1710–1718 *Nucleic Acids Research, 1996, Vol. 24, No. 9*

Differences in mutagenesis during minus strand, plus strand and strand transfer (recombination) synthesis of the HIV-1 *nef* gene *in vitro*

Weimin Wu[1], Chockalingam Palaniappan[1], Robert A. Bambara[1,3] and Philip J. Fay[1,2,*]

Departments of [1]Biochemistry and [2]Medicine and [3]The Cancer Center, University of Rochester, Rochester, NY 14642, USA

Received January 10, 1996; Revised and Accepted March 11, 1996

ABSTRACT

We have developed an HIV *nef–Escherichia coli lacZ* fusion system *in vitro* that allows the detection of low frequency mutations, including frameshifts, deletions and insertions. A portion of the *nef* gene that encompasses a hypervariable region was fused in-frame with a downstream *lacZ* α peptide coding region. The resulting *lacZ* α peptide fusion protein remained functional. Any frameshift mutations in the *nef* insert would put the downstream *lacZ* α peptide gene out of frame, eliminating α complementation. With this system we compared the error rates of frameshift mutations that arise during DNA-directed and RNA-directed DNA synthesis. Results showed that DNA-directed and RNA-directed DNA synthesis did not contribute equally to the generation of mutations. DNA-directed DNA synthesis generated frameshift mutations at a frequency ~10-fold higher than those arising from RNA-directed DNA synthesis. RNA-directed DNA synthesis in the presence of acceptor templates showed an increase in mutation rate and differences in the mutation spectrum. The enhancement of mutation rate was caused by the appearance of mutations at three new locations that correlated with likely recombination sites. Results indicate that recombination is another source of mutations during viral replication.

INTRODUCTION

The type 1 human immunodeficiency virus (HIV-1), just as other retroviruses, exhibits extensive genome heterogeneity (1–3). The average base substitution mutation rates for some retroviruses have been determined to be ~0.5 base substitutions/retroviral genome/replication cycle (4–7). This results in the generation of 'quasi-species', a mixture of closely related but genetically distinct genomes, in the infected population and in infected individuals (2). This hypermutability permits evasion of immunosurveillance and rapid emergence of drug-resistant mutants. Therefore, it has been a major impediment to the development of both an effective vaccine against viral coat protein and the design of successful chemotherapeutic agents for the treatment of AIDS. Consequently, study of the mechanism of HIV-1 mutagenesis is very important in the development of antiviral strategies.

Various mutations can occur during three phases of the HIV-1 replication cycle. (i) In the HIV reverse transcriptase (RT)-directed synthesis phase the virally encoded polymerase converts the viral RNA genome into double-stranded DNA, which is then integrated into the host genome as a provirus. (ii) Subsequently, host DNA polymerases replicate the integrated proviral DNA in each cell cycle. (iii) Host RNA polymerase II transcribes proviral DNA into viral genomic RNA. Because of the high fidelity of cellular DNA replication (10^{-9}–10^{-10} substitutions/bp) (8–10), its contribution to the high mutation rate of HIV-1 is not likely to be significant. The fidelity of host RNA polymerase II is unknown. However, it has been demonstrated that prokaryotic RNA polymerases are highly accurate (6,11–13). Therefore, the majority of mutations are probably introduced during HIV RT-directed DNA synthesis.

Retroviruses replicate through a DNA intermediate (14,15). The viral genomic plus strand RNA is first used as a template for minus strand DNA synthesis by the viral RT. The RNA template is degraded during the process by RNase H activity residing at the C-terminus of the RT. Plus strand DNA synthesis then follows, using minus strand DNA as the template and generating a double-stranded DNA version of the viral genome. During reverse transcription two primer strand transfers are required for generation of full-length proviral DNA (16). In addition, a high frequency of recombination has also been observed from within internal regions of the viral genome (17–19). The high mutation rate exhibited in the course of reverse transcription is partly attributed to the error-prone character of RT. One reason for the poor fidelity of RT is that it lacks a 3′→5′ exonuclease capable of proof-reading, i.e. removing incorrectly added nucleotides. The fidelity of several purified RTs has been measured in cell-free systems, with observed misincorporation rates of from 0.5×10^{-4} to 1.5×10^{-4} (20). In addition, it has been found that existing errors can be repositioned and new errors generated through a recombination process (21,22). Errors made in the process of RT-catalyzed polymerization, plus alterations resulting from recombination, are likely to be the major determinants of HIV-1 hypermutability.

Among the mutations generated during reverse transcription, deletion has been observed to occur frequently. In a variety of retroviruses examined deletion mutants arise spontaneously after repeated undiluted passage (23–26). However, it has not been determined whether minus strand and plus strand DNA synthesis

* To whom correspondence should be addressed

contribute equally to the generation of deletion mutations during viral replication. To address this question we created a novel assay system based on *lacZ* α complementation to separately examine DNA-directed plus and RNA-directed minus strand DNA synthesis on a *nef* segment. *nef* is a non-structrual gene of HIV located near the 3′-end of the HIV-1 genome. Substantial sequence polymorphism has been detected among *nef* genes cloned from various isolates of HIV-1 (27). An imperfect duplication of an adjacent downstream sequence has been reported to appear often around position 23 of the Nef amino acid sequence (28). For our work a 69 nt segment of the *nef* gene encompassing the variable region was fused in-frame with a downstream *lacZ* α peptide coding region. The resulting *lacZ* α peptide fusion protein remained functional. Any frameshift mutations in the *nef* insert would put the downstream *lacZ* α peptide gene out of frame, eliminating α complementation. Employing this system we observed a difference in generating frameshift mutations by HIV-1 RT depending on the strand copied. DNA-dependent DNA synthesis exhibited more frequent production of frameshift mutations compared with RNA-dependent DNA polymerization. A mutation spectrum of RNA-directed DNA synthesis was also generated in the presence of a recombination acceptor RNA template. The presence of the acceptor template allowed strand transfer synthesis, accompanied by an increased rate of mutation and appearance of mutations in new positions. These results suggest that strand transfer is another source of genomic variation.

