The EMBO Journal vol.9 no.5 pp.1583 - 1593, 1990

Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent

Miria Ricchetti and Henri Buc

Institut Pasteur, Unité de Physicochimie des Macromolécules Biologiques, (URA1149 du CNRS), 75724 Paris Cedex 15, France

Communicated by H.Buc

Kinetics of incorporation of correct and incorrect deoxynucleotides by three reverse transcriptases have been followed, by gel assay, on a series of DNA templates, including part of the HIV-1 *gag* DNA minus strand. Insertion kinetics for the properly matched nucleotide at a given place on the template vary strongly from one enzyme to the next. No significant correlation is found between the site-specific Michaelis constants, while the maximal velocities are more closely connected. For a given reverse transcriptase these parameters are strongly influenced by the DNA sequence. A systematic evaluation of the frequencies of misincorporation was then performed at 46 positions. Again great variability was found, precluding a very accurate evaluation of an average misincorporation frequency for a given enzyme and a given mismatch. Qualitatively however, HIV-1 reverse transcriptase is certainly not more error-prone in this assay than the other enzymes assayed. The patterns of misincorporations were again very dependent on the enzyme used to replicate a given template. The variability of the *gag* sequence observed *in vivo* among various HIV-1 isolates was compared with the patterns of misincorporations obtained *in vitro* on the same sequence with HIV-1, AMV and MoMLV reverse transcriptases. A fair agreement was found with the pattern observed in the polymerization directed by the HIV-1 reverse transcriptase. The correlation is less important in the two other cases. However some specific changes observed *in vivo* cannot be accounted for by our misincorporation assay, even when performed with the homologous enzyme, suggesting that an important class of mismatches can only be generated during reverse transcription of the RNA strand. Additional data, using a complementary DNA (positive) strand as a *gag* template support this hypothesis.

Key words: genomic variation/HIV-1/mismatch frequency/reverse transcriptase

Introduction

RNA viruses evolve at such a fast pace that one cannot even speak of an RNA genome for a given species undergoing continuous cycles of replication (for a review see for example Steinhauer and Holland, 1987). This variability has been estimated to be around $10^{-2}-10^{-3}$ substitutions per site per year in the case of the v-*mos* gene of Moloney murine sarcoma virus (Gojobori and Yokoyama, 1985) and 2×10^{-3} per site per year in the case of HIV-1 (Coffin, 1986). More recently, refinements in the technology have allowed estimates of the mutation rate per *replication cycle* to be made. Using denaturing gradient gel electrophoresis, Leider *et al.* (1988) arrived at frequencies as high as 1.4×10^{-4} mutation per nucleotide per replication cycle for Rous sarcoma virus. A two-step assay led Temin and his coworkers to propose a frequency of 2×10^{-5} substitutions per base pair per replication cycle for spleen necrosis virus (Dougherty and Temin, 1988).

Huge differences in evolution rates for RNA and DNA genomes are attributed to the different accuracy of their replication machinery. The enzyme responsible for genome replication in retroviruses is reverse transcriptase (RT), a trifunctional enzyme which synthesizes DNA on RNA as well as on DNA templates and which degrades the RNA through its RNase H activity. Reverse transcriptases are devoid of proof-reading mechanisms, a well accepted explanation for their high inaccuracy, exemplified *in vitro* by their great frequencies of misincorporation of deoxynucleotides on DNA templates. In the case of AMV the *in vitro* error frequency of its RT (Battula and Loeb, 1974) was found to be consistent with the *in vivo* mutation rate: 3×10^{-4} per site per virus passage (Coffin *et al.*, 1980). The same comparison has not yet been performed with the HIV-1 genome. Very recently an analysis of *in vivo* variations of a part of the *gag* and *env* genes yielded an HIV-1 mutation rate of one base per genome per replication cycle (Goodenow *et al.*, 1989). Scarce data are available *in vitro*, and only for the DNA−DNA replication step. Purified HIV-1 RT was found to introduce misincorporations at a rate of 1/2000 − 1/4000 as judged from the output of three different assays (Preston *et al.*, 1988). Using a ΦX174 reversion test, Weber and Grosse (1989) measured a frequency of 1/7400 per bp. These estimates lead to frequencies in the range of 2.5 − 10 mutations per genome per replication cycle.

Besides the classical problems of viability encountered when frequencies of misincorporation measured *in vitro* are compared with variabilities of progeny, several unsolved questions arise in the case of reverse transcriptases. Firstly, whether errors at the DNA polymerase level or those observed during reverse transcription of the RNA template are more representative of the pattern of changes observed in the progeny of a parental clone. Secondly, whether the primary determinant for the *in vitro* variability is the nature of the templates (DNA or RNA), or the nature of the reverse transcriptases, or both. Finally, if variability depends strongly on the nature of both the enzyme and the template, what is the size of the pool of experiments required to estimate average error frequencies. This is specially important in order to assess whether or not HIV-1 RT is less faithful than others (cf. Preston *et al.*, 1988; Roberts *et al.*, 1988; Weber and Grosse, 1989).

