Kinetics of formation of hypoxanthine containing base pairs by HIV-RT: RNA template effects on the base substitution frequencies

Michael R. Valentine and John Termini*

Division of Biology, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010, USA

Received as resubmission October 9, 2000; Revised and Accepted January 4, 2001

ABSTRACT

Hypoxanthine (H), the deamination product of adenine, has been implicated in the high frequency of A to G transitions observed in retroviral and other RNA genomes. Although H·C base pairs are thermodynamically more stable than other H·N pairs, polymerase selection may be determined in part by kinetic factors. Therefore, the hypoxanthine induced substitution pattern resulting from replication by viral polymerases may be more complex than that predicted from thermodynamics. We have examined the steady-state kinetics of formation of base pairs opposite template H in RNA by HIV-RT, and for the incorporation of diTTP during first- and second-strand synthesis. Hypoxanthine in an RNA template enhances the $k_{app}$ for pairing with standard dNTPs by factors of 10–1000 relative to adenine at the same sequence position. The order of base pairing preferences for H in RNA was observed to be H·C >> H·T > H·A > H·G. Steady-state kinetics of insertion for all possible mispairs formed with diTTP were examined on RNA and DNA templates of identical sequence. Insertion of diTTP opposite all bases occurs 2–20 times more frequently on RNA templates. This bias for higher insertion frequencies on RNA relative to DNA templates is also observed for formation of mispairs at template A. This kinetic advantage afforded by RNA templates for mismatches and pairing involving H suggests a higher induction of mutations at adenines during first-strand synthesis by HIV-RT.

INTRODUCTION

Deamination of bases in DNA and RNA is of great interest owing to the fact that conversion of a normal hydrogen bond donor (amine) to an acceptor (carbonyl) alters the base coding properties of nucleic acids. Thus, deamination events in DNA or RNA, if left unrepaird, could result in genetic alteration depending upon the identity of the nucleoside inserted opposite the deaminated bases by various polymerases.

In the special case of RNA genomes, the deamination of adenine to produce hypoxanthine (H) [the corresponding nucleoside of hypoxanthine is inosine] appears to be a common mechanism for hypermutagenesis. The evidence for H in RNA genomes derives principally from base substitutions observed in the corresponding cDNA libraries. The occurrence of A to G transitions is taken as evidence for deamination, owing to the well known thermodynamic preference for H·C base pairing (1). This transition in RNA cannot be readily explained by any other adenine modification. Although H·C pairs are known to be favored thermodynamically, H may pair with other bases, suggesting a wider variety of possible base substitutions upon replication of H-containing RNA. This base pairing degeneracy is best illustrated by the fact that H in tRNA wobble positions can pair with uracil, adenine or cytidine in mRNA. In light of this promiscuity, H has been proposed as an ‘inert’ base whose inclusion in hybridization probes could allow for the recognition of ambiguous nucleotides in complex genomic libraries (1,2).

In all, more than 15 published reports have used cDNA analysis as evidence for the presence of hypoxanthine in RNA (reviewed in 3). In the case of one retrovirus (avian leukosis virus), 30–50% of all adenines in a single clone were found to be modified, giving rise to A to G transitions (4). Similar results have been described for another avian retrovirus, Rous-associated virus 1 (5). RNA hypermutation ascribed to the presence of H in RNA has been described for the matrix mRNA of measles virus (6), the RNA genomes of respiratory syncytial virus (7), parainfluenza virus 3 (8), vesicular stomatitis virus (9) and polyomavirus (10). Examination of base substitutions in HIV env and TAR sequences using the database and alignment programs provided by the Los Alamos HIV Sequence Database (see Materials and Methods) similarly reveals a preponderance of adenine base substitutions. Among all possible base substitutions which could occur in these sequences, A to G transitions are overrepresented to the order of ~20–30%.

Hypoxanthine in RNA thus appears to be widely distributed, and there are several contributing pathways which can account for its occurrence. The discovery of a ubiquitous enzyme with deaminating activity specifically for adenine in dsRNA, the double-stranded RNA adenosine deaminase (dsRAD; 3) has been shown to be important in both the extensive deaminations leading to hypermutagenesis as well as site-specific H modification. Significantly, the TAR sequence of HIV has been shown to be
a substrate for the dsRAD protein (11). Deaminated adenines may also occur in RNA as a result of ITP incorporation during RNA synthesis. The concentration of ITP in human blood cells has been found to be ≥50 µM (12). This may be compared to average CTP levels, which are typically ~100–200 µM. Spontaneous hydrolysis of adenine in RNA would also contribute to the presence of H. The rate for spontaneous hydrolytic purine deamination is ~10-fold slower than that for deamination of pyrimidines (13–15), yet this pathway could contribute significantly since RNA is not repaired.

Purines in general are highly susceptible to deamination by nitric oxide (NO) and related oxidants (16–19). This reaction is significantly since RNA is not repaired. Purines and deaminations are shown in Table 1. The template sequence corresponds to nucleotides 962–996 of the hypervariable region of the envelope gene (env) of the WMJ1 strain of HIV-1 (25). The DNA template was identical in sequence to the wild-type env gene, substituting thymidine for uracil. Primers were used for the single nucleotide ‘standing start’ polymerase assay and terminate immediately before C17, G19, A(H)20 or U(T)25. Oligonucleotides were purified by 20% polyacrylamide (1.5 mm) gel electrophoresis (PAGE). DNA(RNA) was visualized by UV shadowing, excised from the gel, and recovered using the crush and soak method. DNA primers were radiolabeled by incubating 100–200 pmol of primer with 10 µM ATP, 10 µCi [γ-32P]ATP and 10 U of T4 polynucleotide kinase (NEB) in buffer provided by the manufacturer. The labeled primers were purified by phenol-saturated chloroform extraction. Unincorporated nucleotides were removed by passage through G25 spin columns (Pharmacia) and recovered in 100–150 µl of ddH2O.