MATERIALS AND METHODS

Materials

The bacterial strain DH5α was obtained from Life Technologies Inc. The recombinant form of HIV-1 RT (p66/p51 heterodimer) was generously provided by the Genetics Institute (Cambridge, MA). The enzyme had a sp. act. of ~40 000 U/mg. One unit of RT is defined as the amount required to incorporate 1 nmol dTTP into nucleic acid product in 10 min at 37°C using poly(rA)·oligo(dT) as primer–template. To make templates for primer extension and strand transfer reactions the *nef* clone J14 was used as the starting material, which was prepared as described before (22). T7 RNA polymerase, placental RNase inhibitor, DNase I (RNase-free), RNase A (DNase-free), *Taq* DNA polymerase, T4 DNA ligase, *Escherichia coli* DNA polymerase I Klenow fragment, rNTPs, dNTPs and restriction enzymes *Xba*I, *Xho*I, *Sal*I and *Bgl*II were purchased from Boehringer Mannheim. The DNA sequencing kit was supplied by US Biochemical Corp. Plasmid vectors pBluescript II SK(+) phagemid and pGEM 7zf(–) were from Stratagene and Promega respectively. The DNA primers used in the primer extensions were synthesized by Genosys. [α-^{35}S]dATP was provided by Amersham Corp. X-Gal and IPTG were from Life Technologies Inc. Other chemicals were from Sigma.

Preparation of RNA and DNA templates used in the reactions

To make RNA and DNA templates pBSM13(*nef* J14) was constructed by subcloning *nef* J14, a 649 nt long segment of the *nef* gene, into pBSM13(+) as described before (22). Then an *Xba*I restriction site was created at positions 32–37 in the *nef* coding region in pBSM13(*nef* J14), generating plasmid pWM as the starting material.

To generate the RNA template, pWM was linearized with *Bgl*II in the *nef* coding region and transcribed by T7 RNA polymerase *in vitro*. The resulting RNA is 265 nt in length. Run-off transcription was carried out according to the *Promega Protocols and Applications Guide*. The transcription products were subjected to 8% denaturing polyacrylamide gel electrophoresis. The 265 nt RNA transcript was excised and eluted from the gel slice.

To make the acceptor RNA template used in the standard strand transfer assay *nef* J14 with the *Xba*I insert (*nef* J14*) was isolated from pWM by *Eco*RI and *Hin*dIII cleavage and subcloned into pGEM7Zf(–) [pGEM7Zf(*nef* J14*)]. The plasmid was cleaved with *Xho*I and underwent run-off transcription by T7 RNA polymerase. The *Xho*I site is upstream of the *Bgl*II site used to linearize plasmid pWM for generation of the primer elongation template as mentioned above. The full-length 149 nt transcription product was purified as described above. The acceptor template contained a 100 nt region homologous to the primer elongation template.

The DNA template used for DNA-directed DNA polymerization was generated by an asymmetric polymerase chain reaction (PCR). First a 170 nt double-stranded DNA template was generated through 35 cycles of PCR (1 min at 94°C, 1 min at 58°C and 1 min at 72°C) utilizing the primers 5′-GTCAGAATTCATGGGTGGCAAGTGGTCA-3′ and 5′-CAGCATTGTTAGCTGCTGTA-3′. The double-stranded DNA product was purified from the gel. The single-stranded DNA template used in primer extension was produced by using the double-stranded DNA product as template for 35 cycles of asymmetric PCR amplification (1 min at 94°C, 1 min at 48°C and 1 min at 72°C). In the asymmetric PCR reaction, the amount of primer 5′-CAGCATTGTTAGCTGCTGTA-3′ was 100 pmol and that of primer 5′-GTCAGAATTCATGGGTGGCAAGTGGTCA-3′ 2 pmol. The single-stranded DNA template was finally purified from a 6% non-denaturing polyacrylamide gel.

Construction of the vector used in the *LacZ* assay

Plasmid pBluescript SK(+) was modified by elimination of the *Xho*I site in the multiple cloning site. pBluescript SK(+) was linearized with *Xho*I. The cohesive ends were then filled in using *E.coli* DNA polymerase I Klenow fragment. The blunt ends were ligated by T4 DNA ligase, generating pBluescript SK(+)*.

Hybridization of the DNA primer to the template

The DNA primer was incubated with the template at a 4:1 (primer/template) molar ratio of 3′-termini in 10 mM Tris–HCl, pH 7.5, 1 mM EDTA and 80 mM KCl. The hybridization mixture was heated to 70°C for 10 min and then slowly cooled to room temperature.

Primer extension and the standard strand transfer assay

Both RNA- and DNA-directed primer extensions were carried out in a final volume of 20 µl. Eight units of RT was preincubated with 2 nM (in template termini) primer–template at 37°C for 3 min in 50 mM Tris–HCl, pH 8.0, 20 mM KCl and 1 mM dithiothreitol. The reaction was started by adding $MgCl_2$ and dNTPs to final concentrations of 6 mM and 50 µM respectively. For DNA-directed DNA polymerization after 60 min of primer extension the reaction was terminated by phenol extraction. The reaction mixture was digested with *Xba*I and *Xho*I and the restriction fragment was ligated into pBluescript SK(+)*. For RNA-directed DNA polymerization the RNA moiety of the reverse transcription products was removed

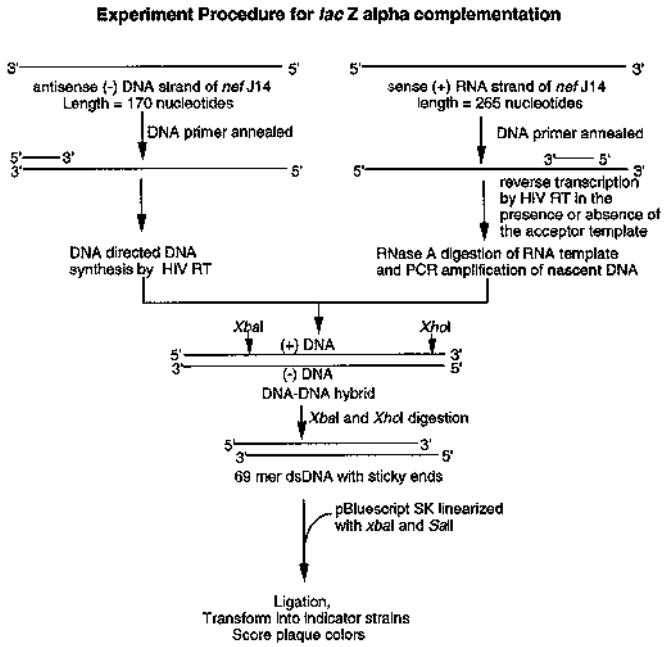

Figure 1. Experimental procedure. Shown in the figure is the experimental approach used for *lacZ* α complementation. A more detailed description is given in the text.

with 0.5 μg RNase A. The remaining nascent single-stranded (–) DNA was amplified by PCR. The resulting double-stranded DNA was digested and ligated as mentioned above.

