Université d'Ottawa · University of Ottawa

Recombinant Human Immunodeficiency Virus Reverse Transcriptases: Activity and Fidelity

A thesis submitted to the School of Graduate Studies at the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Science, Department of Microbiology and Immunology, Faculty of Medicine.

By Abdul-Rahman Zeibdawi, B.Sc.

ABSTRACT

Sequence variation in the human immunodeficiency virus type 1 (HIV-1) results, in part, from inaccurate replication by the enzyme reverse transcriptase (RT). Many forms of HIV-1 RT purified from virions have been found to be enzymatically active. I have examined the fidelity of both the DNA-dependent DNA polymerase and the RNA-dependent DNA polymerase activities of five different forms of recombinant HIV-1 RT. The enzymes were expressed in insect cells using the baculovirus as an expression vector. The DNA synthesis fidelity assay was based on the frequency with which mutations were introduced into the *lacZ* gene of Escherichia coli (E.coli). RNA and DNA of the *lacZ* reporter sequence were used as templates for DNA synthesis and the frequency of mutation was determined by the frequency of appearance of light blue or colorless colonies of E.coli on reporter plates. My results show that these enzymes are active and are error-prone during synthesis in vitro with DNA and RNA templates. Overall, fidelity of synthesis from an RNA template is lower than that from a DNA template. Sequence analysis of mutants generated with the two substrates reveal that base substitutions are random, whereas deletions and insertions are base specific. 'Hot spots' are observed with the mature form (p66/p51) of RT but not with the precursor forms of RT. In both DNA and RNA-dependent DNA synthesis, the mature form of the RT (p66/p51) has the highest error rate and the precursors have significantly lower error rates. In addition, the HIV-1 integrase domain appears to confer higher fidelity to the mature p66/p51 polymerase. Since it has been reported that many forms of

HIV-1 RT have been found to be enzymatically active, I propose that the RT precursor also plays a significant role in the replication of the virus and that the hypermutability observed with the virus is a result of multiple factors which include RT and selection for surviving mutants. A better understanding of the primary function of the HIV-1 RT precursor could lead to the discovery of more effective RT inhibitors.

ACKNOWLEDGEMENTS

I wish to acknowledge the invaluable support and guidance of my graduate supervisor, Dr. Francisco Diaz-Mitoma. I thank Dr. M.T. Aye for his support. I also thank my co-supervisor, Dr. Yu-Wen Hu, for his advice and discussions and for providing the recombinant viruses; Dr. Thomas A. Kunkel for providing the M13mp2 bacteriophage and host bacteria; and to Evan Balaskas for his technical help and support. Finally, this undertaking was made that much easier with the support and encouragements of my best friend Barbara.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AA.	Amino Acid
AIDS.	Acquired Immunodeficiency Syndrome
ABI	Applied Biosystems
AZT	Zidovudine
BSA	Bovine Serum Albumin
BP	Base Pair
cDNA	Complementary DNA
dH2O	Distilled H2O
DNA	Deoxyribonucleic acid
DTT	Dithiothreotol
ENV	Envelope
E.coli	Escherichia coli
EDTA	Ethylenediamine-tetraacetic Acid
EtBr	Ethidium Bromide
GAG	Group Antigen
HIV	Human Immunodeficiency Virus
In	Integrase
IPTG	Isopropylthio-B-D-galactoside
KDa	Kilo Dalton
Kb	Kilo Base

LTR	Long Terminal Repeat
LB	Luria-Bertani Medium
MuLV	Murine LV
NP40	Nonidet P-40
NIH	National Institute Of Health
POL	Polymerase
pORF	Protein Open Reading Frame
PCR	Polymerase Chain Reaction
PBS	Phosphate Buffer Saline
PMSF	Phenylmethyl-Sulfonyl Fluoride
PAGE	Polyacrylamide Gel Electophoresis
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
SDS	Sodium Dodecylsulfate
TE	Tris-EDTA
TBE	Tris-borate/EDTA
UV	Ultra Violet
YT	Yeast Trypton
X-Gal	5-Bromo-4-chloro-3-indolyl-B-D-galactoside

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Acquired immunodeficiency syndrome (AIDS) was first discovered in 1981 in young, promiscuous male homosexuals and injection drug users (1). AIDS is caused by the human retroviruses, human immunodeficiency virus types one and two (HIV-1, HIV-2) (2, 3, 4, 5), and is characterized by marked weight loss, opportunistic infections and unusual types of malignancies resulting from a compromised immune system (1). The predominant mode of transmission of HIV is sexual transmission, but blood-to-blood contact via contaminated needles and syringes used by injection drug users (1, 6), or the infusion of contaminated blood and blood products (prior to the implementation of HIV screening tests) has also been responsible for a significant number of cases.

HIV-1 infection has reached epidemic proportions throughout the world, and has quickly become ubiquitous (1). Cases have now been reported in bisexual and heterosexual populations of both males and females. As of December 1995, the World Health Organization reported that 1,291,810 AIDS cases had been identified worldwide and that half of these cases are in America alone. In Canada, as of December 1995, 11,192 AIDS cases had been reported to the Federal Center for AIDS. It is expected that by the turn of the century, 44 million people will be HIV-positive; including 45,000 Canadians. The young, gay male populations are expected to account for the majority of these new cases (7).

HIV has been classified as a lentivirus within the family of RNA retroviruses (2, 8, 9, 10, 11). Morphologically, the virus has a bullet-shaped core consisting of genomic RNA and viral proteins, including reverse transcriptase (p66/p51), integrase (p31) and nucleocapsid protein (p24). Surrounding this core, is an envelope derived from the cellular membrane of the host cell, along with virally encoded gp120 and gp41 proteins associated with a myristylated protein, p17 (Figure 1) (12, 13). The genome, like the genome of all known retroviruses, is formed by two homologous RNA molecules that are physically linked within the viral particles (14). The genetic structure of HIV-1, along with the various gene products are shown schematically in Figure 2. The gag and pol genes are initially translated as a gag-pol polyprotein precursor (p170) and a gag protein precursor (p155). The gag-pol precursor is then processed into three mature proteins: protease protein (p10), reverse transcriptase (p66 and p51) including RNase H, and integrase protein (p31) (15).

HIV-1 has been found to exhibit extensive genomic heterogeneity (16, 17, 18, 19, 20, 21). All viral isolates constitute a mixed population of closely related, yet genetically distinct genomes (22). Nucleotide sequences of the viral isolates suggest that the mutation rate is at least 10^{-3} nucleotide substitutions per site per year for the env gene and 10^{-4} for the gag gene (16, 23). These values are about 1 million-fold higher than for most DNA genomes (24, 25). This hypermutability is likely a significant factor contributing to the virus's success, perhaps accounting for the enhanced ability of the virus to evade the host's immune system (26, 27, 28, 29). As

Figure 1: schematic representation of the immature and mature HIV-1 particle structure. Gag-pol precursor and gag precursor are found in the immature HIV-1 particle. The gag-pol precursor is processed during the maturation stages to yield gag proteins, protease, reverse transcriptase, and integrase, which are found in the mature particle form (diagram after National Institute of Health, 1994).

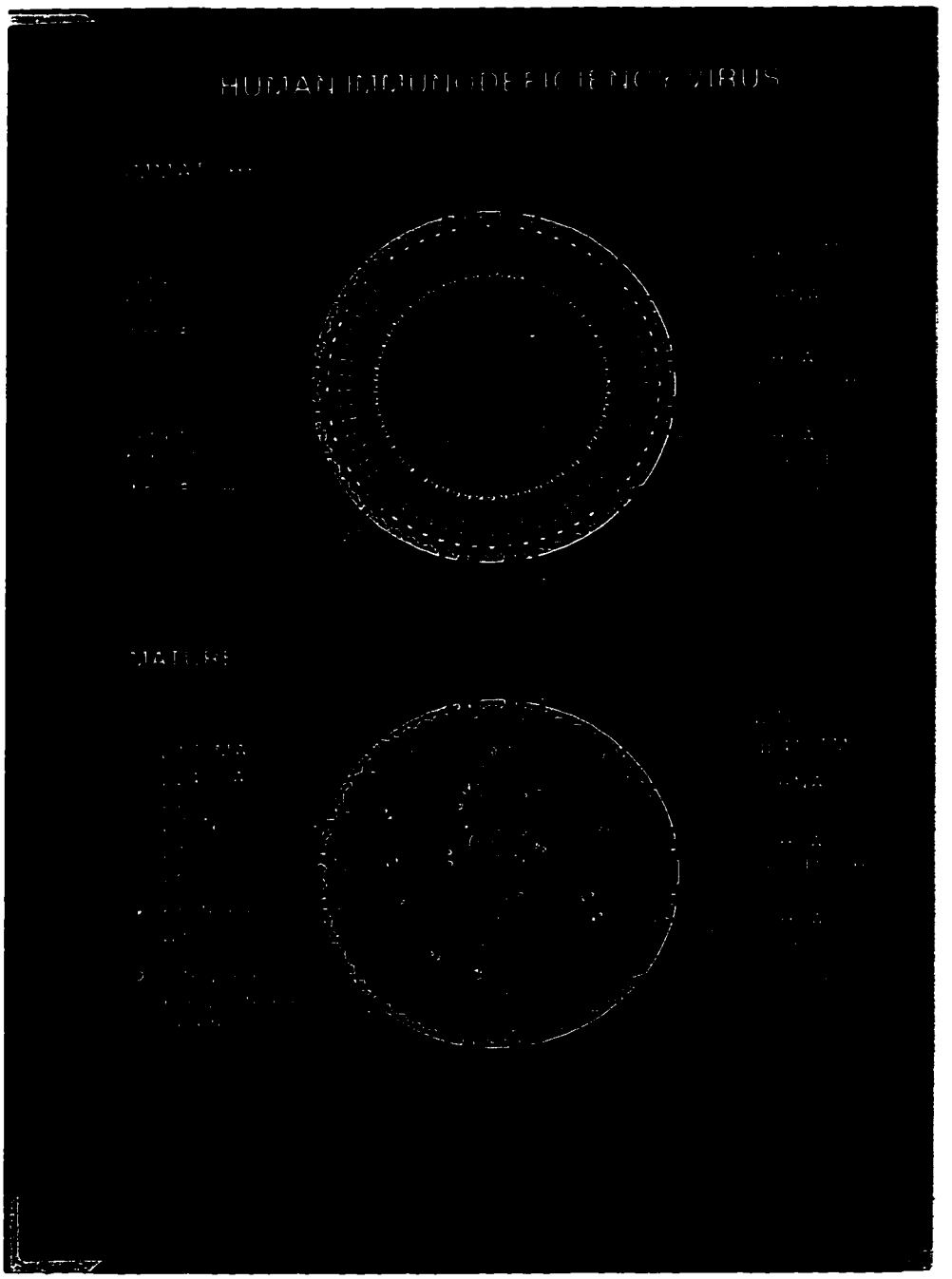
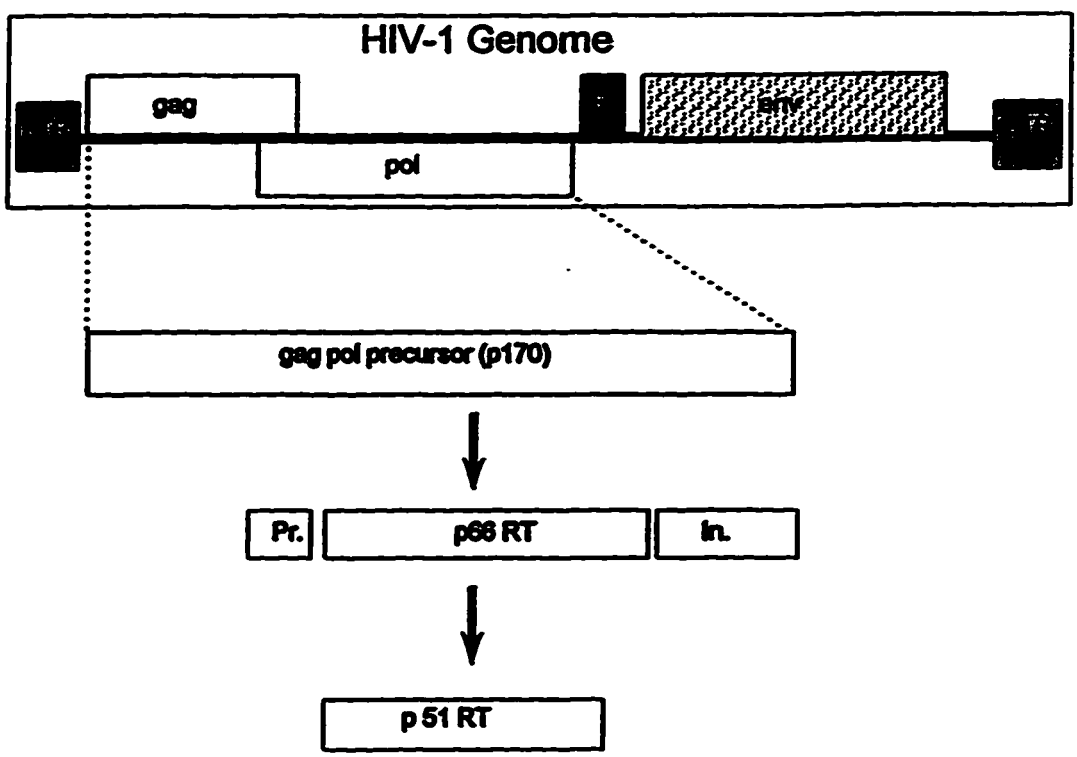


Figure 2: diagram of the genetic structure and steps in the processing of gag-pol precursor protein in the HIV-1 particle. The long terminal repeats (LTR) and the three major genes of HIV-1 are shown. Gag-pol (p170) proteolytic processing leads to proteins p66 RT, p51 RT and (p31) integrase.


HIV-1 Genome
gag
pol
env
gag pol precursor (p170)
Pr.
p66 RT
In.
p 51 RT

a result, this hypermutability could present a significant impediment to the development of an effective vaccine against the viral coat protein (30, 31, 32, 33).

DNA sequence analysis of the env gene from different viral isolates demonstrates that mutations include predominantly single base substitutions, short deletions and insertions (34, 35, 36, 37). These mutations could arise during any of several steps in the virus's life cycle. Three different replication systems operate during the life cycle of the virus (15, 38): reverse transcription, which polymerizes DNA, first using viral RNA as a template, then using the complementary short DNA; cellular DNA polymerase, which replicates the integrated viral DNA; and RNA polymerase II, which transcribes the proviral DNA into the RNAs that are packaged into virions. Recent evidence (20, 21,) has implicated the reverse transcriptase as the primary agent responsible for the high frequency of mutation of the virus.

RT is a multifunctional enzyme, possessing both RNA- and DNA-dependent DNA polymerase and RNase H activities. The RT of HIV-1 and other retroviruses is responsible for the conversion of the viral RNA to double-stranded DNA in the host cell cytoplasm (15) First, it synthesizes the complementary strand DNA from the RNA template, then the enzyme copies this DNA into double-stranded DNA and degrades the RNA template through its RNase H activity (15). The resultant DNA is then integrated into the host cell genome (provirus DNA) (39). Misincorporation by HIV RT could occur during synthesis from both DNA or RNA templates.

DNA-based and RNA-based misincorporation by the mature form of RT has been identified by several labs. Studies using several systems have demonstrated that with HIV-1 RT, there is a decreased fidelity relative to other retroviral RTs while copying DNA templates (16, 20, 21, 40,). The most likely explanation is that this is a result of the enzyme's ability to efficiently elongate mismatched 3' termini of DNA (41). The estimated error rate in vitro of HIV RT copying a DNA template was found to be about 1 in 5,000 polymerized nucleotides (16, 21, 20, 40, 42).

To date, tremendous effort has been directed towards the development of anti-HIV drugs that target the reverse transcriptase (26, 30, 43, 44). In fact, many of the drugs currently employed against HIV are nucleoside analogs (44) (such as 3'-azido-3'dideoxythymidine or AZT; recently renamed zidovudine) that interfere with HIV replication by specifically inhibiting incorporation of nucleotides by the viral RT and leading to termination of synthesis (45). In clinical trials, therapy with zidovudine has been shown to prolong survival (44, 46, 47). Other drugs (such as Nevirapine) are non-nucleoside inhibitors (31, 38) which specifically recognize the three dimensional structure of the DNA and RNA binding groove of the p66 subunit of the enzyme between residues 181 and 188 (43, 48), and hence lead to inhibition of synthesis by the enzyme (43). Unfortunately, the efficiency of nucleoside and non-nucleoside drugs has been very limited because resistant HIV-1 strains appear rapidly during therapy (31, 32, 33, 49, 50, 51, 52). This resistance has been recognized to be primarily a result of one or multiple mutations in the reverse transcriptase gene (27, 28, 29).

Current research has focused on the p66/p51 heterodimer (Figure 3) as the mature natural form of the RT. Effort has been expended on this form of the enzyme because it is mainly found in the mature form particle (15, 38, 53). However, many other forms of HIV-1 reverse transcriptase purified from virions, have been found to be enzymatically active (37, 53, 54, 55). Lori et al. (54), demonstrated that the two major forms of RT (p66 and p51) are independently active and that a precursor p165 was also enzymatically active. Hu et al. (55) had constructed HIV-1 precursors in vitro and were able to demonstrate polymerase activity in many forms of the enzyme.

The rational design of drugs against HIV-1 has been hampered by a lack of detailed knowledge on the precursor proteins in the immature virus. The identification of the role played by the precursor in the replication process and/or in the evasion of host immunity is pivotal for the determination of whether or not these precursor proteins constitute viable targets for antiviral therapy. This could ultimately have a direct impact on the current direction of drug therapy research.

In 1988 three labs reported that HIV RT, at least in part, was responsible for the hypermutability of the AIDS virus (20, 21, 40). None of these labs, however, looked at the fidelity of precursor-form RT. Instead, they all used either the bacterially expressed mature form RT from E.coli or purified mature form HIV RT.

Figure 3: schematic representation of a molecular modeling of the p66/p51 heterodimer reverse transcriptase bound to nucleic acid. The heterodimer consists of one 66 KDa domain (p66; represented in blue and green) and one 51 KDa domain (p51; represented in red). The p66/p51 hetrodimer resembles a right hand: finger subdomain, which interacts with DNA or RNA template (upper left corner); palm subdomain, which holds DNA or RNA (shown in black) in place; and thumb subdomain, which binds the template and initiate synthesis. RNase H of the p66 domain is represented in green (TasMol version 2 beta 2).

Takeuchi et al. (40) used natural HIV RT from two independent isolates of HIV, HTLV-IIIb and HIV[GUN-1]. These enzymes were purified from the culture fluids by column chromatography on DEAE-cellulose and phosphocellulose. They determined the fidelity of DNA synthesis by these reverse transcriptases by measuring the rates of misincorporation of dCTP in the place of dTTP in cell-free DNA synthesis with polyadenylic acid as the template. The fidelity of HIV RT was found to be about one-third (1/30,000) of that of the reverse transcriptases of other retroviruses (1/90,000).

Roberts et al. (20) used both mature natural form and bacterially expressed recombinant (E. coli infected with a plasmid expressing the cloned gene) HIV-1 RT in their assays. To measure the DNA synthesis fidelity they developed a bacteriophage based assay system (M13mp2 fidelity assay system). In this system, the fidelity of in vitro DNA synthesis is determined for a 250-base target sequence in the *lacZ*α gene. Error scores are determined by mutations causing loss of this non-essential gene function (α-complementation). This forward mutational assay is thus capable of detecting frame-shift, deletion, duplication, and complex errors in addition to a large number of different base substitution errors at many sites. Roberts et al. found that HIV enzyme isolated from virus particles or from E. coli was exceptionally inaccurate, having an average error rate of 1/1700. This led them to conclude that the enzyme was the least accurate RT described to date, one-tenth as accurate as the polymerase isolated from avian myeloblastosis virus or murine leukemia viruses, which have average error rates of 1/17,000 and 1/30,000 respectively. In addition, DNA

sequence analysis of mutations generated by HIV polymerase, demonstrated that base substitution, addition, and deletion errors were all produced.

Preston et al. (21) concurrently reported the DNA-dependent error rate of purified HIV reverse transcriptase by using three different assays. They used a bacteriophage assay (ϕx174 *amber*3 assay) which measures the reversion frequency of an amber mutation. This assay focuses on base substitution errors that restore an essential gene function and measures these events at one, two or three template nucleotides. They found that HIV RT introduced base-substitution errors in DNA from the bacteriophage 0x174 amber3 at estimated frequencies of 1/2000 to 1/4000. This rate corresponds to approximately five to ten errors per HIV genome per round of replication in vivo. Analysis of misincorporation rates demonstrated that mismatches were produced with a specificity of A:C>>A:G>>A:A.

Since RT utilizes both RNA and DNA templates, the error rate of synthesis with an RNA template is also of importance. Up until 1992, the error rate generated by the RNA-dependent DNA synthesis had not been determined. Since then, however, many studies have attempted to determine the error rate of the RT with an RNA template (41, 56, 57, 58, 59, 60). In most cases, the data obtained with RNA templates were found to be rather heterogenous. Furthermore, unequal HIV RT error-rates were detected between RNA and DNA templates (59). In 1991, Hubner et al. (58) measured the fidelity of synthesis with RNA templates, and concluded that this

synthesis is less accurate than with a DNA template. However, Boyer et al. (59), in a similar study, developed a system based on the M13mp2 fidelity assay system and found the opposite, namely that the fidelity is several-fold higher with RNA templates as opposed to DNA templates. In Boyer's study, sequence analysis of mutants revealed that fidelity with RNA was >10-fold higher for substitution and minus-one nucleotide errors. Their conclusion was that the errors likely resulted from template-primer slippage. Furthermore they found that HIV RT synthesis with an RNA template-DNA primer was error-prone during incorporation of the first two nucleotides, perhaps due to aberrant enzyme-substrate interactions as synthesis initiates. In contrast to both of these studies, other groups (41, 60) have found that the fidelity of synthesis with an RNA template is comparable to the fidelity of synthesis with a DNA template.