During first-strand DNA synthesis by HIV-1 reverse transcriptase (HIV-RT), the presence of unrepaired H in template RNA will give rise to base substitutions which will be propagated during subsequent DNA replication cycles. Thus, the fidelity of first-strand synthesis governs, to an appreciable extent, the creation of genetic diversity in retroviruses. What this base substitution pattern might look like and to what extent it would reflect base substitutions at adenines observed in clinical isolates has never been examined. Results presented here represent our efforts to model the formation of base pairs involving H using a kinetic, rather than a thermodynamic approach, in order to assess potential pairing interactions during the replication of HIV RNA which could contribute to retroviral mutagenesis.

Steady-state kinetic parameters were determined for the formation of base pairs opposite a uniquely placed H within an RNA template. We have also examined the incorporation of dITP using RNA and DNA templates in order to explore the mutation-inducing potential of H introduced from the nucleotide precursor pool during first- and second-strand DNA synthesis by HIV-RT. In contrast to dUTP, which is rapidly degraded in cells, dITP has been shown to persist in dNTP precursor pools (22) and to be efficiently incorporated during replication (23), suggesting a role in the generation of base substitutions.

MATERIALS AND METHODS

Deoxynucleotide triphosphates

All dNTPs except for deoxyinosine triphosphate were purchased from Pharmacia as 100 mM Ultrapure stocks. Deoxyinosine triphosphate was purchased from Sigma. [γ-32P]ATP was purchased from ICN. All dNTP stock solutions (2X) were made up in ddH2O.

RNA and DNA oligonucleotides

Modified RNA templates were synthesized by Dharmacon, Inc. (Boulder, CO) using 2′-O-bis(2-acetoxxyethoxy)methyl (ACE) phosphoramidite technology (24). The template sequences and primers used for steady-state kinetic deter-

minations are shown in Table 1. The template sequence corresponds to nucleotides 926–996 of the hypervariable region of the envelope gene (env) of the WMJ1 strain of HIV-1 (25). The DNA template was identical in sequence to the wild-type env gene, substituting thymidine for uracil. Primers were used for the single nucleotide ‘standing start’ polymerase assay and terminate immediately before C17, G19, A(H)20 or U(T)25. Oligonucleotides were purified by 20% polyacrylamide (1.5 mm) gel electrophoresis (PAGE). DNA(RNA) was visualized by UV shadowing, excised from the gel, and recovered using the crush and soak method. DNA primers were radiolabeled by incubating 100–200 pmol of primer with 10 µM ATP, 10 µCi [γ-32P]ATP and 10 U of T4 polynucleotide kinase (NEB) in buffer provided by the manufacturer. The labeled primers were purified by phenol-saturated chloroform extraction. Unincorporated nucleotides were removed by passage through G25 spin columns (Pharmacia) and recovered in 100–150 µl of ddH2O.

| Primers | Template
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>5′-CCTCTCGTAAAATATGGC-3′</td>
</tr>
<tr>
<td>DNA</td>
<td>5′-CCTCTCGTAAAATATGGCC-3′</td>
</tr>
</tbody>
</table>

| Table 1. DNA primers and RNA and DNA templates |

The templates are derived from a hypervariable region of the extra-cellular region of the envelope gene of the WMJ-1 strain of HIV-1. Template positions that accept incoming bases in primer extension assays are in bold and denoted by an asterisk.

Purification of HIV-RT

Plasmid pHIV-RT(His)Prot, containing the genes for C-terminal hexahistidine tagged HIV-RT and HIV protease was the generous gift of Dr Paul Boyer (NCI). Co-expression of these genes in *Escherichia coli* results in cleavage of the HIV-RT C-terminus. The larger 51 kDa fragment resulting from this cleavage dimerizes with uncleaved HIV-RT to provide a p66/p51 heterodimer (26).

The plasmid was transformed into BL21(DE3)pLysS cells (Novagen, Madison, WI) by electroporation using a Bio-Rad Gene Pulser according to the manufacturer’s instructions, using chloramphenicol and ampicillin selection. The transformed cells were then grown up in 6 l broth (10 g bactotryptone, 5 g yeast extract, 10 g NaCl, pH 7.5, 25 mg/l chloramphenicol, 20 mg/l ampicillin) to an A600 of 0.6. This was followed by induction with isopropylthiogalactopyranoside (1.5 mM). Cell suspensions were incubated for 5 h at 37°C, harvested by centrifugation and stored at ~80°C. The yield of cells was ~5 g/l. Cells (~10 g) were resuspended in 20 ml of 50 mM NaPO4, pH 8.0,
50 mM NaCl, 1.5 mM phenylmethylsulfonyl fluoride and 0.2 mg/ml lysozyme. The lysate was then centrifuged at 85 × 10^3 g for 1 h. The supernatant was diluted with an equal volume of 66 mM NaPO₄, pH 6.8 and 300 mM NaCl, and loaded at a flow rate of 1 ml/min onto a 5 ml Ni-NTA superflow (Qiagen) column equilibrated with 100 ml of 50 mM NaPO₄, pH 7.0, 300 mM NaCl and 10 mM imidazole. Elution was carried out according to the standard protocol, and the fractions were analyzed by SDS–PAGE. Fractions containing p66 and p51 were pooled, concentrated to 1 ml using Centricon®-30 concentrators (Amicon). Dithiothreitol (DTT) was added to a final concentration of 1 mM, and samples were loaded onto a 300 ml Sephacyrl-300 (Pharmacia) gel filtration column equilibrated with 50 mM Tris, pH 8.0, 150 mM NaCl and 1 mM DTT. Fractions containing p66 and p51 were pooled, and the sample was dialyzed overnight against 50 mM Tris, pH 8.0, 2 mM EDTA and 50% glycerol. DTT was added to a concentration of 1 mM. The final enzyme concentration was determined by using an extinction value (ε₂₈₀) of 520 M⁻¹ cm⁻¹ (27). The unit activity was determined by incorporation of [α-³²P]dTTP (0.5 mM, 0.4 Ci/mmol) into a poly(rA)-pt(dT)₁₂₋₁₈ template (0.4 mM) in a 50 mM Tris, pH 7.2, reaction buffer containing 12.5 mM MgCl₂ and 40 mM KCl. Incorporation of radioactive nucleotides into acid insoluble product was determined by trichloroacetic acid precipitation.