DNA sequencing

The inserts in the isolated mutants were sequenced by the dideoxynucleotide chain termination method, as described in the sequencing manual from US Biochemical Corp.

Data analysis

The error rate for a particular type of mutation was determined by multiplying the percentage of that type of mutation by the frequency of all mutations and then dividing by both the probability of expressing errors in all the scored colonies and by the number of nucleotides examined. The formula is (number of specific type of mutations/total number of mutations) × (mutants/total colonies) × (1/probability of expressing errors) × (1/nucleotides examined). For the assay examining the fidelity of DNA-directed DNA synthesis only one strand expressed the mutations introduced by HIV-1 RT in the final double-stranded DNA. Therefore, the probability of expressing errors in all the scored colonies during DNA-directed DNA synthesis is 1/2. Since the system does not allow selection of most of the base substitution mutations, it is not appropriate to calculate the error rate of this type of mutation using the formula given above.

RESULTS

lacZ assay system

A novel *lacZ* assay system based on α complementation was designed to examine the fidelity of plus and minus strand DNA synthesis (Fig. 1). To study the fidelity of DNA-directed plus strand DNA synthesis a 170 nt minus strand DNA encompassing the variable region of the *nef* gene was generated by asymmetric PCR, as described in Materials and Methods. Primer extension on the DNA template was then carried out with HIV-1 RT. The final products were cleaved with *Xba*I and *Xho*I, releasing a 69 nt DNA fragment that contained the hypervariable region in *nef*. The DNA fragment was cloned into the modified vector pBluescript SK(+), fused in-frame with a downstream *lacZ* α peptide coding region. The resulting *lacZ* α peptide remained functional. When the plasmid was transformed into DH5α cells containing the partially defective *lacZ* ΔM15 gene product complementation generates a functional β-galactosidase protein. In the presence of X-gal and IPTG the colonies containing the plasmids with inserts are blue. Mutations occurring within the examined DNA fragment, such as frameshift, deletion, insertion and base substitution, that could generate a termination codon could put the downstream *lacZ* α peptide gene out of frame, disrupting α complementation. This produces colonies that appear white on plates containing X-gal and IPTG. Plasmids isolated from the white colonies were subjected to DNA sequencing and the spectrum of mutations was generated.

To examine the fidelity of RNA-directed minus strand DNA synthesis a 265 nt plus strand RNA template was obtained by transcription *in vitro*. Reverse transcription generated minus strand DNA. The RNA template was removed by treatment with RNase A and the remaining single-stranded (–) DNA was amplified by 20 cycles of PCR. The resulting double-stranded DNA was cleaved with the restriction enzymes *Xba*I and *Xho*I and the resultant fragment inserted into the vector as described above. The plasmids were transformed into the indicator strains. Inserts in the white colonies were analyzed by DNA sequencing to generate the mutation spectrum. For investigation of strand transfer-related errors reverse transcription was carried out in the presence of RNA acceptor templates. The recombinant molecules were isolated and amplified by PCR. The same 69 nt fragment and derivatives containing mutations were released by restriction enzyme digestion. The detection of mutations was carried out as described above. The spectrum of mutations was compared with that of RNA-directed DNA synthesis in the absence of acceptor template.

The mutation spectrum of DNA-directed DNA polymerization

For DNA-directed DNA synthesis, 97 mutations were detected from 61 white or light blue colonies that were isolated and sequenced (see Tables 1 and 2). Among the mutant colonies, 22 out of 61 contained more than one mutation. Out of 97 mutations 70 were frameshift mutations. The rest were base substitution errors. The total frameshift error rate was 3.63×10^{-4}, as calculated by the procedure described in Materials and Methods. Mutations were divided into two catagories. The first, designated slippage mutations, were presumed to be formed by temporary repositioning of the primer terminus on the template. These include expansion or contraction of a homopolymeric segment by 1 or 2 nt or a single nucleotide deletion in a heteropolymeric region. This type of mutation composed 56% of the total mutations (see Table 2) and displayed an error rate of 2.82×10^{-4} (Table 2). By both definition and natural frequency most of the slippage mutations clustered in runs of a common base, as shown in Figure 2. One of the hot spots for frameshift errors was 5′-GGGAAA-3′ at template positions 18–23 (Fig. 2), which corresponded to a strong pause site during plus strand DNA synthesis. The slippage mutations included +1, –1 and –2 and 41 of them were –1 or –2 frameshifts. Among the slippage mutations, 56% were single mutations. The rest appeared as two changes in a

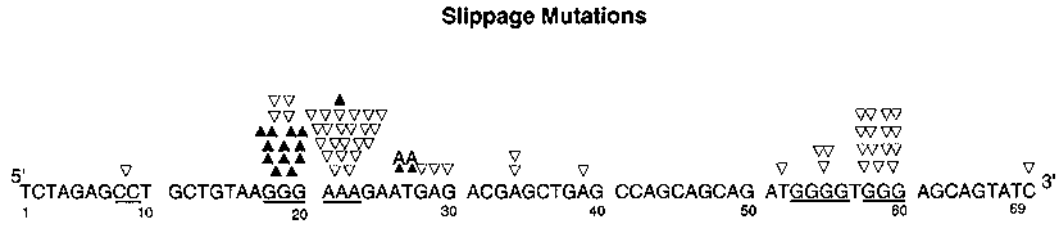

Figure 2. Spectrum of slippage mutations made by HIV-1 RT during DNA-directed DNA synthesis. Shown in the figure is the examined 69 nt sequence corresponding to positions 33–101 of the HIV-1 *nef* gene (+) strand. It represents the 5′→3′ nascent DNA strand copied from the (–) DNA template. Slippage mutations (–1, –2, +1 and +2) are displayed above the wild-type sequence. The +1 additions are shown as ▲ and +2 additions as ▲▲. The –1 and –2 deletions are presented as ▽ and ▽▽ respectively. When a frameshift occurs within a homopolymeric sequence it is not possible to assign the mutation to an individual base. Therefore, the run of sequence is underlined and the symbols are centered above it. The numbers underneath the sequence show the positions of the examined insert.