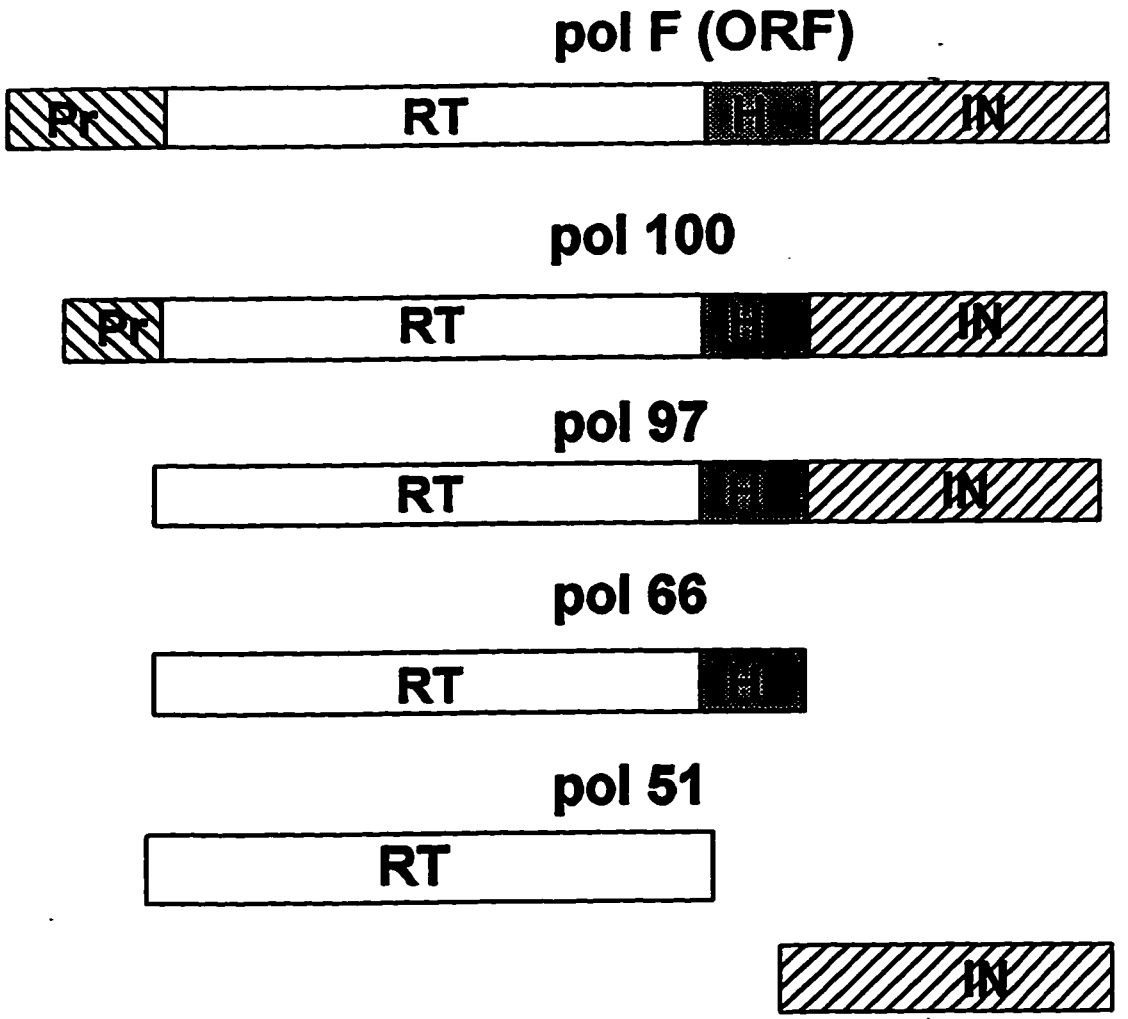
The research discussed above with regard to fidelity with RNA and DNA templates, assumes that the mutation rate of HIV-1 natural enzyme(s) is comparable to that of purified or bacterially expressed mature p66/p51 RT. To address this, Mansky and Temin (57) developed a system that allowed them to study mutations that occurred during a single round of HIV replication *in vivo*. In summary they found the *in vivo* mutation rate of HIV to be lower than the error rate of purified reverse transcriptase by a factor of 20. In their opinion, this difference was likely due to several factors, including the association of viral and nonviral accessory proteins during reverse transcription, the influence of cellular mismatch repair mechanisms,

and/or differences between the reverse transcriptase produced *in vivo* with that assayed *in vitro*.

Understanding the basis for the hypermutability of the virus by determining the role of the reverse transcriptase precursor in the virus's life cycle may prove to be crucial for the management of acquired immunodeficiency syndrome. To my knowledge after a comprehensive search of the literature, the RNA- and DNA-based DNA synthesis fidelity of the precursor has not been determined. Hence, this is the first study that compares the fidelity of recombinant precursors to the fidelity of the mature form p66/p51 and p66 and p51 separately, and compares them to that of the published reports. Regions that could be responsible for the observed differences will also be identified by comparing error rates from partially cleaved enzymes. The logical approach to this task was to express the precursor constructs (Figure 4) using the baculovirus system and employ the M13mp2 assay system (developed by Kunkel) (61) to determine the fidelity of these enzymes.

The M13mp2 DNA substrate (62), contains a 390 nucleotide gap in the *lacZ*α gene. Polymerization errors produced during *in vitro* gap-filling DNA synthesis were measured after transfection of E. coli strain MC1061 with the filled product (mutations being scored as light blue or colourless plaques resulting from loss of β-galactosidase activity (61)).

Figure 4: diagram of different forms of the recombinant pol gene products. The gene structure of different recombinant proteins used in this study is shown: pol F, complete reverse transcriptase precursor which includes both the protease and the integrase subdomains (1016 a.a.); pol 100, reverse transcriptase precursor with partially cleaved protease (912 a.a.); pol 97, the entire protease subdomain has been removed from the precursor (849 a.a.); pol 66, as in pol 97, and the entire integrase region has been removed (563 a.a.); pol 51, the RNase H subdomain has been removed from the pol 66 construct (439 a.a.); In (pol 31), construct that contains the entire open reading frame of the integrase subdomain.


pol F (ORF)
Pr
RT
H
IN
pol 100
Pr
RT
H
IN
pol 97
RT
H
IN
pol 66
RT
H
pol 51
RT
IN

In this study the main objectives were to determine the DNA-based and RNA-based mutation rate and type of mutations of five different recombinant HIV-1 reverse transcriptases and to compare them to published values for that of the mature form of RT. In this regard, the specific objectives were:

1) to implement and optimize the fidelity assay using bacteriophage M13mp2 and a compatible host cell, through a collaboration with Dr. Kunkel (who supplied the required phage and bacteria);

2) to test the recombinant proteins expressed using the baculovirus expression system for DNA polymerase and reverse transcriptase activity;

3) to use enzymatically active recombinant HIV-1 precursor proteins in the M13mp2 fidelity assay system in order to obtain DNA-based DNA synthesis and RNA-based DNA synthesis mutation rates for these enzymes;

4) to compare the mutation frequencies for each of the recombinant enzymes to published frequencies of the mature p66/p51 heterodimer (which is believed to be the active enzyme and the sole enzyme for reverse transcription);

5) to characterize the types of mutations encountered by sequencing and compare these to published reports for the mature form of RT (p66/p51);

6) to examine the mutation rates generated by the truncated forms (p100 and p97) in an attempt to predict the region or motif responsible for changing the error rate of the p66/p51 heterodimer; and

7) to construct the subunit (protease or integrase) and determine its effect on the mature protein p66/p51 in order to confirm the region predicted in objective 6.

It is hoped that in accomplishing these objectives, it can be shown that the precursor RT is enzymatically active and possesses a different error rate than the mature form RT for HIV-1. In so doing, perhaps further research into the viability of drug therapies aimed at precursor RT could be stimulated.

MATERIALS AND METHODS

Bacterial Strains

Escherichia coli strains NR9099 [Δ(*pro-lac*), *rec*A, *ara*, *thi*⁻/F'(*pro*AB, *lac*I_qZ ΔM15)] for growth M13 phage, MC1061 [*hsd*R, *hsd*M^+, *ara*D, Δ(*ara*, *leu*), Δ(*lac*IPOZY)], *gal*U, *gal*K, *str*A], CSH50 [Δ(*pro-lac*), *ara*⁻,*thi*⁻/F'(*tra*D36, *pro*AB, *lac*IqZΔM15)] and wild-type M13mp2 were obtained from Thomas A. Kunkel, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. Recombinant polymerase polyproteins polF, pol100, pol97, pol66 and pol51 were obtained from Yu-Wen Hu, Canadian Red Cross, National Office, Ottawa, Ontario. NIH reverse transcriptase (HIV-1 BH10 reverse transcriptase, # 454, 2 units/ug protein) was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID.

Nuclear and Cytoplasmic Preparations of Polymerase Polyproteins

Either Hi5 or Sf9 cells were infected with the recombinant baculoviruses containing pol F, pol 100, pol 97, pol 66 and pol 51 (supplied by Dr. Yu-Wen Hu, Canadian Red Cross). The flasks were harvested 72 hours post-infection; each in the same manner. The cells were loosened from the flask and removed to a sterile 50 ml tube. The suspension was spun down on a Beckman GP centrifuge for 10 minutes, at 1000 rpm and ambient temperature. The supernatant was removed (stored at 4 °C) and the pellet washed in 35 ml cold phosphate buffered saline (PBS) and centrifuged as

above. The wash supernatant was discarded, the pellet washed again in 20 ml cold PBS and centrifuged as above. The supernatant was again discarded and the pellet was resuspended in 500 ul cold PBS, transferred to a microcentrifuge tube and centrifuged for 7 minutes, at 4000 rpm and 4 °C, on an IEC Centra MP-4R centrifuge. The supernatant was discarded and 700 ul of cold cytoplasmic buffer was added (40 mM Tris pH 7.5, 0.5 mM EDTA, 0.5% (v/v) NP40, 0.1 mM NaCl, 400 ug/ml PMSF). The mixture was incubated for 20 minutes at 4 °C on a rocker then centrifuged in the IEC centrifuge for 7 minutes, at 4000 rpm and 4 °C. The supernatant (cytoplasmic fraction) was transferred to a sterile tube and an equal volume of glycerol was added before being stored at -20 °C. The pellet was washed once with 1 ml of cold PBS and centrifuged in the IEC centrifuge as before. The supernatant was discarded and 700 ul of nuclear buffer was added (40 mM Tris pH 7.5, 0.5 mM EDTA, 0.1% (v/v) TritonX-100, 0.1 mM NaCl, 400 u/ml PMSF). The mixture was sonicated for 25-30 seconds on ice using a 2 mm probe attachment on a Cole-Parmer Ultrasonic homogenizer (nuclear fraction). An equal amount of glycerol was added before being stored at -20 °C.

Measurement of Recombinant Protein

Total cell lysate containing different recombinant proteins was measured using the Bio Rad protein assay. On a 96-well plate containing 40 ul of Bio Rad dye reagent, 160 ul of diluted sample (1/10 dilution) or standard protein (1.57 ug/ml to 12.5 ug/ml of bovine serum albumin (BSA) in distilled water (dH_20)) were added and

mixed. Blank wells containing 160 ul dH_20 and 40 ul dye reagent were used as negative controls. Duplicates were performed for all reactions. The reaction mixtures were incubated for 10 minutes at room temperature and were read at 595 nm using a multi-scan plate reader (V max; Molecular Devices). Based on the concentrations obtained from the Bio Rad protein assay, equal concentrations of total protein were electrophoresed on a denaturing 6% polyacrylamide gel at 100 volts for 3 hours. Concentration of the recombinant protein was then estimated using an ultrasonic XL densitometer (LKB).

Preparation of Double-stranded M13mp2 DNA

Bacteriophage M13mp2 contains a segment of the E. coli *lac* operon within the intergenic region of the M13 phage. M13mp2 phage was plated on minimal plates (described below), using E. coli NR9099 as a host strain. A single plaque was added to 1/2 liter of 2 x YT medium (containing, per liter, 16 g Bacto-Tryptone, 10 g of yeast Extract, 5 g of NaCl, pH 7.4) containing 5 ml of an overnight culture of E. coli NR9099. M13mp2-infected cells were grown overnight at 37 °C with vigorous shaking in a controlled environment incubator shaker (series 25; Scientific Co., Inc.). Bacteria were harvested by centrifugation at 5000 x g for 30 minutes (Sorvall RC-5 superspeed refrigerated centrifuge (SS34); DuPont Instruments) and replicative form DNA was prepared by the method of Birnboim and Doly (63). The DNA was resuspended in 200 ul of Tris-EDTA (TE) buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

Preparation of single-stranded M13mp2 DNA.

M13mp2 phage were precipitated from the clear culture supernatant (described previously) by the addition of polyethylene glycol 8000 to 3% and NaCl to 0.5 M. The phage was peleted at 5000 x g (Sorvall RC-5 (SS34)) for 30 minutes at 4 °C and the phage pellet was resuspended in phenol extraction buffer (100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1 mM EDTA). The single-stranded phage DNA was purified by phenol, followed by two phenol:chloroform:isoamyl alcohol (25:24:1) extractions. The DNA was precipitated with ethanol and resuspended in 500 ul TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

Preparation of Gaped M13mp2 DNA

Double-stranded M13mp2 was digested at 37 °C with PvuI and PvuII restriction endonucleases (GibcoBRL). The digested DNA was subjected to electrophoresis in a 0.8% agarose gel in Tris-borate EDTA (TBE) buffer (0.5 x: 0.045 M Tris-borate, 1 mM EDTA (5 x: 54 g Tris base, 27.5 boric acid, 20 ml of 0.5 M EDTA (pH 8.0)) containing 0.5 mg/ml ethidium bromide. DNA was illuminated with UV light, and the 6.8 kilobase fragment was excised with a scalpel blade. DNA from the excised band was purified and concentrated using the GeneClean procedure (GeneClean II kit; BIO 101, (Bio/Can Scientific)) according to the manufacturer's procedure. The DNA fragment was then suspended in 50 ul TE buffer. The gaped double-stranded circular DNA molecule was prepared by mixing 100 ug each of the digested fragment and the single-stranded circular viral DNA (7196 bases) in 30 mM NaCl, 30 mM sodium

citrate. The mixture was heated to 95 °C for 15 minutes, cooled on an ice bath, and incubated at 65 °C for 30 minutes. The DNA was then subjected to electrophoresis in a 0.8% agarose gel and the gaped DNA molecule was excised and DNA was purified by the GeneClean II kit method as above.

DNA Synthesis Reaction

The DNA synthesis reaction (50 mM potassium phosphate (pH 7.5), 6.7 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 200 uM dATP, 200 mM dCTP, 200 mM dGTP, 200 mM dTTP) contained 100 ng of gaped circular M13mp2 DNA and 1 unit of enzyme) in 30 ul. In each reaction, the unit definition was that of the enzyme supplier, as determined in the supplier's laboratory before shipment. The reaction mixture contained either pol F, pol 100, pol 97, pol 66, pol 51, NIH RT, MuLV or AMV enzymes and was incubated at 37 °C for 90 minutes, and then terminated by addition of EDTA to 15 mM. The filled double-stranded DNA was resuspended in an equal volume of phenol, followed by a chloroform extraction. The DNA was precipitated with ethanol and resuspended in TE buffer and stored. 2 ul of each reaction were mixed with 1 ul loading buffer (6 x; 0.25% bromophenol blue, 0.25% xylene cyanol, 40% (w/v) sucrose in water) in 10 ul reaction volume and subjected to electrophoresis in a 0.8% agarose gel in TBE buffer. Electrophoresis was at a constant 80 volts for 16 hours. Bands were then illuminated with UV light and photographed.

Transcription Reaction

The transcription reaction was performed in 50 ul reaction (40 mM Tris-HCl (pH 8.0), 25 mM NaCl, 8 mM $MgCl_2$, 2 mM spermidine-$(HCl)_3$, 5 mM DTT, 0.4 mM each of dATP, dCTP, dGTP and dUTP) containing 2 ug of Fsp I-linearized M13mp2 DNA with the T7 RNA polymerase promoter, and 10 units of T7 RNA polymerase (Boehringer Mannheim). The reaction was incubated for 120 minutes at 37 °C. The remaining DNA template was digested with RNase-free DNase (GibcoBRL; Life Technologies, 1 unit per ug) for 15 minutes at 37 °C. The RNA was then purified with phenol, followed by two phenol:chloroform:isoamyl alcohol (25:24:1) extractions, precipitated with ethanol, redissolved in 50 ul of RNase-free water, and stored at -70 °C. 2 ul were mixed with 1 ul of 10 x sequencing gel-loading buffer (98% deionized formamide, 0.025% xylene cyanol FF, 0.025% bromophenol blue) in 10 ul reaction volume and subjected to electrophoresis on 6% denaturing polyacrylamide gel in 1 x TBE buffer. Electrophoresis was at 250 Volts for 5 hours. Bands were then illuminated with UV light and photographed.

cDNA Synthesis Reaction

cDNA synthesis was performed in 50 ul reaction (50 mM Tris-HCl (pH 8.3), 40 mM KCL, 6 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 500 uM each of dATP, dCTP, dGTP and dTTP) containing 1 ul RNase inhibitor (Promega; 10 u/ul), 1 pmol of RNA molecules (100 ng) primed with a 2-fold molar excess of a 15-mer DNA oligonucleotide (5' CGG TGC GGG CCT CTT 3'), and enzyme. The RNA was first

heated to 65 °C for 5 minutes before it was added to the synthesis reaction. The reaction mixture was incubated at 37 °C for 120 minutes and stopped by adding EDTA to 15 mM. The mixture was heated to 80 °C for 5 minutes and cooled on ice. The remaining RNA was digested with RNase A and RNase T1 (RNace-It cocktail; 10,000 u/ml, Stratagene) for 1 hour at 37 °C. The cDNA was purified with phenol followed by two phenol:chloroform:isoamyl alcohol (25:24:1) extractions, precipitated with 2 volumes of ethanol and resuspended in 30 ul sterile water. Next, 2 ul were mixed with 1 ul of sequencing gel-loading buffer in 10 ul reaction volume and subjected to electrophoresis on 6% denaturing polyacrylamide gel in 1 x TBE buffer. Electrophoresis was at 250 Volts for 5 hours.

Phosphorylation of cDNA

The 5' end of the cDNA was phosphorylated in 50 ul reaction (50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 1.5 mM spermidine, 1 mM ATP) containing 1 unit of T4 Polynucleotide Kinase (GibcoBRL) and cDNA. Reactions were incubated for 1 hour at 37 °C and were terminated by heating to 65 °C for 10 minutes.

Hybridization of cDNA Fragment to Gaped DNA

cDNA was mixed with gaped M13mp2 DNA (5:1) in 30 mM NaCl, 30 mM sodium citrate, heated to 70 °C and slowly cooled to room temperature. Hybridization products were subjected to electrophoresis in 6% denaturing polyacrylamide gel in 1 x TBE buffer and electrophoresed at constant 250 volts for 5 hours and visualized by

autoradiography.

Preparation of MC1061 Cells For Electroporation

A liter of 2 x YT (per liter, Bactotryptone, 16 g; Bactoyeast extract, 10 g; NaCl, 10 g; pH 7.4) medium was inoculated with 1/100 volume of a fresh overnight culture of E. coli MC1061. The cells were grown at 37 °C with vigorous shaking to an optical density at 600 nm of 0.5 to 0.8. Cultures were chilled on ice for 30 minutes, and the cells were transferred to a cold 1 liter centrifuge bottle and pelleted by centrifugation at 4000 x g (Sorval RC-5; GSA rotor) for 30 minutes at 4 °C. The pelleted cells were resuspended in 1 liter of ice cold dH_20, pelleted again at 2200 x g (Sorval RC-5; GSA rotor) for 20 minutes at 4 °C and resuspended in 500 ml ice cold dH_20. Cells were pelleted again as above and resuspended in 20 ml ice cold 10% glycerol, transferred to a 50 ml conical tube and pelleted at 3000 x g for 15 minutes at 4 °C. The cell pellet was resuspended in 3 ml of ice cold 10% glycerol and divided into 80 ul aliquots in 1.5 ml microcentrifuge tubes. Cells were then frozen rapidly in a dry ice/ethanol bath and stored at -70 °C.

Electroporation and Plating

The products from the gap-filling DNA synthesis reactions (DNA-based DNA synthesis reaction) and the cDNA hybridization reactions (RNA-based DNA synthesis reaction) were used to transform the MC1061 cells (described above). In a cold 1.5 ml polypropylene tube, 80 ul of the cell suspension were mixed with 4 ul of DNA

in TE. The mixture was mixed well and was placed on ice for 1 minute. The mixture of cells and DNA was transferred to a cold electroporation cuvette (0.1 cm). Transformation was carried out by electroporation at 18 KV/cm field strength (1.8 V, 5 milliseconds) using *the E. coli* pulser (BIO-RAD). Immediately following electroporation, 1 ml SOC medium (Appendix 1) at room temperature was added to the cells and the cells were quickly but gently resuspended with a pasteur pipette. The cell suspension was then transferred to a polypropylene tube, placed on ice for at least 3 minutes, then plated onto minimal agar plates within 30 minutes of electroporation to avoid the release of M13mp2 viral particles while the cells were still in suspension. For most reactions 1 to 10 ul was used in soft agar (described below).