In the present paper, comparative data on the catalytic properties of three reverse transcriptases, acting as DNA

© Oxford University Press

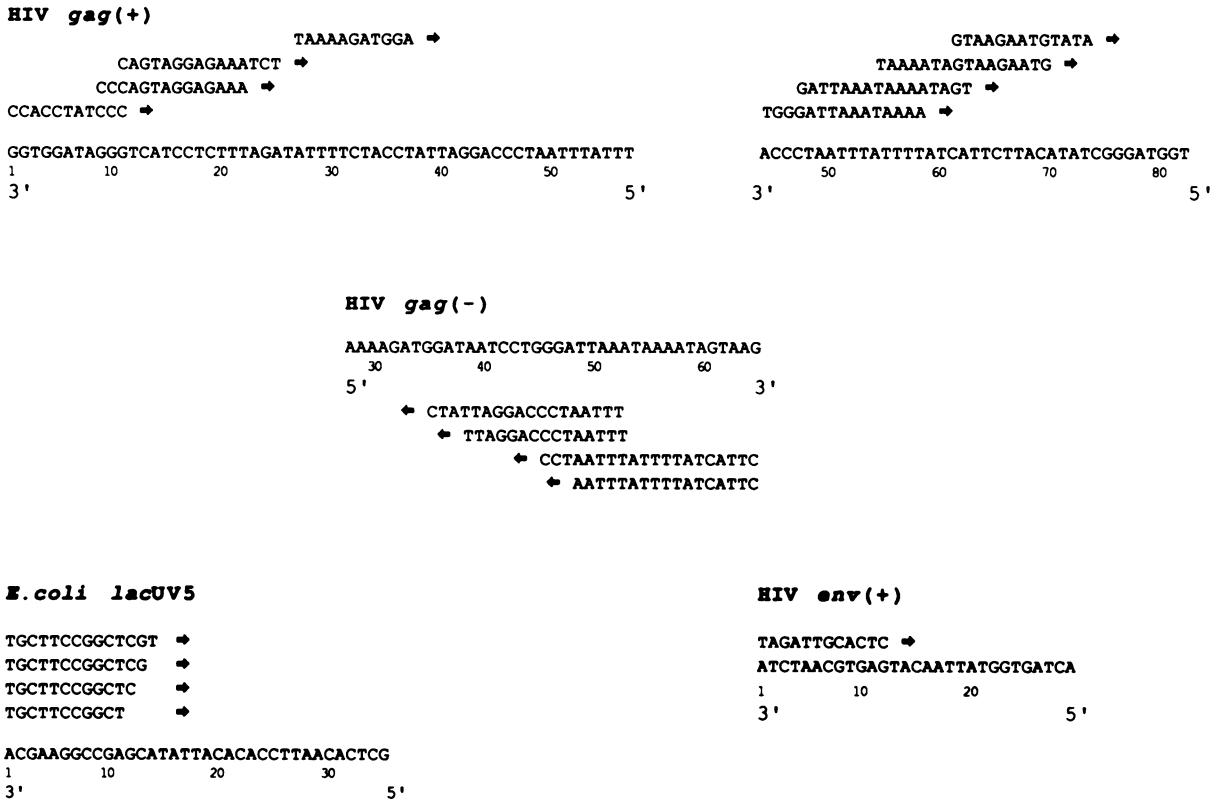

Fig. 1. DNA templates and primers used. Arrows indicate the direction of DNA synthesis. The symbols (+) or (−) indicate which strand of the HIV-1 genome is synthesized.

polymerases, are presented. Reverse transcriptases from avian myeloblastosis virus (AMV), from Moloney murine leukaemia virus (MoMLV), and from human immunodeficiency virus 1 (HIV-1) were compared, first for the incorporation of correct nucleotides during primer extension performed on several DNA templates; second, patterns and frequencies of mismatches were compared again on the same sequences. A DNA segment coding for p25 *gag*, previously used by Goodenow *et al.* (1989) to assess the internal variability among highly polymorphic HIV-1 isolates, was included in the assays. Our results could therefore be compared with the pattern of substitutions and the relative frequencies of various transitions and transversions estimated from their data.

