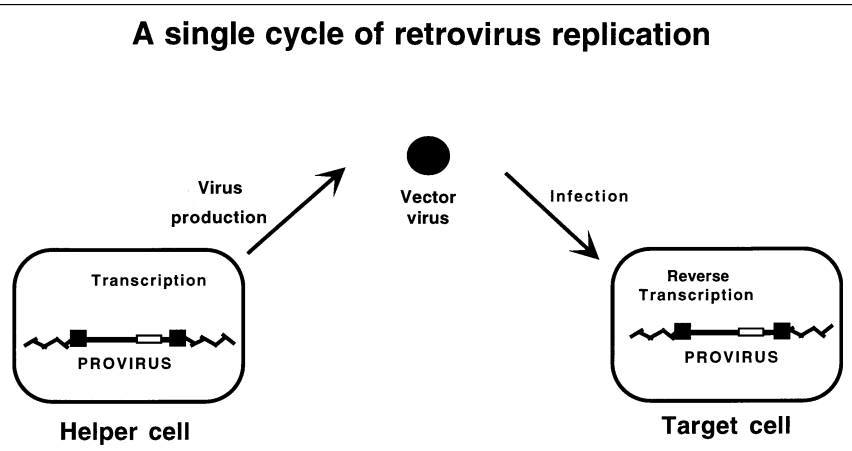

Fig. 1. A single cycle of retrovirus replication. The steps going from the retroviral vector provirus in the helper cell to the retroviral vector provirus in the target cell constitute a single cycle of retrovirus replication. These steps include transcription of the proviral DNA by the cellular transcription machinery, packaging of the viral RNA, release of viral particles, infection of target cells, reverse transcription and integration of newly synthesized viral DNA to generate a vector provirus.

provirus (i.e. the integrated retroviral DNA) in the virus-producing helper cell to the provirus in the infected target cell was defined as a single round of virus replication. Analysis of proviruses from infected target cell clones indicated a mutation frequency of 5×10^{-3} mutations per cycle. Reversion of an amber codon introduced into the *neo* gene allowed the calculation of a reversion rate for base pair substitutions (Fig. 2). The base pair substitution rate was calculated to be 2×10^{-5} mutations per base pair per replication cycle (Dougherty & Temin, 1988).

A more elegant approach for detecting mutations during SNV replication was the use of the *lacZα* peptide gene as a reporter gene for mutations and the blue/white colony colour selection method for identifying mutant proviruses in *E. coli* (Pathak & Temin, 1990*a*, *b*). In this approach, the *in vivo* forward mutation rates for various types of mutations were calculated in a single cycle of retrovirus replication. Forward mutation rates were able to be detected because the *lacZα* peptide gene was not placed under selective pressure during replication. The major types of mutations found were base pair substitutions, frameshifts, simple deletions and deletions with insertions. The overall *in vivo* mutation rate of SNV in this system was determined to be 1×10^{-5} mutations per target base pair per replication cycle. Mansky & Temin (1994) extended these studies to BLV. BLV was found to have a mutation rate of 5×10^{-6} mutations per target base pair per replication cycle. A similar distribution of mutation types was found with BLV relative to that of SNV, indicating that the a common property of reverse transcriptase is responsible for all these error processes. Temin (1993) speculated that this common property was the strand-transfer process.

Comparison of the error rates of purified enzymes from different retroviruses indicated that the HIV reverse transcriptase was more error prone that other reverse transcriptases (Bebenek & Kunkel, 1993). This led to predictions that the HIV mutation rate was comparable to the error rate of purified reverse transcriptase (Nowak, 1990). Mansky & Temin (1995) tested the hypothesis that the mutation rate for HIV-1 is comparable with the error rate of purified HIV-1 reverse transcriptase by determining the *in vivo* mutation rate of HIV-1 in a single cycle using the *lacZα* gene as a mutational target. The mutation rate of HIV-1 in this system was determined to be 3×10^{-5} mutations per target base pair per cycle (Mansky, 1996*a*; Mansky & Temin, 1995). The most commonly detected mutations were base substitution mutations (G to A and C to T transition mutations) and frameshift mutations (-1 frameshifts in runs of T's and A's). This mutation rate is maximally 5% of the reported error rates for purified HIV-1 reverse transcriptase with a sense-strand RNA and DNA template in cell-free studies, indicating that the mutation rate of HIV-1 is less than that predicted by the measured fidelity of purified HIV-1 reverse transcriptase (Boyer *et al.*, 1992).