Plating of Transformed Cells

Plating was performed by adding an appropriate volume of the electroporated cells (to yield about 400 plaques per plate in 1 to 10 ul) to a tube at 48 °C containing 2.5 ml of 0.8% soft agar (in 0.9% NaCl), 2.5 mg X-gal, 250 ug IPTG and 0.4 ml of mid log phase culture of E. coli CSH50. This mixture was poured onto plates containing 30 ml 1.5% agar in minimal medium (M9 plates: Appendix 2) and allowed to solidify. The plates were inverted and incubated for 18-20 hours at 37 °C, followed by an additional 24-48 hours incubation at room temperature. To prepare a large number of plates, the mixture of soft agar, IPTG and X-gal was prepared in a flask and kept at 48 °C; an aliquot of the electroporated cells was then added to the flask, followed by the addition of an aliquot of the CSH50 cells. The mixture was vortexed

briefly and 3.5 ml aliquots of the soft agar mixture were pipetted onto the minimal plates. This procedure allowed large numbers of plates (e.g., 30 plates at a time) to be prepared and plated easily.

Scoring Mutants

Lighter blue or colourless plaques are produced as a result of inactivation of the α-complementation resulting from mutation in the *lac* DNA in M13mp2 phage. Mutants were scored after 18-20 hours of incubation at 37 °C followed by 24 hours of incubation at room temperature. Colourless plaques were assigned a value of 0 and wild-type plaques a value of 3 (i.e., dark blue). Plaques of intermediate colour were assigned values based on colour, of 1 (light blue), or 2 (medium blue). Mutants (i.e., light blue or colourless plaques) were removed from the agar plate with a pasteur pipet and expelled into a tube containing 1 ml of 0.9% NaCl. A typical plaque contains about 10^9 to 10^{10} plaque forming units. The M13mp2 phage were allowed to diffuse into the 0.9% NaCl for 1 hour and the mixture was vortexed and stored at 4 °C. Some mutant phenotypes, as indicated by colour, were confirmed by streaking onto an agar lawn of CSH50 cells (as with the plating procedure) with a strip of sterile 3M Watman paper. Similarly, wild-type phage were streaked on an overlapping region of the same plate. For some mutant plaques, the presence of mutations was confirmed by DNA sequence analysis.

Sequencing of Mutant Plaques

To prepare DNA for sequencing, plaques were picked from the plates, being careful to avoid other plaques, and placed into a tube containing 100 ul dH_20. M13mp2 phage DNA was directly amplified from the plaque suspension using polymerase chain reaction (PCR) and two primers specifically designed to span the *lacZ* gene (sense primer: 5'-GCT CAC TCA TTA GGC ACC C, antisense primer: 5'-CCA CTG CAA GCT ACC TGC). A 2 ul aliquot of the plaque suspension was added to PCR buffer (100 mM Tris-HCl pH 8.2, 50 mM KCl, 2.5 mM $MgCl_2$, 200 uM of each of dATP, dCTP, dGTP, and dTTP) containing 20 pmoles of each primer and 2.5 units of Taq polymerase (GibcoBRL) in a 50 ul final volume. The reaction was first heated to 95 °C for 3 minutes to lyse the M13mp2 phage and denature the DNA followed by 30 cycles of denaturation at 95 °C for 18 seconds; annealing at 63 °C for 15 seconds; and extension at 72 °C for 35 seconds. Amplification products were analyzed on 1% agarose gel in TBE buffer containing 0.5 mg/ml ethidium bromide. DNA was briefly illuminated with UV light, the expected DNA fragment excised with a scalpel blade, and placed in a 1.5 ml microcentrifuge tube. DNA from the excised band was purified and concentrated using the GeneClean II kit according to the manufacturer's procedure. The DNA was then suspended in 20 ul TE buffer. An aliquot equivalent to 500 ng of the amplified and GeneClean purified DNA was used in the sequencing reaction. A TaqDyeDeoxy Terminator cycle sequencing kit (Applied Biosystems) and each of the sense and antisense primers were used for sequencing. Sequencing products were analyzed on the Applied Biosystems sequencing system

model 373A (ABI).

Amplification of the Integrase Domain

The HIV-1 Integrase domain was amplified from 8E5/LAV cells (containing one copy of integrated proviral HIV-1 DNA per genome) using two primers that span the integrase open reading frame (sense primer: 5'-ACG GAT CCT AAT GGA ATA GAT AAG G, antisense: 5'-ATC TGC AGC CCT AGC TTT CCC TG). A volume equivalent to 100 ng of genomic DNA was added to the amplification reaction (100 mM Tris-HCl pH 8.2, 50 mM KCl, 2.5 mM $MgCl_2$ 200 uM of each of dATP, dCTP, dGTP, dTTP, 20 pmoles of each primer and 2.5 units of Taq polymerase (GibcoBRL) in a 50 ul final volume. The reaction was first heated to 95 °C for 3 minutes to denature the DNA followed by 30 cycles of denaturation at 95 °C for 18 seconds, annealing at 56 °C for 18 seconds and extension at 72 °C for 1 minute. Amplification products were analyzed on a 0.8% agarose gel in TBE buffer containing 0.5 mg/ml ethidium bromide. The gel was briefly illuminated with UV light, and the 920 base pair DNA fragment was excised with a scalpel blade. DNA from the gel fragment was purified using the GeneClean II procedure. The DNA was then suspended in 25 ul of TE buffer.

Cloning of the Integrase Domain in the TA Vector:

GeneClean purified PCR product (described above) was first ligated to a PCR II vector (3.9 kb) using the TA cloning kit (Invitrogen). White colonies represent

PCR II vectors with the integrase gene inserted in the TA cloning site while blue colonies represent wild-type PCR II vectors without the integrase gene. Five white colonies and one blue colony (negative control) were picked from the agar plate and plasmid DNA was prepared by the method of Birnboim and Doly (63). DNA was then resuspended in 50 ul of TE buffer. DNA was digested at 37 °C with BamHI and PstI (GibcoBRL) for 3 hours and subjected to electophoresis and GeneClean purification as previously described. DNA was used for cloning into BamHI and PstI digested pVL1393 cloning vector (Baculovirus expression system (Pharmingen).

Subcloning of the Integrase Domain

A volume equivalent to 500 ng of pVL1393 vector DNA was digested with BamHI and PstI restriction endonucleases. Digested DNA was purified using the GeneClean II kit as previously described. An aliquot (100 ng) of this DNA was mixed with an equal amount of BamHI and Pst I digested integrase DNA (described in the previous section) and T4 DNA ligase in ligation buffer (66 mM Tris-HCl (pH 7.6), 6.6 mM $MgCl_2$, 10 mM DTT, 66 uM ATP) containing 20 units of T4 DNA ligase (GibcoBRL) in a 20 ul volume and incubated at 16 °C overnight. DH5-α competent cells (GibcoBRL) were transformed (by heat shock procedure) with 20 ng of the ligated DNA in a water bath at 45 °C for 45 seconds and immediately placed on ice. The cells were then suspended in 459 ul of SOC medium (Appendix 1), shaken at 37 °C and plated on LB plates (Appendix 3) containing 50 ug/ml of Ampicillin. PCR was used to detect those colonies that contained the recombinant pVL1393. Colonies

were picked from the plate, suspended in 30 ul of dH_20, vortexed vigorously for 5 seconds and spun briefly. An aliquot (20 ul) of this mixture was added to 3 ml of LB medium which was then incubated in an orbital shaker (250 rpm) overnight at 37 °C. The remaining suspension (10 ul) was used in the PCR. Amplification conditions were as described above for amplification of the integrase gene . PCR positive colonies were used to isolate the recombinant DNA. Double stranded DNA was prepared by the method of Birnboim and Doly (69) and was resuspended under sterile conditions in 50 ul of filtered sterile water. The presence of the integrase gene in the BamHI and PstI cloning site of pVL1393 was confirmed by restriction analysis (with BamHI and PstI) and by sequencing as described previously.

Expression of the Integrase Doamin in Insect Cells

The baculovirus system (Baculogold kit; Pharmingen) was employed to express the integrase gene. A volume containing 500 ng of the recombinant pVL1393 transfection vector was used in the cotransfection step according to protocols provided by the supplier. After three rounds of viral amplification in Sf9 insect cells, the insect cells were harvested and cell lysate (containing the recombinant protein) was prepared as previously described.

Detection of Integrase Activity

Two oligonucleotides (A+B) were used in this assay. Oligonucleotide A was 5'-end labeled with ^{32}P ATP using the 5' DNA terminus labelling system (GibcoBRL).

Oligonucleotide A (5'-ATG TGG AAA ATC TCT AGC AGT 3') and oligonucleotide B (5'-TAC ACC TTT TAG AGA TCG TCA 3') were hybridized together by heating to 60 °C and cooling slowly to room temperature. To the integrase reaction buffer (25 mM Tris-HCl pH 7.5, 12 mM $NaCl_2$, 10 mM 2-Mercaptoethanol, 10% glycerol and 7 mM of either $MnCl_2$ or $MgCl_2$) 50 ng of double-stranded oligonucleotides (A/B) and 2 ul of recombinant enzyme (100 ng/ul) were added, and the reaction mixture was adjusted to final volume of 16 ul. Samples were incubated at 37 °C for 1 hour and then the reaction terminated by the addition of EDTA to 15 mM. An aliquot (5 ul) was mixed with 1 ul of sequencing loading buffer in 10 ul reaction volume and was heated to 95 °C for 5 minutes before loaded on a 10% denaturing polyacrylamide gel. Electophoresis was at 800 volts for 2 hours. Bands were visualized by autoradiography.

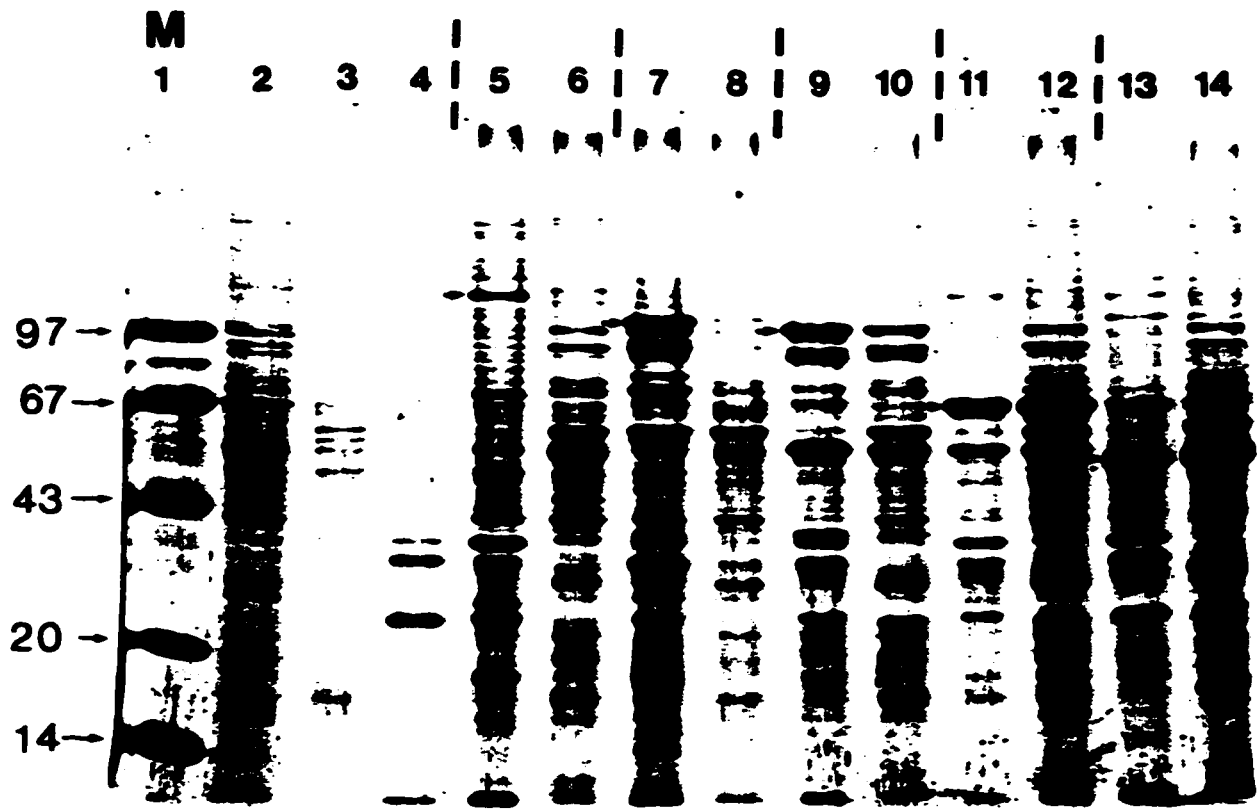
RESULTS

Expression and Detection of Different Forms of the pol Gene Product in Insect Cells

The levels of expression of the HIV-1 pol gene products were analyzed by SDS PAGE (Figure 5). The pol F (PORF) construct contains the full-length pol open reading frame starting from nucleotide 2087 to 5131. This construct contains coding sequences of viral-specific protease, reverse transcriptase, RNase H, and endonuclease/ integrase. The resulting pol F construct should generate a product of 1016 amino acids (Figure 4). As expected, this size band was detected on the gel. Staining with Coomassie Blue (Figure 5, lanes 5 and 6 respectively) indicates that the precursor polyprotein is present predominantly in the nuclear fraction of the insect cell lysate, and to a very small degree, if any, in the cytoplasmic fraction. Further, Western blot analysis using infected cell lysate and HIV+ antisera was performed, indicating 66 kDa, 51 kDa, and 34 kDa proteins present on the gel (results not shown).

The pol 100 (p100) construct starts from the first AUG codon of the pol open reading frame, In contrast to pol F, 103 amino acids are deleted at the amino terminus of the protease. This construct spans nucleotide 2396 to 5131 and contains genetic information for 912 amino acids (Figure 4). After expression in the baculovirus system, I found that large quantities of unprocessed 100 kDa protein accumulated in the Hi5 cells, specifically in the nuclear fraction (Figure 5, lane 7).

Figure 5: expression of the pol genes in Hi5 cells by recombinant baculovirus. Hi5 cells were infected with recombinant baculovirus and were harvested 3 days after infection. Cytoplasmic and nuclear fractions of the cell lysate were electrophoresed in 8% denaturing polyacrylamide gel. Proteins were made visible by Coomassie Blue staining. Lanes: 1, 14-97 KDa protein molecular weight markers; 2, uninfected Hi5 cell lysate; 3 and 4, nuclear and cytoplasmic fractions respectively of Hi5 cells infected with wild-type virus. Lanes 5, 7, 9, 11, and 13 represent nuclear fractions from pORF, p100, p97, p66 and p51 respectively. Lanes 6, 8, 10, 12 and 14 represent cytoplasmic fractions from pORF, p100, p97, p66 and p51 respectively. The arrows indicate the location of the recombinant proteins.


M
1
2
3
4
5
6
7
8
9
10
11
12
13
14
97
67
43
20
14

The pol 97 construct spans nucleotides 2588 to 5131, and has the entire viral protease sequence deleted (Figure 4). This construct should produce a protein of 849 amino acids. As expected, a 97 kDa protein was observed on a Coomassie Blue stained gel (Figure 5, lanes 9 and 10). As seen from the gel, similar amounts of this protein accumulated in both the cytoplasmic and nuclear fractions of the insect cell lysate.

The pol 66 construct spans nucleotides 2588 to 4273, and has a deletion of both protease and endonuclease/integrase sequences. This construct should express a protein of 563 amino acids (Figure 4). Large amounts of a 66 kDa protein were observed in both the nuclear and the cytoplasmic fraction of the cell lysate of the Hi5 cells (Figure 5, lanes 11 and 12).

The pol 51 construct starts at nucleotide 2588 and ends at 3864, and contains only the reverse transcriptase portion of the pol open reading frame. This protein has the entire protease coding sequence, the endonuclease/integrase coding sequence, and a large portion of the RNase H coding sequence deleted. This construct codes for 439 amino acids. In Figure 5, lanes 13 and 14, it can been seen that the expected 51 kDa protein was highly expressed in the Hi5 cells.

Activity of DNA Synthesis with Different Enzymes

These experiments were designed to determine which of the recombinant

proteins have RNA-dependent DNA polymerase activity as measured by the intensity and the length of the ^{32}P labeled product. Autoradiography of DNA synthesis reaction products from cytoplasmic and nuclear fractions, following electrophoresis on 8% denaturing polyacrylamide gel (Figure 6), reveal that longer extension products are produced with nuclear fractions than with cytoplasmic fractions. Higher activity, as detected by the intensity of the smear, was detected with nuclear p97 and p100 than with cytoplasmic p97 and p100 (Figure 6, lane 3 and 4). Conversely, stronger activity was detected with cytoplasmic p51 and p66 than with nuclear p51 and p66. Background activity with wild-type cytoplasmic and nuclear fractions is shown (Figure 6, lane 5) and can be compared to the MuLV positive control (Figure 6, lane 6).

The M13mp2 DNA-based DNA Synthesis Assay

The general procedure used to assay the accuracy of *in vitro* DNA synthesis is shown in Figure 7. The end product of this procedure, is a gaped molecule in which the gap contains the target sequence. The result of the hybridization of the two homologous DNA molecules, as analyzed by electrophoresis in an agarose gel, is illustrated in Figure 8. After the hybridization of the circular single-stranded DNA (Figure 8, lane 1) and the homologous truncated linear double-stranded M13mp2 (Figure 8, lane 4), hybridization product is produced (Figure 8, lanes 2 and 3). Two bands that were not present in the starting DNA are observed and represent the

Figure 6: gel analysis of RNA-dependent DNA synthesis activity with different recombinant enzymes. Cytoplasmic and nuclear fractions from infected Hi5 cells were used to study DNA synthesis from an RNA template corresponding to the *lac Z* sequence. After synthesis aliquots from the reactions were loaded on 8% denaturing polyacrylamide gel, which was then dried and exposed to X-ray film for 18 hours. Lanes: 1, DNA synthesis with p51; 2, p66; 3, p97; 4, p100;, 5, wild-type; 6, MuLV.

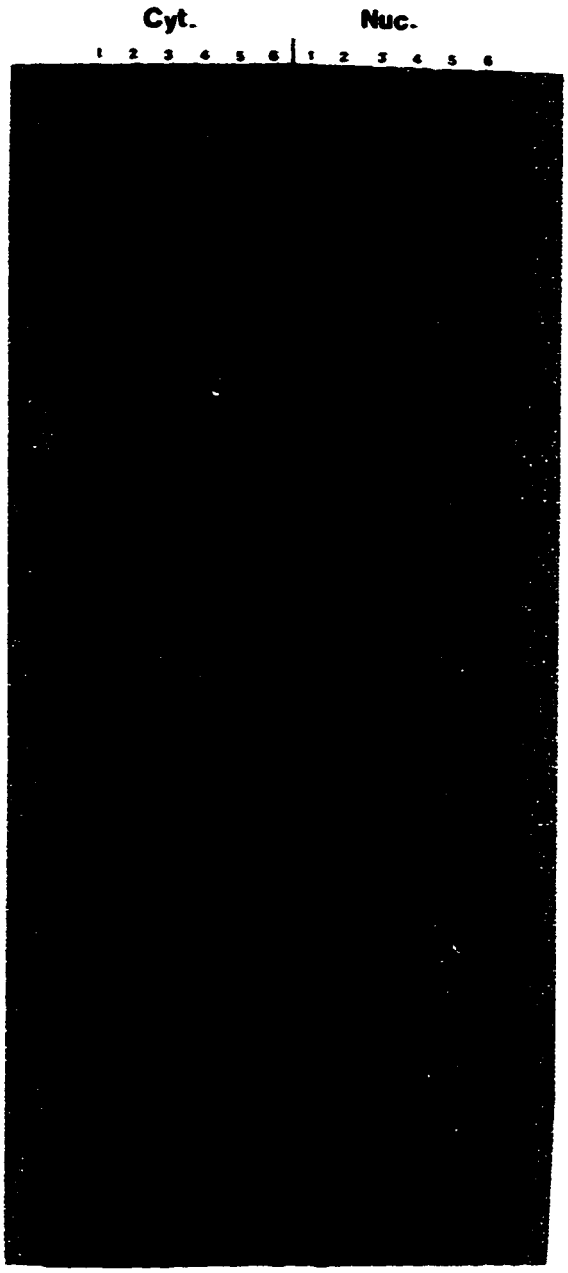

Cyt.
Nuc.
1 2 3 4 5 6
1 2 3 4 5 6

Figure 7: experimental outline of the DNA-based DNA synthesis fidelity assay. Two restriction enzymes, PvuI and PvuII, were used to cut the M13mp2 vector and generate the gaped molecule. The two enzymes excised a 390 base pair DNA fragment. A 390 nucleotide gap was generated which extends from position 174 (5' end) to -216 (3' end, where +1 represents the start of transcription of the *lacZ* alpha sequence). This molecule was denatured and used for hybridization to a circular undigested single-stranded M13mp2 (7196 nt). The molecule contained a single-stranded gap that extends from position 174 to -216. The direction of DNA synthesis within this gap is right to left (i.e., 3' at +174 to 5' at -216). The substrate was then used to transfect E. coli. Mutations were scored by plating onto reporter M9 plates (Appendix 2). The DNA sequence of several mutants were then confirmed. Asterix (*) indicates the step in which mutations are introduced.