**Primer extension reactions**

The primer extension protocol was adapted from the method of Boosalis et al. (28). Primer/template annealing was accomplished by combining 200 nM primer, 300 nM template in 20× buffer (1 M Tris, pH 7.2, 250 mM MgCl₂, 400 mM KCl), 2 mM DTT and −0.5 µCi labeled primer in 75 µl. This mixture was heated to 90°C, allowed to cool slowly to ∼40°C and then placed on ice. Next, 25 µl of 1 µM HIV-RT was added, and 4 µl of enzyme/template/primer mixture was aliquoted into each of a set of 0.5 ml Eppendorf tubes. Individual reaction tubes were warmed to 37°C and 4 µl of prewarmed 2× dNTP solution was added to each tube. Reactions proceeded for 5 min and were quenched with 16 µl of 90% formamide, 20 mM EDTA and 0.05% each of xylene cyanol FF and bromphenol blue, before placing on ice. The reactions were analyzed using 20% PAGE (17 × 15 cm). Gels were dried and exposed to a phosphorimager screen. The images were visualized using a PhosphorImager (Molecular Dynamics).

**Data analysis**

The amount of radioactivity in each band was determined using the ImageQuant software package (Molecular Dynamics). The velocity of each reaction was determined using the following equation:

\[
v = I_v[\text{primer}]/[I_0 + 1/2v_0] + t\]

where \(v\) is velocity in pM/s, \(I_0\) is the intensity of the unextended primer band, \(I_v\) is the intensity of the extended primer band (or the sum of all such bands if there are more than one), and \(t\) is reaction time in seconds. The velocity data were plotted using direct linear plots (29) and \(K_4\) and \(V_{max}\) were determined for each base pairing reaction. The frequency of insertion for a mispair (\(f_{ins}\)) is defined at \(k_{2app}^{(M)}/k_{2app}^{(WC)}\), the ratio of apparent second order rate constants for a mispair and the corresponding Watson–Crick base pair.

**Base substitution analyses of HIV sequences**

Sequence analysis was conducted using the database tools available at the HIV Sequence Database site (http://hiv-web.lanl.gov). Consensus sequences were generated using 18 complete env subtype A and B sequences (∼4 kb). For construction of the TAR consensus sequence, 409 sequence entries were used. Alignments were achieved using Sequencher™3.0.1 or Microsoft® Excel 98. The most common base at a given position was defined as the consensus sequence base. Sequences were compared to their respective consensus sequences using the Hypermut program. This returns a list of the differences between each sequence and the consensus. These were sorted by base changes and tabulated. Sequence differences resulting from insertions and deletions were omitted. Specific base substitutions expressed as percentages of all changes are provided in Table 4 for env subtypes A, B and TAR sequences.

**RESULTS**

**HIV-RT insertion of pyrimidine dNTPs opposite H or A in RNA**

The sequences of primers and templates used in these studies are presented in Table 1. The 35mer RNA template, based upon a region of the HIV-1 hypervariable domain of the env glycoprotein, was uniquely substituted with hypoxanthine. The steady-state parameters that describe the insertion of all four unmodified nucleotides at this site were determined. The corresponding values obtained for RNA templates containing adenine at this same position were also measured. Thus, the mutation-inducing potential resulting from deamination of A to H in a particular sequence context could be evaluated. In order to compare the potential for mispairs at H relative to A, the frequency of insertion values (\(f_{ins}\)) were calculated and are presented graphically in Figure 1. These values are the apparent second-order rate constants (\(k_{2app}^{(H)} = k_{cat}/K_M\)) for each possible base pair involving H₂₀ and A₂₀ in RNA divided by \(k_{cat}/K_M\) for the reaction A₂₀ + dTTP. A complete set of kinetic data for all reactions involving RNA templates is compiled in Table 2.
Table 2. HIV-RT steady-state kinetic data for RNA templates