A similar observation between the cell-free error rate of purified reverse transcriptase and the *in vivo* mutation rate has been noted for M-MLV by Varela-Echavarria *et al.* (1992), who used a amber codon reversion assay to determine the rate of base substitution mutations occurring during the replication of an M-MLV vector and compared this rate to the fidelity of purified M-MLV reverse transcriptase in a cell-free system. The base substitution rate at a single locus was found to be 2×10^{-6} mutations per base pair per cycle. The *in vivo* rate of base substitution of the M-MLV vector was 30-fold lower than the rate of formation of T·G mispairs observed in the cell-free system. The differences in the M-MLV mutation rate found by Varela-Echavarria *et al.* (1992) compared with those determined by Monk *et al.* (1992) may be due to differences in experimental conditions, which could influence the selective forces acting during virus replication.

The SNV mutation rate system with the *lacZα* peptide gene as a reporter has been used to assess the contributions of the early (RNA transcription and minus-strand synthesis) and late stages (plus-strand synthesis and DNA repair) in going from the provirus in the helper cell to the provirus in the target cell. The *lacZα* peptide gene was inserted into the long terminal repeat of the SNV vector and proviruses were recovered from infected cells containing either one or both long terminal repeats. Recovered proviruses with both long terminal repeats yielded a mutant phenotype only when the *lacZα* peptide genes in both long terminal repeats were mutated, which would occur during the early stage. Recovered proviruses with only the original long terminal repeat containing the *lacZα* peptide gene identified mutant phenotypes regardless of at what stage the mutations occurred. Therefore, mutant frequencies of proviruses recovered containing either two or one long terminal repeats allowed the measurement of the early-stage and total mutation rates, respectively. The results of Kim *et al.* (1996) indicated that the early and late stages had mutation rates that were within 2-fold of each other. During this study, two mutants with A to G hypermutations were recovered. Kim *et al.* suggested that these hypermutations could involve the action of a double-stranded RNA adenosine deaminase.

Retrovirus genes that influence the mutation rate

Reverse transcriptase fidelity clearly plays the major role in determining the rate at which mutations occur during the process of reverse transcription. However, accessory replication proteins may influence the accuracy of reverse transcription (Mansky, 1997). This is suggested by the difference between the *in vivo* mutation rate of HIV-1 and the error rate of purified HIV-1 reverse transcriptase. To test the hypothesis that viral proteins may contribute to the accuracy of reverse transcription, the HIV-1 *vpr* gene was mutated and the influence of *vpr* on the rate of mutation in a single cycle determined using the *lacZα* gene as a mutational target (Mansky, 1996*b*). The mutation rate of a vpr^- vector mutant

was as much as 4-fold higher than that of the *vpr*$^+$ parental vector, supporting the conclusion that the *vpr* gene partially accounts for the lower than predicted *in vivo* mutation rate of HIV-1. The *vpr* gene encodes a 96 amino acid, non-structural protein that is associated with virus particles (Cohen *et al.*, 1990), and requires the p6 region of the *gag* gene for incorporation (Kondo *et al.*, 1995; Lu *et al.*, 1993, 1995; Paxton *et al.*, 1993). Specific roles of *vpr* include the nuclear localization of the preintegration complex and prevention of cell proliferation during chronic infection (He *et al.*, 1995; Heinzinger *et al.*, 1994; Jowett *et al.*, 1995; Planelles *et al.*, 1995; Rogel *et al.*, 1995). The HIV-1 Vpr protein has been found to interact with cellular proteins (Refaeli *et al.*, 1995; Zhao *et al.*, 1994), including the DNA repair enzyme uracil DNA hydroxylase (UNG) (Bouhamdan *et al.*, 1996). The observation that *vpr* can influence the mutation rate indicates that accessory proteins are involved in reverse transcription and can influence the accuracy of the process. It remains to be seen whether accessory proteins influence not only reverse transcription accuracy, but efficiency. Also, the influences of cellular proteins on reverse transcription remains to be elucidated.