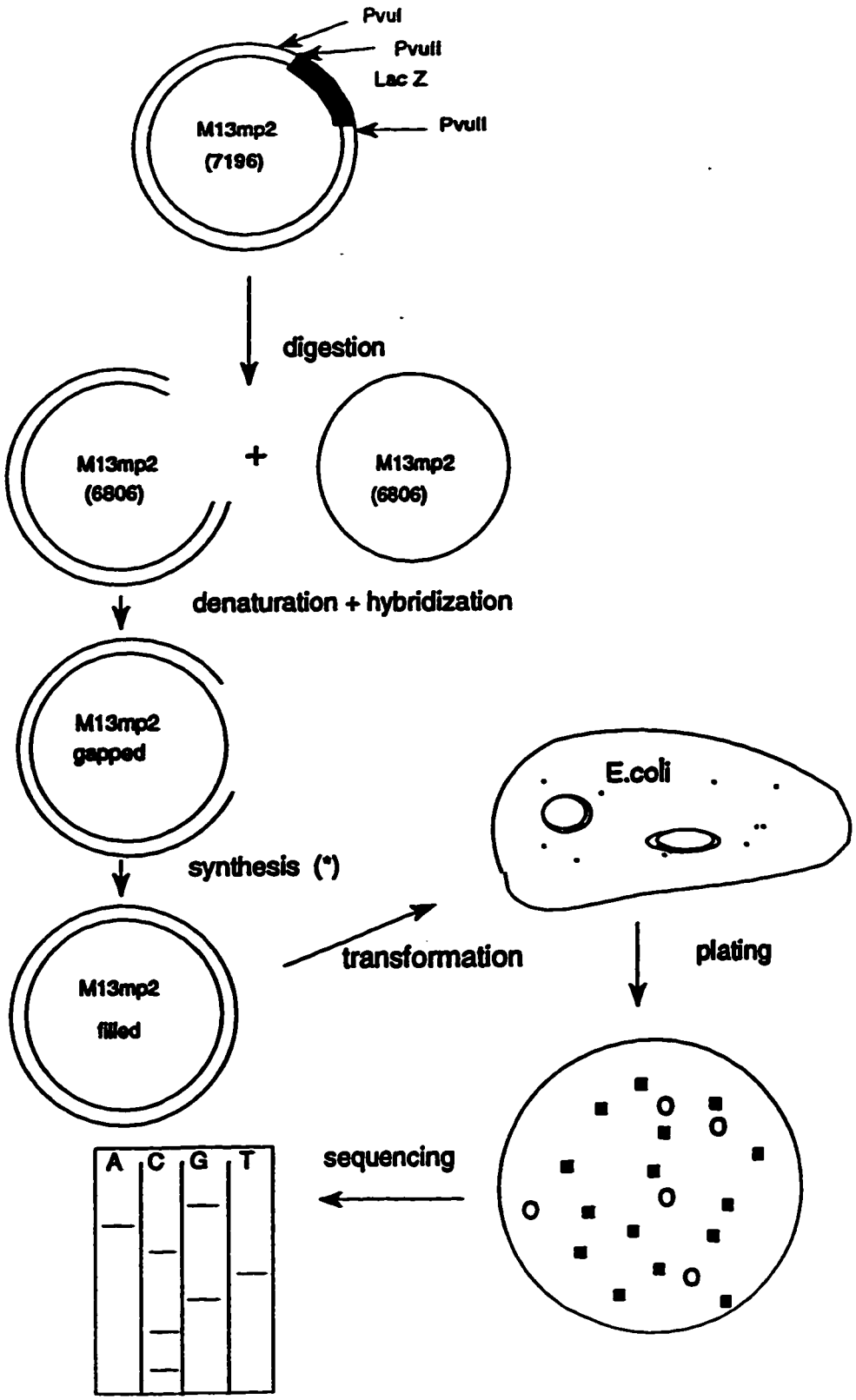
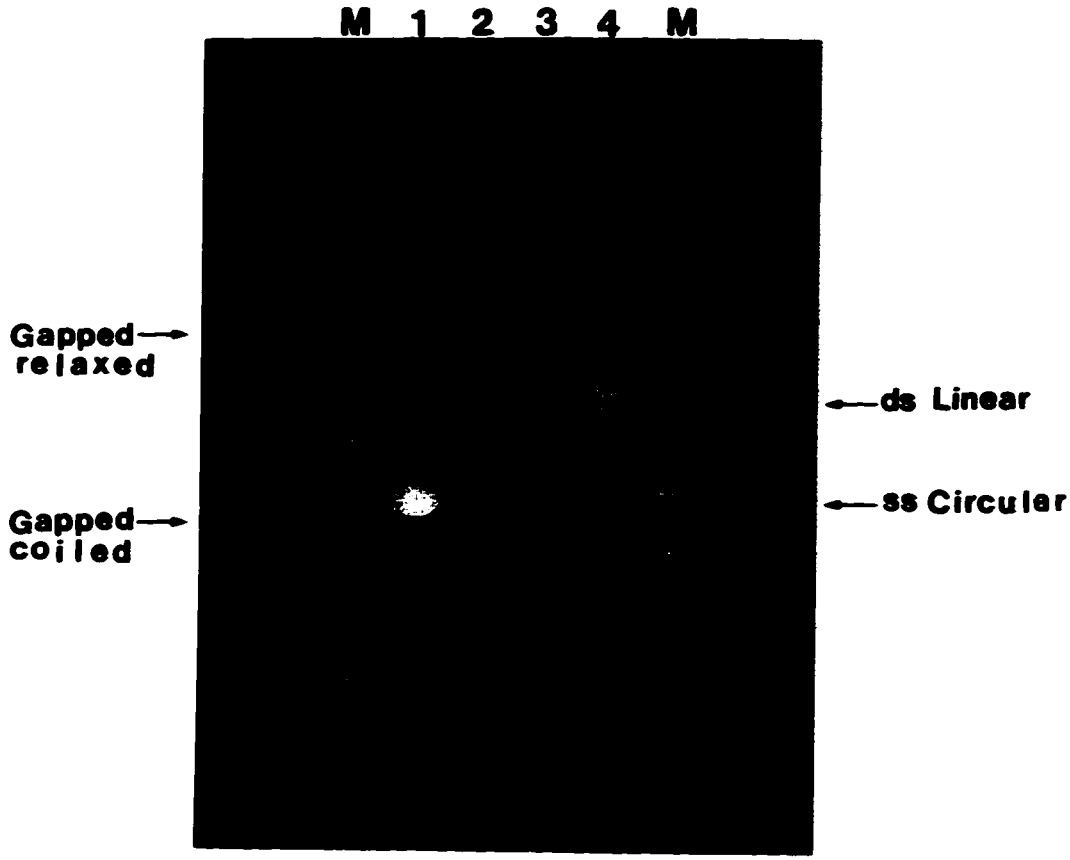

PvuI
PvuII
Lac Z
PvuII
M13mp2
(7196)
digestion
M13mp2
(6806)
+
M13mp2
(6806)
denaturation + hybridization
M13mp2
gapped
synthesis (*)
M13mp2
filled
transformation
E.coli
plating
sequencing
A
C
G
T

Figure 8: gel analysis of the gaped M13mp2 molecules. After annealing, the gaped molecules were analyzed by electrophoresis in a 0.8% agarose gel followed by staining with ethidium bromide. Lanes: M, 1 Kb DNA ladder (GibcoBRL); 1, 1 ug of single-stranded circular DNA; 2, product of hybridization using 1 ug of circular DNA and 1 ug of double-stranded DNA; 3, product of hybridization using 1 ug of circular DNA and 0.5 ug of double-stranded DNA; 4, 2 ug of double-stranded linear DNA.


M 1 2 3 4 M
Gapped relaxed
Gapped coiled
ds Linear
ss Circuler

newly generated gaped molecules. These gaped molecules were filled by a single cycle of *in vitro* DNA synthesis using one of the recombinant reverse transcriptases. A portion of the product was then analyzed by electrophoresis and autoradiography to assure complete synthesis (Figure 9). The two well delineated bands that are observed represent filled DNA molecules. The remainder of the DNA was used to infect cells and assay for α-complementation. Certain errors during the *in vitro* DNA synthesis result in altered plaque phenotypes. The exact nature of the errors can be determined by DNA sequence analysis of the single-stranded viral DNA (Figure 7).

The RNA-based DNA Synthesis Assay

Above I described the assay that scores for errors in the *lacZ* gene during DNA-dependent gap-filling synthesis. To examine fidelity with an RNA template of this same sequence, I adapted the assay as shown in Figure 10. The T7 transcription promoter was placed adjacent to the *lacZ*α gene (61). After digestion of double-stranded DNA with restriction endonuclease Fsp I, and upon transcription the desired RNA template is produced. After transcription by T7 RNA polymerase, a transcript of 313 nucleotides (from position -118 through +195) is expected. Figure 11 demonstrates that the run-off transcription reaction produced a product of the expected length, as compared to RNA molecular weight standard. Using the desired recombinant reverse transcriptase (Figure 4), cDNA synthesis was initiated producing a cDNA molecule that spans the *lacZ*α target (position -84 through 174). This same region was also used for scoring errors with the gaped DNA template. Errors were

Figure 9: gel analysis of the DNA synthesis gap-filling experiment. The hybridization product (Figure 8) was filled using p51 and NIH RT (lanes 1 and 2 respectively) and ^{32}P dCTP. Molecules were analyzed in a 0.8% agarose gel followed by drying and autoradiography for 2 hours. Arrows indicate relaxed (upper arrow) and coiled (lower arrow) gaped molecules that were filled by the enzymes.

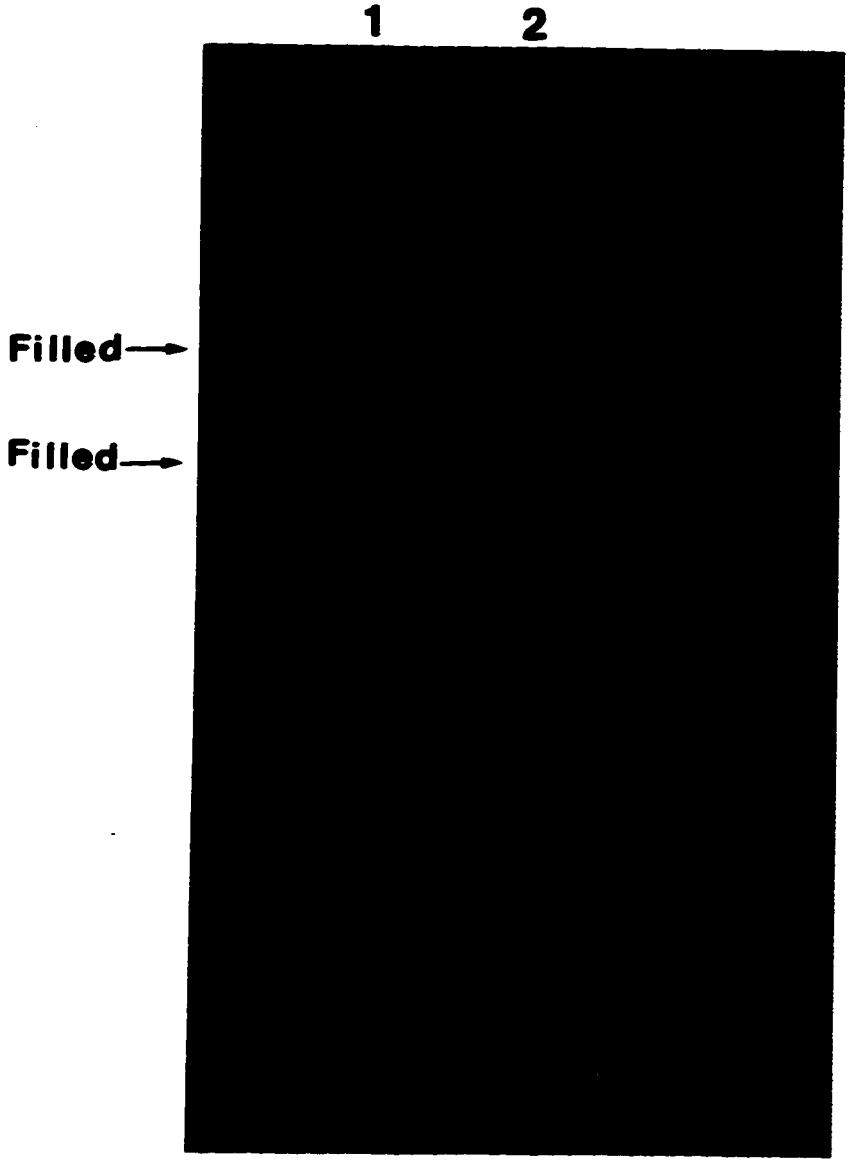
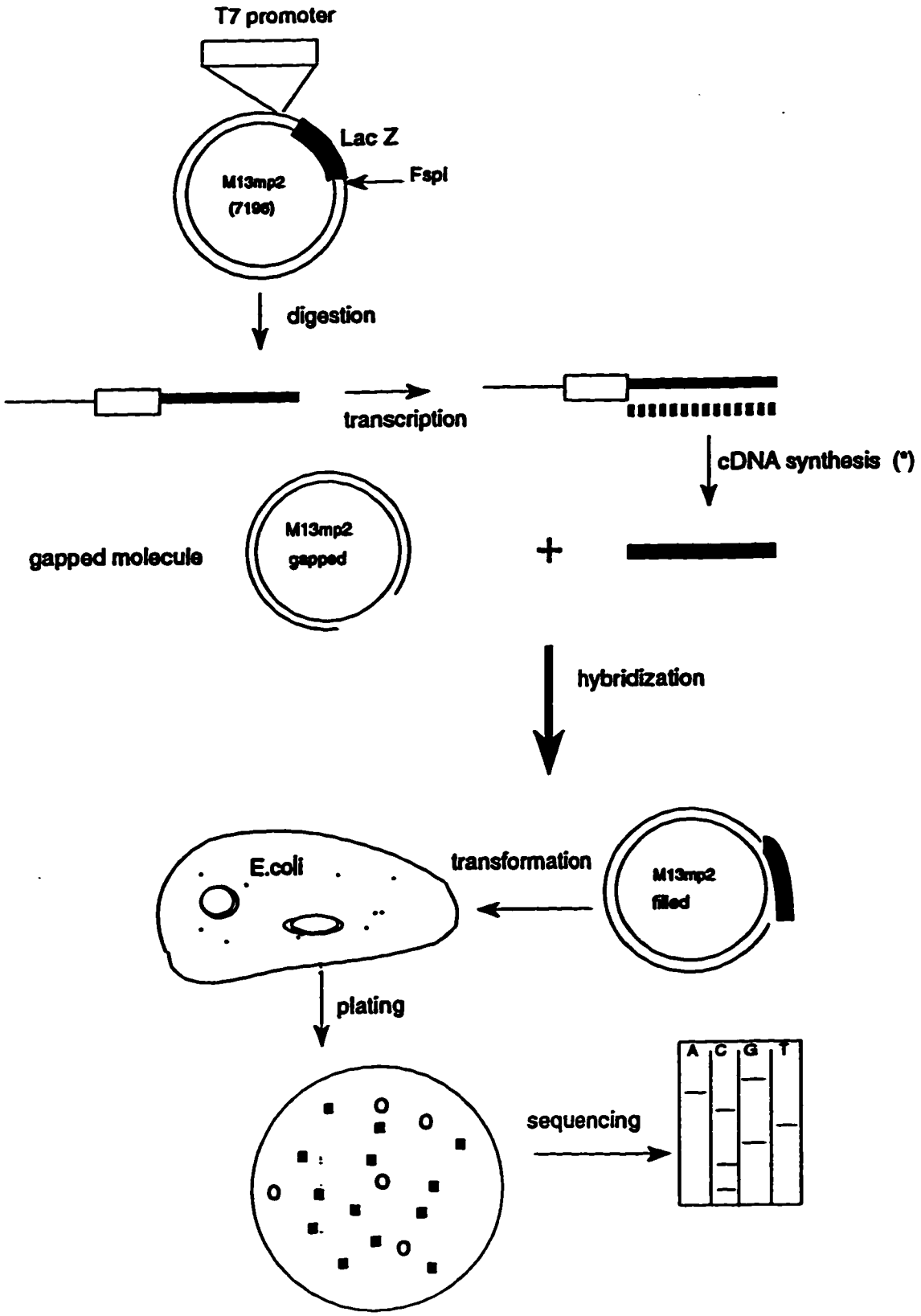

1
2
Filled
Filled

Figure 10: experimental outline of the RNA-based DNA synthesis fidelity assay. The T7 promoter was placed in the M13mp2 vector. After digestion of the vector with the restriction enzyme Fsp I, transcription using T7 RNA polymerase produced a 313 nucleotide RNA fragment. Using this transcript, reverse transcriptase and a 15-mer DNA primer, cDNA was produced. The remaining RNA was digested with RNase A and RNase T1 and the single-stranded cDNA was hybridized to the gaped substrate (Figure 8). The substrate was then used to transfect competent E. coli. Mutations were scored by plating onto reporter plates. Some plaques' phenotype were confirmed by sequencing. The asterix (*) indicates the step in which mutations are introduced.


T7 promoter
Lac Z
M13mp2
(7196)
FspI
digestion
transcription
cDNA synthesis (*)
gapped molecule
M13mp2
gapped
+
hybridization
E.coli
transformation
M13mp2
filled
plating
sequencing
A
C
G
T

recovered by hybridizing the cDNA strand to the gaped circular DNA substrate (as with the DNA-based fidelity assay) and transferring this into the CSH50 host strain (as with the DNA-based fidelity assay). After transcription by T7 RNA polymerase, a transcript of 313 nucleotides (from position -118 through +195) is expected to be generated. Figure 11 demonstrates that the run-off transcription reaction produced a product of the expected length, as compared to RNA molecular weight standard.

The cDNA product was then hybridized to the circular M13mp2 substrate that contains a 390-nucleotide gap. The products of hybridization, as analyzed by electrophoresis in an acrylamide gel, are illustrated in Figure 12. The upper band in lanes 1, 3, 5 and 7 shows the position for a gaped molecule with the 5' end labeled cDNA molecule (generated from different recombinant reverse transcriptases) bound to its gaped region. The lower cDNA band represents the labeled and unbound cDNA molecules. In lanes 2, 4, 6 and 8, the gaped molecule and the cDNA molecule were mixed together, but were not heated and cooled slowly to allow hybridization. In these lanes, the upper band (cDNA bound M13mp2) is absent and only the lower molecular weight cDNA can be seen.

Analysis of DNA- and RNA-based DNA Synthesis Fidelity

The products of the DNA-based gap-filling were used in transfection and plating experiments (Figure 13 and Figure 14). As can be seen, the majority of the plaques are dark blue with the exception of few which are light blue or colourless.

Figure 11: gel analysis of transcription product. Transcription product was analyzed by gel electrophoresis an a 1% agarose RNA gel followed by staining with ethidium bromide. Lanes: 1, RNA molecular weight marker III (Boehringer Mannheim); 2, RNA product.

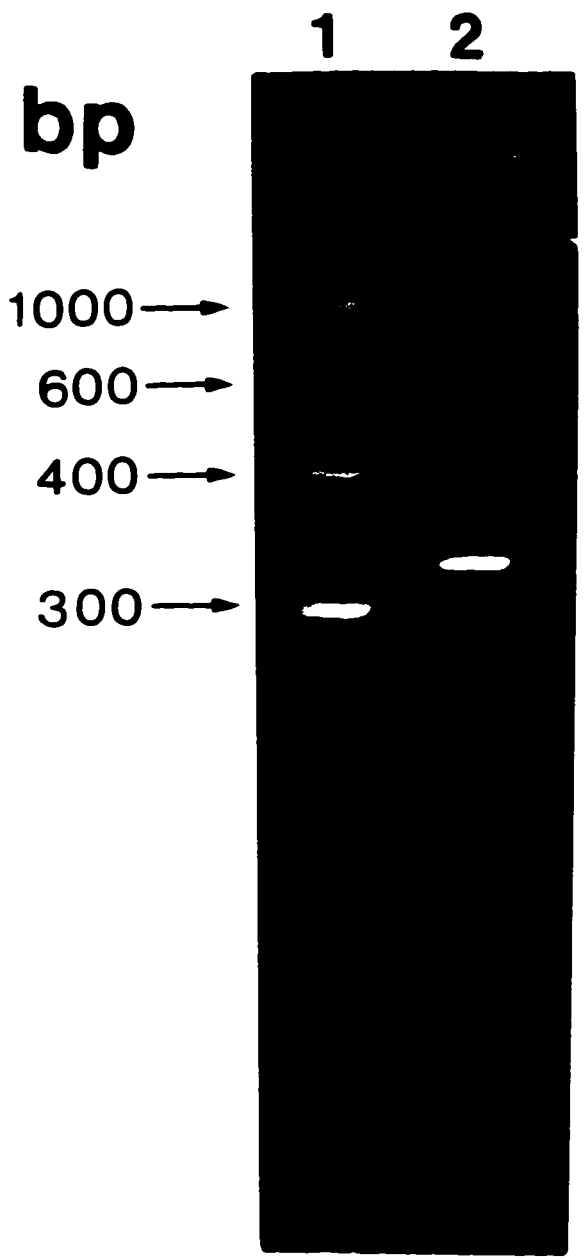

1
2
bp
1000
600
400
300

Figure 12: gel analysis of hybridization product. The product of hybridization of cDNA to gaped M13mp2 DNA were analyzed by electrophoresis on a 5% polyacrylamide gel followed by autoradiography. Lanes: 1, 3, 5 and 7 are ^{32}P dCTP labelled cDNA generated by reverse transcription with p66/p51, p97, p100 and NIH RT respectively, after hybridization to gaped circular M13mp2. Lanes: 2, 4, 6 and 8 are ^{32}P dCTP labelled cDNA generated by reverse transcription with p66/p51, p97, p100 and NIH RT before hybridization to gaped circular M13mp2.

