The EMBO Journal vol.14 no.18 pp.4622–4627, 1995

# Human immunodeficiency virus type 1 reverse transcriptase $_t$G:T mispair formation on RNA and DNA templates with mismatched primers: a kinetic and thermodynamic study

**Monica Sala[1], Simon Wain-Hobson[1,2] and Francis Schaeffer[3]**

[1]Unité de Rétrovirologie Moléculaire and [3]Unité de Physicochimie des Macromolécules Biologiques (URA1149 du CNRS), Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, France

[2]Corresponding author

**The relationship between human immunodeficiency virus (HIV) type 1 reverse transcriptase $_t$G:T mispair formation and base pair stability was investigated using DNA and RNA templates with 15 bp matched or mismatched DNA primers. $_t$G:T mispair formation during primer elongation was undetectable on $_t$DNA–DNA duplexes but occurred with a frequency of $10^{-4}$ on matched $_t$RNA–DNA duplexes. The frequency increased to $7.0\times10^{-4}$ and $1.3\times10^{-3}$ on $_t$RNA–DNA duplexes with $_t$G:T mismatches located 6 and 9 bp beyond the polymerization site. From $K_m$ values at 37°C, the free energy change upon dissociation ($\Delta G°_{37}$) of the $_t$G:T mispair increased from matched to mismatched $_t$RNA–DNA duplexes by 0.36–1.21 kcal/mol. $\Delta G°_{37}$ for a correct $_t$G:C pair decreased by 0.06–1.00 kcal/mol. In comparison with DNA–DNA duplexes, thermal melting measurements on RNA–DNA duplexes demonstrated smaller enthalpy ($\Delta\Delta H°$ = –17.7 to –28.1 kcal/mol) and entropy ($\Delta\Delta S°$ = –59.3 to –83.4 cal/mol/K) components. A strong entropy–enthalpy compensation resulted in small free energy differences ($\Delta\Delta G°_{37}$ = 0.8 to –2.2 kcal/mol). Thus, although DNA–DNA and RNA–DNA duplexes are of comparable stability in solution, the RNA–DNA duplex presents more facile base pair opening and higher conformational flexibility. The release of helical strain at constant helix stability in RNA–DNA duplexes may facilitate base mispairing during reverse transcription, particularly in the context of lentiviral G→A hypermutation.**

*Keywords*: base pairing free energy differences/HIV-1 RT polymerase errors/mispair stability/retrotranscription/RNA–DNA duplexes

## Introduction

Human immunodeficiency virus (HIV) type 1 reverse transcriptase (RT) lacks a 3′ exonucleolytic activity. Consequently, RT-catalysed polymerization errors occur during both RNA- and DNA-dependent DNA synthesis. One of the more remarkable phenomena associated with reverse transcription is G→A hypermutation in which hundreds of template G residues may be monotonously substituted by A (Pathak and Temin, 1990; Vartanian *et al.*, 1991, 1994). The phenomenon is most apparent during reverse transcription of virion RNA with substitutions displaying a distinct preference for the GpA and GpG dinucleotide contexts. Only very rarely has G→A hypermutation been observed during DNA-directed DNA synthesis (Vartanian *et al.*, 1991). The phenomenon is a consequence of reverse transcription in the presence of highly biased intracellular [dCTP]/[dTTP] ratios, resulting in the incorporation of dTTP in lieu of dCTP opposite templated rG, and may be achieved *in vitro* using any RNA template, RT and biased deoxypyrimidine triphosphate concentrations (Martinez *et al.*, 1994).

The influence of mispairing within the template–primer upon subsequent RT error is usually of little importance given that retroviral mutation rates are in the order of $0.3–3.5\times10^{-5}$ per base per round of replication. However, given the sheer density of transitions encountered among G→A hypermutated genomes, occasionally every G within a stretch of 20–30 bp may be substituted, there is a possibility that *de novo*-formed $_t$G:T mispairs within the template–primer may exacerbate the phenomenon. RNA–DNA hybrids are structurally distinct from DNA–DNA or RNA–RNA duplexes. In solution, NMR studies suggest that in RNA–DNA helices, the RNA strand adopts an A conformation while the DNA strand behaves in a polymorphic manner, adopting either A or B conformations as a function of sequence and environment (Salasar *et al.*, 1993; Gonzàlez *et al.*, 1994, 1995). Within the RT complex, a double-stranded DNA template–primer is constrained within the polymerization site in an A conformation which, following a 45° kink at ~9 bp from this site, relaxes into a B–DNA structure (Jacobo-Molina *et al.*, 1993). Nevertheless, the precise conformation of an RNA–DNA hybrid within the RT cleft is unknown. Furthermore, little is known as to the thermodynamic properties of RNA–DNA duplexes in general (Hall and McLaughlin, 1991; Ratmeyer *et al.*, 1994; and references therein).