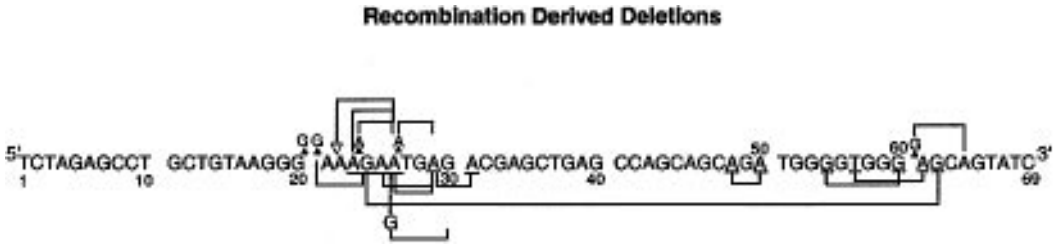

Figure 3. Spectrum of recombination-derived deletion mutations made by HIV-1 RT during DNA-directed DNA synthesis. The sequence shown is the same as that in Figure 2. The mutations displayed are simple deletions >2 nt. The nucleotides at the 5′ and 3′ deletion junctions are underlined. The deleted internal sequence is indicated by a bracket. The numbers underneath the sequence indicate the positions of the inserted DNA.

sequence having two adjacent triplets of a repeated nucleotide, 5′-GGGAAA-3′ (primer sequence), encompassing positions 18–23 in Figure 2. These multiple mutants were apparently generated by a dislocation mechanism (29). A transient slippage between the primer and the template occurred at the template sequence 5′-GGG-3′ at positions 18–20 (Fig. 2), which introduced an error. Realignment between the primer and the template generated a mispaired primer–template terminus. Extension of the mispaired terminus caused the second error in the 5′-AAA-3′ sequence at positions 21–23. This result shows that two adjacent runs of sequence can generate a hot spot for hypermutation.

Table 1. Mutation frequencies for reaction with DNA and RNA templates

Reaction carried out by reverse transcriptase	No. of colonies		Mutation frequency
	Total	Mutant	
DNA directed DNA synthesis	3695	61	1.6×10^{-2}
RNA directed DNA synthesis in the absence of acceptor	3237	12	0.37×10^{-2}
RNA directed DNA synthesis in the presence of acceptor	7583	57	0.75×10^{-2}

The second catagory, designated recombination-derived deletion mutations, were presumed to be generated by an intrastrand recombination mechanism. They include deletions of ≥2 nt and changes involving deletion and addition of nucleotides. Fourteen such mutations were isolated (Fig. 3), which exhibited an error rate of 0.71×10^{-4}. The largest deletion was 37 nt. Out of 14 deletion mutations, 12 were made by removing a short sequence between 2 nt having the same base. Furthermore, 10 of the deletion mutations started from within a stretch of sequence 5′-GGGAAAGAA-3′, where HIV-1 RT paused extensively during plus strand DNA polymerization. In four cases frameshift or base substitution errors occurred at the 5′-end of the deletion junction. Then the misincorporated nucleotide usually formed a single correct base pair with another nucleotide downstream on the template strand to serve as a functional primer terminus. Extension of the misaligned primer terminus deleted the internal sequence. These results suggest that the mechanism of deletion formation is similar to that of the strand transfer reaction carried out by HIV-1 RT.

Complex deletions

We detected two complex deletions in the mutant collection during DNA-directed DNA synthesis. The mutations exhibited a deletion of 21 nt, accompanied by a 36 nt insertion. As shown in Figure 4, the complex deletion was generated by two sequential intrastrand template switchings. When the growing point of the nascent strand reached a palindromic sequence 5′-TCTAGA-3′ (underlined in Fig. 4) the primer looped back to form three base pairs, providing a functional self-primed primer–template. This was the first template switch. The previously synthesized nascent DNA strand was then used as a new template, on which synthesis continued for 38 nt until the primer reached another palindromic sequence 5′-GAATTC-3′. Subsequently base pairing in the newly synthesized DNA was disrupted, allowing annealing of the nascent DNA back to the original (–) strand DNA template. Continued polymerization generated this specific type of mutation. Obviously, the likelihood of generating a complex deletion is much lower than that of making a frameshift or simple deletion. Nevertheless, detection of this kind of mutation still reflects the unique property of HIV-1 RT to carry out strand transfer.

Table 2. Frame shift error rates with DNA and RNA templates

Error	DNA directed DNA synthesis		RNA directed DNA synthesis		RNA directed DNA synthesis in the presence of acceptor	
	Occurence	Error rate[a]	Occurence	Error rate	Occurence	Error rate
Slippage mutation	54	2.82×10^{-4}	6	2.8×10^{-5}	35	6.2×10^{-5}
Recombination derived deletion	14	0.71×10^{-4}	–	–	–	–
Complex deletion	2	0.1×10^{-4}	–	–	–	–
Base substitution	27	–[b]	6	–[b]	30	–[b]
Total	97	3.63×10^{-4}[c]	12	2.8×10^{-5}	65	6.2×10^{-5}

[a]Error rate is determined by multiplying the percentage of each type of mutation by mutant frequency (shown in Table 1) and then dividing by the probability of expressing errors in all the scored colonies and also by the number of nucleotides examined.
[b]The system does not allow the selection of base substitution mutations. It is inappropriate to use the equation given above.
[c]After substracting the background PCR error rate, the total error rate for DNA directed DNA synthesis in 3.0×10^{-4}, as shown in the text.