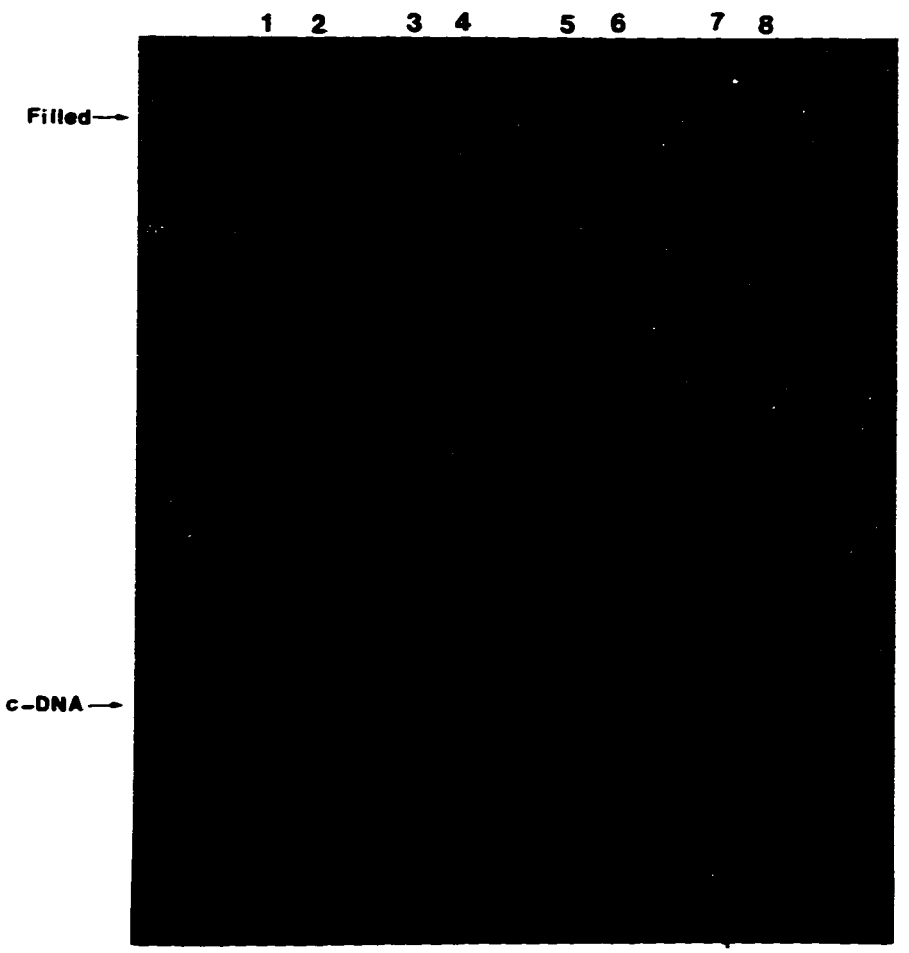

1 2 3 4 5 6 7 8
Filled
c-DNA

Figure 13: scoring of plaques with the DNA based fidelity assay with p66 RT. A typical plate used for scoring mutation frequencies. Following transformation cells were plated on minimal plates with X-gal, IPTG and host bacteria (CSH50). The plates were incubated for 18-20 hours at 37 °C followed by an additional 24-48 hours incubation at room temperature. Wild-type plaques are seen as dark blue and mutant plaques are seen as either light blue or colourless (shown by the arrows).

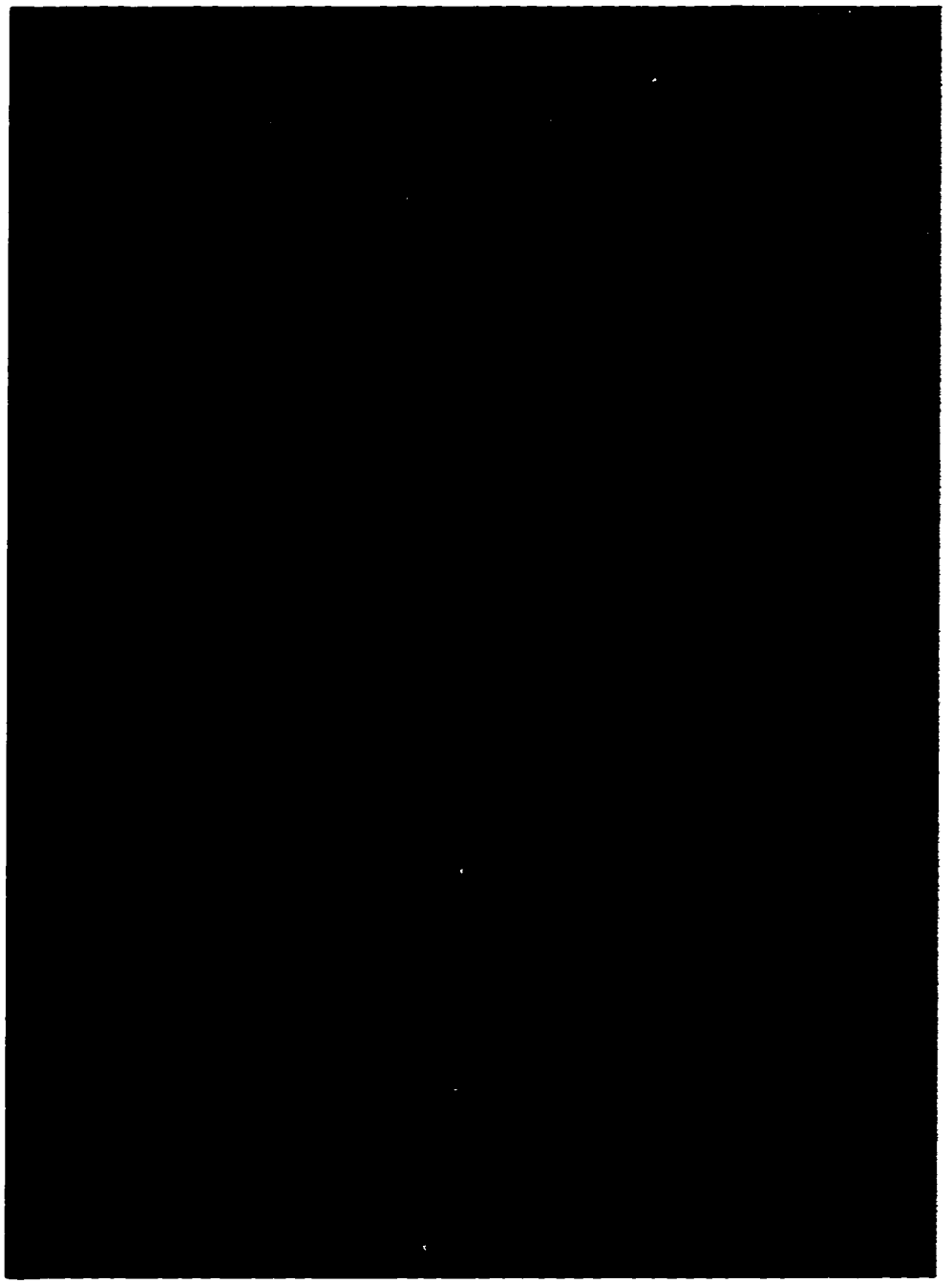
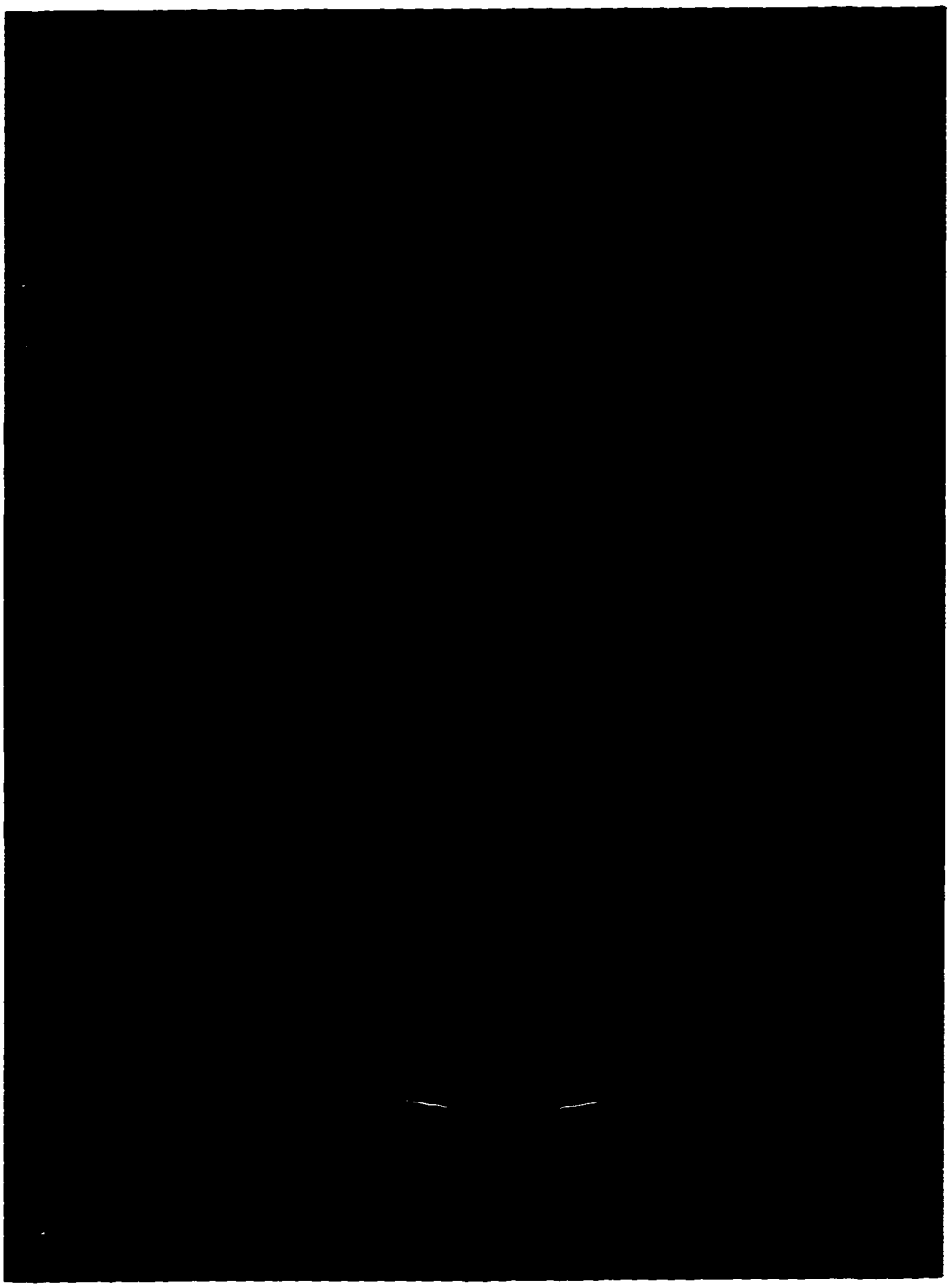


Figure 14: confirmation of plaques using the streak method. A typical plate used for confirmation of the phenotype of borderline plaques. These plaques were picked and expelled in 1 ml of 0.9% NaCl. Plaques were streaked, using a strip of 3M Watman paper, on minimal plates with X-gal, IPTG and host bacteria (CSH50). The plates were incubated for 18-20 hours at 37 °C. Three different phenotypes can be seen: dark blue (wild-type), light blue and colourless (mutant phenotype).

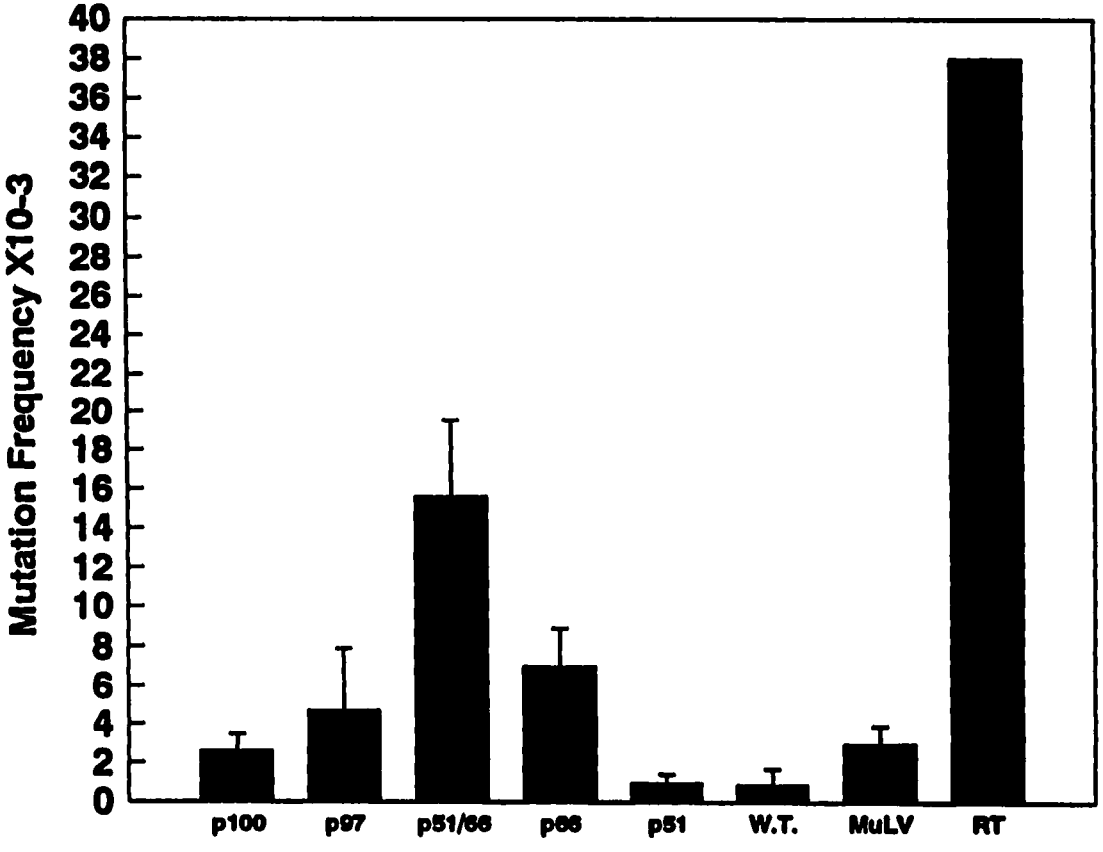


Phenotypes of selected plaques were confirmed by further analysis. A typical scoring plate is shown in Figure 14. The three different phenotypes that are typically observed on a plate are represented by dark blue, light blue or colourless plaques. The DNA-based fidelity of the recombinant HIV-1 polymerases, as well as the MuLV and recombinant RT (NIH) controls, are shown in Figure 15. In most cases, the mutation frequency is an average of three independent experiments (unless otherwise stated, see Appendix 4), each performed separately using different aliquots of DNA substrates and MC1061 competent cells. In addition, some phenotypes were confirmed by further colour analysis of the plaques as described in the Methods and Materials section.

The mutation frequency was calculated by dividing the number of all mutant phenotypes by the total number of plaques. When DNA sequence analysis was performed, light blue and colourless plaques were distinguished. The mutation frequency was found to be 1.0×10^{-3} and 7.0×10^{-3} for DNA copied by p51 and p66 respectively. These rates are in contrast to the values reported in the literature for mature form RT of 38×10^{-3} (59). The nature of this difference (i.e., several-fold lower mutation frequencies) prompted an examination of the mutation frequencies of p51 and p66 when present together, in equal ratios, in a mixture (p66/p51) (see Figure 15). The mutation frequencies obtained under these conditions, were at least two-fold higher than for each enzyme alone. For DNA copied by recombinant p97 (4.8×10^{-3}) and p100 (2.6×10^{-3}), fidelities are about three and six times higher respectively, than the value for p66/p51 (15.6×10^{-3}). The background frequency of

Figure 15: DNA-based DNA synthesis mutation frequencies of recombinant gag proteins. Recombinant proteins were used in the DNA-based DNA synthesis fidelity assay. Filled M13mp2 DNA was used to transform host cells which were then scored for their colour. The mutation frequency (Y-axis) is the number of light blue or colourless plaques divided by the total number of plaques. See Appendix 4 for results.

DNA-Based DNA Synthesis Mutation Rate



mutation was determined by infecting cells with wild type baculovirus rather than with the recombinant baculoviruses mentioned above. The fidelity for the background thus determined, was found to be 0.9 x 10^{-3}; significantly higher than those of the precursor proteins (except for p51, which yielded a mutation frequency of 1.0 x 10^{-3} when plated alone).

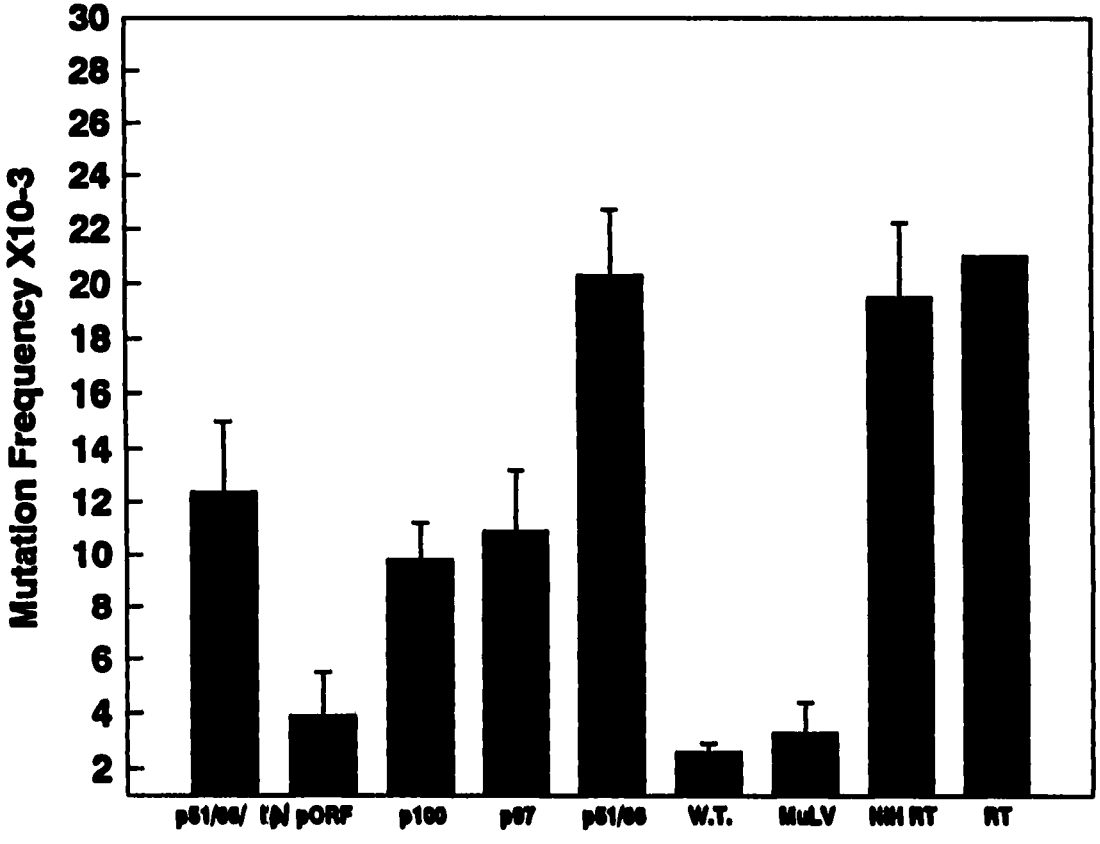
Transfection with the products of RNA-dependent DNA synthesis reactions yielded in general, several-fold higher mutation frequencies than transfection with products of DNA-dependent DNA synthesis reactions. Results presented in Figure 16 reveal substantial differences in error rates with RNA and DNA templates of the same sequence. The mutation frequency value calculation is the same as that for the DNA-based fidelity assay (see Appendix 5). The NIH RT was employed in the RNA-based DNA synthesis assay as a control experiment for the fidelity assay. The mutation frequency obtained for the NIH RT of 19.5 x 10^{-3}, was quite similar to the value of 21 x 10^{-3} reported by Boyer et al. (59). Furthermore, these values were also comparable to the fidelity value obtained for the recombinant p66/p51 heterodimer (20.3 x 10^{-3}). In all cases, lower mutation rates were obtained for the recombinant polymerase precursors, where pORF was found to be the most accurate polymerase (3.9 x 10^{-3}), and p97 the least accurate (10.9 x 10^{-3}).

Sequence Analysis of Mutants

Since loss of α-complementation can result from several different types of

Figure 16: RNA-based DNA synthesis mutation frequencies with recombinant gag proteins. Recombinant proteins were used in the RNA-based DNA synthesis fidelity assay. Filled M13mp2 DNA was used to transform host cells which were then scored for their colour. The mutation frequency (Y-axis) is the number of light blue or colourless plaques divided by the total number of plaques. See Appendix 5 for results.

RNA-Based DNA Synthesis Mutation Rate



events, mutants were analyzed by DNA sequencing to determine the exact nature of the errors made by the recombinant polymerase proteins. For this purpose, a collection of randomly selected, independent mutants were isolated and scored for their intensity of blue colour in direct comparison to wild-type M13mp2 blue plaques (i.e. secondary plating was performed as described in Materials and Methods). Light blue or colourless plaques were analyzed by sequencing the entire 258-base α-complementation sequence present within the single-stranded gap (nucleotides -84 through +174).

The mutational spectra for HIV-1 RT, p66/p51, p97 and p100 are presented in Figure 17. Two features of the error spectra are immediately apparent. First, the distribution of mutants is not random. Thus, although there are 113 template nucleotides that permit detection of base substitution errors and frameshift errors (deletions and insertions), most of the mutants are found at a small number of template positions. Second, all the mutant plaques analyzed contained either single-base substitutions or single-base frameshifts. None of the mutant plaques that were analyzed contained two or more adjacent deletions or insertions. Further analysis of the mutants indicates that phenotypically silent and non-silent mutations were nonpreferential and the distance between individual mutations on the α-complementation sequence is random.