To ascertain the influence of a template–primer with preformed mismatches on $_t$G:T mispair formation, a series of DNA and RNA template–DNA primer oligonucleotide duplexes of homologous and mismatched sequences was examined by both kinetic and thermodynamic analyses (Schaeffer *et al.*, 1982; Aboul-ela *et al.*, 1985; Boosalis *et al.*, 1987; Ricchetti and Buc, 1990). It was possible to show that (i) HIV-1 RT $_t$G:T mispair formation was readily detectable using an RNA but not a DNA template, (ii) this formation was favoured by pre-existing $_t$G:T mismatches located within the 18 bp nucleic acid binding cleft of RT (Jacobo-Molina *et al.*, 1993) and (iii) the different polymerization behaviour of the enzyme on homologous RNA and DNA templates may be related to the very different melting properties of homologous DNA–DNA and RNA–DNA helices.


**Fig. 1.** RNA and DNA templates analysed. Only RNA template–DNA primer duplexes are given. DNA template–DNA primer duplexes are homologous. RNA (R) and DNA (D) templates were annealed with matched DNA primer (M) or with DNA primers with a T opposite template G at positions –4 (primer T1), –7 (primer T2) and –4 and –7 (primer T1T2). Template–primer duplexes are referred to as RM or DM, RT1 or DT1, RT2 or DT2, and RT1T2 or DT1T2.

## Results

### *Nucleotide insertion kinetics*

Both RNA and DNA templates were analysed with four matched and mismatched primers, the RNA–DNA hybrids being shown in Figure 1. The $_t$G:T mismatches within the template–primer were located at positions –4 and –7. The CCGCC motif at the 3′ end of the templates was incorporated to reduce fraying which might hinder enzyme–template–primer recognition. The kinetics of incorrect dTTP and correct dCTP insertion opposite template G at position +2 were determined as a function of nucleotide concentration (Boosalis *et al.*, 1987; Ricchetti and Buc, 1990).

Under the experimental conditions used, T misincorporation opposite G+2 could not be detected for any of the four DNA template–primer pairs (data not shown). Conversely, under the same conditions, relatively high rates of T misincorporation were observed with the four RNA template–primer hybrids (Table IA). Up to 4- to 10-fold increases in the relative misinsertion rate ($V_{max}/K_m$) were observed between matched and $_t$G:T mismatched RNA template–primers resulting mainly from large reductions in $K_m$. C insertion kinetics at G+2 were measured at the smallest concentration of dTTP with respect to $K_m$ for T misinsertion at +2, which ensured $V_{max}$ insertion kinetics of T opposite A+1 (Figure 1 and Table I). The C insertion rate remained roughly constant across the four RNA–DNA hybrids (Table IB) as variations in $K_m$ and $V_{max}$ virtually cancelled each other out. Taken together, the data show that $_t$G:T mismatches at positions –4 and –7 act in opposite ways on $K_m$ and $V_{max}$ for the insertion of the correct C and incorrect T nucleotides. The misincorporation frequency $f$ ($f = f_k^{-1}f_v$, where $f_k = K_m$[mismatch]/$K_m$[match] and $f_v = V_{max}$[mismatch]/$V_{max}$[match]) is enhanced some 4- to 13-fold when preformed mismatches are incorporated into the DNA primer (Table IC). This result is mainly a function of the relative $K_m$ values for the match and mismatch ($f_k$) which act in concert.

The ratio of $K_m$ values for the insertion of a nucleotide in the matched and mismatched $_t$RNA–DNA duplexes was used to estimate the differences in the standard free energy change upon dissociating the nucleotide from the respective enzyme–$_t$RNA–DNA duplex–substrate complexes:

$$\Delta\Delta G^\circ_{kinetics} = \Delta G^\circ_{37}(RX) - \Delta G^\circ_{37}(RM) = RT \ln[K_m(RM)/K_m(RX)], \quad (1)$$

where X = T1, T2, T1T2 (Figure 1). This assumes that the binding of a dNTP to the $_t$RNA–DNA in the enzyme–$_t$RNA–DNA complex reaches equilibrium and that $K_m$ is the dissociation constant for the (enzyme–$_t$RNA–DNA duplex)–dNTP complex (Petruska *et al.*, 1988). $\Delta\Delta G^\circ_{kinetics}$ values obtained in this way are given in Table II and show that $_t$G:T mismatches at –4 and particularly at –7 in the $_t$RNA–DNA duplex decrease the stability of correct nucleotide ($\Delta\Delta G^\circ_{kinetics} < 0$) while increasing the stability of the incorrect nucleotide ($\Delta\Delta G^\circ_{kinetics} > 0$) in the polymerization site of the elongating enzyme. Absolute values of $\Delta\Delta G^\circ_{kinetics}$ for both the $_t$G:C match and $_t$G:T mismatch increased in the order RT1 < RT2 < RT1T2 (Table II).