Determining rates of recombination

Many studies have indicated that recombination also occurs at high rates. Thus, recombination is a major variable in creating mutations during retrovirus replication. Many early studies analysed recombination during mixed infection of cells with two viruses (Alevy & Vogt, 1978; Blair, 1977; Kawai & Hanafusa, 1972; Linial & Brown, 1979; Vogt, 1971; Wyke & Beamand, 1979). Such studies found that the frequency of recombination exceeded 10% of the progeny. In these studies, recombination was measured after multiple rounds of replication. Two models, forced-copy choice and strand displacement-assimilation, have been proposed for the mechanism of recombination (Coffin, 1979; Hunter, 1978; Skalka *et al.*, 1982). The work of Hu & Temin (1990*a*, *b*) first described a system to measure the rate of retrovirus recombination in a single round of replication. In this system, an SNV-based retroviral vector that expresses both the *neo* gene and the *hygro* gene was used. Two frameshift mutations were introduced into the vector such that different marker genes were inactivated in different constructs (Fig. 2). So, while the two vectors would be highly homologous, each would confer resistance to different drug selection (either G418 for *neo* or hygromycin for *hygro*, but not for both). Each vector would be propagated separately, and then used to coinfect helper cells. Helper cell clones with dual resistance would contain both vectors. Virus from doubly infected helper cell clones would then be used to infect permissive target cells. Since retrovirus particles encapsidate two molecules of viral genomic RNA, this method allows the formation of heterozygous virus particles containing two different genomic RNAs in one round of replication. Recombination events that occur in the 1 kb region between the mutations introduced into the drug-resistance markers can be detected. Two outcomes from recombination occur (i.e. recombinants with two functional drug-resistance genes or recombinants with two defective genes), but only one-half of the recombinants (i.e. those with two functional genes) can be scored in the assay. Therefore, the rate of recombination is twice the measured rate. The measured rate of recombination is determined by comparing the titre of the parental viruses (resistant to a single drug) with the titre of the recombinant viruses (resistant to both drugs). With this approach, the measured rate of recombination was found to be 2% per kilobase per replication cycle (Hu & Temin, 1990*a*, *b*), indicating a rate of recombination of 4% per kilobase per replication cycle. Recombination is not proportional to marker distance (Anderson *et al.*, 1988*a*). Recombinants appeared to be formed by the mechanism of forced-copy choice. Further studies by Jones *et al.* (1994) have indicated that SNV

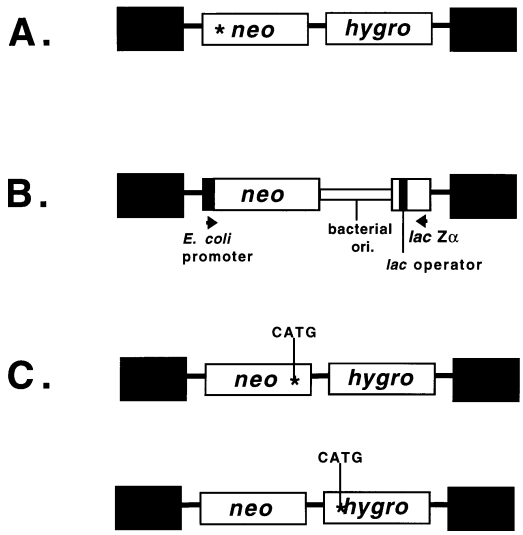

Fig. 2. Examples of retroviral vectors used for determination of mutation and recombination rates. All vectors are shown in the proviral DNA form. (A) SNV-based vector used for initial mutation rate studies in a single cycle of retrovirus replication (Dougherty & Temin, 1988). The asterisk indicates the location of an amber codon introduced into the *neo* gene. Reversion to the wild-type during replication allows the determination of a reversion rate of mutation. (B) SNV shuttle vector for determining the forward rate of mutation using the *lacZα* peptide gene as a reporter in one cycle of replication (Pathak & Temin, 1990*a*). Base pair substitution, frameshift and deletion mutations can be detected using such a vector. (C) SNV vector pair used for determining the homologous recombination rate in one cycle of replication (Hu & Temin, 1990*a*). Insertion of 4 bases (CATG) into the *neo* and *hygro* genes (indicated by the asterisks) allowed for detection of recombination events in the 1 kb region between the mutations. The black rectangular boxes at the ends of the vectors represent the long terminal repeats. Thick horizontal black lines indicate viral sequence. White boxes are foreign coding sequences, as indicated. The *E. coli* promoter driving expression of the *neo* gene (conferring resistance to kanamycin in bacteria), the bacterial origin of replication (for replication of the vector proviral DNA as a plasmid in *E. coli*), the *lac* operator sequence and the *lacZα* peptide gene region are indicated. *neo*, neomycin phosphotransferase resistance gene; *hygro*, hygromycin resistance gene.

undergoes, on average, one additional or aberrant strand transfer every round of replication.

Homologous recombination that results in deletions can occur by intramolecular or intermolecular template switching. The rates of intramolecular and intermolecular switching were directly compared by Hu *et al.* (1997) with two retroviral vectors containing direct repeats flanked by two different sets of restriction site markers. Heterozygous virions containing RNAs from the two parental vectors were made and recombinant viruses were subsequently generated. The rates of switching (intramolecular or intermolecular) were determined by analysis of the markers flanking the direct repeats in recombinant and nonrecombinant proviruses. Hu *et al.* found that intramolecular template switching was much more efficient than intermolecular template switching and concluded that the deletion of direct repeats occurred primarily through intramolecular template switching. The results from these experiments also indicated that recombination occurs within distinct virus subpopulations and that recombination exhibits high negative interference (i.e. the selection for one recombination event increases the probability that a second recombination event will occur). High negative interference can occur by correlated template switching events during minus-strand DNA synthesis (Anderson *et al.*, 1988*b*).