The specificity (i.e., type of nucleotide involved) of single-base substitution and

Figure 17: single-base mutations with RNA template from position +40 to +171. Base substitutions are indicated by the letter representing the new base. Frame-shift mutations are represented by a solid triangle and base deletions are represented by an open arrow. The position of the DNA primer used in the reverse transcription reactions is indicated by the underlined DNA sequence.

Single-base mutations with an RNA template

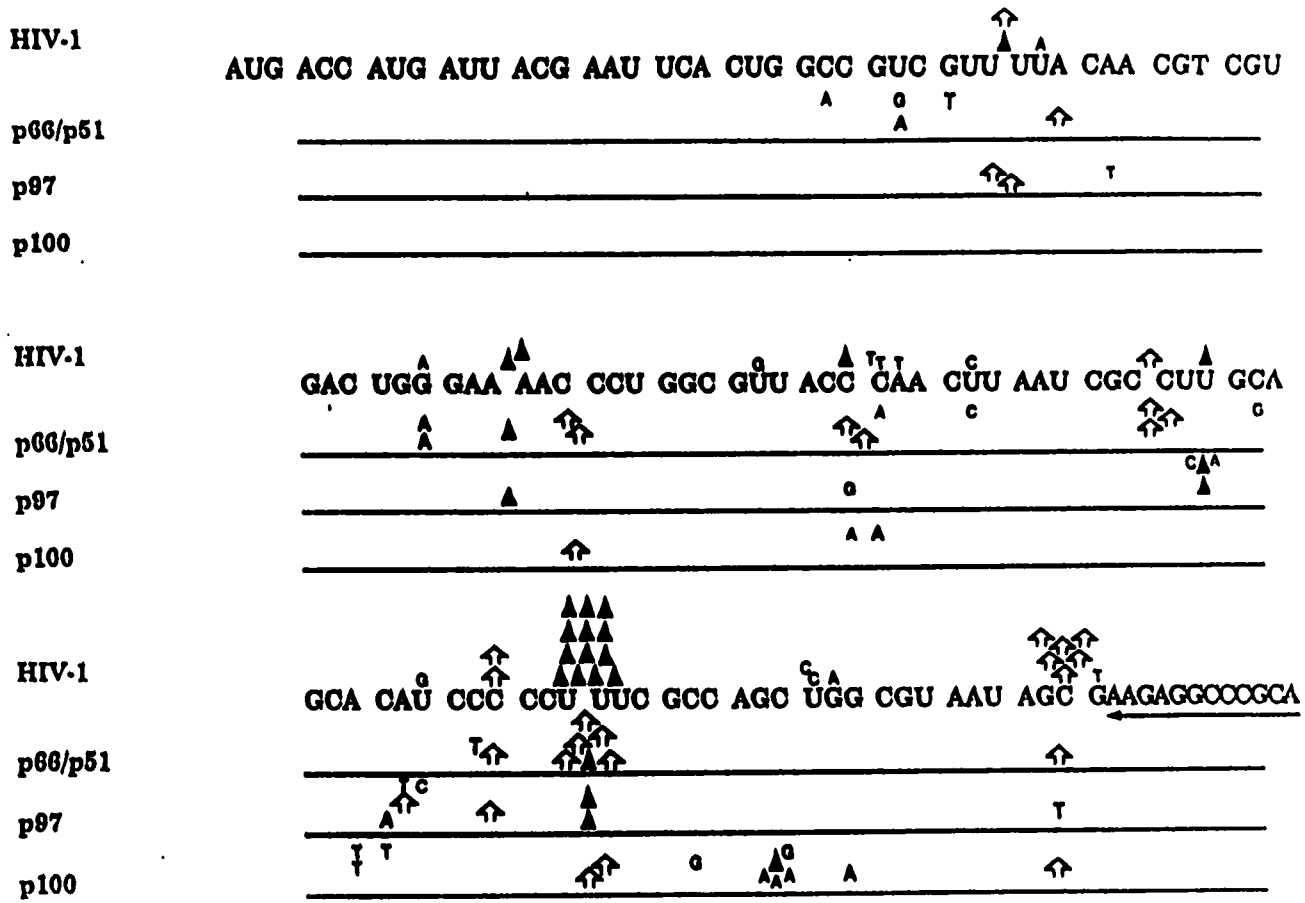


Table 1: base substitution and frameshift mutations with RNA template. Sequence results described in Figure 17 were further analyzed. The number of plaques with A, C, G and T base substitutions are reported for pol 66/51, pol 97, pol 100 and mature HIV-1 RT expresses in E. coli (59).

Base Substitution and Frame Shift Mutation Rates with RNA Template

ERROR

Base Substitution	pol51/66	pol97	pol100	HIV-1
A---C				
A---G				
A---T		2	2	1
C---A	2	1	5	
C---G	1	1	2	
C---T	1	1	1	2
G---A	2		1	2
G---C				
G---T	1			3
T---A	1	1		1
T---C	1	2		3
T---G	1			2

Table 2: base deletion and base insertion mutations with RNA template. Sequence results in Figure 17 were further analyzed. The number of plaques with base substitutions and base deletions are reported for pol 66/51, pol 97, pol 100 and mature reverse transcriptase expressed in E. coli .

Deletion/Insertion	↑	▲	↑	▲	↑	▲	↑	▲
A	1	1		1		1		6
C	9		2		2		32	2
G								
T	5	1	2	4	2		3	44

single-base frame shifts was also analyzed (Tables 1 and 2). From Table 1, all mutations were produced except for A.C and A.G. The distribution for the rest of the mutations was random for all the enzymes that were tested. As for the deletion and insertion mutations, only three of the four possible bases (A, C or T) were either deleted or inserted in the α-complementation gene. In contrast to NIH RT mutations, most (23/31) of the mutations, for the recombinant proteins that were identified were deletion mutations. Similarly, no substitution or frameshift mutations were observed in any of the mutant plaques at the first base (G) after the primer.

Integrase Activity Analysis

Polyprotein precursors pORF, p100 and p97 contain the integrase protein. The apparent decrease in mutation frequency with these recombinant polyprotein precursors led us to synthesize the integrase protein *in vitro* and determine its effect on the fidelity when tested in combination with the mature form protein. The integrase was designed in such a way that it starts from the AUG codon of the integrase ORF. The strategy adopted to construct and express the integrase sequence is depicted in Figure 18. Two primers (sense (4270 to 4284) ACG GAT CCT AAT GGG AAT AGA TAA GG and antisense (5182 to 5168) ATC TGC AGC CCT AGC TTT CCC TG) were designed to amplify the 912 nucleotide sequence. This construct should code for a 286 amino acid integrase protein similar to the natural protein (Figure 4). As with the reverse transcriptase polyprotein constructs, the baculovirus expression system was used to generate the recombinant integrase protein. Polyacrylamide gel

Figure 18: experimental outline of the construction and expression of the integrase sequence. Two primers were designed to span the entire ORF of the integrase sequence (286 a.a.). The restriction enzyme BamHI was incorporated in the +ve sense primer and PstI was incorporated in the -ve sense primer. After amplification by PCR, the DNA product was purified by GeneClean II and ligated to a TA vector (Invitrogen). The same restriction endonucleases were used to release the insert DNA which was then ligated to the baculovirus expression vector pVL1393. This recombinant DNA was used in the cotransfection experiment. Cell lysates were analyzed for integrase activity (Figure 19 and 20). The RNA-based DNA synthesis fidelity was measured using the integrase and the p66/p51 heterodimer (Figure 16).

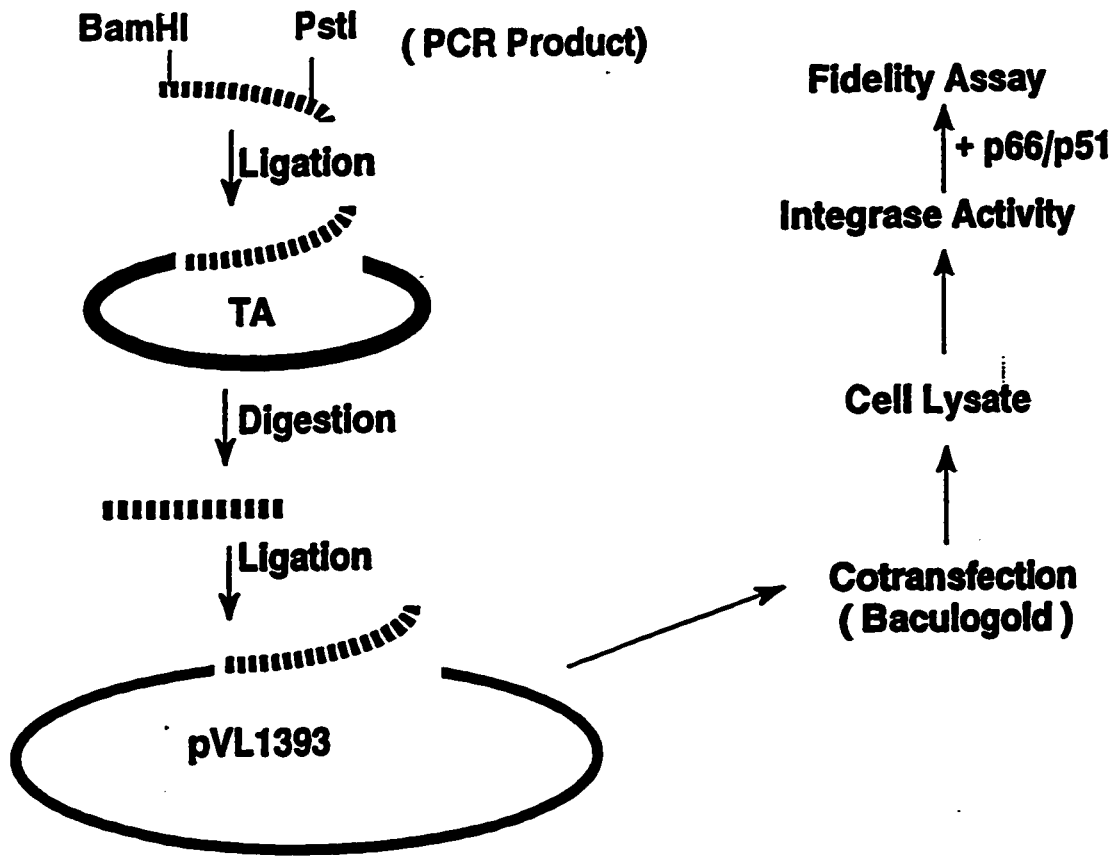
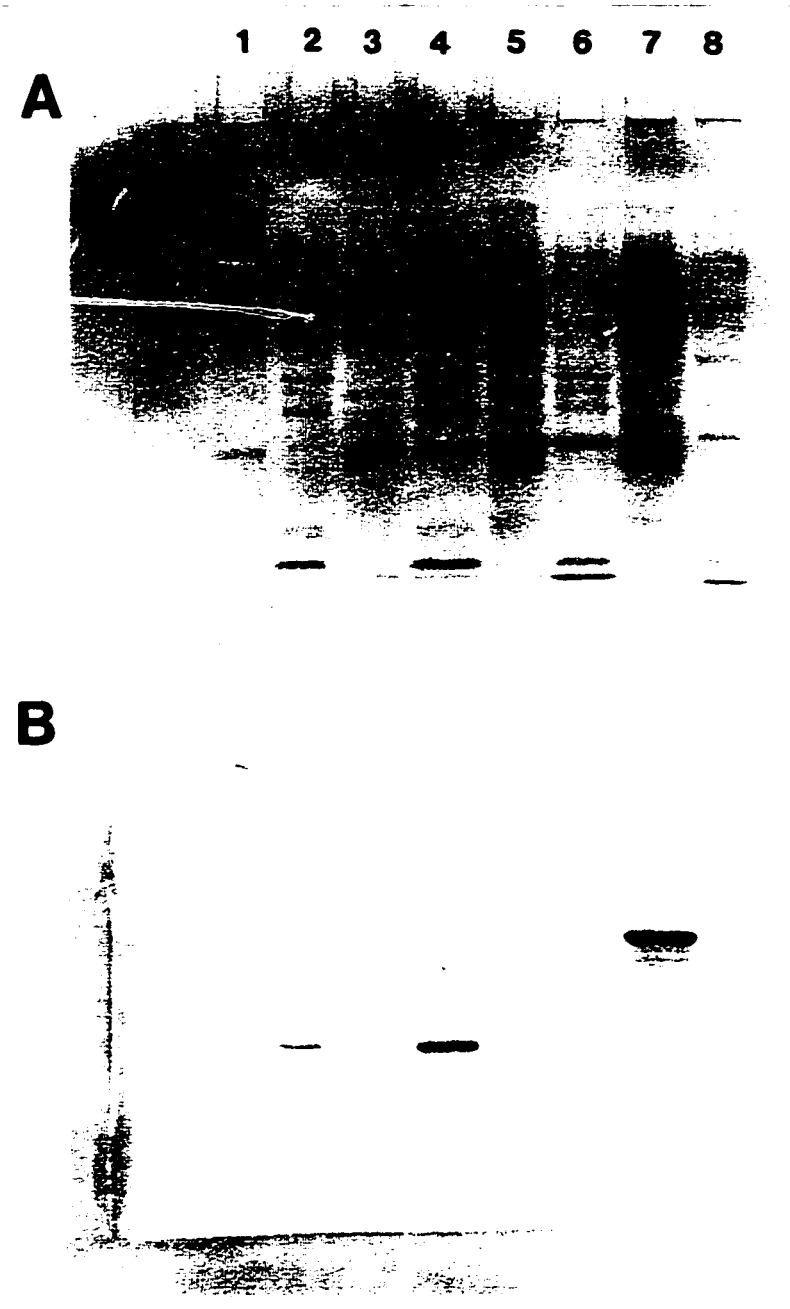

BamHI
PstI
(PCR Product)
Ligation
TA
Digestion
Ligation
pVL1393
Cotransfection
(Baculogold)
Cell Lysate
Integrase Activity
+ p66/p51
Fidelity Assay

Figure 19: expression of the integrase gene in Hi5 cells by recombinant baculovirus. Hi5 cells were infected with recombinant baculovirus and were harvested 3 days after infection. Cytoplasmic and nuclear fractions of the cell lysate were electrophoresed in an 8% denaturing polyacrylamide gel. Proteins were made visible with Coomassie Blue staining (A), and with Western blot using HIV-1 antiserum (B). (A) lanes: 1 and 2, cytoplasmic and nuclear fractions respectively from cells expressing the integrase gene 48 hours after infection; 3 and 4, cytoplasmic and nuclear fractions respectively from cells expressing the integrase gene 72 hours after infection; 5 and 6, cytoplasmic and nuclear fractions respectively of cells expressing integrase gene with a frame shift mutation 72 hours after infection; 7, cytoplasmic fraction from cells expressing p66 gene 48 hours after infection; 8, molecular weight marker. (B) Western blot of a duplicate gel.


1 2 3 4 5 6 7 8
A
B

analysis and Coomassie Blue staining of the cytoplasmic and nuclear fractions (Figure 19A)revealed that the protein is predominantly localized in the nuclear fraction of the insect cell lysate; especially 48 hours after transfection. These results were confirmed with Western blot analysis using HIV^+ antiserum (Figure 19B).

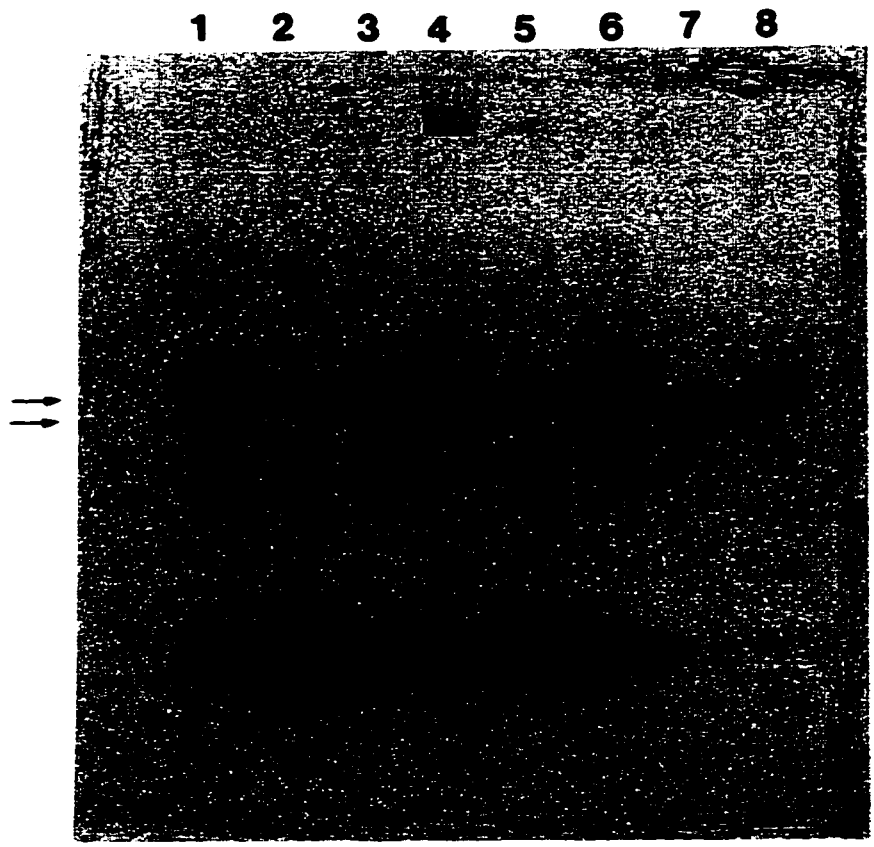
Analysis of the Recombinant Integrase Activity

To determine whether this recombinant integrase protein retains the activity of native integrase. The protein was tested for its ability to remove two nucleotides from the 3' termini of double stranded DNA (Figure 20). Shorter DNA fragments can then be detected by gel analysis. The result of digestion with the recombinant enzyme, as analyzed by gel electrophoresis and autoradiography, is seen in Figure 20. Recombinant integrase (p31) had integrase activity (optimum activity at 1:10 dilution) as indicated by the presence of the lower band (Figure 20 , lanes 4, 5 and 6). Only reactions with $MnCl_2$ exhibited integrase activity when compared with the control activity. Reactions with $MgCl_2$ did not exhibit any integrase activity as indicated by the lack of the lower (19-mer) band (Figure 20, lanes 1, 2 and 3).

Mutation Frequency of Recombinant p66/p51/In Protein

Transfection with the products of RNA-dependent DNA synthesis filled by p66/p51/In yielded almost two-fold lower mutant frequency than the p66/p51 heterodimer alone (Figure 16).

Figure 20: products of the HIV-1 integrase cleavage reaction. The products of the integrase activity assay were electrophoresed on a 10% polyacrylamide gel and visualized by autoradiography. In lanes 1-3, the reaction buffer contained $MgCl_2$ and in lanes 4-6, the reaction buffer contained $MnCl_2$. Lanes 7 and 8 are ^{32}P-ATP labelled 18-mer and 21-mer oligonucleotide markers respectively. Lanes 1 and 5, reaction with integrase; 2 and 6, reaction with p66; 3 and 6, reactions performed in the presence of p97.


1
2
3
4
5
6
7
8

DISCUSSION

The data in this report include both the fidelity and specificity of recombinant precursor proteins of HIV-1. The methodology used in this work (M13mp2 assay system) has already been successfully employed in many different studies (20, 64, 56). The notable difference, is that in order to measure RNA-based and DNA-based fidelities, previous studies all relied on measuring the fidelity of the mature form of RT. I addressed the question of whether naturally occurring precursor proteins (53, 54, 65) exhibit any differences in fidelity as compared to the mature form of RT, and if so, whether they have a higher or lower degree of fidelity.

For the success of this study it was important to obtain recombinant proteins in biologically active forms. Several unique features of the baculovirus expression system made it the system of choice for this study. This expression system features post translational modification of proteins including: phosphorylation, glycosylation, myristolation, as well as high levels of protein expression without the need for CO_2. Using this system, I was able to successfully obtain high levels of protein expression for four of the five constructs (Figure 5). The pORF construct, however, produced a somewhat lower amount of protein than the other constructs (Figure 5, lanes 5 and 6). There is a possibility that some toxicity was exhibited by the viral protease on the Hi5 cells, thus impeding their growth cycle and hence affecting the level of expression of the pORF (personal communication with Dr. Yu-Wen Hu, Canadian Red Cross, 1996).

Precursor proteins such as pORF, p100 and p97 were predominantly concentrated in the nucleus of infected insect cells (Figure 5, lanes 5, 7 and 9) whereas smaller proteins of smaller constructs (p51 and p66) were present in both fractions, but predominantly in the cytoplasmic fraction of Hi5 cells (Figure 5, lanes 12 and 14). Explanations for this interesting observation are hypothetical, however, the observation that a high molecular weight nucleoprotein preintegration complex containing precursor RT protein, integrase, p17, and cDNA is found in the nuclei of HIV-1-infected cells (66), supports the theory that binding of these enzymes to other protein(s) or DNA occurs, leading to the translocation of this complex to the nucleus through a nuclear targeting sequence present in the associated protein(s). The possibility that a nuclear targeting sequence is present in the C-terminus of the polyprotein precursor, perhaps in the integrase domain, cannot be ruled out.