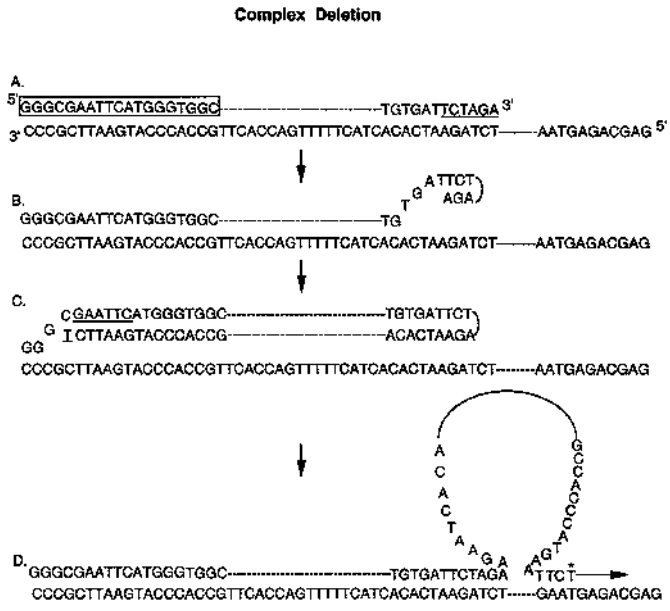

Figure 4. Complex deletion generated during DNA-directed DNA synthesis. The process is shown whereby a complex deletion with an insertion was generated. Only a portion of the template sequence is presented. The primer for DNA-directed DNA synthesis is boxed. The two palindromic sequences mentioned in the text are underlined. The dashed lines indicate the sequences not shown. (**A**) The growing point of the nascent strand reaches a palindromic sequence 5′-TCTAGA-3′ (underlined). (**B**) The primer loops back to form three base pairs that can be utilized as a functional primer template. (**C**) Template switching occurs, using the previously synthesized nascent DNA strand as a template. Synthesis continues for 38 nt until the primer reaches the sequence 5′-GAATTC-3′ (underlined). (**D**) Disruption of base pairing in the newly synthesized DNA allows a rearrangement involving annealing of the nascent DNA with the downstream DNA template. Continued polymerization on the DNA template fixes the mutation, generating a 21 nt deletion accompanied by a 39 nt insertion. The asterisk indicates a base substitution at the deletion junction.

Base substitutions

Although the procedure we employ does not specifically select for base substitution errors (Fig. 5), 27 were observed in the mutant collection. Three were detected because they generated termination codons in-frame that interrupted expression of the downstream *lacZ* α peptide gene. The rest were co-isolated with frameshift mutations. The sites of double mutations were generally well separated, suggesting that they were independently generated.

To test directly whether detection of these co-isolated base substitutions was related to the generation of frameshift mutants, we examined the frequency of base substitutions in blue colonies. DNA in these colonies did not have frameshift mutants, but could have undetected base substitutions. Five base substitutions were isolated from 14 blue colonies. This frequency (5/14) was close to the frequency of co-isolated base substitutions (27/97) presented above. This result indicates that the co-isolated base substitutions were not related to the frameshift mutants. They were generated by independent polymerization errors.

Experimental background mutations for DNA-directed DNA synthesis

Since the DNA template used in the reaction was generated by a total of 60 cycles of PCR (see Materials and Methods), it was necessary to examine the mutation background on the DNA template introduced by *Taq* DNA polymerase. The single-stranded DNA template used in the experiments was converted to double-stranded DNA through another 15 cycles of PCR. The DNA sample that eventually resulted from a total of 75 cycles of PCR was then subjected to the *lacZ* mutation assay. Fifteen single mutations were isolated from 3603 clones. The error rate after 75 cycles of PCR was calculated as 6×10^{-5}/nt, a value that is ~6-fold lower than the total error rate measured during DNA-directed DNA synthesis by HIV RT. Since extra cycles of PCR were applied to the experimental DNA to perform this control, the actual experimental PCR background error rate is even lower. This result was consistent with a previous determination that *Taq* polymerase exhibited 100–1000-fold greater discrimination against mismatch extension compared with avian myeloblastosis and HIV-1 RTs (30). About half of the PCR mutations detected were one base deletions clustered on 5′-GGG-3′ at positions 58–60 on the template, partially overlapping the mutation spectrum generated by HIV-1 RT (Fig. 2) in the same region. The rest were randomly distributed deletions and base substitutions that did not overlap the mutation spectrum generated by HIV-1 RT. The most important observation was that no deletion >1 nt was detected in the PCR mutation collection. This was consistent

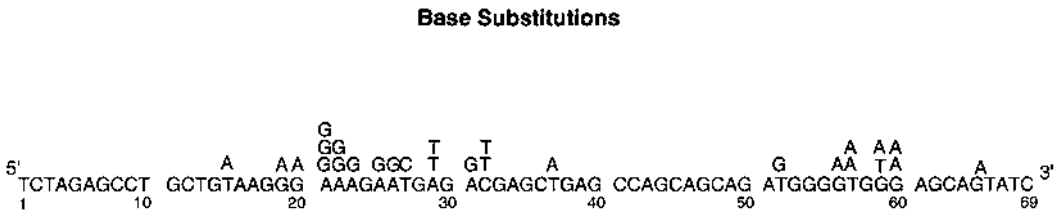

Figure 5. Spectrum of base substitutions produced by HIV-1 RT during DNA-directed DNA synthesis. The sequence shown is the same as that in Figure 2. The letters above the wild-type bases indicate the base substitution mutations found in the nascent DNA copied from the (–) strand DNA template. The numbers underneath the sequence indicate the positions of the inserted DNA.

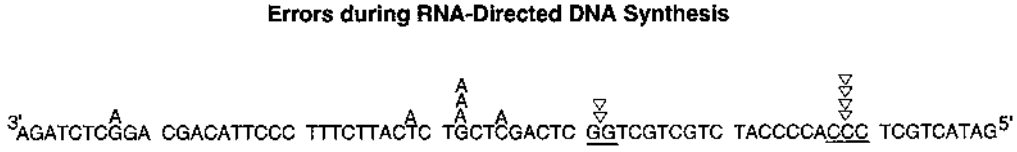

Figure 6. Spectrum of mutations made by HIV-1 RT during RNA-directed DNA synthesis. The examined 69 nt sequence corresponding to positions 33–101 of the HIV *nef* gene (–) strand is shown in the direction from 3′ to 5′, which represents the nascent DNA copied from (+) RNA template. The letters above the wild-type sequence indicate the base substitutions detected. Deletion mutations within a run of sequence are symbolized by ∇. The run of sequence is underlined.

with previous data showing that PCR did not generate deletion mutations > 1 nt (31). After subtracting the background error rate during PCR the total error rate of DNA-directed DNA synthesis was 3.0×10^{-4}. Overall, these results indicate that 80–90% of the frameshift mutations isolated during DNA-directed DNA synthesis were introduced by HIV-1 RT, rather than by *Taq* DNA polymerase.