Nonhomologous recombination can also occur, but at a lower rate compared to that of homologous recombination. Recombination occurs between viral sequences and nonhomologous host sequences, resulting in transduction. The formation of the highly oncogenic retroviruses occurs in this fashion. Recombination can also occur between nonhomologous sequences in the retroviral genome, resulting in rearrangements and deletions (Pathak & Temin, 1990*b*). In a single round of retrovirus replication, Zhang & Temin (1993) determined the rate of nonhomologous recombination to be about 0·1–1% of the rate of homologous recombination. The mechanism of nonhomologous recombination is similar to that of homologous recombination with the exception of the selection of templates. Such mechanisms are likely to be similar to the mechanisms proposed for the creation of mutation events called deletion with insertion, which were first reported by Pathak & Temin (1990*b*).

Mutation, selection, evolution and disease progression

An important goal in studying the genetic diversity of retroviruses is to gain a better understanding of the correlation between retrovirus variation and disease progression. The relatively high rates of mutation (and recombination) of retroviruses are certainly major factors involved in rapid evolution. In the case of HIV-1, the high rate of replication is also very important (Coffin, 1995).

Although a variety of different approaches have been utilized to measure mutation rates of retroviruses, it is apparent that these rates are very high. In general, these rates range from about 10^{-4} to 10^{-6} mutations per base pair per replication cycle. Recombination rates are also high, occurring at a rate of about 4%. Although relatively limited, the range of known retrovirus mutation rates indicates that retroviruses have somewhat similar intrinsic rates of mutation. If these rates are considered in the light of variation of natural viral sequences under selection, it appears that variables other than the mutation rate determine the variability of retrovirus populations. Additional variables that could influence diversity include the mode of replication (i.e. replication as a provirus by cell DNA replication or replication as a virus via reverse transcription) and modulation of the mutation rate by accessory replication proteins (such as HIV-1 Vpr) (Mansky, 1997).

Variants of reverse transcriptase could also influence diversity. For example, it has been proposed that the M184V mutation in HIV-1 reverse transcriptase, which confers drug resistance to 2′,3′-dideoxy-3′-thiacytidine (3TC), increases the fidelity of the purified enzyme and decreases HIV-1 diversity (Wainberg *et al.*, 1996). Additional data regarding this observation will help verify and extend these studies. Because of the relatively high level of HIV-1 diversity, it is likely that variants (reverse transcriptase, protease, etc.) that confer drug resistance would pre-exist in the population. However, these mutations must have some slightly deleterious effect on virus replication, otherwise they would exist in the wild-type virus. Since these types of drug resistant mutant would not replicate as well as wild-type, it is possible that this reduced level of replication would allow for successful clinical treatment, and favour a strategy for preventing disease progression that consists of antivirals that limit replication of the most resistant mutant and not just drugs that affect wild-type virus replication. It has been suggested that the long-term changes in virus load in response to drug therapy are influenced by the immune response, the killing of uninfected $CD4^+$ cells, and the differential efficacies of drugs in different cell populations (Bonhoeffer *et al.*, 1997*a*, *b*).

A highly error-prone reverse transcriptase could have a significant influence on virus diversity. Growth of HIV-1 or retroviral vectors in cell culture has led to the characterization of isolates with a large number of mutations, and these isolates have been described as multiple mutants or hypermutants (Mansky & Temin, 1995; Pathak & Temin, 1990*a*; Vartanian *et al.*, 1991). Some of these hypermutants have contained G to A mutations (Vartanian *et al.*, 1991), and are thought to occur due to imbalances in nucleotide pools (Martinez *et al.*, 1994; Vartanian *et al.*, 1997). However, it is not clear how biologically relevant these hypermutants, which may be created by a hypermutable reverse transcriptase, are in regard to pathogenesis and disease progression or what selective advantage would be conferred by such a mutable virus.

If the above mentioned variables significantly influence the level and rate of virus replication, then they could have

implications for diversity. The low half-life of an HIV-1-infected cell and the high rate of HIV-1 replication illustrates how a high replication rate can be a driving force for creating diversity in retrovirus populations (Coffin, 1995). However, such variables are all acted upon by selective forces, and ultimately it is selection that would shape the changes in virus diversity and evolution. Coffin (1992) argued that no HIV-1 mutation can be considered to be neutral.