<table>
<thead>
<tr>
<th>Template base</th>
<th>dNTP</th>
<th>$K_M$ (µM)</th>
<th>$V_{max}$ (pM/s)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>Second order rate (s$^{-1}$M$^{-1}$)</th>
<th>$f_{ins}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_{20}$ A</td>
<td>14</td>
<td>60</td>
<td>$4.8 \times 10^{-1}$</td>
<td>$3.4 \times 10^{-1}$</td>
<td>$5.0 \times 10^{-1}$</td>
<td></td>
</tr>
<tr>
<td>H$_{20}$ C</td>
<td>0.024</td>
<td>206</td>
<td>$2.1 \times 10^{-3}$</td>
<td>$8.7 \times 10^{-4}$</td>
<td>$1.3 \times 10^{-1}$</td>
<td></td>
</tr>
<tr>
<td>H$_{20}$ G</td>
<td>83.000</td>
<td>150</td>
<td>$1.2 \times 10^{-3}$</td>
<td>$1.4 \times 10^{-4}$</td>
<td>$2.1 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>H$_{20}$ T</td>
<td>0.590</td>
<td>88</td>
<td>$7.0 \times 10^{-4}$</td>
<td>$1.2 \times 10^{-5}$</td>
<td>$1.8 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$ H</td>
<td>530.000</td>
<td>110</td>
<td>$8.8 \times 10^{-4}$</td>
<td>$1.7 \times 10^{-5}$</td>
<td>$2.4 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$ G</td>
<td>0.043</td>
<td>104</td>
<td>$8.3 \times 10^{-4}$</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$8.0 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>C$_{17}$ H</td>
<td>92.000</td>
<td>76</td>
<td>$6.1 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$8.3 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>G$_{19}$ H</td>
<td>125.000</td>
<td>44</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$2.8 \times 10^{-5}$</td>
<td>$5.9 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$ A</td>
<td>97.000</td>
<td>51</td>
<td>$4.1 \times 10^{-4}$</td>
<td>$4.2 \times 10^{-5}$</td>
<td>$6.2 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>C$_{17}$ C</td>
<td>6.200</td>
<td>32</td>
<td>$2.6 \times 10^{-4}$</td>
<td>$4.1 \times 10^{-5}$</td>
<td>$6.1 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$ G</td>
<td>120.000</td>
<td>16</td>
<td>$1.3 \times 10^{-4}$</td>
<td>$1.1 \times 10^{-5}$</td>
<td>$1.6 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>C$_{17}$ A</td>
<td>41.000</td>
<td>60</td>
<td>$4.8 \times 10^{-4}$</td>
<td>$1.2 \times 10^{-5}$</td>
<td>$4.9 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>G$_{19}$ A</td>
<td>78.000</td>
<td>18</td>
<td>$1.4 \times 10^{-4}$</td>
<td>$1.8 \times 10^{-5}$</td>
<td>$2.3 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>C$_{17}$ G</td>
<td>0.099</td>
<td>30</td>
<td>$2.4 \times 10^{-4}$</td>
<td>$2.4 \times 10^{-5}$</td>
<td>$2.4 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>G$_{19}$ C</td>
<td>0.039</td>
<td>39</td>
<td>$3.1 \times 10^{-4}$</td>
<td>$8.0 \times 10^{-5}$</td>
<td>$8.0 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$ T</td>
<td>0.052</td>
<td>44</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$6.8 \times 10^{-5}$</td>
<td>$6.8 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>U$_{25}$ A</td>
<td>0.140</td>
<td>84</td>
<td>$6.7 \times 10^{-4}$</td>
<td>$4.8 \times 10^{-5}$</td>
<td>$4.8 \times 10^{-5}$</td>
<td></td>
</tr>
</tbody>
</table>

1Numbers correspond to template base positions from Table 1.

2Frequency of insertion ($f_{ins}$) is defined in the legend to Figure 1.

Inspection of Figure 1 reveals a dramatic preference for H·C base pair formation by HIV-RT. The insertion of dCTP opposite H$_{20}$ was found to occur 2100 times more frequently than insertion of dCTP opposite A$_{20}$. Relative to formation of the normal Watson–Crick base pair at this position (A$_{20}$ + dTTP), the H$_{20}$ + dCTP pairing occurs 13 times more frequently. In contrast, the incorporation of dCTP opposite A$_{20}$ occurs 164-fold less frequently than the Watson–Crick pairing. Comparison of $k_{2app}$ values reveals that this reaction is favored even over the formation of G·C pairs by a factor of ~10. The data in Table 2 demonstrate that this is due to both lower $K_M$ and increased $V_{max}$ values. In contrast to the situation observed with dCTP, the kinetic discrimination for insertion of dTTP at H$_{20}$ versus A$_{20}$ is not very large. Examination of the apparent second order rate constants for the reactions r(A/H)$_{20}$ + dTTP reveals that the H·T pairing is only ~5-fold less efficient than formation of the Watson–Crick base pair.

HIV-RT insertion of purine dNTPs opposite H or A in RNA

Comparison of $f_{ins}$ values for the reactions at r(A/H)$_{20}$ with dGTP reveals that H·G pairs are preferred, by a factor of 13. This preference is influenced primarily by a substantially larger $V_{max}$ (~10-fold) rather than a significant lowering of $K_M$ (~1.4-fold). Insertion of dATP opposite rH$_{20}$ by HIV-RT is favored over the rA$_{20}$ + dATP reaction by a factor of three, due solely to differences in $K_M$. Thus, the data in Figure 1 and Table 2 reveal that substitution of H for A in template RNA significantly increases the probability for formation of base pairs leading to transversions.

H-N base pair formation efficiency: participation as template base or dITP

For any given H-N base pair, the efficiency of formation may vary depending upon whether H participates as an RNA template base or as dITP from the nucleoside triphosphate precursor pool. These situations are not kinetically equivalent. Steady-state kinetic differences for individual base pair combinations involving hypoxanthine were examined within this context (rH + dNTP versus rN + dITP). The apparent second order rate constants for H-N combinations are grouped in this manner and are plotted in Figure 2. It can be seen that, for any given base pair combination, the reaction in which H is in the RNA template proceeds with greater catalytic efficiency. Alternatively, these results may be viewed as disfavoring reactions which require the binding and orienting of dITP at the polymerase active site. In the case of H·A pairs, insertion of dATP opposite template H is favored ~20-fold relative to insertion of dTTP opposite A. The differences in second order rate constants are primarily reflected in a nearly 40-fold difference in $K_M$, while the $k_{cat}$ values differ by <2-fold.

In the case of H·G, the reaction with template H is only modestly favored by a 2-fold greater $k_{cat}$, while the $K_M$ values are nearly equivalent. A larger ‘RNA template effect’ is observed for H·Pyridine pairings. The H·C pairing reaction favors H in the template by nearly 5-fold, due to enhanced $k_{cat}$ and lowered $K_M$. The much larger variation observed for the second order rate constants for the rH + dTTP and rU + dITP reactions (~400-fold) may reflect, in part, differences between thymine and uracil. It is difficult, however, to rationalize this large
effect on the basis of the 5-methyl group alone. The $K_M$ for the favored reaction with H in the template is $\sim 200$-fold lower, while the $k_{cat}$ advantage is small, only 2-fold (Table 2). Thus, for every possible pairing involving H catalyzed by HIV-RT, H in the RNA template provides a more favorable opportunity for non-standard pairing. The magnitude of the differences in $k_{cat}/K_M$ values between rH + dNTP versus rN + dITP undoubt-
edly reflect, to some extent, the influence of nearest-neighbor bases. Nonetheless the values for reactions with H in the RNA template are greater in every case, regardless of the exact sequence environment. This suggests that the RNA template effect overrides insertion preferences due to sequence context.