**Table I.** Insertion kinetics of correct and incorrect nucleotides opposite template G at +2 on matched and mismatched DNA-primed RNA templates[a]

| Primer | $K_m$ (μM) | $V_{max}$ (nM/min) | $V_{max}/K_m$ (rel) |
|---|---|---|---|
| **A $_t$G:T mismatch, RNA template** | | | |
| M | 525.9 ± 103.7 | 0.30 ± 0.02 | 1.0 |
| T1 | 292.5 ± 120.5 | 0.56 ± 0.07 | 3.6 |
| T2 | 81.0 ± 44.3 | 0.43 ± 0.04 | 9.5 |
| T1T2 | 74.2 ± 45.0 | 0.19 ± 0.03 | 4.5 |
| **B $_t$G:C match, RNA template** | | | |
| M | 0.049 ± 0.019 | 0.27 ± 0.02 | 1.0 |
| T1 | 0.054 ± 0.022 | 0.15 ± 0.01 | 0.5 |
| T2 | 0.166 ± 0.067 | 0.70 ± 0.08 | 0.8 |
| T1T2 | 0.248 ± 0.096 | 1.55 ± 0.14 | 1.1 |
| **C** | | | |
| Duplex | $f$ ($f_k^{-1}f_v$) | $f_k$ | $f_v$ |
| RM | $1.0\times10^{-4}$ | 10 733 | 1.1 |
| RT1 | $7.3\times10^{-4}$ | 5417 | 3.7 |
| RT2 | $1.3\times10^{-3}$ | 488 | 0.6 |
| RT1T2 | $4.0\times10^{-4}$ | 299 | 0.1 |

[a]Results are the average of three independent experiments ± standard deviation. The misinsertion frequency, $f$, is the ratio of $V_{max}/K_m$ values for T and C nucleotides, where $f_k = K_m(T)/K_m(C)$, $f_v = V_{max}(T)/V_{max}(C)$ and $f = f_k^{-1}f_v$.

**Table II.** Estimated stability differences of correct C and incorrect T nucleotides in the RT polymerization catalytic site on matched and mismatched duplexes[a]

| Duplexes | $\Delta\Delta G^\circ_{kinetics}$ at G+2 (kcal/mol) | |
|---|---|---|
| | rG:dC | rG:dT |
| RM–RT1 | –0.06 | 0.36 |
| RM–RT2 | –0.75 | 1.15 |
| RM–RT1T2 | –1.00 | 1.21 |

[a]Values of $\Delta\Delta G^\circ_{kinetics}$ were obtained using Equation 1. Data are ±0.49 kcal/mol, the maximum error calculated from the errors in $K_m$ (Table I).

**Table III.** Results of melting studies of DNA and RNA–DNA duplexes[a]

**A**

| Duplex | $\Delta S°$ (cal/mol/K) | $\Delta H°$ (kcal/mol) | $\Delta G°_{37}$ (kcal/mol) |
|---|---|---|---|
| DM | –327.2 | –119.8 | –18.3 |
| DT1 | –257.0 | –93.5 | –13.4 |
| DT2 | –288.6 | –104.4 | –14.9 |
| DT1T2 | –219.6 | –78.5 | –10.3 |
| RM | –243.8 | –91.7 | –16.1 |
| RT1 | –185.4 | –70.3 | –12.8 |
| RT2 | –214.1 | –80.0 | –13.6 |
| RT1T2 | –160.3 | –60.8 | –11.1 |

**B**

| Comparison | $\Delta\Delta S°$ (cal/mol/K) | $\Delta\Delta H°$ (kcal/mol) | $\Delta\Delta G°_{37}$ (kcal/mol) |
|---|---|---|---|
| DM–RM | –83.4 | –28.1 | –2.2 |
| DT1–RT1 | –71.6 | –23.2 | –0.6 |
| DT2–RT2 | –74.5 | –24.4 | –1.2 |
| DT1T2–RT1T2 | –59.3 | –17.7 | +0.8 |
| DM–DT1 | –70.2 | –26.3 | –4.9 |
| DM–DT2 | –38.6 | –15.4 | –3.4 |
| DM–DT1T2 | –107.6 | –41.3 | –8.0 |
| RM–RT1 | –58.4 | –21.4 | –3.3 |
| RM–RT2 | –29.7 | –11.7 | –2.5 |
| RM–RT1T2 | –83.5 | –30.9 | –5.0 |

[a]Errors in $\Delta S°$, $\Delta H°$ and $\Delta G°_{37}$ were ±2%, while those in $\Delta\Delta S°$, $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ were ±4%.

### *Thermodynamic stability of DNA–DNA and RNA–DNA duplexes*

Thermal melting measurements were made with the DNA–DNA and RNA–DNA duplexes shown in Figure 1 using the elongation reaction buffer. Standard thermodynamic parameters defining duplex stability were evaluated by the van't Hoff method and are given in Table IIIA. Differences in $\Delta G°_{37}$ values for strand dissociation, i.e. $\Delta\Delta G°_{37}$, showed that matched DNA–DNA duplex was more stable than matched RNA–DNA duplex by –2.2 kcal/mol (Table IIIB). A single $_t$G:T mismatch decreased the difference in stability, while the incorporation of two $_t$G:T mismatches resulted in the RNA–DNA duplex being slightly more stable ($\Delta\Delta G°_{37}$ = +0.8 kcal/mol). Not surprisingly, $_t$G:T mismatches decreased the stability of both DNA–DNA and RNA–DNA duplexes with respect to the matched primer, although the extent of destabilization conferred by a $_t$G:T mismatch was smaller by ~1.4 kcal/mol for the three RNA–DNA duplexes (Table IIIB).