Analysis of the specific types of mutational events that occur in retrovirus populations (e.g. HIV-1) can provide some insight into the selective forces acting. The proportion of synonymous codon sites changed is considered to be largely a function of the number of replication cycles the population has undergone, with the exception of areas that are conserved because of RNA secondary structure or because of translation in many reading frames. If the proportion of synonymous codon changes is high relative to the number of non-synonymous codon sites that have changed, then this may be an indication that selection for replication fitness is occurring. On the other hand, if the proportion of nonsynonymous codon sites is high relative to the proportion of synonymous codon site changes, then this would suggest a more random accumulation of mutations and that selection is acting in a more diverse manner by host factors (e.g. the immune system selecting for escape mutants). This second scenario has been noted in several analyses of HIV-1 (Balfe *et al.*, 1990; Holmes *et al.*, 1992; Lukashov *et al.*, 1995; Seibert *et al.*, 1995; Simmonds *et al.*, 1990; Wolinsky *et al.*, 1996).

Models to explain how the immune system provides positive selection for variation have been proposed. The antigenic diversity threshold theory states that virus load causes disease, immune responses reduce virus load, and virus evolution during infection weakens the effect of the immune response to HIV-1, thus increasing the virus load and leading to rapid disease progression (Nowak, 1995; Nowak *et al.*, 1991). A central point to this model is the dynamic threshold condition that specifies whether a population of HIV-1 in an individual is controlled by the immune response. The aspect of the model that indicates that the capability of the immune response is exceeded once antigenic diversity increases beyond a certain threshold is controversial because the relatively high mutation rate and rapid turnover rate of HIV-1 are major factors that determine the high rate of virus evolution (Nowak *et al.*, 1996). Future experiments will certainly help to address the various aspects of this model.

In the antigenic diversity model, as in other models that have been proposed to describe HIV-1 diversity (Coffin, 1995), the assumption has been made that selection and mutation are the only important forces influencing variation, and that virus evolution is treated deterministically, with the size of the virus population in an infected individual being infinite. Leigh Brown (1997) suggested that stochastic models, rather than deterministic models, may better describe HIV-1 populations. In deterministic models, the size of the population is infinite and the relative fitness of different allelic variants directly determines their abundance in the progeny after virus replication. In stochastic models, the population is finite and a difference in fitness is a difference in probability. This allows for a class of selectively neutral mutations that have a fate determined entirely by random genetic drift (Leigh Brown, 1997; Leigh Brown & Richman, 1997). Most synonymous mutations are usually considered to be in this class. However, Coffin (1995) indicated that the level of synonymous mutations in HIV-1 is determined by a balance between the forward mutation rate and the rate of removal by selection (a deterministic model). Several observations have been used in support of stochastic models. One is the unpredictable rate at which alleles become fixed (e.g. the rate at which zidovudine-conferring resistance alleles appear in spite of a consistent replicative advantage) (Boucher *et al.*, 1992; Cleland *et al.*, 1996). The methionine to leucine substitution at codon 41 in reverse transcriptase confers the highest levels of zidovudine resistance, but is not observed in some patients, and therefore is not expected in deterministic models because they are generally not considered to be biologically relevant due to the failure of deterministic models to allow for sampling effects (Wright, 1969). Further experimental data will help to clarify the most appropriate models to describe HIV-1 populations.

In conclusion, it is clear that the mutation rate has a significant influence on retrovirus genetic variation. However, relative virus fitness predictably determines the ultimate diversity of a population. Therefore, an understanding of the combined selective pressures that produce variants, along with the nature of the variants produced, is central to the correlation of virus mutation rate and virus evolution to immune escape, drug resistance and disease progression.

I thank Shiaolan Yang for helpful comments and discussion of the manuscript. I am supported by Public Health Service grant GM56615.

References

Alevy, M. C. & Vogt, P. K. (1978). Ts pol mutants of avian sarcoma viruses: mapping and demonstration of single cycle recombinants. *Virology* **87**, 21–33.

Anderson, J. A., Bowman, E. H. & Hu, W.-S. (1988*a*). Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation. *Journal of Virology* **72**, 1195–1202.

Anderson, J. A., Teufel, R. J., Yin, P. D. & Hu, W.-S. (1988*b*). Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *Journal of Virology* **72**, 1186–1194.

Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Brown, A. J. L. (1990). Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. *Journal of Virology* **64**, 6221–6233.

Bebenek, K. & Kunkel, T. A. (1993). The fidelity of retroviral reverse transcriptases. In *Reverse Transcriptase*, pp. 85–102. Edited by A. M.

Skalka & S. P. Goff. Plainview, NY: Cold Spring Harbor Laboratory Press.