Table 3. HIV-RT steady-state kinetic data for DNA template

<table>
<thead>
<tr>
<th>Template base</th>
<th>dNTP</th>
<th>$K_M$ (µM)</th>
<th>$V_{max}$ (µM/s)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>Second order rate (s$^{-1}$M$^{-1}$)</th>
<th>$f_{ins}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A$_{20}$</td>
<td>H</td>
<td>210.000</td>
<td>56</td>
<td>$4.5 \times 10^{-4}$</td>
<td>2.1 $\times 10^0$</td>
<td>6.7 $\times 10^{-3}$</td>
</tr>
<tr>
<td>C$_{17}$</td>
<td>H</td>
<td>0.082</td>
<td>210</td>
<td>$1.7 \times 10^{-3}$</td>
<td>2.0 $\times 10^0$</td>
<td>1.0 $\times 10^0$</td>
</tr>
<tr>
<td>G$_{19}$</td>
<td>H</td>
<td>550.000</td>
<td>100</td>
<td>$8.0 \times 10^{-4}$</td>
<td>1.4 $\times 10^0$</td>
<td>4.1 $\times 10^{-4}$</td>
</tr>
<tr>
<td>T$_{25}$</td>
<td>H</td>
<td>34.000</td>
<td>68</td>
<td>$5.4 \times 10^{-4}$</td>
<td>1.6 $\times 10^0$</td>
<td>2.6 $\times 10^{-4}$</td>
</tr>
<tr>
<td>A$_{20}$</td>
<td>A</td>
<td>23.000</td>
<td>17</td>
<td>$1.36 \times 10^{-4}$</td>
<td>5.9 $\times 10^6$</td>
<td>1.8 $\times 10^{-4}$</td>
</tr>
<tr>
<td>A$_{20}$</td>
<td>C</td>
<td>52.000</td>
<td>95</td>
<td>$7.6 \times 10^{-4}$</td>
<td>1.5 $\times 10^0$</td>
<td>4.1 $\times 10^{-4}$</td>
</tr>
<tr>
<td>A$_{20}$</td>
<td>G</td>
<td>78.000</td>
<td>35</td>
<td>$2.8 \times 10^{-4}$</td>
<td>3.6 $\times 10^0$</td>
<td>1.0 $\times 10^0$</td>
</tr>
<tr>
<td>C$_{17}$</td>
<td>A</td>
<td>5.400</td>
<td>35</td>
<td>$2.8 \times 10^{-4}$</td>
<td>5.2 $\times 10^1$</td>
<td>1.4 $\times 10^{-4}$</td>
</tr>
<tr>
<td>G$_{19}$</td>
<td>A</td>
<td>34.000</td>
<td>43</td>
<td>$3.4 \times 10^{-4}$</td>
<td>1.0 $\times 10^0$</td>
<td>1.7 $\times 10^{-4}$</td>
</tr>
<tr>
<td>C$_{17}$</td>
<td>G</td>
<td>0.075</td>
<td>190</td>
<td>$1.5 \times 10^{-3}$</td>
<td>2.0 $\times 10^9$</td>
<td></td>
</tr>
<tr>
<td>G$_{19}$</td>
<td>C</td>
<td>0.026</td>
<td>104</td>
<td>$8.3 \times 10^{-4}$</td>
<td>3.2 $\times 10^0$</td>
<td></td>
</tr>
<tr>
<td>A$_{20}$</td>
<td>T</td>
<td>0.056</td>
<td>250</td>
<td>$2.0 \times 10^{-3}$</td>
<td>3.6 $\times 10^9$</td>
<td></td>
</tr>
<tr>
<td>T$_{25}$</td>
<td>A</td>
<td>0.027</td>
<td>205</td>
<td>$1.6 \times 10^{-3}$</td>
<td>6.1 $\times 10^9$</td>
<td></td>
</tr>
</tbody>
</table>

*Numbers correspond to template base positions from Table 1.

**Figure 2.** Apparent second order rate constants describing the formation of heteroduplex H·N base pairs and the corresponding N·H inversion pair. The $k_{cat}$ values for reactions rH + dNTP and rN + dITP for a given N are grouped by color.

**Insertion of dITP on DNA and RNA templates within the same sequence context**

The formation of base pairs involving hypoxanthine by HIV-RT was also examined on a DNA oligonucleotide identical in sequence to the RNA template except for replacement of U with T, and A for H at position 20 (Table 1). All possible mispairs on the DNA template involving H were examined using dITP (Table 3). These reactions could be compared directly to the corresponding insertion reactions of dITP on RNA templates (Table 2). The complete steady-state kinetic values measured for all reactions using the DNA template are presented in Table 3.

For ease of comparison, $f_{ins}$ for the reaction of dITP at specific positions on DNA and RNA templates are displayed graphically in Figure 3. The values of $f_{ins}$ measured for insertion of dITP opposite template RNA bases are significantly greater than those determined opposite bases in the DNA template for the same sequence positions. For example, the $f_{ins}$ for dITP insertion opposite rG$_{19}$ catalyzed by HIV-RT is $\sim 20$-fold greater than that for insertion opposite dG$_{19}$. The fidelity for any ‘non-standard’ pairing event occurring is defined as $1/f_{ins}$. Thus, insertion of dITP at rG$_{19}$ can be expected to occur approximately once every 1000 replication events, whereas insertion of dITP at dG$_{19}$ within the same sequence context would occur once in 22,000 cycles.