The effects of sequence context on G:T mismatch stability were smaller for RNA–DNA than for DNA–DNA duplexes. T1 and T2 primer-encoded $_t$G:T mismatches are in UGA (TGA) and UGC (TGC) contexts, respectively (Figure 1). The difference in $\Delta G°_{37}$ values between T1 and T2 duplexes were 1.5 kcal/mol for DNA duplexes, yet only 0.8 kcal/mol for RNA–DNA duplexes (Table IIIA). The fact that the sum of $\Delta\Delta G°_{37}$ values for the two single G:T mismatches did not equal $\Delta\Delta G°_{37}$ for T1T2 suggests that the effects of the two mismatches on helix stability were not entirely independent (Table IIIB).

### *Enthalpy and entropy compensation*

As can be seen from Figure 2A, differences in enthalpy ($\Delta\Delta H°$) and entropy ($\Delta\Delta S°$) contributions between DNA–

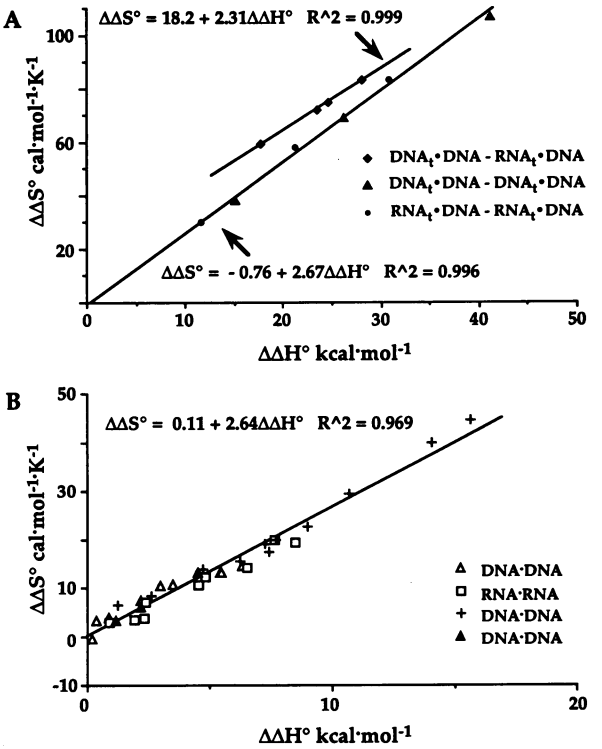


**Fig. 2.** Entropy–enthalpy compensation upon melting of DNA–DNA, RNA–RNA and RNA–DNA duplexes. (**A**) Raw data for the series of DNA–DNA and RNA–DNA comparisons given in Table IIIB, and independently the comparison of the same primer with either a DNA or an RNA template (Table IIIB). (**B**) Entropy–enthalpy compensation taken or derived from the literature and normalized as $\Delta\Delta S°$ and $\Delta\Delta H°$ per bp doublet. (+) Internal DNA pairs and mispairs flanked by A:T (Aboul-ela *et al.*, 1985); (▲) terminal DNA mispairs flanked by G:C (Petruska *et al.*, 1988); (△) DNA base pair doublets (Breslauer *et al.*, 1986); (□) RNA base pair doublets (Freier *et al.*, 1986). In (A) and (B), $\Delta\Delta S°$ and $\Delta\Delta H°$ are plotted with opposite signs. For internal DNA pairs and mispairs, $\Delta\Delta S° = [\Delta S°(dCA_7G\text{–}dCT_7G) - \Delta S°(dCA_3XA_3G\text{–}dCT_3YT_3G)]/2$, where X and Y represent one of the four bases, and $\Delta\Delta H°$ is calculated in the same manner. For base pair doublets in DNA, $\Delta\Delta S° = \Delta S°(XY/X'Y') - \Delta S°(GA/CT)$, and in RNA, $\Delta\Delta S° = \Delta S°(XY/X'Y') - \Delta S°(AU/AU)$, where $XY/X'Y'$ represent one of the nine base pair doublets; $\Delta\Delta H°$ is calculated in the same manner.