Blair, D. G. (1977). Genetic recombination between avian leukosis and sarcoma viruses. Experimental variables and the frequencies of recombination. *Virology* **77**, 534–544.

Bonhoeffer, S., Coffin, J. M. & Nowak, M. A. (1997*a*). Human immunodeficiency virus drug therapy and virus load. *Journal of Virology* **71**, 3275–3278.

Bonhoeffer, S., May, R. M., Shaw, G. M. & Nowak, M. A. (1997*b*). Virus dynamics and drug therapy. *Proceedings of the National Academy of Sciences, USA* **94**, 6971–6976.

Boucher, C. A., O'Sullivan, E., Mulder, J. W., Ramautarsing, C., Kellam, P., Darby, G., Lange, J. M., Goudsmit, J. & Larder, B. A. (1992). Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *Journal of Infectious Diseases* **165**, 105–110.

Bouhamdan, M., Benichou, S., Rey, F., Navarro, J.-M., Agostini, I., Spire, B., Camonis, J., Slupphaug, G., Vigne, R., Benarous, R. & Sire, J. (1996). Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *Journal of Virology* **70**, 697–704.

Boyer, J. C., Bebenek, K. & Kunkel, T. A. (1992). Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proceedings of the National Academy of Sciences, USA* **89**, 6919–6923.

Burns, D. P. & Desrosiers, R. C. (1991). Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. *Journal of Virology* **65**, 1843–1854.

Cleland, A., Watson, H. G., Robertson, P., Ludlam, C. A. & Brown, A. J. (1996). Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **12**, 6–18.

Coffin, J. M. (1979). Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *Journal of General Virology* **42**, 1–26.

Coffin, J. (1992). Genetic diversity and evolution of retroviruses. *Current Topics in Microbiology and Immunology* **176**, 143–164.

Coffin, J. (1995). HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489.

Coffin, J. M. (1996). *Retroviridae*: the viruses and their replication. In *Fields Virology*, 3rd edn, pp. 1767–1847. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

Coffin, J. M., Tsichlis, P. N. & Barker, C. S. (1980). Variation in avian retrovirus genomes. *Annals of the New York Academy of Sciences* **354**, 410–425.

Cohen, E. A., Dehni, G., Sodroski, J. G. & Haseltine, W. A. (1990). Human immunodeficiency virus *vpr* product is a virion-associated regulatory protein. *Journal of Virology* **64**, 3097–3099.

Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J. C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., VandePol, S., Steinhauer, D., DePolo, N. & Holland, J. (1985). The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance – a review. *Gene* 1–8.

Dougherty, J. P. & Temin, H. M. (1988). Determination of the rate of base-pair substitution and insertion mutation in retrovirus replication. *Journal of Virology* **62**, 2817–2822.

Duran-Reynals, F. (1942). The reciprocal infection of ducks and chickens with tumor-inducing viruses. *Cancer Research* **2**, 343–369.

Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S. & Parks, W. P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**, 1548–1553.

He, J., Choe, S., Walker, R., Di Marzio, P., Morgan, D. O. & Landau, N. R. (1995). Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G_2 phase of the cell cycle by inhibiting $p34^{cdc2}$ activity. *Journal of Virology* **69**, 6705–6711.

Heinzinger, N. K., Bukrinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M.-A., Gendelman, H. E., Ratner, L., Stevenson, M. & Emerman, M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proceedings of the National Academy of Sciences, USA* **91**, 7311–7315.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A. & Leigh Brown, A. J. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of HIV-1 within a single infected patient. *Proceedings of the National Academy of Sciences, USA* **89**, 4835–4839.

Hu, W.-S. & Temin, H. M. (1990*a*). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proceedings of the National Academy of Sciences, USA* **87**, 1556–1560.

Hu, W.-S. & Temin, H. M. (1990*b*). Retroviral recombination and reverse transcription. *Science* **250**, 1227–1233.

Hu, W.-S., Bowman, E. H., Delviks, K. A. & Pathak, V. K. (1997). Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *Journal of Virology* **71**, 6028–6036.

Hunter, E. (1978). The mechanism for genetic recombination in the avian retroviruses. *Current Topics in Microbiology and Immunology* **79**, 295–309.

Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. (1991). The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* **185**, 217–228.

Jones, J. S., Allan, R. W. & Temin, H. M. (1994). One retroviral RNA is sufficient for synthesis of viral DNA. *Journal of Virology* **68**, 207–216.

Jowett, J. B. M., Planelles, V., Poon, B., Shah, N. P., Chen, M.-L. & Chen, I. S. Y. (1995). The human immunodeficiency virus type 1 *vpr* gene arrests infected T cells in the $G_2 + M$ phase of the cell cycle. *Journal of Virology* **69**, 6304–6313.