The base pairing reaction with dITP at rC$_{17}$ was catalyzed 8-fold more readily than the reaction at dC$_{17}$. In the case of the DNA template reaction, the insertion of dITP at dC$_{17}$ was characterized by an apparent second order rate constant which was virtually identical to that measured for the insertion of dGTP at the same position.

The insertion of dITP at rA$_{20}$ is preferred over insertion at dA$_{20}$ by $\sim 4$-fold. The least discrimination in favor of RNA templates was observed for the insertion of dITP at (U/T)$_{25}$,
where the reaction at U was favored by ~2-fold. These comparisons reveal that the incorporation of dITP from nucleotide precursor pools by HIV-RT is more likely to occur during first-strand DNA synthesis from the retroviral RNA template rather than during copying of the cDNA (second-strand DNA synthesis).

Examination of the individual components of the $f_{in}$ values, the apparent second order rate constants ($k_{2app}$), reveals that some of the enhanced values of $f_{in}$ for the reactions of dITP on the RNA template reflect lower values of $k_{2app}$ for the Watson–Crick base pairing reactions. For example, the reactions of dITP at C17 are characterized by identical $k_{2app}$ values on both DNA and RNA ($2 \times 10^4 \text{s}^{-1}\text{M}^{-1}$, Tables 2 and 3). However, the $k_{2app}$ for the insertion of dGTP opposite C17 (the denominator component of $f_{in}$ for the dITP + C17 reaction) is nearly 10-fold lower on the RNA template. Thus, the lower efficiency of formation of C17-G pairs on RNA enables dITP to compete more efficiently for insertion with dGTP, increasing the probability of insertion. Watson–Crick base pair formation on the DNA template rather than during copying of the cDNA (second-strand DNA synthesis).

A similar explanation appears to hold for the 4-fold increased probability in favor of insertion of dITP at rA20 relative to dA20. The $k_{2app}$ values are virtually identical for the respective reactions with dITP (~2 $\text{s}^{-1}\text{M}^{-1}$), yet the reaction forming A20-T base pairs is substantially less efficient on the RNA template (Tables 2 and 3). Comparison of $f_{in}$ values shows that the insertion of dITP opposite rG19 relative to dG19 is favored in the former case by 2-fold. In this example, the $k_{2app}$ for the rG19 + dITP reaction is ~5-fold greater than the value describing dITP insertion at dG19. The rG19 + dCPT reaction has a 4-fold lower value of $k_{2app}$ compared to the reaction on DNA, which contributes to the large increase in insertion fidelity. Interestingly, we observe in all cases for each position within this template sequence that the $k_{2app}$ values for Watson–Crick base pair formation catalyzed by HIV-RT are invariably lower (4–13-fold) on RNA compared to DNA templates.

**Base substitution analyses of HIV sequences**

In order to better analyze the base substitution trends in HIV-1 which could potentially arise via the intermediacy of hypoxanthine, a comprehensive analysis of env subtypes A and B was undertaken using the ‘Search and Align’ feature of the HIV Sequence Database site (http://hiv-web.lanl.gov). Frequencies of base substitutions could be compared with biases predicted from our in vitro data. This database includes updated HIV-1 sequence information in addition to a number of tools for genome analysis. Analyses were carried out within distinct subtypes of env since these are normally considered to have evolved independently. Thus, in theory the base substitution patterns within different subtypes might be dissimilar. The TAR sequence was also analyzed for base substitutions since it is a known double-stranded element in the HIV genome, and has been shown to be a substrate for the dsRAD (11). Base substitution frequencies for all three sequences are provided in Table 4.

<table>
<thead>
<tr>
<th>Base change</th>
<th>env Subtype A (%)</th>
<th>env Subtype B (%)</th>
<th>TAR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A to C</td>
<td>9.6</td>
<td>8.3</td>
<td>1.9</td>
</tr>
<tr>
<td>A to G</td>
<td>23.4</td>
<td>20.4</td>
<td>27.4</td>
</tr>
<tr>
<td>A to T</td>
<td>4.8</td>
<td>5.5</td>
<td>4.2</td>
</tr>
<tr>
<td>C to A</td>
<td>6.0</td>
<td>5.3</td>
<td>1.9</td>
</tr>
<tr>
<td>C to G</td>
<td>2.6</td>
<td>2.0</td>
<td>2.9</td>
</tr>
<tr>
<td>C to T</td>
<td>12.0</td>
<td>10.4</td>
<td>9.7</td>
</tr>
<tr>
<td>G to A</td>
<td>18.8</td>
<td>21.7</td>
<td>15.2</td>
</tr>
<tr>
<td>G to C</td>
<td>4.0</td>
<td>3.5</td>
<td>1.7</td>
</tr>
<tr>
<td>G to T</td>
<td>3.4</td>
<td>2.9</td>
<td>10.9</td>
</tr>
<tr>
<td>T to A</td>
<td>4.5</td>
<td>5.9</td>
<td>5.4</td>
</tr>
<tr>
<td>T to C</td>
<td>9.9</td>
<td>7.0</td>
<td>11.9</td>
</tr>
<tr>
<td>T to G</td>
<td>5.0</td>
<td>7.2</td>
<td>6.9</td>
</tr>
</tbody>
</table>

Examination of Table 4 reveals that the most frequent base substitution found among all sequences is A to G. In the TAR sequence, this accounts for as much as 30% of all base changes. The phenomenon of G to A hypermutation (30) is also observed in the analyses of these sequences, accounting for ~20% of base changes in the env sequences. The extent of this substitution is lower in TAR sequences (15%). Other types of base substitutions are more represented in TAR than in the env subtypes. For example, G to T substitutions are substantially higher in TAR (~11 versus ~3%) than in env subtypes. Other substitutions, for example T to A, do not vary significantly among the analyzed sequences. It is interesting to note that the relative ratio of A to C and A to T substitutions in the TAR sequence (1:2) resembles the ratio of insertion frequencies for dGTP and dATP opposite H by HIV-RT. This is not the case for the env subtypes, where A to C changes are more numerous than A to T (1.5–2:1).
DISCUSSION