DNA and RNA–DNA duplexes of homologous sequences (Table III) are correlated directly, indicating a relationship of the form:

$$\Delta\Delta G° = \Delta\Delta H° - T\Delta\Delta S°. \quad (2)$$

The large $\Delta\Delta H°$ values are nearly cancelled out by large $T\Delta\Delta S°$ values, explaining the fact that $\Delta\Delta H°$ are 13- to 37-fold larger than the corresponding free energy ($\Delta\Delta G°_{37}$, Table IIIB). The strong entropy–enthalpy relationship described above concerns both matched and mismatched DNA–DNA and RNA–DNA duplexes. This relationship is all the more striking because Tinoco and colleagues have shown that $\Delta\Delta S°$ is proportional to $\Delta\Delta H°$ for normal base pairs and mispairs in a variety of DNA–DNA duplexes (Petruska *et al.*, 1988). Sequence-induced variations of $\Delta\Delta S°$ and $\Delta\Delta H°$ for RNA–RNA duplexes were calculated from published nearest-neighbour values (Freier *et al.*, 1986). Figure 2B shows that the same proportionality ($\Delta\Delta S°/\Delta\Delta H°$) was noted for both DNA–DNA and RNA–RNA duplexes. The fact that the slopes in Figure 2A and B were virtually the same suggests that the enthalpy–entropy compensation already observed for DNA–DNA

duplexes may be extended to all forms of nucleic acid helix. The slope of $\Delta\Delta S°/\Delta\Delta H°$ for the DNA–DNA and RNA–DNA duplexes of homologous sequences (the upper line in Figure 2A) reflects the difference in stability of a $_t$G:T mismatch in the two types of helix.

## Discussion

### *Prior retrotranscription errors affect fidelity*

The kinetic analysis showed that prior retrotranscription errors found within the primer may have a profound effect upon the fidelity of subsequent base incorporation. Normally such a situation is of academic interest given an average HIV-1 mutation rate of ~$3.5\times10^{-5}$ per base per replication cycle (Mansky and Temin, 1994). However, the phenomenal G→A transition frequencies (up to 60% of all Gs may be substituted), as well as the use of the *in vitro* reaction (Martinez *et al.*, 1994) to hypermutagenize any RNA sequence at will, with its application to protein evolution *in vitro*, prompted this investigation.

A striking difference of RT fidelity on the DNA and RNA templates was evident. T misincorporation appears almost as an 'all or nothing' phenomenon, absent on DNA templates ($f$ = 0) yet occurring at high frequency ($f$ = $10^{-4}$–$1.3\times10^{-3}$) on RNA templates. The sequence context and/or the buffer chosen to mimic physiological conditions may contribute to this difference. The comparison of HIV-1 fidelity on RNA and DNA templates has yielded mixed results. Depending upon the experimental system used, RNA-dependent DNA polymerization was more (Hüber *et al.*, 1992), equally (Ji and Loeb, 1992) or less (Boyer *et al.*, 1992) error prone than DNA-dependent DNA synthesis. The above results are more in keeping with those of Hüber *et al.* (1992), although once again the considerable differences in technique and sequence preclude hasty conclusions being drawn.

Kinetic discrimination in favour of T misinsertion on $_t$G:T mismatched RNA templates was based much more on $K_m$ ($f_k$ values, Table IC) than on $V_{max}$ values ($f_v$ values, Table IC). From $K_m$ values in Table I and Equation 1, $\Delta G°_{37}$ for dissociation of the $_t$G:T mispair is 0.36–1.21 kcal/mol higher, and $\Delta G°_{37}$ for dissociation of the $_t$G:C pair is 0.06–1.00 kcal/mol lower, in elongating complexes formed with $_t$G:T mismatched $_t$RNA–DNA duplexes than in the elongating complex formed with matched $_t$RNA–DNA duplex ($\Delta\Delta G°_{kinetics}$, Table II). The opposite, but qualitatively equal, variations in stability of the $_t$G:C pair and $_t$G:T mispair observed with each mismatched $_t$RNA–DNA duplex (Table II) strongly suggested that the variation in the stability of correct and incorrect base pairs in the RT polymerization site resulted from the same phenomenon, i.e. the transmission to the enzyme substrate binding site of base pair instability induced by distal $_t$G:T mismatches in the $_t$RNA–DNA duplex.

### *Melting properties of homologous DNA and RNA–DNA*

Template–primer duplex stability was determined using the same buffer as that employed in kinetic analyses. Table III reveals that DNA and RNA–DNA duplexes in solution display drastically different melting behaviours. Expressing the data in Table III in terms of mean values per base pair showed that base pairs in homologous DNA and RNA–DNA duplexes are of approximately equal stability as the $\Delta\Delta G°_{37}$ values vary from +0.05 to –0.15 kcal/mol. However, the energy spent and the number of degrees of freedom gained upon base pair dissociation were considerably larger in the DNA duplex than in the RNA–DNA duplex ($\Delta\Delta H°$ = 1.2–1.9 kcal/mol and $\Delta\Delta S°$ = 4.0–5.6 kcal/mol/K). In other words, the RNA–DNA duplex in solution presents facile base pair opening and a high conformational flexibility with respect to the more rigid DNA duplex (it is assumed here that solvent contributions to $\Delta\Delta S°$ are negligible). These differences may explain the greater stability of the G:T mispair [$\Delta\Delta G°_{37}$(G:T) ~1.4 kcal/mol], and the lower influence of nearest neighbours on G:T mispair stability observed in the RNA–DNA duplex. The release of helical strain at constant helix stability may favour the insertion of non-Watson–Crick base pairs in the RNA–DNA duplex. The different solution behaviour of DNA and RNA–DNA duplexes resulted from the relationship between entropy and enthalpy contributions to $\Delta\Delta G°$ (Table III): $\Delta\Delta H°$ was nearly cancelled out by $-T\Delta\Delta S°$ to give a small $\Delta\Delta G°$ (Equation 2). First indicated for DNA molecules (Petruska *et al.*, 1988), this study shows that entropy–enthalpy compensation in aqueous solution holds for RNA–DNA hybrids (Figure 2A) as well as for RNA–RNA helices (Figure 2B).