The mutation spectrum for RNA-directed DNA synthesis

Compared with the frameshift error rate of DNA-directed DNA synthesis, the error rate for the generation of frameshift mutations during RNA-directed DNA synthesis was ~10-fold lower (see Table 2). Among the 3237 clones, only 12 single mutations were isolated (Fig. 6). Frameshift errors comprised half of the mutants. The rest were base substitutions. Interestingly, no deletion > 1 nt was isolated. The results demonstrate that the efficiency of generating frameshift mutations was much lower during minus compared with plus strand DNA synthesis.

The mutations isolated during RNA-directed DNA synthesis could have originated from three potential sources: transcription errors generated by T7 RNA polymerase, reverse transcription errors caused by HIV RT and PCR errors by *Taq* DNA polymerase. To estimate the error rate generated by T7 RNA polymerase, we compared the error rates of HIV RT and MuLV RT. We observed that the error rate of MuLV RT is 6-fold lower than that of HIV RT. Since the maximum error rate of T7 RNA polymerase should not exceed the total error rate during reverse transcription by MuLV RT, this indicates that the errors induced by T7 RNA polymerase are no larger than 17% of the total errors generated during reverse transcription by HIV RT. We also determined the error rate by *Taq* DNA polymerase after 20 cycles of amplification of the template directly derived from plasmids. The result showed that the error rate generated by PCR amplification was 10-fold lower than that of HIV RT. So the sum of the errors induced by T7 RNA polymerase and *Taq* DNA polymerase were ~27% of the total errors detected during RNA-directed DNA synthesis by HIV RT. About 70% of the mutations were generated solely by HIV RT. At this level of mutagenesis the adjusted error rate (1.96×10^{-5}) during RNA-directed DNA synthesis is still much lower than the error rate of DNA-directed DNA synthesis.

The difference in error rate could lie in a fundamental difference by which the RT interacts with RNA compared with DNA templates. Even if this were so, the error rates on either template might be influenced by reaction conditions that affect the strength of protein–nucleic acid interactions, such as ionic strength. Our standard assays on RNA and DNA templates were carried out at the same ionic strength. To examine the possibility that ionic strength influences mutagenesis on RNA or DNA templates we titrated the concentration of salt in the reactions. For DNA-directed DNA synthesis, the reactions were carried out at 20, 80 and 120 mM KCl. In each case the mutation rate, defined as the proportion of white colonies in the sum of the white and blue colonies, remained unchanged. For RNA-directed DNA synthesis the reaction was carried out at 20 and 80 mM KCl. Similarly, the mutation rate was not affected by the change in salt concentration. Evidently the differences in mutation rate observed during RNA-directed versus DNA-directed DNA polymerization catalyzed by HIV-1 RT are not readily altered by changes in ionic strength.

Spectrum of mutations induced by RNA-directed DNA synthesis in the presence of acceptor templates

The mutation rate caused by RNA-directed DNA synthesis was measured in the presence of acceptor templates. The system applied here allowed detection of a very low frequency of mutation at the recombination junction. As shown in Table 2, the total error rate for minus strand DNA synthesis increased 2-fold in the presence of acceptor templates. Compared with the mutation spectrum of RNA-directed DNA synthesis in the absence of acceptor templates mutations appeared at several new positions. Most of them were clustered in five stretches of sequence (boxed in Fig. 7). Since the donor template in the strand transfer assay was similar to that used before in determining the recombination sites, except for insertion of an *Xba*I site (22), we could predict the likely sites of primer strand

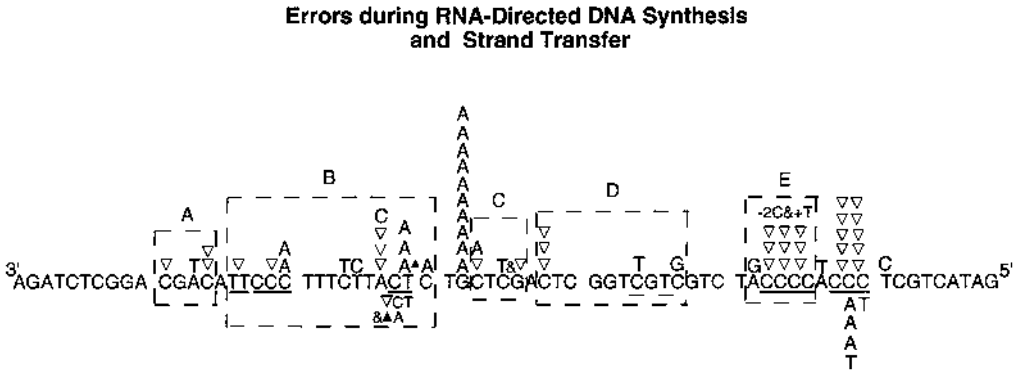

Figure 7. Spectrum of mutations produced by RNA-directed DNA synthesis in the presence of acceptor templates. The sequence shown is the same as that in Figure 6. Letters above the wild-type sequence indicate base substitutions. ∇ represents a deletion mutation. ▲ indicates an insertion mutation. Compared with the mutation spectrum of RNA-directed DNA synthesis in the absence of acceptor templates, mutations occurred at several new positions where transfer was supposed to take place. These new places, labeled A–E, are boxed with a dashed line.

transfer occurring in the assay here. Three of the clusters of mutations in Figure 7, boxed and labeled as regions A, D and E, correlated with zones R5, R3+R2 and the region downstream of R1 where primer strand transfer had previously been found to occur (22). Regions B and C did not fall into previously localized transfer zones. However, the mutations observed in the sequence element TAGC in region B could be caused by a different mechanism, misincorporation-driven transfer, in which the transfer is induced by misincorporation rather than pausing during DNA synthesis. This mechanism has been described by Temin (33). Especially interesting was a stretch of sequence CCCC, labeled region E in Figure 7. Ten mutations appeared in the area, including –1 and –2 slippage mutations. This contrasted sharply with the observation that no mutation was detected in the same position during RNA-directed DNA synthesis in the absence of acceptor templates. We have already observed from previous experiments that the enzyme shows a strong pause around this area (22). Inclusion of this site in the homology region during strand transfer caused a great increase in strand transfer efficiency (data not shown). This suggests that the new mutations detected in this area in the presence of acceptor templates were caused during the strand transfer event. Runs of the same sequence appear to be more error prone in terms of generation of transfer-related mutations if they are present at a recombination junction. These results indicate that recombination can be another source of mutation during viral replication.