Results

The comparative study of three reverse transcriptases was performed on the DNA templates given in Figure 1. Location of the various primers is also indicated. As explained in the Materials and methods, template was in excess over primer and the enzyme concentration was limiting. The elongation of a 5′ labelled primer by the enzyme was followed after addition of the four possible deoxynucleotides, for a fixed amount of time. The products of the reaction were separated by electrophoresis. From the patterns obtained (cf. Figure 2), conditions fulfilling initial velocity measurements were defined and the ratio of product(s) to substrate determined, basically as indicated in Boosalis *et al.*, 1987 (cf. also the Materials and methods). Under the conditions chosen, the reaction was linear with respect to time, and proportional to the enzyme concentration.

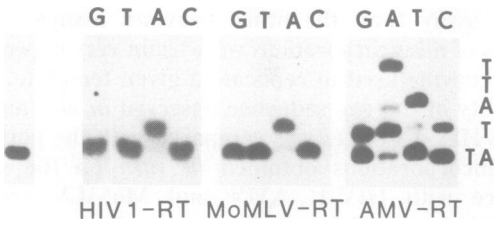

Fig. 2. Gel assay for deoxynucleotide incorporation opposite to the T residue at position 15 of the *lac* sequence by three reverse transcriptases. The sequence given at the right represents the template, hybridized to a primer initially terminated by a T. The nucleotide (250 μM) used in each assay is given at the top of each lane. In the case of both HIV-1 RT and MoMLV RT the primer is elongated only in the presence of dATP. When the reaction is performed with AMV RT, the incorporation of dATP is followed by one additional elongation step corresponding to an A_t:A misincorporation and by two correct pairings. When dTTP is added, a first misincorporation event is produced (T_t:T) followed by one correct pairing. dCTP and dGTP are also misincorporated opposite to the T residue.

Incorporation of a correct deoxynucleotide

Figure 3 illustrates how the initial velocity of elongation of a primer of 14 nucleotides hybridized to the HIV-1(+1) *gag* template varies in response to the concentration of dCTP (the second nucleotide to be incorporated) when the three reverse transcriptases are used in separate assays. The apparent Michaelis constants derived from the hyperbolic curves vary drastically while the maximal velocities are of the same order of magnitude. In general, those differences do persist when a different primer is used to shift the position of incorporation with respect to the 3′ end of the primer. Hence, the differences are enzyme specific. Other examples

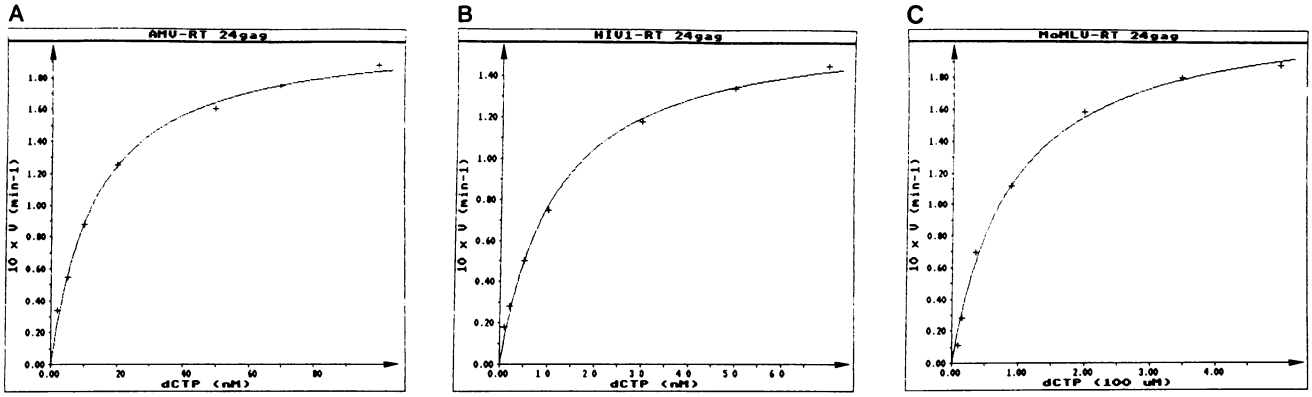

Fig. 3. Rate of dCMP incorporation as a function of substrate concentration. G residue at position 24 of the *gag*(−) template. Enzymes: (**a**) AMV RT; (**b**) HIV-1; (**c**) MoMLV RT.

are given in Tables I to III. One can observe firstly that MoMLV RT incorporates nucleotides with K_m values in the micromolar range while for AMV RT and for HIV-1 RT the values are usually in the nanomolar range with rare cases in the micromolar range. Secondly for a given enzyme, the kinetic constants are very much dependent on the position of the nucleotide, with a much broader dispersion for the apparent Michaelis constants than for the maximal velocities. The distribution for two subsets of data, where the pool size was large enough (dATP and dTTP) was compared to the overall distribution in the case of AMV RT and of HIV-1 RT; it appeared that the nature of the nucleotide incorporated was not the major factor in generating the broad distribution observed for the Michaelis constants. This finding agrees with the recent study of Mendelman *et al.* (1989) and Bebenek (1989); it implies that the vicinal DNA sequences are of utmost importance in generating this diversity. Before attempting to delineate the structural determinants for these effects, we first looked for a possible correlation between Michaelis constants on the one hand and between maximal velocities on the other for the three enzymes considered. We took these sets of data two by two and peformed a correlation analysis.