Overall, the level of expression of these proteins was relatively high, therefore facilitating the next step, i.e. the examination of the activity associated with these recombinant enzymes. Results from enzyme activity studies demonstrate that both recombinant enzymes p51 and p66 and recombinant precursor enzymes (p97and p100) have reverse transcriptase activity (Figure 6). This confirms the findings of other researchers (54, 55, 67, 68, 69) that different forms of pol gene products, in addition to the mature form RT, are enzymatically active. However, in contrast, other researchers were unable to demonstrate reverse transcriptase activity with p51 (70, 71), recombinant pol gene products with deletions in the carboxy terminus (72), and a 98

KDa precursor expressed in E. coli (73). My findings, as compared to theirs, may be a result of differences between prokaryotic and eukaryotic (such as baculovirus expression system) postranslational modifications causing differences in the 3-dimensional structure of the same enzyme. Furthermore, proteases co-purified with recombinant enzymes, i.e., insect-cell's proteases, might be less effective in degrading the enzymes and aborting their activity than proteases co-purified in prokaryotic expression systems (55).

Overall, nuclear fractions of these enzymes showed higher enzymatic activity, and produced longer extension products than cytoplasmic fractions (Figure 6). This perhaps indicates a higher affinity and longer retention of the nuclear enzymes for the substrate. Since I used equal concentration of proteins in these experiments (an equivalent of 50 ng of each protein were used in every experiment), one might conclude that nuclear fractions of these enzymes might bear more resemblance to the natural RT with respect to its 'folding', or to its ability to associate with nuclear factors or nucleic acid (DNA or RNA). Increasing the concentration of cytoplasmic p97 and p100 proteins did not result in higher activity, indicating that activity is independent of cytoplasmic protein concentration. Based on these results, only nuclear fractions were used in the fidelity experiments.

The next set of experiments was aimed at measuring the frequency and specificity of mutations generated *in vitro* by recombinant enzymes. The M13mp2

cell-free assay system (64) was chosen for fidelity studies because it has several major advantages over other assays for mutational frequencies *in vitro*. The M13mp2 assay scores for loss of *lacZ* gene function (Figure 7 and 10). This function is non essential for the cell and therefore, a wide variety of mutations can be tolerated and detected (61, 62, 64). Furthermore, the substrate DNA is relatively easy to produce in large quantities from a small culture, in double-stranded and single-stranded forms. This facilitates routine DNA sequence analysis of mutants. Moreover, the M13mp2 vector produces a dark blue colour on M9 reporter plates (Figures 13 and 14) which allows a wide spectrum of mutants to be scored. By using a cell-free assay (for HIV-1 RT fidelity) that mimics some of the replicative steps in the viral life cycle, I was able to identify errors introduced by the enzyme during DNA-dependent as well as RNA-dependent DNA polymerization in the *lacZ* region.

Overall, fidelity of synthesis with DNA and RNA templates using commercial and the recombinant mature form of RT was low, as compared to other reverse transcriptases (MuLV) as previously determined (Figure 15, 16). Based on theses results alone, one would conclude that if the mature form of RT is the only enzyme used by the virus, then the frequency of errors induced by HIV-RT could account for the high mutability and rapid evolution of the HIV genome, and as a result, efforts should continue to focus on targeting the mature enzyme. The next series of experiments was designed to address this issue.

Results of fidelity with DNA templates are shown in figure 15. It is important to mention that the term 'mutation frequency' used in this study is not synonymous with the term 'error rate' usually used by some researchers. The frequency of mutation was measured by dividing the observed number of light blue or colourless mutant plaques (scored as 0, 1 or 2) by the total number of plaques on each plate (scored as 0, 1, 2 or 3). The error rate, on the other hand, takes into account background mutations. It is noteworthy to mention that I did not calculate the error rate of these enzymes. The main objective was to compare the mutation frequencies of these enzymes and therefore, for the sake of comparison, an absolute value (i.e., error rate) was not required. The mutation rate of DNA-based DNA synthesis with p66/p51 is one mutant for each 16,000 nucleotides incorporated. This value is about 2.5-fold greater than that obtained by Kunkel et al. (20, 59) (1/38,000). The most likely potential sources for the discrepancy that come to mind, include: possible differences resulting from processing of proteins expressed with eukaryotic systems, as opposed to those expressed with prokaryotic systems; differences in technical aspects of the experimental method that could produce slight, but perhaps additive effects on the results; or such things as source of reagents for the minimal media, or differences between one batch of reagent and another. More likely, however, the preponderance of the discrepancy may be related to the method of scoring 'border-line' plaques, i.e., those which were somewhat difficult to classify as either light blue or dark blue. In my experiments, I was quite deliberate about maintaining consistency in my scoring method to minimize any discrepancies that could be derived from this. In order to

determine the relative contribution of these various factors, commercially available MuLV RT was tested and the results compared to those of Boyer et al. (20, 59). The similarity of results obtained for the commercially available enzyme, indicates that variations in experimental technique or reagents bear little, if any, responsibility for the discrepancies found for the p66/p51 enzyme. In this light, results seem to be dependent upon the selection of cell system for expression of the protein (eukaryotic vs. prokaryotic), and the consistency in scoring border-line plaques. Therefore, these factors should be taken into consideration when making comparisons with other studies.

The mutation frequency observed with DNA templates for the precursors p97 and p100 (4.8×10^{-3} and 2.6×10^{-3} respectively) were 3-fold lower than for the mature form p66/p51 (15.6×10^{-3}) (Figure 15). This indicates that the RT precursors are more accurate than the mature form of RT and hence, sequences in the N-terminus (protease subdomain) or in the C-terminus (integrase subdomain) may be implicated in the higher fidelity of these proteins. Surprisingly, p51 and p66, when measured separately, showed higher fidelity (1.0×10^{-3} and 7.0×10^{-3}, respectively) than when measured together (15.6×10^{-3}) in the polymerization reaction. These results were totally unexpected and led us to conclude that conformational changes generated by the formation of the heterodimer (see Figure 3 for computer model of p66/p51 heterodimer) causes the enzyme to make more errors than if each enzyme was in its monomer form. Further work needs to be done in this regard if the process of

dimerization is to be fully understood. It is perhaps possible that by inhibiting this process, hypermutability by the virus could be diminished leading to a virus population more susceptible to the immune response.

To ensure that the frequency of background mutations were significantly lower than the mutation frequency obtained with the recombinant proteins, cell lysates from wild-type baculovirus infected cells were used in this assay as background controls. HIV RT generated mutants for the recombinant p66/p51, at a frequency of 15.6 X 10^{-3}. This frequency was about 20 times above the spontaneous background mutation frequency of uncopied DNA (0.87 x 10^{-3}) (Figure 15), demonstrating that errors are in fact the result of DNA polymerization by the recombinant RT polyproteins *in vitro*. This value, in turn, was about 2.5 times more accurate than that reported with recombinant E. coli RT by Boyer et al. (59). This discrepancy is most likely the result of the differences between the prokaryotic and eukaryotic systems used for expression.

The results also indicated that p51 is 7-fold more accurate than p66 (1.0 x 10^{-3} versus 7.0 x 10^{-3}). This result could implicate the RNase H subdomain in conferring lower fidelity to the p66 domain. RNase function (degradation of RNA from RNA/DNA hybrid) has been well characterized (15). However, the specificity of RNase H activity has been the subject of debate (74, 75, 76). Most of the results of these studies lean toward the conclusion that retrovirus' RTs display 3' to 5' endonuclease activity (74, 76) accounted for by the RNase H subdomain, and lack 3'

to 5' exonuclease activity (20). It is possible that through a number of events, including endonuclease cleavage and subsequent filling, that higher numbers of errors result. This would explain the higher mutation rate observed with p66 over p51. More work is needed, however, to determine effect on fidelity.

Literature values for HIV-1 RT error rates with a DNA template using M13mp2 or ϕx174 fidelity assays are 10^{-5} to 10^{-3} respectively, while the error rate of the major host-cell polymerases (DNA polymerase α, β or ϵ), are at least one order of magnitude lower (15, 60). This is also true with RNA polymerases from E. coli as well as T3 or T7 bacteriophages which have been measured and found to be highly accurate (77). The lower spontaneous rate of mutation in prokaryotes, results in part from proofreading of DNA synthesis errors by the 3' -- 5' exonuclease activity associated with prokaryotic DNA polymerase (60, 77). Therefore, the most error-prone polymerization steps in the HIV life cycle are likely to be viral reverse transcriptase and/or plus-strand DNA synthesis, both catalyzed by HIV RT (15). However, after viral replication, cellular mismatch repair systems could play a part during DNA replication, resulting in an underestimation of errors (15).

Genetic hypermutability of the HIV-1 is the result of the mutation rate and the number of replication cycles (57). It has been suggested that HIV-1 completes about 300 or more cycles per year (78), and the average lifetime of an HIV-1 infected cell is less than 1 to 2 days (32, 79). Hence, it is possible that selection of mutations may

have a greater influence on the genetic variation of HIV-1 than the mutation rate (57). This hypothesis is supported by the fact that mutation rates determined in this study were found to be lower than similar rates reported in the literature.

Mansky and Temin (57) studied whether the *in vivo* rate of mutation during reverse transcription is as high as predicted by cell-free studies and found that mutation rate is about 20-fold lower than the mutation rate of purified HIV-1 reverse transcriptase, with the same target sequence. This fidelity indicates that HIV-1 reverse transcription in vivo is not as error prone as predicted from cell-free studies with purified RT. Hence, fidelity of purified HIV-1 RT may not accurately reflect the level of genetic variation in a natural infection. I propose that this difference is largely dependent on the contribution to the error rate by the precursor proteins.

Up to this point I have described the fidelity of the DNA-dependent DNA polymerase activity of the different forms of RT. Although such studies are relevant to second-strand synthesis, the fidelity of reverse transcription (RNA-dependent DNA polymerase activity) is also relevant to retroviral mutagenesis. By modifying the M13mp2 assay, I was able to examine HIV-1 RT fidelity with heteropolymeric RNA of the *lacZ* gene. This assay, therefore, permitted a direct comparison of the fidelities of recombinant RT with RNA and DNA templates of the same sequence. For this study, this was a major advantage over other methods.

The results of this assay indicated that the overall fidelity of reverse transcription from an RNA template was lower than that from a DNA template (Figure 15 vs. Figure 16). This difference in fidelities suggests that minus-strand and plus-strand replication errors may not contribute equally to the final mutation rate and sequence diversity. The mutation rate of HIV-1 p66/p51 (20.3 x 10^{-3}) was comparable to that reported for mature RT and NIH RT (21 x 10^{-3}) and RT expressed in E. coli (19.5 x 10^{-3}) (Figure 16). The mutation frequency observed with the precursors p97 and p100 (10.9 x 10^{-3} and 9.3 x 10^{-3} respectively) were 2-fold lower than for the mature form p66/p51. This is interesting in light of the fact that similar results were obtained with the DNA-based mutation frequency, where the mutation frequency for precursors were 3-fold lower than for the mature form p66/p51. The mutation frequency of precursor pORF (3.9 x 10^{-3}) was 2.5-fold lower than for the p97 and p100 precursors and 5-fold lower than for the mature form p66/p51. This, therefore, supports the idea mentioned previously, that the N-terminal sequences of the protease subunit might also contain sequences that influence fidelity or proofreading activity. It is interesting to note that no associated exonuclease activity or proofreading activity has been reported with these enzymes (40).

As with the DNA template, wild-type baculovirus infected Hi5 cell lysate was used in the assay to measure background mutations in the template DNAs. Only when background levels were low, were results considered to be meaningful. The background mutation frequency was low, as expected, at 2.6 x 10^{-3}. In addition,

mutation rates measured from reactions lacking the Hi5 cell lysate (i.e., DNA template and reaction components), were extremely low; producing no mutants on all 20 plates used in the experiment (about 400 plaques per plate) (results not shown). The low background mutation rate, coupled with the extremely low blank reaction rate, precludes that any appreciable contribution to the mutation rates of the recombinant proteins is derived from either of these sources, or for that matter, from the transcription reaction with T7 polymerase. In other words, there is only one possible source for the mutants that were recovered with the RNA template, that being, the recombinant proteins themselves.

In addition to the frequency of mutation, I was also interested in examining the types of mutation generated by these enzymes. Just as with commercial RT, errors were found to be non-randomly distributed (41, 59, 56, 77) (Figure 17). Some errors were specific to only one enzyme, whereas others were common to all enzymes (Figure 17). This observation has also been reported with RT from HIV-1 and MuLV (41). It is likely that these enzyme-specific mutations could be a contributing factor for the differences in the mutation rates with the different recombinant enzymes (Figure 16). One predominant 'hot spot' was located in a region consisting of 5 C's followed by 3 U's (Figure 17). It is possible that such a sequence run could lead to more frequent template-primer slippage (22, 16, 59, 61, 69), or to a dislocation mechanism, wherein a substitution is initiated by template-primer slippage (16, 61). Nucleotide runs of dU's and dA's have been known to cause strong pauses during the

copying of a DNA template (80). Furthermore, certain sequences of about 6 nucleotides are excessively frequent in the hypervariable regions of the env gene (81, 17) and the error spectra of these sequences have been found to differ; indicating that base composition of sequences is of the utmost importance in affecting error production by RT. In fact, strong sequence dependence mispair formations (between primer and template) by HIV-1 RT has been reported, using the same M13 based mutation assay (61). Additions could have been generated by HIV-1 RT or they could reflect T7 RNA polymerase errors. The latter possibility is suggested by the fact that additions at the homopolymeric runs are known to be produced during transcription.

One area of high base substitution was only observed with p100 (Figure 17), 14 bases upstream of the 3' end of the primer, and yet other 'hot spots' were excluded by this enzyme. The additional N-terminal sequences of this enzyme may have substantially affected the enzyme error spectra. Two distinct steps might have lead to base substitution mutations: nucleotide misinsertions and elongation from 3'-terminal DNA mispairs (61).

Another region, two bases upstream of the primer was found to be error-prone, particularly for commercial RT (Figure 17). This type of mutation has been previously reported (59) and contributes to the RT errors and not transcription errors (59). One can predict that incorporation of the initial nucleotides may not be very accurate, possibly reflecting aberrant enzyme-template-primer interactions. It is also

possible that the probability of template-primer dissociation is substantially greater after the first few nucleotide incorporations, than after subsequent incorporations.

All base substitutions were observed except for A to C and A to G (Table 1). If we assume conditions for the insertion of correct and incorrect nucleotides is the same, then the mutation spectra observed is a direct measure of the accuracy of the different RTs. My results contradict other reports (82), where all possible substitutions were observed. One possible explanation is that the number of plaques sequenced and analyzed by this study was not large enough to observe a complete distribution of substitutions. This is especially likely if A to C and A to G substitutions are not as common. (In one study, however, A to C and A to G substitutions were found to dominate the error spectra (21)). Deletions and insertions occurred only at A, C and T but not G (Table 2). This observation has not been previously documented in the literature. For this reason, it may be interesting to perform future studies which focus on this aspect. An in-depth look at error mechanisms through the sequencing of a large number of plaques would perhaps, remove any ambiguity and better define results reported so far.

A question which arose during this study was whether the integrase sequence motif contributed to the higher fidelity observed with recombinant HIV-1 precursors when compared to mature form RT (Figure 16). In order to address this question, I adapted the baculovirus approach (Figure 18) to produce recombinant integrase protein.

Using this approach, I constructed and expressed the integrase protein as shown by Coomassie Blue staining in Figure 19A (lanes 1 to 4) and by Western blot results in Figure 19B (lanes 1 to 4). Results from Figure 19A and 19B indicate that protein is predominantly localized in the nucleus of infected insect cells, most prominently, 48 hours after transfection. A similar situation for the precursor proteins was already observed. This result was very interesting because it offers an explanation for why precursor RT proteins pORF, p100 and p97 (all with the integrase subdomain) were mainly localized in the nucleus of insect cells, whereas, p66 and p51 (both lacking the integrase subdomain) were localized in the cytoplasm of insect cells (Figure 5). This explanation includes the possibility that the integrase protein contains a motif responsible for nuclear localization. Bushman et al. (83), using several deletion mutants, concluded that the integrase subdomain displays several functions in addition to the well known function of 3' processing and strand transfer (84, 85). Future studies could include the identification of the exact sequence responsible for nuclear targeting.

Our next effort was to try to determine whether the protein had retained its enzymatic activity (3' processing), i.e., the ability of the enzyme to remove two nucleotides (5'-GT-3') from the 3' termini of double-stranded DNA in the presence of $MnCl_2$ (86). This activity would produce a DNA molecule with two cohesive ends each with a two-base protruding end (3'-AC-5'). This test (Figure 20), clearly indicated that the recombinant integrase protein has retained its natural structure and

was enzymatically active. From Figure 20, DNA molecules in the presence of integrase and $MnCl_2$ (lane 4, 5 and 6) produced the 2-base shorter DNA fragment. These results also demonstrate that $MnCl_2$ and not MgCl2 is required for the enzymatic function.

The mutation rate obtained with p66/p51 in the presence of the recombinant integrase (12.4×10^{-3}) was about 2-fold less than that of p66/p51 lacking the integrase (20.3×10^{-3}) (Figure 16). Hence it is possible that the integrase subdomain is responsible for the reduced mutation rate observed with precursor proteins. These results exclude the possibility that the protease subdomain was responsible. Better understanding of integrase/RT interaction could shed some light on the mechanisms responsible for the higher fidelity observed.

One can speculate that since the integrase is known to enhance DNA binding to protein, and has been found to be part of the pre-integration complex (87), several possibilities could be arrived at: 1) the integrase binds the heterodimer and enhances the interaction between the protein and the template-primer, 2) the integrase protein may induce conformational changes in the heterodimer leading to lower mutation frequencies, or, 3) the integrase binds the template-primer dimer inducing conformational changes that lead to higher fidelity of synthesis with the heterodimer. This last possibility is highly probable since there is evidence that integrase has nonspecific affinity for DNA which has been mapped to the C-terminus of the protein

(88, 89). Furthermore, the nucleoprotein (preintegration) complex contains viral RNA and DNA in association with RT, viral matrix protein and integrase (87). Clearly, these questions open up avenues for further possible study in this area.

In summary, the M13mp2 fidelity assay system is an excellent system to measure DNA synthesis errors from both DNA and RNA templates. High levels of protein expression were obtained with most of the constructs. The data reported in this study suggest that different forms of HIV-1 reverse transcriptase precursors are enzymatically active, and have lower mutation rates than that of the mature form RT, and that the HIV-1 integrase of the precursors may confer the higher observed fidelity. Overall, fidelity of synthesis from an RNA+ template was lower than that from a DNA template. Errors encountered by the enzymes include base substitutions, deletions and insertions. 'Hot spots' were observed with mature form of RT but not with the precursors.

The work described above has provided additional clues with respect to the role of the reverse transcriptase precursor in the process of HIV-1 pathogenesis. Although these experiments utilize recombinant constructs and not purified, naturally occurring precursor, the observations made may permit some inferences to be made to the issue of viral hypermutability and anti-reverse transcriptase drug design. The methodology utilized in the fidelity studies may also prove useful in the study of fidelity of other enzymes.

Further research in the area of RT inhibition may take two primary directions. In the direction of nonnucleoside inhibitors, there is work to be done to find drugs that can recognize and inhibit the function of precursor and mature form of RT. By achieving this goal, it is hoped that reverse transcription can be completely abolished before resistance occurs. In the area of precursor role in the replication systems and maturation of the virus, more work needs to be done at the basic level to try to understand the process of heterodimer formation (including ways of inhibiting this process), and the role of the integrase subdomain in increasing fidelity.