The overrepresentation of adenine in the HIV genome (~60%; 31) suggests that even modest levels of deamination of this base could potentially contribute to hypermutagenesis. We have examined the steady-state kinetics of formation of base pairs involving the adenine deamination product, hypoxanthine, by HIV-RT during first-strand DNA synthesis on RNA templates. Kinetic analyses of base pairing events involving hypoxanthine have not previously been described. By substituting H for A at a specific site in an oligoribonucleotide template, the mispairing potential in RNA due to a single deamination event within a defined sequence context could be examined in detail. The frequencies with which dITP from nucleoside precursor pools could be incorporated at various positions during first-strand synthesis on RNA templates and on DNA templates were also examined within identical sequence contexts.

The hierarchy of kinetic insertion preferences for dNTPs opposite H in RNA by HIV-RT determined in our studies was found to be H·C >> H·T > H·A > H·G. The order of thermodynamic stabilities of base pairs containing H in DNA has been determined to be H·C > H·A > H·T > H·G (1). According to thermodynamic melting studies H·C is less stable than an A·T base pair within the same sequence context of RNA. However, our data indicate that formation of an H·C base pair occurs 13-fold more frequently than an A·T base pair in RNA with identical flanking sequences. If the same stability trends found in DNA hold for the heteroduplex primer/template complexes, it would appear that thermodynamic considerations may not provide a reliable indicator of insertion frequencies. Thermodynamic data from Martin et al. (1) suggest that H·T base pairs should be the least stable among all four possible H-containing base pairs. However, we find that H·T base pair formation is characterized by relatively high insertion frequencies, surpassed only by H·C base pairs (Fig. 1). This is true not only for heteroduplex base pairs formed on RNA templates (rH20 + dTTP), but for reactions on DNA templates as well (dT25 + dTTP; Fig. 3).

Echols and Goodman (32) have demonstrated that base pair thermodynamic stability does not usually predict base pairing preferences by DNA polymerases. Additional thermodynamic studies on H-containing heteroduplexes will be required to evaluate what role, if any, thermodynamic stability plays in base pairing preferences when RNA is the template. Echols and Goodman have also formulated a principle of geometric selection, dictated by base pair geometry, to explain the observed order of insertion frequencies by polymerases (32,33).

The greater importance of base pair geometry relative to simple thermodynamic considerations in rationalizing frequencies of mispair formation is further supported in our studies by examining homopurine pairing events. Somewhat surprisingly, homopurine base pairs may perturb overall helix geometry in a negligible manner. However, the overall contribution of such pairs to thermodynamic stability may be destabilizing. For example, A·G pairs in RNA produce large helical perturbations, whereas H·A pairs are accommodated with little deviation from the normal A helix dimensions (34). Yet A·G pairs contribute a greater thermodynamic stability to duplexes than H·A pairs. We observe an ~30-fold preference for H20·A over A20·G pairing by HIV-RT in our studies, in spite of the greater thermodynamic stability of the latter pair.

A similar rationale may hold for the observation of preferential A·A over A·G base pair formation in RNA templates. The rA20 + dATP reaction is favored over rA20 + dGTP by a factor of ~4-fold (Table 2). The NMR structure of the A·A pair shows remarkably little deviation from the normal B configuration in DNA, and effective stacking within the duplex is observed (35). This is in spite of the fact that in the anti/anti configuration, only a single hydrogen bond (NH2·N1) may be formed. Although there is a greater thermodynamic contribution to duplex stability by A·G base pairs (36), geometric discrimination by HIV-RT may favor the A·A pairing arrangement. The A·A pair is also favored over A·G in DNA homoduplexes, although to a lesser extent.

For each H·N base pair, we have shown that reactions where H is in the RNA template invariably proceed with a greater efficiency than when participating as the incoming triphosphate. This suggests that in retrovirial first-strand synthesis, deamination of a A in the template is potentially more mutagenic than deamination of dATP in the nucleotide triphosphate precursor pool. This variation may be due in part to differences in the polarity of stacking for a given base pair with respect to the helix, depending upon whether it is in the 3′ or 5′ strand. The geometrical and thermodynamic properties of each configuration are expected to be non-equivalent. Partitioning between the relative contributions of each of these effects will require NMR or crystallographic studies of H·N pairs in both possible orientations in heteroduplexes in conjunction with calorimetric studies.

The notion that hypoxanthine pairs like guanine in DNA is supported by the fact that the apparent second order rate constants for the dC17 + dITP and dC17 + dGTP reactions are the same (2 × 10^4 M^-1 s^-1). The respective contributions of K_M and V_max to this value of k_on are nearly identical for both reactions. This is not the situation, however, when the template strand is RNA. Here, the rC17 + dITP reaction is characterized by an apparent second order rate constant which is nearly 10-fold greater than for rC17 + dGTP. Thus, when RNA is the template, the condition of first-strand synthesis, HIV-RT appears to favor the formation of H·C even over G·C pairs within the same sequence context. This large preference for H·C formation in heteroduplexes can also be observed by comparing k_on/K_M values for the rH20 + dCTP reaction with the formation of G·C pairs at independent sites (different sequence contexts). This reaction is 11- and 36-fold more efficient than the rG19 + dCTP and rC17 + dGTP reactions, respectively (Table 2).

The structures of all possible base pairs involving hypoxanthine are shown in Figure 4. Crystal structures of H·C-containing duplexes in both B and Z DNA (36,37) reveal an overall geometry similar to that of G·C pairs. The two hydrogen-bond distances in the H·C base pair in DNA are nearly equivalent to the corresponding bonds in G·C pairs within the same sequence context (36). The local helix geometry of H·C pairs in B form DNA resembles that of a polyadenine tract in DNA, possessing large propeller twist angles and a narrow minor groove. Whether a similar perturbation exists in RNA–DNA heteroduplexes is currently unknown.