No significant correlation whatsoever was found when considering the pairs of observed Michaelis constants (Table IV, line 5), the correlation coefficients being close to zero for AMV RT and MoMLV RT or HIV RT and MoMLV RT, and equal to 0.38 for the last pair. By contrast, a correlation is indeed observed for the corresponding maximal velocities (Table IV, line 10; 29 common positions in the sequences are taken into account). The data appear to be proportional to each other, as deduced from a linear regression analysis (data not shown). A simple experiment illustrates this finding. Pausing patterns were observed by the gel electrophoresis technique with the three enzymes at increasing concentrations of dNTPs (cf. Figure 4). They are uniquely determined by the incubation time (which is constant here) and by the rate constants for the incorporation of consecutive nucleotides. As shown in lanes h and i, at high substrate concentrations the patterns look the same while they are completely different at low substrate concentration (lanes a−f). Hence, when translocations take place at their maximal velocities, a correlation exists between the three systems. This correlation vanishes as soon as the apparent Michaelis constants for substrate incorporation come into play. It is concluded that the three enzymes do 'see' the DNA sequences differently. The kinetic parameters for incorporation of the correct nucleotide are always strongly influenced by the sequence of the template but the sequence affects the Michaelis constants observed with each enzyme differently. (The structural origin of this difference is further analysed in the Discussion).

Mismatches created by the three reverse transcriptases

The same sequences were used to determine the frequency of misincorporation during primer elongation (see Figure 2). In total 46 positions were tested. For positions that are not located at +1, not all the nucleotides could be assayed for their ability to create mismatches (for the second base of a doublet, misincorporation can only be assayed after the creation of a mismatch at the preceding position). As a result, only slightly more than 100 mismatches could be checked (cf. Figure 5). A subset of data is given in Tables I−III, where both the apparent Michaelis constants and the maximal velocities were accurately measured (referred to as K'_m and V'_{max} respectively for the misincorporated nucleotide). In most cases, however, only the ratio of the initial velocity to the substrate concentration was obtained. For a substrate concentration low enough with respect to the apparent Michaelis constant of the misincorporated nucleotide, this quantity is a good estimate of V'_{max}/K'_m. The frequency of misincorporation, f, is calculated assuming no accumulation of intermediates in a branched competitive pathway (Fersht, 1977). When no misincorporation was actually observed, we could however estimate at each position the maximal frequency of misincorporation f*, as explained in the Materials and methods. Because the values of V_{max}/K_m are of the same order of magnitude for HIV-1 RT and AMV RT, these values are comparable in these two cases. It is substantially larger, by two to three orders of magnitude, for MoMLV RT, a factor that restricts for this last enzyme the 'window' of misincorporation frequencies that can be observed.

For any of the three enzymes, the main contribution to f comes from the increase in the value of the Michaelis constant when going from the correct to the incorrect nucleotide. In the cases tested, the ratio of the turn-over numbers, V'_{max}/V_{max}, is never smaller than 0.1 and averages of 0.71 for HIV-1 RT and 0.43 for AMV RT. On the whole, however, if the average values of the distributions of f are compared, the reverse transcriptase from HIV-1

Tables I–III. Kinetic parameters for the incorporation of correct and incorrect nucleotides on a restricted set of sequences