Kawai, S. & Hanafusa, H. (1972). Genetic recombination with avian tumor viruses. *Virology* **49**, 37–44.

Kim, T., Mudry, R. A., Rexrode, C. A. & Pathak, V. K. (1996). Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. *Journal of Virology* **70**, 7594–7602.

Kondo, E., Mammano, F., Cohen, E. A. & Gottlinger, H. G. (1995). The $p6^{gag}$ domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *Journal of Virology* **69**, 2759–2764.

Leider, J. M., Palese, P. & Smith, F. I. (1988). Determination of the mutation rate of a retrovirus. *Journal of Virology* **62**, 3084–3091.

Leigh Brown, A. J. (1997). Analysis of HIV-1 *env* gene sequences reveals evidence for a low effective number in the viral population. *Proceedings of the National Academy of Sciences, USA* **94**, 1862–1865.

Leigh Brown, A. J. & Richman, D. D. (1997). HIV-1: gambling on the evolution of drug resistance? *Nature Medicine* **3**, 268–271.

Linial, M. & Brown, S. (1979). High-frequency recombination within the *gag* gene of Rous sarcoma virus. *Journal of Virology* **31**, 257–260.

Lu, Y.-L., Spearman, P. & Ratner, L. (1993). Human immunodeficiency

virus type 1 viral protein R localization in infected cells and virions. *Journal of Virology* **67**, 6542–6550.

Lu, Y.-L., Bennett, R. P., Wills, J. W., Gorelick, R. & Ratner, L. (1995). A leucine triplet repeat sequence $(LXX)_4$ in $p6^{gag}$ is important for Vpr incorporation into human immunodeficiency virus type 1 particles. *Journal of Virology* **69**, 6873–6879.

Lukashov, V. V., Kuiken, C. L. & Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *Journal of Virology* **69**, 6911–6916.

Mansky, L. M. (1996*a*). Forward mutation rate of human immunodeficiency virus type 1 in a T-lymphoid cell line. *AIDS Research and Human Retroviruses* **12**, 307–314.

Mansky, L. M. (1996*b*). The mutation rate of human immunodeficiency virus type 1 is influenced by the *vpr* gene. *Virology* **222**, 391–400.

Mansky, L. M. (1997). Accessory replication proteins and reverse transcription accuracy: implications for retroviral genetic diversity. *Trends in Genetics* **13**, 134–136.

Mansky, L. M. & Temin, H. M. (1994). Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. *Journal of Virology* **68**, 494–499.

Mansky, L. M. & Temin, H. M. (1995). Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than predicted from the fidelity of purified reverse transcriptase. *Journal of Virology* **69**, 5087–5094.

Martinez, M. A., Vartanian, J.-P. & Wain-Hobson, S. (1994). Hypermutagenesis of RNA using human immunodeficiency virus type 1 and biased dNTP concentrations. *Proceedings of the National Academy of Sciences, USA* **91**, 11787–11791.

Monk, R. J., Malik, F. G., Stokesberry, D. & Evans, L. H. (1992). Direct determination of the point mutation rate of a murine retrovirus. *Journal of Virology* **66**, 3683–3689.

Nowak, M. (1990). HIV mutation rate. *Nature* **347**, 522.

Nowak, M. A. (1995). AIDS pathogenesis: from models to viral dynamics in patients. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **10**(Suppl.), S1–S5.

Nowak, M. A., Anderson, R. M., Boerlijst, M. C., Bonhoeffer, S. & May, R. M. (1996). HIV-1 evolution and disease progression. *Science* **274**, 1008–1009.

Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F. W., Goudsmit, J. & May, R. M. (1991). Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963–969.

O'Rear, J. J., Mizutani, S., Hoffman, G., Flandt, M. & Temin, H. M. (1980). Infectious and noninfectious recombinant clones of the provirus of SNV differ in cellular DNA and are apparently the same in viral DNA. *Cell* 423–430.

Overbaugh, J., Rudensey, L. M., Papenhausen, M. D., Benveniste, R. E. & Morton, W. R. (1991). Variation in simian immunodeficiency virus *env* is confined to V1 and V4 during progression to simian AIDS. *Journal of Virology* **65**, 7025–7031.

Pathak, V. K. & Temin, H. M. (1990*a*). Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proceedings of the National Academy of Sciences, USA* **87**, 6019–6023.

Pathak, V. K. & Temin, H. M. (1990*b*). Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proceedings of the National Academy of Sciences, USA* **87**, 6024–6028.