The different melting behaviour of DNA and RNA–DNA duplexes displayed in Table III is in excellent agreement with recent solution NMR studies of RNA–DNA duplexes (Salasar *et al.*, 1993; Gonzàlez *et al.*, 1994, 1995). RNA–DNA duplexes were found to adopt a strongly heteronomous duplex conformation in which the RNA strand remained in the A conformation while the DNA strand behaved in a polymorphic manner, adopting either A or B conformations as a function of sequence and environment (Salasar *et al.*, 1993; Gonzàlez *et al.*, 1994, 1995). In the same conditions, DNA duplexes remain in the B format. These structural variations suggest a high conformational flexibility of the RNA–DNA hybrid compared with double-stranded DNA (Gonzàlez *et al.*, 1994, 1995). Indeed, this is observed in Table III.

### *Kinetic and melting data comparison*

Kinetic and melting data could be compared as follows. Stability differences between matched and mismatched RNA–DNA duplexes ($\Delta\Delta G°_{37}$, Table IIIB) were, when expressed as a mean per base pair, ~3-fold smaller than the estimated variations in stability of the $_t$G:C pair in the RT polymerization site, i.e. between RM and RT1T2, $\Delta\Delta G°_{kinetics}$ = –1.0 kcal/mol for the rG:dC pair (Table II) and $\Delta\Delta G°_{37}$ = –0.36 kcal/mol (Table IIIB). Such a difference may result from the loss of a stacking interaction at the 3′ end of the primer. However, if the enzyme binding cleft is hydrophobic (Jacobo-Molina *et al.*, 1993), base stacking interactions will increase in the bound template–primer duplex, giving rise to increased $\Delta\Delta H°$ values by T1 and T2 mismatches. Bound to the enzyme, the template–primer duplex is constrained, leading to lower values of $\Delta\Delta S°$. An increase in $\Delta\Delta H°$ or a decrease in $\Delta\Delta S°$ by a factor of 1.5 would be sufficient to bring the solution values of $\Delta\Delta G°_{37}$ up to the estimated values of $\Delta\Delta G°_{kinetics}$ in the RT polymerization site (Table II). A factor <1.5 can be considered if both $\Delta\Delta H°$ and $\Delta\Delta S°$

vary within the confines of the polymerization site. The present melting experiments do not provide sufficient free energy data to compare with $_t$G:T mispair stabilization in the RT polymerization site (Table II).

### Biological implications

Recently, it has been shown that a single amino acid substitution in the p66 palm (Glu89Gly) and finger (Leu74-Val) subdomains altered remarkably the discrimination of the modified nucleotides ddG and ddI respectively (Boyer *et al.*, 1994). In particular, the Glu89 mutation contacts the sugar–phosphate backbone of the template strand in the double-stranded part of the template–primer 3–4 bp from the active site. Together with the results presented here, it becomes apparent that the RT polymerization activity may be influenced by long-range protein–template–primer contacts and the melting properties of nucleic acid helices in direct contact with the enzyme, at least over the 9 bp between the polymerization site and the T2 mismatch.

It is possible that these results will shed some light on the phenomenon of G→A hypermutation. G:T mismatches in the template–primer duplexes 6–9 bp distal to the polymerization site exacerbated T misinsertion 4- to 13-fold (Table IC). The $K_m$ value discrimination could be enhanced *in vivo* by locally imbalanced intracytoplasmic dNTP concentrations. The current observations pertain to steady state kinetics in which [template–primer] > [RT]. However, in the replication complex there is a 30- to 50-fold molar excess of RT. The highly biased [dCTP]/[dTTP] pools conducive to G→A hypermutation slow down reverse transcription considerably (Martinez *et al.*, 1994). It is probable that such conditions, as well as the molar excess of enzyme, would enable considerable cycling of the RT and misincorporation of dT opposite rG, which itself would enhance the probability of further dT misincorporation. This ensemble might help to explain the remarkably high density of G→A transitions typical of hypermutated HIV-1 genomes.

## Materials and methods

### HIV-1 RT, DNA and RNA oligonucleotides

The heterodimeric p66/p51 HIV-1 RT was purified from an overproducing strain of *Escherichia coli* (Müller *et al.*, 1989). DNA and RNA templates and DNA primers were synthesized according to standard methods and purified by high performance liquid chromatography. Single-strand concentrations were determined using DNA and RNA single-strand extinction coefficients at 260 nm, calculated from published monomer and dimer coefficients (Richards, 1975).