Table I. HIV-1 RT

Template sequence	Position	K_m (nM)	V_{max} (min^{-1})	f (× 10^4)		
				TS	TV_1	TV_2
TTTĀ	23 *gag*	45 ± 7	0.21 ± 0.01	1	<0.02	1.5
TTTĀ	59 *gag*	106 ± 36	0.07 ± 0.01	<0.1	<0.1	2.6
GATĀ	27 *gag*	740 ± 450	0.12 ± 0.02	30	<0.6	35
CATĀ	16 *lac*	54 ± 24	0.14 ± 0.02	–	<0.03	–
TAGĀ	25 *gag*	7.7 ± 2.3	0.20 ± 0.01	–	–	<0.04
GGGĀ	78 *gag*	10 ± 1	0.20	34	<0.03	<0.03
TACĀ	70 *gag*	133 ± 22	0.17 ± 0.01	<0.08	<0.08	22
TTAT̄	60 *gag*	106 ± 7	0.10	<0.1	<0.1	<0.1
AGAT̄	26 *gag*	6.7 ± 0.2	0.07	2	<0.1	<0.01
TCAT̄	63 *gag*	8.2 ± 0.7	0.15	0.64	0.07	<0.005
TCTT̄	67 *gag*	6.4 ± 1.3	0.20	0.8	–	–
TTCT̄	66 *gag*	4.7 ± 1.5	0.20 ± 0.02	1.8	–	–
TTAḠ	24 *gag*	11.3 ± 1.7	0.16 ± 0.01	0.24	0.1	<0.01
ATCḠ	75 *gag*	35 ± 6.5	0.20	0.14	33	<0.01
TATC̄	74 *gag*	90 ± 21	0.13 ± 0.01	5	0.2	<0.01
ATTC̄	65 *gag*	7.5 ± 0.7	0.20	0.3	0.06	<0.06

Table II. AMV RT

Template sequence	Position	K_m (nM)	V_{max} (min^{-1})	f (× 10^4)		
				TS	TV_1	TV_2
TTTĀ	23 *gag*	12 ± 2.0	0.16 ± 0.01	0.15	<0.007	12
CATĀ	16 *lac*	7.8 ± 3.0	0.17 ± 0.02	–	0.12	–
GATĀ	27 *gag*	200 ± 24	0.22 ± 0.01	3	0.5	0.9
TAGĀ	25 *gag*	24 ± 5	0.20 ± 0.01	–	–	0.25
TACĀ	70 *gag*	570 ± 169	0.14 ± 0.02	<0.4	<0.4	<0.4
GGCĀ	14 *lac*	12 ± 2.0	0.18 ± 0.01	<0.006	<0.006	<0.006
ATAT̄	17 *lac*	6.7 ± 0.6	0.20	1.4	1.2	–
AGAT̄	26 *gag*	3.0 ± 1.7	0.10 ± 0.01	0.03	<0.01	<0.01
TCAT̄	63 *gag*	28 ± 11	0.13 ± 0.02	4.1	0.16	0.14
GCAT̄	15 *lac*	7.2 ± 1.0	0.15 ± 0.01	5	10	<0.02
TTAḠ	24 *gag*	13 ± 2	0.21 ± 0.01	0.08	<0.06	<0.06
ATCḠ	75 *gag*	7.6 ± 2.9	0.22 ± 0.03	0.32	0.93	<0.003
TATC̄	74 *gag*	1460 ± 460	0.18	350	12	<0.3
ATTC̄	65 *gag*	8.4 ± 2.4	0.22 ± 0.01	0.28	0.11	0.34
GAGC̄	13 *lac*	1770 ± 290	0.09	<1.7	<1.7	<1.7

Table III. MoMLV RT

Template sequence	Position	K_m (μM)	V_{max} (min^{-1})	f (× 10^4)		
				TS	TV_1	TV_2
CATĀ	16 *lac*	0.5 ± 0.3	0.18 ± 0.03	–	–	–
TTTĀ	23 *gag*	1.3 ± 0.2	0.14 ± 0.01	<	<	8
AGCĀ	14 *lac*	18 ± 41	0.08 ± 0.01	<	<	<
AGAT̄	26 *gag*	2 ± 0.1	0.09	<	<	<
GCAT̄	15 *lac*	35 ± 12	0.17 ± 0.01	<	<	<
TTAḠ	24 *gag*	80 ± 27	0.22 ± 0.02	<	8	<
TATC̄	74 *gag*	10 ± 4.6	0.16 ± 0.01	<	80	<

The two first columns give the sequence of the template located 3′ upstream of the position examined (overlined character) and its location with respect to the constructs given in Figure 1. Columns 3 and 4 give the kinetic parameters for the correct incorporation. When standard deviations are not given, they are lower than the last digit given for the average. The frequencies of mismatches are listed in the last three columns. TS corresponds to transitions T_t:G, A_t:C, C_t:A and G_t:T respectively. TV_1 stands for homologous transversions T_t:T, A_t:A, C_t:C and G_t:G, and TV_2 for heterologous purine-purine or pyrimidine-pyrimidine mispairings (T_t:C, A_t:G, C_t:T and G_t:A). The dashes correspond to misincorporations that have not been assayed. The < sign stands for frequencies lower than the detection level.

is not more error-prone than the other two (overall average value of $f = 1.9 \times 10^{-4}$, compared to 4.9×10^{-4} for AMV RT and to $\sim 10^{-3}$ for MoMLV RT; cf. Table IV, line 11). This average value however depends strongly on the template analysed. If one considers only the positions where *in vivo* data are available (70 mismatches assayed instead of 105), averages increase for HIV-1 RT from $1.9-2.4 \times 10^{-4}$ and for AMV RT from $4.9-5.8 \times 10^{-4}$; cf. Table IV, line 12.