Paxton, W., Connor, R. I. & Landau, N. R. (1993). Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of *gag* and mutational analysis. *Journal of Virology* **67**, 7229–7237.

Planelles, V., Bachelerie, F., Jowett, J. B., Haislip, A., Xie, Y., Banooni, P., Masuda, T. & Chen, I. S. Y. (1995). Fate of the human immunodeficiency type 1 provirus in infected cells: a role for *vpr*. *Journal of Virology* **69**, 5883–5889.

Refaeli, Y., Levy, D. N. & Weiner, D. B. (1995). The glucocorticoid receptor type II complex is a target of the HIV-1 vpr gene product. *Proceedings of the National Academy of Sciences, USA* **92**, 3621–3625.

Rogel, M. E., Wu, L. I. & Emerman, M. (1995). The human immunodeficiency virus type 1 *vpr* gene prevents cell proliferation during chronic infection. *Journal of Virology* **69**, 882–888.

Rous, P. & Murphy, J. B. (1913). Variations in a chicken sarcoma caused by a filterable agent. *Journal of Experimental Medicine* **17**, 219–231.

Seibert, S. A., Howell, C. Y., Hughes, M. K. & Hughes, A. L. (1995). Natural selection on the gag, pol and env genes of HIV-1. *Molecular Biology and Evolution* **12**, 803–813.

Simmonds, P., Balfe, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990). Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *Journal of Virology* **64**, 5840–5850.

Skalka, A. M., Boone, L., Junghans, R. & Luk, D. (1982). Genetic recombination in avian retroviruses. *Journal of Cellular Biochemistry* **19**, 293–304.

Temin, H. M. (1960). The control of cellular morphology in embryonic cells infected with Rous sarcoma virus *in vitro*. *Virology* **10**, 182–197.

Temin, H. M. (1989). Is HIV unique or merely different? *Journal of Acquired Immune Deficiency Syndromes* **2**, 1–9.

Temin, H. M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proceedings of the National Academy of Sciences, USA* **90**, 6900–6903.

Temin, H. M. & Rubin, H. (1958). Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. *Virology* **6**, 669–688.

Varela-Echavarria, A., Garvey, N., Preston, B. D. & Dougherty, J. D. (1992). Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase in vitro. *Journal of Biological Chemistry* **267**, 24681–24688.

Vartanian, J.-P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. (1991). Selection, recombination, and G to A hypermutation of human immunodeficiency virus type 1 genomes. *Journal of Virology* **65**, 1779–1788.

Vartanian, J.-P., Plikat, U., Henry, M., Mahieux, R., Guillemot, L., Meyerhans, A. & Wain-Hobson, S. (1997). HIV genetic variation is directed and restricted by DNA precursor availability. *Journal of Molecular Biology* **270**, 139–151.

Vogt, P. K. (1971). Genetically stable reassortment of markers during mixed infection with avian tumor viruses. *Virology* **46**, 947–952.

Wain-Hobson, S. (1996). Running the gamut of retroviral mutation. *Trends in Microbiology* **4**, 135–141.

Wainberg, M. A., Drosopoulos, W. C., Salomon, H., Hsu, M., Borkow, G., Parniak, M. A., Gu, Z., Song, Q., Manne, J., Islam, S., Castriota, G. & Prasad, V. R. (1996). Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. *Science* **271**, 1282–1285.

Weiss, R., Teich, N., Varmus, H. & Coffin, J. (1985). *RNA Tumor Viruses*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Willems, L., Thienpont, E., Kerkhofs, P., Burny, A., Mammerickx, M. &

Kettmann, R. (1993). Bovine leukemia virus, an animal model for the study of intrastrain variability. *Journal of Virology* **67**, 1086–1089.

Wolinsky, S. M., Korber, B. T. M., Neumann, A. U., Daniels, M., Kuntsman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D., Safrit, J. T. & Koup, R. A. (1996). Adaptive evolution of HIV-1 during the natural course of infection. *Science* **272**, 537–542.

Wright, S. (1969). Evolution and the genetics of populations. In *The Theory of Gene Frequencies*, pp. 512. Chicago: University of Chicago Press.

Wyke, J. A. & Beamand, J. A. (1979). Genetic recombination in Rous sarcoma virus: the genesis of recombinants and lack of evidence for linkage between pol, env and src genes in three factor crosses. *Journal of General Virology* **43**, 349–364.

Zhang, J. & Temin, H. M. (1993). Rate and mechanisms of non-homologous recombination during a single cycle of retroviral replication. *Science* **259**, 234–238.

Zhao, L.-J., Mukherjee, S. & Narayan, O. (1994). Biochemical mechanism of HIV-1 Vpr function. *Journal of Biological Chemistry* **269**, 15577–15582.