### Primer labelling and primer–template annealing

Primer 5′ termini were labelled with [$\gamma$-$^{32}$P]dATP (10 Ci/mol, 10 mCi/ml, Amersham) and T4 polynucleotide kinase (Pharmacia). $\gamma$-$^{32}$P-labelled primers were annealed in 50 mM NaCl and 50 mM Mg(aspartate)$_2$ for 2 min at 90°C. Samples were cooled slowly to 20°C. The amounts of active DNA–DNA and RNA–DNA duplex (>95% hybridization of the radiolabelled primer) were determined by an *in vitro* elongation assay, adding both HIV-1 RT and dNTPs in excess (Ricchetti and Buc, 1990) to the reaction buffer described below. Consequently, the template–primer concentration ratio was adjusted for DNA duplexes to 1:1, being equivalent to a 2.5 μM final concentration in the hybridization solution, with 2.2 μM of cold primer and 0.3 μM of $\gamma$-$^{32}$P-labelled primer. For RNA–DNA duplexes, the template–primer concentration ratio was adjusted to 1:10, the primer being 1 μM (0.7 μM of cold and 0.3 μM of $\gamma$-$^{32}$P-labelled primer). RNA was handled under standard RNase-free conditions. Strand breakage was not detectable during hybridization and polymerization reactions. Prior to the elongation reaction assay, duplexes and hybrids were diluted 10-fold in the following buffer: 50 mM NaCl, 50 mM Mg(aspartate)$_2$, 500 mM HEPES, pH 7.0 (final concentrations).

### HIV-1 RT steady-state kinetics and gel electrophoresis

Steady-state kinetics were performed (Ricchetti and Buc, 1990) using the following modifications. To approach physiological conditions, a HEPES (pH 7.0) buffer replaced Tris–HCl to avoid contaminating traces of calcium which inhibit HIV-1 RT polymerase activity (data not shown). The concentration of $Mg^{2+}$ was increased, while $[K^+] > [Na^+]$, reflecting the intracellular environment. The weaker anions ($CH_3COO^-$ and $Asp^-$) virtually replaced $Cl^-$. Polyethylene glycol (PEG) 6000 was added to reduce the reaction volume and favour enzyme–template–primer complex formation. Thus the final buffer (10 μl reaction volume) was 50 mM HEPES, pH 7.0, 15 mM NaCl, 15 mM Mg(aspartate)$_2$, 130 mM $KCH_3COO$, 1 mM dithiothreitol and 5% PEG 6000. The molar ratio of template–primer and enzyme in the reaction mixture was estimated to be 20:1, sufficient to saturate the enzyme. The reaction was initiated by first equilibrating HIV-1 RT with annealed template–primer in the absence of dNTPs (10 min at 37°C), followed by the addition of appropriate dNTPs at various concentrations. The reaction was carried out for 40 s at 37°C, which ensured steady-state kinetics. The reaction was stopped by the addition of 10 mM EDTA in 90% formamide and the sample was electrophoresed (Ricchetti and Buc, 1990). $\gamma$-$^{32}$P radioactive band intensities were measured using an Applied Biosystems PhosphorImager. Band intensities were fitted to a Michaelis–Menten equation using the UltraFit Macintosh program (Version 1.03, Biosoft). Kinetic parameters were determined as described previously (Ricchetti and Buc, 1990).

### Thermodynamic analysis

Absorption versus temperature profiles were recorded for the eight template–primer duplexes (Schaeffer *et al.*, 1982). Equivalent amounts of template and primer strands were annealed in the same buffer used in the kinetic study. The absorption profiles were recorded at 260 nm with a heating rate of 0.2°C/min from 2 to 98°C. These conditions were found to ensure complete reversibility of the strand separation reaction (data not shown), and the precise determination of the initial and final slopes of the absorption profiles necessary for normalization. Melting temperatures ($T_m$) were determined as described previously (Aboul-ela *et al.*, 1985). For each duplex, $T_m$ measurements were made at total strand concentrations ($C_t$) in the range 0.7–12.4 μM. $\Delta H°$ and $\Delta S°$ were determined by a linear least-squares fit to a van't Hoff plot (Marky and Bresslauer, 1987):

$$T_m^{-1} = (R/\Delta H°)\ln(C_t) + (\Delta S° - R\ln 4)/\Delta H°, \quad (3)$$

assuming an all or none transition, and that $\Delta H°$ and $\Delta S°$ are independent of temperature (R = the gas constant). The errors in $T_m$ were ±0.2°C. Plots were linear for the DNA and RNA–DNA duplexes studied, with least-squares fit coefficients $R^2 \geq 0.99$ (data not shown). The $\Delta G°_{37}$ values, the standard free energy at 37°C, were derived using $\Delta G°_{37} = \Delta H° - 310.15\Delta S°$.