The sensitivity of $\bar{f}$ to the nature of the sample results from the very broad distribution of the values of f (cf. Figures 6A and B, where distributions are given for HIV-1 RT and for AMV RT). For these two enzymes the mismatch frequencies range from 5×10^{-6} to 3×10^{-2}. For MoMLV RT the distribution of observed mismatches is narrower but this is simply due to our inability to observe low values of f.

The relative occurrence of the twelve possible types of

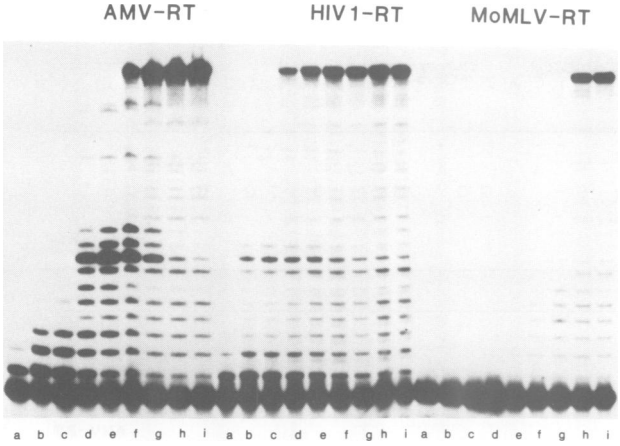

Fig. 4. Gel assay for the incorporation of consecutive nucleotides from position 26 to position 52 on an HIV-1 *gag* template (for sequence see Figure 1). At time zero the same concentration of each of the four deoxyribonucleotides was added and the reaction terminated after 10 min of incubation. Increasing amounts of dNTPs were used (**lanes a–i** indicate respective concentrations of 2.5 nM, 12.5 nM, 25 nM, 125 nM, 250 nM, 1.2 μM, 2.5 μM, 25 μM and 250 μM).

mismatches was then analysed for each reverse transcriptase. No striking difference is observed. In our sample, errors that could be accounted for by a 1 bp frameshift are not specifically frequent for any of the three enzymes tested. There is no clear tendency to favour an enrichment in AT or in GC content. Transitions generally dominate over transversions. A_t:G mismatches are however specifically abundant during polymerization with HIV-1 RT. This is illustrated in Figure 7, where a template lacking G can be copied by addition of dATP and dGTP alone. Increasing concentrations of dGTP allow HIV-1 RT to elongate successively past positions A68, A70 and A72. Thus incorporation of dGTP allows elongation of the primer beyond the mismatch and synthesis stops only at G75.

In fact, the ability of HIV-1 RT to elongate easily a primer where a mismatch has previously been introduced appears to be the second specific feature of this enzyme. Nineteen mismatches were examined in this respect and all could be elongated within 10 min by the incorporation of the correct subsequent nucleotide, added at saturating concentration. Furthermore, at specific positions HIV-1 RT appeared to incorporate up to four mismatches in a row (for example A_t:G and T_t:G mismatches are found one after the other at positions *gag* 70 to 73). The two other enzymes have difficulty in resuming a faithful incorporation after a misincorporation, in 25% of the cases for AMV RT and 50% of the cases for MoMLV RT, when using the test previously mentioned. For example, both failed to elongate after creating an A_t:A mismatch. Other specific examples are the difficulty of AMV RT to elongate after an T_t:C mismatch generated at position 71 *gag* or of MoMLV RT to continue polymerization after A_t:C mismatch created at position 27 *gag*. In no case did we observe the appearance of more than two mismatches in a row with these two enzymes.

Contrary to the average frequency of creation of a given type of mismatch, their localization within the sequence is significantly different. This feature, already apparent in Figure 5 where mismatches occurring at given positions are labelled with different characters depending on their frequency, is fully confirmed by a linear regression analysis. This analysis was performed on the 105 pairs of error frequencies, observed with HIV-1 RT (f_y) and with AMV RT (f_x); it shows a very poor overall correlation coefficient (0.06) the best fit being given by:

$$f_y = 0.01\ f_x + 1.94 \times 10^{-4}$$

A more detailed analysis, on each class of mismatches in our sample, confirmed that the errors made by these two enzymes are not in general significantly correlated (cf. Table V as an example). Among the twelve subclasses, only the A_t:C mismatches ($r = 0.78$ for nine common assays) show some significant correlation. Such an extensive analysis was not performed with MoMLV RT for the reasons given above, but as shown in Figure 5 and Table IV, the same conclusion could be drawn on the overall results. In conclusion, as for correct incorporations, the three enzymes read the DNA sequence in a different manner, and have difficulties at different points during elongation. The determinants in the DNA sequence which lead to this differential behaviour will be examined in the discussion.