DISCUSSION

Utilizing a novel *nef–lacZ* α complementation assay system we have compared the error rates for frameshift mutations during DNA-directed plus strand DNA synthesis and RNA-directed minus strand DNA synthesis. Results showed that the contribution of RNA-directed and DNA-directed DNA synthesis to the generation of frameshift mutations was asymmetrical. DNA-directed DNA synthesis appeared to produce frameshift mutations at a frequency that was 10-fold higher than RNA-directed DNA synthesis. Since our system did not allow specific selection of base substitution mutations, we cannot compare the relative contributions of RNA-directed and DNA-directed DNA synthesis to this type of mutation.

We did not use RNA and DNA templates of the same sequence in our assay because we tried to simulate viral genome replication *in vivo*. Utilization of the same sequence for RNA and DNA templates has the advantage of direct comparison, but the change in template from RNA to DNA has numerous consequences. For example, positions of synthesis pausing, which are different on RNA versus DNA templates, correlate with positions where mutations are generated (33). HIV-1 RT tends to pause in runs of rGs and rCs on an RNA template, but in runs of dAs and dTs on a DNA template (34). In general, a homopolymer run with high termination probability correlates with high frameshift frequency (33). Therefore, even if the same sequence of RNA and DNA templates were used, the pattern of pause sites and their effects on fidelity on the DNA and RNA templates would still be different. Consequently the comparison of the error rate of polymerization at one particular position on RNA and DNA templates of the same sequence is not particularly more informative than a comparison of large regions of varying sequence. The final conclusion would still have to be drawn from the total error rate, similar to the situation when DNA and RNA templates of the complementary sequences were used, as in the work presented here.

It was found in our experimental system that DNA-directed DNA synthesis was more prone to generate frameshift mutations than RNA-directed DNA synthesis. This might be caused by different interactions between HIV-1 RT and the primer–template during DNA-directed and RNA-directed DNA synthesis. Previous data have shown that the dissociation rate constant of HIV-1 RT for a DNA-primed DNA template was 5–10-fold higher than the DNA-primed RNA template, while the association rate constant for both types of substrates was similar (35,36). This indicates that HIV-1 RT binds with less stability to a DNA-primed DNA template in comparison with a DNA-primed RNA template. Affinity measurements are consistent with the observation that processivity during DNA-directed plus strand DNA synthesis is lower (37). Lower processivity also correlates with a higher error rate (33). During dissociation of the enzyme from the primer-template or reassociation after dissociation there could be distortion of the primer–template, which would lead to separation of the nascent primer from the template. Reformation of base pairing might be imperfect, but still provide a functional primer–template for further polymerization, generating misalignment-driven mutation. For HIV-1 RT its intrinsic property of catalyzing strand transfer would further promote this type of mutation during DNA-directed DNA

synthesis. Such a process is similar in mechanism to intrastrand template switching between regions with limited homology.

The rate of base substitution during DNA-directed DNA synthesis detected in our assay system (5×10^{-3}), estimated by dividing the percentage of base substitution errors in the blue colonies isolated by the number of nucleotides examined, is ~10-fold higher than the 6×10^{-4} determined previously *in vitro* (38). It has been shown that the mutation rate is dependent on local nucleotide sequence. Mutation rates at two different locations can vary by 10–100-fold (39). The difference detected here could be due to a concentration of mutation hot spots within the examined fragment.

We observed that RNA-directed minus strand DNA synthesis did not contribute significantly to slippage mutations. This indicates that a misaligned primer–template was formed or extended with lower efficiency on the DNA-primed RNA template than on the DNA-primed DNA template. That presumption is partially supported by the evidence that A·C mispair formation with a DNA primer was ~10 times lower on a RNA compared with a DNA template of the same sequence (7). It was determined that the extension efficiency for the mismatched primer terminus was >5 orders of magnitude lower than that of the matched one (40). Therefore, formation and extension of a mismatched primer terminus is an inefficient process on DNA-primed RNA. The situation is likely to be the same for formation and extension of a misaligned primer terminus.

One might also predict that on a mispaired or misaligned DNA-primed RNA template, during the time that the enzyme is struggling to make the extension, the RNase H activity of HIV-1 RT would degrade the RNA template, preventing further extension of the primer terminus. This would make the generation of frameshift mutations on the RNA template less frequent.

Another consideration is that the stability of RNA–DNA hybrids is generally greater than DNA–DNA hybrids (41). This will result in relatively less template–primer slippage with a primed RNA template and lead to less frequent generation of frameshift mutations during reverse transcription. Such an assumption is reasonable, but might not be applicable to the situation of RNA-directed reverse transcription, during which the RNA template is constantly degraded by the RNase H activity of RT. It is unlikely in this situation that RNA–DNA hybrids are more stable than DNA–DNA hybrids. So the observed difference in the generation of frameshift mutations is more likely to be caused by a difference in the interaction between HIV-1 RT and the primer–template during DNA-directed and RNA-directed DNA synthesis.

Slippage mutations were observed to be generated predominantly during DNA-directed DNA synthesis in our assay system *in vitro*. However, this might not reflect the situation *in vivo*. The slippage mutations occurring during plus strand DNA synthesis result in viral DNA with looped out nucleotides that can be subjected to correction by host cell repair systems after integration of the viral DNA into a host chromosome. Accordingly, the actual error rate of this type of mutation observed *in vitro* would be higher than that observed *in vivo*. Furthermore, the actual error rate *in vivo* might depend on the type of host cell and the capabilities of its unique repair system. In the case of HIV this would affect its pathogenesis in different host cells.

The frequency of recombination during viral replication has been determined to be as high as 4% per replication cycle (18). Our results show that homologous recombination can introduce mutations at the recombination junction, suggesting that strand transfer is another source of mutation during viral replication. The enzymatic mechanism for this observation could be that incorporation of the initial nucleotide after binding of HIV RT to the primer terminus is more error prone than subsequent polymerization (12). So while the nascent primer transfers to the second RNA template during recombination, a process requiring reinitiation of synthesis by RT, the initial incorporation might be more error prone, generating mutations at the junction. This brings up a concern for the design of drugs that target HIV-1 RT. Many of these drugs affect the interaction of RT with the primer–template. Therefore, one possibility we need to take into consideration is whether the drug will increase the rate of recombination. Enhancement of the efficiency of recombination will result in an increased error rate, allowing quicker emergence of drug resistance.