The duplexes involved in chain elongation by RT are either heteroduplexes (first-strand synthesis) or all DNA duplexes (second-strand synthesis). Since heteroduplexes have some structural features which distinguish them from DNA
duplexes, it is possible that some of the kinetic differences we observe between first- and second-strand synthesis for a given pairing are due to these features. Heteroduplexes demonstrate a heteronomous configuration, meaning that the RNA strand maintains a C3′ endo (A form) sugar pucker while the DNA strand is found to be C2′ endo (38). However, under certain conditions, DNA homoduplexes may also display heteronomous configurations between strands (39). Thus, the extent to which this may contribute to the kinetic differences is unclear.

Heteroduplexes appear to differ significantly from RNA and DNA homoduplexes in the extent of hydration of the major groove. RNA homoduplexes are the least hydrated. Although more water is accessible to the major groove of DNA homoduplexes, there is more significant hydration in the heteroduplexes (40). This is facilitated by the 2′-OH group, which in the heteroduplex adopts a preferred orientation facing the complementary base, rather than towards C3′ (Φ = 0 to −30°) of the template base. Water molecules hydrogen bonded in this manner in an RNA base pair result in a significant amount of hydration at the C2 position of adenine (40). Analogously, one might anticipate a similar hydration environment at H in heteroduplexes. Water molecules oriented about the C2 position of H may facilitate the incorporation of dCTP by lowering the transition state energy for incorporation via hydrogen bonding interactions with O2 of dCTP (Fig. 4). This may facilitate the rapid adoption of an optimal orientation for polymerization within the RT active site, resulting in the unusually large $k_{\text{app}}$ observed in our studies.

Alignment analyses of three HIV gene sequences revealed an unusual preponderance of A to G transitions. The largest percentage of such changes, nearly 30% were found in the double-stranded TAR sequence. Studies of base substitution patterns for other HIV genes, such as gag, have similarly revealed a high percentage of A to G transitions (41). The majority of these base substitutions are not lethal, suggesting that they may play a role in generating adaptive variants of the virus (42). It is also noteworthy that the base substitution patterns within the two env subtypes are remarkably similar, in spite of the fact that they must have evolved from different ancestors in non-identical environments. This suggests that the mechanistic forces acting to produce base substitutions are similar.

The steady-state kinetics of dITP insertion by HIV-RT on RNA templates were directly compared to insertion on DNA templates in order to evaluate the mutagenic potential of H introduced from the nucleoside precursor pool during first- and second-strand DNA synthesis. Base pairs are formed more readily when RNA is the template for all pairings with dITP. This suggests that during retroviral replication, incorporation of dITP from the nucleoside precursor pool is more likely to occur during first-strand rather than second-strand synthesis. This effect is largest for the G19 + dITP reaction, which is favored in RNA over DNA by a factor of 20. Second-strand DNA synthesis following incorporation of dITP opposite G would most likely pair dCTP with H in the initially synthesized DNA strand. This accomplishes an overall G to C transversion in the retroviral genomic sequence. A similar mechanism could account for some percentage of A to C transversions in the retroviral genome, due to mis-incorporation of dITP opposite A during first-strand synthesis.

The rC17 + dITP reaction is 8-fold more efficient than the rC17 + dGTP reaction. However, since second-strand synthesis by HIV-RT would likely pair the resulting H with dCTP, no mutation results. Incorporation of dITP opposite rC17, rG19, rA20 and U25 occurs, respectively, 8-, 20-, 4- and 2-fold more frequently than for the corresponding DNA template bases. Thus, we would predict that there is a greater probability for incorporation of H from nucleotide precursor pools during first-strand synthesis, since dITP can compete more effectively with unmodified dNTPs when RNA is the template.

The relatively high efficiency of the rH + dTTP reaction implies that the standard method of assaying for the extent of adenine deamination using cDNA transcript analysis may actually underestimate the level of H in RNA. This efficiency may depend to some extent on the exact polymerase used to construct the cDNA. Footprinting methods for the direct detection of H in RNA have recently been developed (43), and should now make possible a more definitive quantitation of the occurrence of H in various RNA genomes.

Independent of the involvement of H, HIV-RT-catalyzed formation of A·C mispairs in RNA can also lead to A to G transitions. It is interesting to note that the rA20 + dCTP reaction has a 15-fold higher $f_{\text{ins}}$ value than the dA20 + dCTP reaction. In fact, relative to mispairs formed on RNA with unmodified bases, the A·C mispair occurs with the highest frequency. However, the value of $f_{\text{ins}}$ for the H·C mispair in RNA is greater than that of A·C by more than three orders of magnitude, thus, one would still predict a greater influence of H in the A to G transition phenomenon.

Tables 2 and 3 contain additional kinetic information on ‘normal’ mispairs with mutation-inducing potential for both DNA and RNA templates. Comparison of $f_{\text{ins}}$ values for A·A and A·G mispairs reveals a 2- and 3-fold greater propensity for formation on RNA templates, respectively. It remains to be determined whether mispairs in general occur with greater frequency on RNA templates when HIV-RT is the replicative polymerase. Since most of the HIV-RT fidelity studies have relied on DNA templates in order to evaluate mismatch potential, an expanded RNA database of mispairs in different sequence contexts will be required to assess this possibility. This would suggest, as we have illustrated for replication involving H, that HIV-RT copying of heteroduplexes is
inherently more error prone, and would imply that the majority of base substitutions likely occur during first-strand synthesis.

ACKNOWLEDGEMENTS

We thank Dr Gerald E. Wenschell for helpful discussions and advice. The assistance of Toni Martinez in the preparation of this manuscript is gratefully appreciated. This work was supported by PHS Grant GM53692 to J.T.

REFERENCES