Correlation with in vivo results

The relevance of these findings could be tested by looking at a particular subset of data, those obtained on p25 *gag*. The results observed by Goodenow *et al.* (1989) could have been due to misincorporations occurring at the DNA polymerization step only; they would have led to the changes given in the upper line of Figure 5. Fourteen of these changes were assayed *in vitro* using either HIV-1 RT or AMV RT in the assay. We asked whether those mistakes grouped here as class I, were often found in the enzymatic test, and whether they were more specifically made when the homologous reverse transcriptase was used. The distribution of their occurrence, within the error frequency scale, is represented by cross-hatched rectangles in the histograms given in Figures 6A and B, while the distribution of the other errors, tested *in vitro* on the same sequence but not found among HIV-1 isolates, is represented by hatched boxes (class II).

Average error frequencies $\bar{f}$, taken over the two classes, are 5.8×10^{-4} for AMV RT and 2.4×10^{-4} for HIV-1 RT, respectively. When class I alone is considered, a more than two-fold increase occurs if polymerization was directed by the homologous enzyme ($\bar{f} = 5.4 \times 10^{-4}$ for HIV-1 RT). In the control experiments, the average decreases from 5.8 to 1.1×10^{-4} (cf. Table IV, lines 12 and 13 which also give the corresponding standard deviation). Thus, the relevant *in vitro* assay detects the errors belonging to class I with an increased frequency, unlike the control test.

The histograms given in Figure 6 illustrate the main causes of these opposite trends. Firstly none of 14 true mistakes are hot spots in the AMV RT assay (cf. upper three columns of the histogram 6B). Secondly at the opposite end of the spectrum, more mismatches are undetected with AMV RT than with the HIV-1 RT enzyme (5 instead of 4). Third in the medium range, when there is not a great difference between the two distributions, errors coming from class I are slightly favoured with HIV-1 RT over AMV RT. At the sequence level (cf. Figure 4), a G_t:T mismatch at position 24 and a T_t:G mismatch at position 66 go undetected or are barely significant in the AMV RT assay, while they are