We did not detect any duplications in our system during either DNA-directed or RNA-directed DNA synthesis. However, an imperfect duplication has often been detected among *nef* genes isolated from various isolates of HIV-1 (28). This indicates that the appearance of such a duplication might occur rarely during viral replication, but have a selective advantage *in vivo*. Alternatively, some cellular factor may promote generation of the duplication.

Mutagenesis in HIV is a complex process that increases the capacity of the virus to evade efforts at therapy. The frequency and nature of mutations are affected by the template used for DNA synthesis. Mutagenesis is also promoted during viral recombination. Cellular repair functions, the use of antiviral drugs and other environmental factors are likely to influence the formation, retention and propagation of mutations. Full understanding of the mechanisms involved is important to our efforts to control HIV infection.

ACKNOWLEDGEMENTS

We thank Drs Jasbir Seehra and John McCoy, representing the Genetics Institute, for the generous gift of HIV-1 RT. We also thank Dr Benjamin M. Blumberg for providing the *nef* J14 gene utilized as the template in the fidelity assay. This work was supported by NIH grant GM 49573 and in part by Core Grant CA 11198 to the University of Rochester Cancer Center.

REFERENCES

1 Coffin,J.M. (1986) *Cell*, **46**, 1–4.
2 Steinheuer,D.A. and Holland,J.J. (1987) *Annu. Rev. Microbiol.*, **41**, 409–433.
3 Saag,M.S., Hahn,B.H., Gibbons,J., Li,Y., Parks,E.S., Parks,W.P. and Shaw,G.M. (1988) *Nature*, **334**, 440–444.
4 Leider,J.M., Paleae,P. and Smith,F.I. (1988) *J. Virol.*, **62**, 3048–3091.
5 Katz,R.A. and Skalka,A.M. (1990) *Annu. Rev. Genet.*, **24**, 409–445.
6 Ji,J. and Loeb,L.A. (1992) *Biochemistry*, **31**, 954–958.
7 Varela-Echavarria,A., Garvey,N., Preston,B. and Dougherty,J.P. (1992) *J. Biol. Chem.*, **267**, 24681–24688.
8 Wabl,M., Burrows,P.D., Gabain,A.V. and Steinberg,C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 479–482.
9 Roberts,J.D., Bebenek,K. and Kunkel,T.A. (1988) *Science*, **242**, 1171–1173.
10 Roberts,J.D., Preston,B.D., Johnson,L.A., Soni,A., Loeb,L.A. and Kunkel,T.A. (1989) *Mol. Cell. Biol.*, **9**, 469–475.
11 Blank,A., Gallant,J.A., Burgess,R.R. and Loeb,L.A. (1986) *Biochemistry*, **25**, 5920–5928.
12 Boyer,J.C., Benenek,K. and Kunkel,T.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6919–6923.
13 Hubner,A., Kruhoffer,M., Grosse,F. and Krauss,G. (1992) *J. Mol. Biol.*, **223**, 596–600.
14 Baltimore,D. (1970) *Nature*, **226**, 1209–1211.
15 Temin,H.M. and Mizutani,S. (1970) *Nature*, **226**, 1211–1213.
16 Gilboa,E., Mitra,S.W., Goff,S. and Baltimore,D. (1979) *Cell*, **18**, 93–100.

17 Hu,W.S. and Temin,H.M. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1556–1560.
18 Hu,W.S. and Temin,H.M. (1990) *Science*, **250**, 1227–1233.
19 Goodrich,D.W. and Duesberg,P.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2052–2056.
20 Skalka,A.M. and Goff,S.P. (eds) (1993) *Reverse Transcriptase.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 Peliska,J.A. and Benkovic,S.J. (1994) *Biochemistry*, **33**, 3890–3895.
22 Wu,W., Blumberg,B.M., Fay,P.J. and Bambara,R.A. (1995) *J. Biol. Chem.*, **270**, 325–332.
23 Shields,A., Witte,O.N., Rothenberg,E. and Baltimore,D. 1978, *Cell*, **14**, 601–609.
24 Coffin,J.M., Tsichlis,P.N., Barker,C.S., Voynow,S. and Robinson,H.L. (1980) *Annls NY Acad. Sci.*, **354**, 410–425.
25 Spindler,K.R., Horodyski,F.M. and Holland,J.J. (1982) *Virology*, **119**, 98–108.
26 Voynow,S.L. and Coffin,J.M. (1985) *J. Virol.*, **55**, 67–78.
27 Blumberg,B.M., Epstein,L.G., Saito,Y., Chen,D., Sharer,L.R. and Anand,R. (1992) *J. Virol.*, **66**, 5256–5264.
28 Shugars,D.C., Smith,M.S., Glueck,D.H., Nantermet,P.V., Seillier-Moiseiwitsch,F. and Swanstrom,R. (1993) *J. Virol.*, **67**, 4639–4650.
29 Kunkel,T.A. (1992), *J. Biol. Chem.*, **267**, 18251–18254.
30 Huang,M.M., Arneim,N. and Goodman,M.F. (1992) *Nucleic Acids Res.*, **20**, 4567–4573.
31 Keohavong,P. and Thilly,W.G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9253–9257.
32 Bebenek,K., Abbots,J., Roberts,J.D., Wilson,S.H. and Kunkel,T.A. (1989) *J. Biol. Chem.*, **268**, 16948–16956.
33 Temin,H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6900–6903.
34 Klarmann,G.J., Schauber,C.A. and Preston,B.D. (1993) *J. Biol. Chem.*, **268**, 9793–9802.
35 DeStefano,J.J., Bambara,R.A. and Fay,P.J. (1993) *Biochemistry*, **32**, 6908–6915.
36 Yu,H. and Goodman,M.F. (1992) *J. Biol. Chem.*, **267**, 10888–10896.
37 Huber,H.E., McCoy,J.M., Seehra,J.S. and Richardson,C.C. (1989) *J. Biol. Chem.*, **264**, 4669–4678.
38 Weber,J. and Grosse,F. (1989) *Nucleic Acids Res.*, **17**,1379–1393.
39 Kunkel,T.A. (1992) *J. Biol. Chem.* 267, 18251–18254.
40 Creighton,S., Huang,MM., Cai,H., Arnheim,N. and Goodman,MF. (1992) *J. Biol. Chem.*, **267**, 2633–2639.
41 Roberts,R. and Crothers,D.M (1992) *Science*, **258**, 1463–1465.