REFERENCES

1. Merson, M.H., 1993, Slowing the Spread of HIV: Agenda for the 1990s, Science, Vol. 260: 1266-1268

2. Gallo R.C., Salahuddin S.Z., Popovic M., Et. al., 1984, Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk of AIDS, Science, Vol. 224: 500-503

3. Barre-Sinoussi F., Chermann J.C., Rey F., et al., 1983, Isolation of a T-lymphocyte retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), Science, Vol. 220, 868-871

4. Pantaleo G., Graziosi C., Faughi A.S., 1993, The Immunopathogenesis of Human Immunodeficiency Virus Infection, The New England Journal of Medicine, Vol. 328, No. 5: 327-336

5. Weiss R.A., 1993, How Does HIV Cause AIDS?, Science, Vol. 260: 1273-1279

6. Friedland G.H. et al., 1987, Transmition of the human immunodeficiency virus, The New England Journal of Medicine, Vol. 317: 1125

7. International Conference on AIDS,(News Report), April 1996, Vancouver

8. Chanock R.M., Hirsch M.S., Melnick J.L., Manoth T.P., and Roizman B., 1990, Fields Virology, Raven Press, New York, Vol. 2: 1529-1570

9. Arya S. et al., 1986, Three novel genes of human T-lymphotropic virus type III: immune reactivity of their products with sera from acquired immune defeciency syndrome patients, Proceedings of the National Academy Science, USA, Vol. 83: 2209

10. Sodroski J. et al., 1986, Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions, Science, Vol. 231: 1549

11. Sodroski J. Et al., 1986, A second post-transcriptional trans-activator gene required for HTLV-III replication, Nature, Vol. 321: 412

12. Zhang H., Zhangg Y., Spicer T.P., Abbott L.Z., Abbott M., and Poiesc B.J., 1993, Reverse Transcription Takes Place within Extracellular HIV-1 Virions: Potential Biological Significance, AIDS Research and Human Retroviruses, Vol. 9, No. 12:

1287-1296

13. Gelderblom H.R., Hausmann E.H.S., Ozel M., Pauli G., and Koch M.A., 1987, Fine structure of LAV/HTLV III and immunolocalization of structure poteins, Virology, Vol., 156: 171-176

14. Greene W.C., 1991, The Molecular Biologiy of Human Immunodeficiency Virus Type 1 Infection, The New England Journal of Medicine, Vol. 324, No. 5: 308-317

15. Goff S.P., 1990, Retroviral Reverse Transcriptase: Synthesis, Structure and Function, Journal of Acquired Immune Deficiency Syndromes, Vol. 3: 817-831

16. Bebenek K., Abbotts J., Roberts J.D., Wilson S.H., and Kunkel T.A., 1989, Specificity and Mechanism of Error-prone Replication by Human Immunodeficiency Virus-1 Reverse Transcriptase, The Journal of Biological Chemistry, Vol. 264, No. 28: 16948-16956

17. Saag M.S., Hahn B.H., Gibbons J., Li Y., Parks E.S., Parks W.P. and Shaw G.M., 1988, Extensive variation of human immunodeficiency virus type-1 in vivo, Nature, Vol. 334: 440-444

18. Coffin J.M., 1986, Genetic variation in AIDS viruses, Cell, Vol. 46: 1-4

19. Myers G., Josephs S.F., Rabson A.B., Smith T.F., and Wong-Staal F., 1987, in: Human Retroviruses and AIDS, 1987 Los Alamos National Laboratory, Los Alamos, NM

20. Roberts J.D., Bebenek K., and Kunkel T.A., 1988, The Accuracy of Reverse Transcriptase from HIV-1, Science, Vol. 242:1171-1173

21. Preston B.D., Poiesz B.J., and Loeb L.A., 1988, Fidelity of HIV-1 Reverse Transcriptase, Science, Vol. 242: 1168-1171

22. Ji J., Hoffmann J., and Loeb L., 1993, Mutagenecity and pausing of HIV reverse transcriptase during HIV plus-strand DNA synthesis, Nucleic Acids Research, Vol. 22, No. 1: 47-52

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

24. Drake J.W., 1969, Genetic control of mutation rates in bacteriophage T4, Nature, Vol. 221: 1132

25. Wabl M., Burrows P.D., von Gabain A., and Steinberg C., 1985, Hypermutation at the immunoglobulin heavy chain locus in a pre-B-cell line, Proceedings of the National Academy of Science, U.S.A., Vol. 82: 479-482

26. Kang I. and Wang J.H., 1994, Design of Structure-based Reverse Transcriptase Inhibitors, The Journal of Biological Chemistry, Vol. 269, No. 16: 12024-12031

27. Richman D., Shih C.K., Lowy I., Rose J., Prodanovich P., Goff S., and Griffin J., 1991, Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture, Proceedings of the National Academy of Science, U.S.A., Vol. 88: 11241-11245

28. Shih C.K., Rose J.M., Hansen G.L., Wu J.C., Bacolla A., and Griffin J.A., 1991, Chimeric human immunodeficiency virus type 1/type 2 reverse transcriptases display reversed sensitivity to nonnucleoside analog inhibitors, Proceedings of the National Academy of Science, U.S.A., Vol. 88: 9879-9882

29. Boyer P.L., Currens M. J., McMahon J. B., Boyd M. R., and Hughes S. H., 1993, Analysis of Nonnucleoside Drug-resistance Variants of Human Immunodeficiency Virus Type 1 Reverse Transcriptase, Journal of Virology, Vol. 67, No. 4: 2412-2420

30. Yeni, Schooley and Hammer, 1993, Antiretroviral and immune-based therapies, AIDS, Vol. 7, suppl. 1: s177-178

31. Merluzzi V.J., Hargrave K.D., Labadia M., Grozinger K., Skoog M., Wu J., Shih C.K., Eckner K., Hattox S., Adams J., Rosenthal A., Faanes R., Eckner R.J., Koup R.A., and Sullivan J.L., 1990, Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor, Science, Vol. 250: 1411-1413

32. Pauwels R., Andries K., Desmyter J., Schols D., Kukla M.J., Breslin H.J., Raeymaeckers A., Gelder J.V., Woestenborghs R., Heykants J., Schellekens K., Hanssen M.A.C., Clercq E.D. and Jassen P.A.J., 1990, Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives, Nature, Vol. 343: 470-474

33. Baba M., DeClercq E., Tanaka H., Ubasawa M., Takashima H., Sekiya K., Nitta I., Umezu K., Nakashima H., Mori S., Shigeta S., Walker r.T., and Miyasaka T., 1991, Potent and selective inhibition of human immunodeficiency virus type-1 (HIV-1) by 5-ethyl-6-phenylthiouracil derivatives through their interaction with the

HIV-1 reverse transcriptase, Proceedings of the National Academy of Science, U.S.A., Vol. 88: 2356-2360

34. Wiley C.A., Rutledge R.A., Dias S., Folks T., Theodore T., Buckler C.E., and Martin M.A., 1986, Cellular localization of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients, Proceedings of the National Academy of Science, U.S.A., Vol. 83: 5038-5042

35. Modrow S., Hahn B.H., Shaw G.M., Gallo R.C., Wong-Staal F., and Wolf H., 1987, computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions, Journal of Virology, Vol. 61: 570-578

36. Goodenow M., Huet T., Saurin W., Kwok S., Sninsky J., and Wain-Hobson S., 1989, Journal of AIDS, Vol. 2: 344-352

37. Vartanian J.P., Meyerhans A., Asjo B., and Wein-Hobson S., 1991, Selection, recombination, and G...A hypermutation of human immunodeficiency virus type-1 genomes, Journal of Virology, Vol. 65: 1779-1788

38. Tarrago-Litvak L., Andreola M.L., Nevinsky G.A., Sharih-Cottin L., and Litvak S., 1994, The reverse transcriptase of HIV-1: from enzymology to therapeutic intervention, The FASEB Journal, Vol. 8: 497-503

39. Onions D., 1991, Integration of Viruses into Chromosomal DNA, Journal of Pathology, Vol. 163: 191-197

40. Takeuchi Y., Nagumo T., and Hoshing H., 1988, Low Fidelity of Cell-Free DNA Synthesis by Reverse Transcriptase of Human Immunodeficiency Virus, Journal of Virology, Vol. 62: 3900-3902

41. Bakhanashvili M. and Hizi A., 1993, The fidelity of the reverse transcriptases of human immunodeficiency viruses and murine leukemia virus, exhibited by the mispair extension frequencies, is sequence dependent and enzyme related, FEBS Letters, Vol. 319, No. 1,2: 201-205

42. Weber J. and Gross F., 1989, Fidelity of human immunodeficiency virus type 1 reverse transcriptase in copying natural DNA, Nucleic Acids Research, Vol. 17: 1379-1393

43. Cohen K.A., Hopkins J., Ingraham R.H., Pargellis C., Wu J.C., Palladino D.E.H., Kinkade P., Warren T.C., Rogers S., Adams J., Farina P.R., and Grob P.M., 1991, Characterization of the Binding Site for Nevirapine (BI-RG-587), a Nonnucleoside

Inhibitor of Human Immunodeficiency Virus Type-1 Reverse Transcriptase, The Journal of Biological Chemistry, Vol. 266, No. 22: 14670-14674

44. Moore R.D., Hidalgo J., Sugland B.W., Chaisson R.E., 1991, Zidovudine and the Natural History of the Acquired Immunodeficiency Syndrome, The New England Journal of Medicine, Vol. 324, No. 20: 1412-1416

45. De Clercq E., 1992, HIV inhibitors targeted at the reverse transcriptase, AIDS Research and Human Retroviruses, Vol. 8: 111-134

46. Larder A. L., Darby G., and Richman D. D., 1989, HIV with Reduced Sensitivity to Zidovudine (AZT) Isolated During Prolonged Therapy, Science, Vol. 243: 1731-1734

47. Tudor-Williams G., St Clair M. H., Mckinney R. E., Maha M., Walter E., Santacroce S., Mintz M., O'Donnell K., Rudoll T., and Vavro C.L., Connor E.M., Wilfert C.M., 1992, HIV-1 Sensitivity to zidovudine and clinical outcome in children, The Lancet, Vol. 339: 15-19

48. Richman D.D., Havlir D., Corbeil J., Looney D., Ignacio Caroline, Spector S.A., Sullivan J., Cheeseman S., Kevin Barringer K., Pauletti D., Shih C.K., Myers M., and Griffin J., 1994, Nevirapine Resistance Mutations of Human Immunodeficiency Virus Type 1 Selected during Therapy, Journal of Virology, Mar: 1660-1666

49. Clair M.H.S., Martin J.L., Tudor-Williams G., Bach M.C., Vavro C.L., King D.M., Kellam P., Kemp S.D., and Larder B.A., 1991, Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase, Science, Vol. 253: 1557-1559

50. Romero D.L., Busso M., Tan C.K., Reusser F., Palmer J.R., Poppe S.M., Aristoff P.A., Downey K.M., So A.G., Resnick L., and Tarpley W.G., 1991, Nonnucleoside reverse transcriptase inhibitors that potently and specifically block human immunodeficiency virus type 1 replication, Proceedings of the National Academy Science, U.S.A., Vol. 88: 8806-8810

51. Nunberg J., Schleif W.A., Boots E.J., O'Brien J., Quintero J.C., Hoffman J.M., Emini E.A., and Goldman M., 1991, Viral resistance to human immunodeficiency virus type 1-specific Pyridinone reverse transcriptase inhibitors, Journal of Virology, Vol. 65: 4887-4892

52. Kellam P., Boucher C. A. B., and Larder A., 1992, Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development

of high-level resistance to zidovudine, Proceedings of the National Academy of Science U.S.A., Vol 89: 1934-1938

53. Ross E.K., Fuerst T.R., Orenstein J.M., O'Neill T., Martin M.A., and Venkatesan S., 1991, Maturation of Human Immunodeficiency Virus Particles Assembled from the gag Precursor Protein Requires In Situ Processing by gag-pol Protease, AIDS Research and Human Retroviruses, Vol. 7, No. 5: 475-483

54. Lori F., Scovassi A.I., Zella D., Achilli G., Cattaneo E., Casoli C., and Bertazzoni U., 1988, Enzymatically Active Forms of Reverse Transcriptase of the Human Immunodeficiency Virus, AIDS Research and Human Retroviruses, Vol. 4, No. 5: 393-398

55. Hu Y. and Kang C., 1991, Enzyme activities in four different forms of human immunodeficiency virus 1 *pol* gene products, Proceedings of the National Academy of Science U.S.A., Vol. 88: 4596-4600

56. Ji J., and Loeb A., 1991, Fidelity of HIV-1 Reverse Transcriptase Copying RNA in Vitro, Biochemistry, Vol. 31: 954-958

57. Mansky L.M. and Temin H.M., 1995, Lower In Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase, Journal of Virology, Vol. 69, No. 8: 5087-8094

58. Hubner A., Kruhoffer M., Gross F., and Krauss G., 1991, Journal of Molecular Biology, Vol. 223: 595-600

59. Boyer J.C., Bebenek K., and 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 Science U.S.A., Vol. 89: 6919-6923

60. Ji J. and Loeb L.A., 1992, Fidility of HIV-1 reverse transcriptase copying RNA in vitroBiochemistry, Vol. 31: 954-958

61. Kunkel T.A., 1985, The Mutational Specificity of DNA Polymerase-beta during in Vitro DNA Synthesis, Journal of Biological Chemistry, Vol. 260, No. 9: 5787-5796

62. Kunkel T.A. and Soni A., 1988, Exonucleolytic Proofreading Enhances the Fidelity of DNA Synthesis by Chick Embryo DNA Polymerase-gamma, Journal of Biological Chemistry, Vol. 263, No. 9: 4450-4459

63. Birnboim H.C. and Doly J., 1979, A repid alkaline extraction procedure for screening recombinant plasmid DNA, Nucleic Acid Research, Vol. 7:1513

64. Kunkel T.A., 1984, Mutational specificity of depurination, Proceedings of the National Academy of Science U.S.A., Vol. 81: 1494-1498

65. Parke J., and Morrow C.D., 1991, Overexpression of the gag-pol Precursor from Human Immunodeficiency Virus Type 1 Proviral Genomes Results in Efficient Proteolytic Processing in the Absence of Virion Production, Journal of Virology, Sept. 1991: 5111-5117

66. Bukrinsky M.E., Sharova N., Dempsey M.P., Stanwick T.L., Bukrinskaya A.g., Haggerty S., and Stevenson M., 1992, Active nuclear import of human immunodeficiency virus type 1 preintegration complexes, Proceedings of the National Academy of Science U.S.A., Vol. 89: 6580-6584

67. Prasad V.R. and Goff S.P., 1989, Linker insertion mutagenesis of the human immunodeficiency virus reverse transcriptase and endonuclease expressed in bacteria: definition of the minimal polymerase domain, Proceedings of the National Academy of Science U.S.A., Vol 86: 3104-3108

68. Chandra A., Gerber T., and Chandra P., 1986, Biochemical heterogeneity of reverse transcriptase purified from the AIDS virus, HTLV-III, FEBS Letters, Vol. 197: 84-88

69. Starnes M.C. and Cheng Y., 1989, Human immunodeficiency virus reverse transcriptase-associated RNase H activity, Journal of Biological Chemistry, Vol. 264: 7073-7077

70. Starnes M.C., Gao W., Ting R.C., and Cheng Y.C., 1988, enzyme activity gel analysis of human immunodeficiency virus reverse transcriptase, Journal of Biological Chemistry, Vol 263: 5132-5134

71. Lightfoote M.M., Coligan J.E., Folks T.M., Fauci A.S., martin M.A. and Venkatesan S., 1986, Structural characterization of reverse transcriprase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus, Journal of Virology, Vol. 60: 771-775

72. Hizi A., McGill C., and Hughes S.H., 1988, Expression of soluble, enzymatically active, human immunodeficiency virus reverse transcriptase in Escherichia coli and analysis of mutants, Proceedings of the National Academy of Science U.S.A., Vol. 85: 1218-1222

73. Leuthardt A., Le Grice S.F.J., 1988, Biosynthesis and analysis of a genetically engineered HIV-1 reverse transcriptase/endonuclease polyprotein in Escherichia coli, Gene, Vol. 68: 35-42

74. DeStefano J., Buiser R. G., Mallaber L. M., Bambare R. A., and Fay P. J., 1991, Human Immunodeficiency Virus Reverse Transcriptase Displays a Partially Processive 3' to 5' Endonuclease Activity, The Journal Of Biological Chemistry, Vol. 66, No. 36: 24295-24301

75. Keller W., and Crouch R., 1972, RNA-primed DNA synthesis in vitro, Proceedings of the National Academy of Science U.S.A., vol. 69: 3360-3364

76. Krug M.S., and Berger S.L., 1989, Ribonuclease H activities associated with viral reverse transcriptase are endonucleases, Proceedings of the National Academy of Science U.S.A., Vol. 86: 3539-3543

77. Ji J., and Loeb L. A., 1994, Fidelity of HIV-1 Reverse Transcriptase Copying a Hypervariable Region of the HIV-1 *env* Gene, Virology, Vol. 199: 323-330

78. Coffin J., 1995, HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy, Cience, Vol. 267: 483-489

79. Ho D.D., Neumann A.U., Perelson A.S., Chen W., Leonard J.M., and Markowitz M., 1995, Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection, Nature (London), Vol. 373: 123-126

80. Huber H.E., McCoy J.M., Seehra J.S., and Richardson C.C., 1989, Human immunodeficiency virus I reverse transcriptase: Template binding, processivity, strand displacement synthesis and template switching, Journal of Biological Chemistry, Vol. 264: 4669-4678

81. Hahn B. H., Gonda M. A., Shaw G. M., Popovic M., Hoxie J. A., Gallo R. C., and Wong-Staal F., 1985, Genomic diversity of the aquired immune deficiency syndrome virus HTLV-III: Different viruses exhibit greatest divergence in their envelope genes, Proceeding of the National Academy of Science U.S.A., Vol. 82:4813-4817

82. Bushman F. D., and Craigie R., 1991, Activities of human immunodeficiency virus (HIV) integration protein *in vitro*:specific cleavage and integration og HIV DNA, Proceedings of the National Academy of Science U.S.A., Vol. 88:1339-1343

83. Bushman F. D., Engelman A., Palmer I., Wingfield P., and Craigie R., 1993, Domains of the Integrase protein of human immunodeficiency virus type 1

responsible for polynucleotidyl transfer and zinc binding, Proceedings of the National Academy of Science U.S.A., Vol.90: 3428-3432

84. Craigie R., Fujiwara T., and Bushman F., 1990, The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*, Cell, Vol. 62:829-837

85. Cannon P. M., Wilson W., Byles E., Kingsman S. M., and Kingsman A. J., 1994, Human Immunodeficiency Virus Type 1 Integrase: Effect on Viral Replication of Mutation at Highly Conserved Residues, Journal of Virology, Vol. 68:4768-4775

86. Katz R.A., Mack J.P.G., Merkel G., Kulkosky J., Ge Z., Leis J., and Skalka A.M., 1992, Requirement for a conserved serine in both processing and joining activities of retroviral integrase, Proceedings of the National Academy of Science U.S.A., Vol. 89: 6741-6745

87. Bukrinsky M. I., Sharova N., McDonald T.L., Pushkarskaya T., Tarpley W.G., and Stevenson M, 1993, Association of integrase, matrix and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection, Proceedings of the National Academy of Science U.S.A., Vol.90:6125-6129

88. Vink C., Van Gent D. C., Elgersma Y., and Plasterk R. H. A., 1991, Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage, Journal of Virology, Vol. 65:4636-4644

89. Woerner A. M., Klutch M., Levin J. G., and Marcus-Sekura C. J., 1992, Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein, AIDS Research and Human Retroviruses, Vol. 8:297-304

90. Roberts J.D., and Kunkel T.A., 1996, Methods to Determine the Fidelity of DNA Replication in Human Cells, unpublished

Appendix 1: SOC Medium

This procedure was adopted from Roberts and Kunkel, 1996 (90) with some modifications. Bactotryptone, 20 g: Bactoyeast extract, 5 g; NaCl, 10 ml of a 1 M solution; and KCl, 2.5 ml of a 1 M solution were added to 970 ml dH_20. This solution was sterilized in an autoclave and then cooled to room temperature. Next, 10 ml of a 2 M Mg^{2+} stock (1 M $MgCl_2$ and 1 M $MgSO_4$, filter sterilized), and 10 ml of 2 M glucose stock (filter sterilized) were then added. The media was filtered through a 0.2 um filter.

Appendix 2: Minimal Plates (M9)

This procedure was modified from Roberts and Kunkel, 1996 (90). First,16 g of Difco agar were added to 1 liter of dH_20 and sterilized in an autoclave. When the agar had cooled to 50 °C, 0.3 ml of 100 mM IPTG, 20 ml of 50 x VB salts (described below), 20 ml of 10% glucose and 5 ml of 1 mg thiamine-HCl were added. Each of these solutions was sterilized either by filtration (0.2 um) or in an autoclave prior to their addition to the 50 °C agar. The solution was mixed well and dispensed into sterile, levelled petri dishes (25 ml/plate).

VB salts (50 x): $MgSO_4.7H_20$, 10 g; citric acid (anhydrate), 100 g; K_2HPO_4, 500 g; $Na_2HPO_4.2H_20$, 75 g. The above salts were dissolved in 1 liter of water, and the solution was sterilized in an autoclave.

Appendix 3: LB plates.

To prepare LB medium (Luria-Bertani medium), add 15 g bacto-agar, 10 g bacto-tryptone, 5 g bacto-yeast extract, and 10 g NaCl to 950 ml of deionized H_20. Shake the solution until the solutes dissolve. Adjust the pH to 7.0 with 5N NaOH, and adjust the volume to 1 liter with dH_20. Sterilize by autoclaving for 20 minutes, and allow the solution to cool to 50 °C. Pour 30 ml aliquots on 90 mm plates. If ampicillin is used, add when the temperature is 50 °C.

Appendix 4: mutation frequencies for reactions with DNA templates.

	MuLV	WT	p51	p66	p51/66	p97	p100
EXPERIMENT 1							
Total plaques	2020	4388	2050	663	8321	3905	7537
Mutants	8	4	3	6	129	15	24
Mut. frequency ($x10^{-3}$)	3.9	.9	1.5	9	15.5	3.8	3
EXPERIMENT 2							
Total plaques	4165	3480	1872	1384	7989	2018	10463
Mutants	9	0	1	7	133	17	20
Mut. frequency ($x10^{-3}$)	2.16	0	.53	5.1	16.6	8.4	1.9
EXPERIMENT 3							
Total plaques	3415	1196	2378	1583	6330	8477	---
Mutants	10	2	2	11	127	19	---
Mut. frequency ($x10^{-3}$)	2.93	1.7	0.84	6.9	20.0	2.2	---
EXPERIMENT 4							
Total plaques	---	---	---	---	13380	---	---
Mutants	---	---	---	---	139	---	---
Mut. frequency ($x10^{-3}$)	---	---	---	---	10.4	---	---
Standard Deviation	.9	0.85	0.49	1.95	3.9	3.2	0.92

Appendix 5: mutation frequencies for reactions with RNA templates.

	NIH RT	MuLV	WT	p51/66	p97	p100	pORF	p51/66/NIH
EXPERIMENT 1								
Total plaques	2812	3366	2453	1050	6068	236	1013	---
Mutants	51	13	6	23	59	26	5	---
Frequency ($x10^{-3}$)	18.14	3.9	2.4	21.9	9.7	11	4.9	---
EXPERIMENT 2								
Total plaques	3220	3951	2448	2690	3802	401	988	---
Mutants	57	16	6	58	36	40	2	---
Frequency ($x10^{-3}$)	17.70	4.0	2.5	21.6	9.5	10.0	2.0	---
EXPERIMENT 3								
Total plaques	2518	483	3049	4092	2505	2508	1068	1863
Mutants	57	1	9	72	34	21	5	22
Frequency ($x10^{-3}$)	22.6	2.0	3.0	17.6	13.6	8.4	4.7	11.8
EXPERIMENT 4								
Total plaques	---	---	---	7580	---	---	---	6965
Mutants	---	---	---	151	---	---	---	111
Frquency ($x10^{-3}$)	---	---	---	19.9	---	---	---	15.9
Standard Deviation	2.7	1.1	.3	2.4	2.3	1.4	1.6	2.6