## Acknowledgements

We would like to thank Dr Goody for the gift of purified HIV-1 p66/51 RT and Catherine Gouyotte for the high performance liquid chromatography purification of the oligonucleotides. This work was carried out with grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA (ANRS). M.S. was supported by fellowships from the Ministero della Sanità (Italy) and the Fondation pour la Recherche Médicale (Sidaction).

## References

Aboul-ela,F., Koh,D., Tinoco,I.,Jr and Martin,F.H. (1985) Base–base mismatches. Thermodynamics of double helix formation for dCA3XA3G+dCT3YT3G (X, Y = A,C,G,T). *Nucleic Acids Res.*, **13**, 4811–4825.

Boosalis,M.S., Petruska,J. and Goodman,M.F. (1987) DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *J. Biol. Chem.*, **262**, 14689–14696.

Boyer,J.C., Bebenek,K. and Kunkel,T. (1992) Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl Acad. Sci. USA*, **89**, 6919–6923.

Boyer,P.L., Tantillo,C., Jacobo-Molina,A., Nanni,R.G., Ding,J., Arnold,E.

and Hughes,S.H. (1994) Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template length; the sensitivity of drug-resistant mutants does not. *Proc. Natl Acad. Sci. USA*, **91**, 4882–4886.

Breslauer,J.K., Frank,R., Blöcker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Nielson,T. and Turner,D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.

Gonzàlez,C., Stec,W., Kobylanska,A., Hogrefe,R., Reynolds,M. and James,T.L. (1994) Structural study of a DNA–RNA hybrid duplex with a chiral phosphorothioate moiety by NMR: extraction of distance and torsion angle constraints and imino proton exchange rates. *Biochemistry*, **33**, 11062–11072.

Gonzàlez,C., Stec,W., Reynolds,M. and James,T.L. (1995) Structure and dynamics of a DNA–RNA hybrid duplex with a chiral phosphorothioate moiety: NMR and molecular dynamics with conventional and time-averaged restraints. *Biochemistry*, **34**, 4969–4982.

Hall,K.B. and McLaughlin,L.W. (1991) Thermodynamic and structural properties of pentamer DNA–DNA, RNA–RNA, and DNA–RNA duplexes of identical sequence. *Biochemistry*, **30**, 10606–10613.

Hüber,A., Kruhoffer,M., Grosse,F. and Krauss,G. (1992) Fidelity of human immunodeficiency virus type 1 reverse transcriptase in copying natural RNA. *J. Mol. Biol.*, **223**, 595–600.

Jacobo-Molina,A. *et al.* (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.

Ji,J. and Loeb,L. (1992) Fidelity of HIV-1 reverse transcriptase copying RNA *in vitro*. *Biochemistry*, **31**, 954–958.

Mansky,L.M. and Temin,H.M. (1994) Lower *in vitro* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.*, **69**, 5087–5094.

Marky,L.A. and Bresslauer,K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.

Martinez,M.A., Vartanian,J.P. and Wain-Hobson,S. (1994) Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl Acad. Sci. USA*, **91**, 11787–11791.

Müller,B., Restle,T., Weiss,S., Gautel,M., Sczakiel,G. and Goody,R.S. (1989) Co-expression of the subunits of the heterodimer of HIV-1 reverse transcriptase in *Escherichia coli*. *J. Biol. Chem.*, **264**, 13975–13978.

Pathak,V.K. and Temin,H.M. (1990) Broad spectrum of in-vitro forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl Acad. Sci. USA*, **87**, 6019–6023.

Petruska,J., Goodman,M.F., Boosalis,M.S., Sowers,L.C., Cheong,C. and Tinoco,I.,Jr (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc. Natl Acad. Sci. USA*, **85**, 6252–6256.

Ratmeyer,L.R., Vinayak,R., Zhong,Y.Y., Zon,G. and Wilson,W.D. (1994) Sequence specific thermodynamic and structural properties for DNA–RNA duplexes. *Biochemistry*, **33**, 5298–5304.

Ricchetti,M. and Buc,H. (1990) Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent. *EMBO J.*, **9**, 1583–1593.

Richards,E.G. (1975) Nucleic acids. In Fasman,G.D. (ed.), *Handbook of Biochemistry and Molecular Biology*. 4th edn. CRC Press, Boca Raton, FL, p. 583.

Salasar,M., Fedoroff,O.Y., Miller,J.M., Ribeiro,N.S. and Reid,B.R. (1993) The DNA strand in DNA–RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry*, **32**, 4207–4215.

Schaeffer,F., Kolb,A. and Buc,H. (1982) Point mutations change the thermal denaturation profile of a short DNA fragment containing the lactose control elements. Comparison between experiment and theory. *EMBO J.*, **1**, 99–105.

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

Vartanian,J.-P., Meyerhans,A., Sala,M. and Wain-Hobson,S. (1994) G→A hypermutation of the HIV-1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl Acad. Sci. USA*, **91**, 3092–3096.

*Received on May 24, 1995; revised on July 3, 1995*