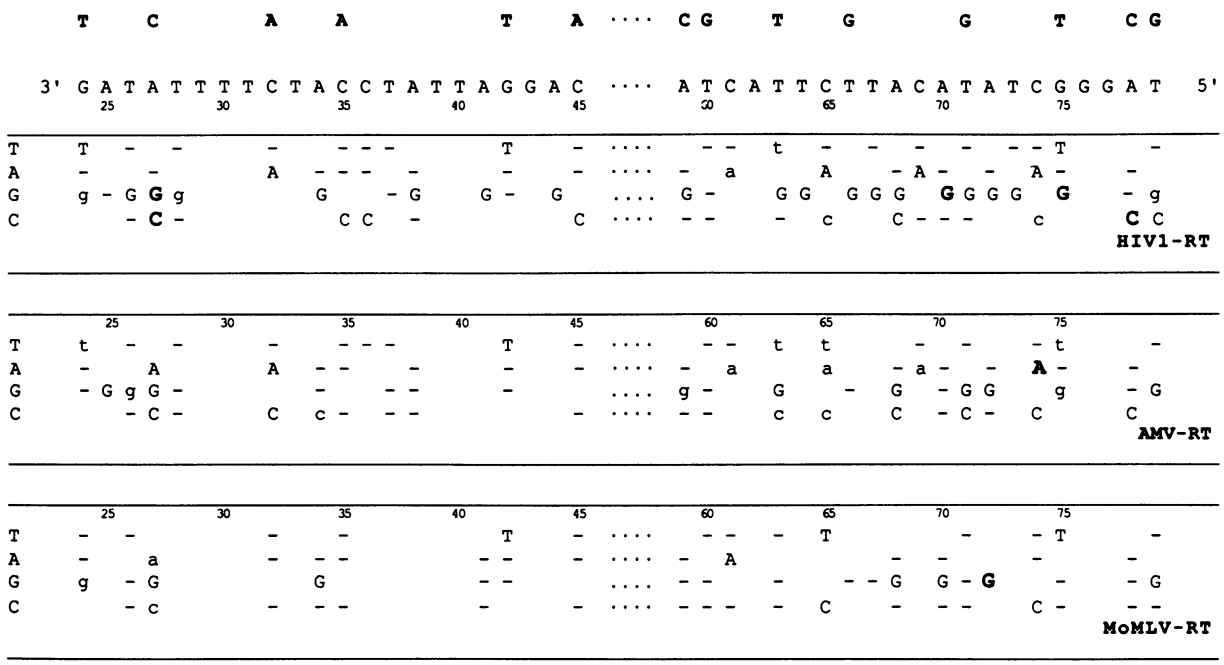

Fig. 5. Variability in the frequency of mismatches detected *in vitro*; comparison with *in vivo* results. In the second line, the p25 *gag*(−) HIV-1 sequence, which has been tested both *in vivo* and *in vitro*, is given (position 25 corresponds arbitrarily to position 1002 of the HIV-1 genome (Myers *et al.*, 1988)). The boldface characters, listed in the upper line, represent the set of misincorporations (class I) which were observed in sequencing the plus strand of various HIV-1 isolates. The lower part of the figure summarizes the error frequencies found *in vitro* with the three reverse transcriptases. Bases in boldface indicate mismatches occurring at a frequency higher than 10 $\bar{f}$, $\bar{f}$ being the arithmetical mean of the misincorporation frequencies for the enzyme considered on this particular set of data (class I + class II). Capital letters are used for mismatches occurring at frequencies ranging between 0.1 $\bar{f}$ and 10 $\bar{f}$, lower case letters for weak misincorporations ($f < 0.1\ \bar{f}$). A hyphen indicates that the corresponding mismatch has not been detected.

Table IV. One- and two-dimensional analysis of the kinetic data

		HIV-1 (1)	AMV RT (2)	MoMLV RT (3)	1−2	1−3	2−3
K_m (nM)	A	27 ± 37	15 ± 11	13 000 ± 16 000			
	T	156 ± 242	119 ± 195	6600 ± 8000			
	G	49 ± 41	734 ± 725	10 000 ± 4600*			
	C	23 ± 12	8.3 ± 3	80 000 ± 27 000*			
	N	76 ± 161	127 ± 339	18 400 ± 25 700	$r = 0.38$	$r = -0.2$	$r = -0.04$
V_{max} (min^{-1})	A	0.164 ± 0.05	0.167 ± 0.04	0.128 ± 0.03			
	T	0.160 ± 0.04	0.175 ± 0.02	0.134 ± 0.04			
	G	0.167 ± 0.03	0.164 ± 0.05	0.158 ± 0.01*			
	C	0.177 ± 0.02	0.211 ± 0.01	0.223 ± 0.03*			
	N	0.167 ± 0.05	0.182 ± 0.03	0.146 ± 0.04	'$r = 0.80$	'$r = 0.78$	'$r = 0.93$
f		1.9 ± 6.2	4.9 ± 34	10 ± 56	$r = 0.06$	$r = -0.01$	$r = 0.30$#
f(I+II)		2.4 ± 6.3	5.8 ± 42		$r = 0.05$		
f(I)		5.4 ± 11	1.1 ± 2		$r = 0.16$		

*A single reliable determination (cf. Table III).
#This value is marginally significant. By linear regression analysis, the best straight line corresponds to f(2) being independent of f(3).
'*r*-values reaching level of significance at $P < 0.05$.
Each set of data (K_m, V_{max}, f) is characterized by its arithmetic mean value and its standard deviation. The letters specify the nature of the correct nucleotide incorporated (the lines labelled N correspond to the overall distribution of values of K_m and V_{max}). Frequencies of misincorporation, f, have been multiplied by 10^4. The distribution of f is taken either over the overall sample (line 11), on the *gag* sequence tested *in vivo* (classes I + II, line 12) or on the class of mismatches found *in vivo* (class I, line 13). The first three columns of numbers show values observed respectively for HIV-1 RT (1), AMV RT (2) and MoMLV RT (3). The last three columns give the correlation coefficients, *r*, between the three distributions for the kinetic parameter considered; values that are significant are indicated.

rather frequent in the HIV-1 RT assay; two mismatches A_t:C 27 and A_t:C 78 are hot spots when the homologous enzyme replaces the control system; two other misincorporations G_t:T 24 and C_t:A 32 are significantly more frequent, while the 8 remaining errors are found at a similar level. MoMLV RT behaves qualitatively as AMV RT on those sequences: no increase in the average frequency of misincorporation when the sample is restricted to class I; detection of only four misincorporations out of fourteen with this *in vitro* enzymatic system.

Figure 8 summarizes in a different way the propensity of HIV-1 RT (full line) and of AMV RT (dotted line) to make *in vitro* errors compatible with the *in vivo* findings. This is a selectivity diagram which should be read from right to left.

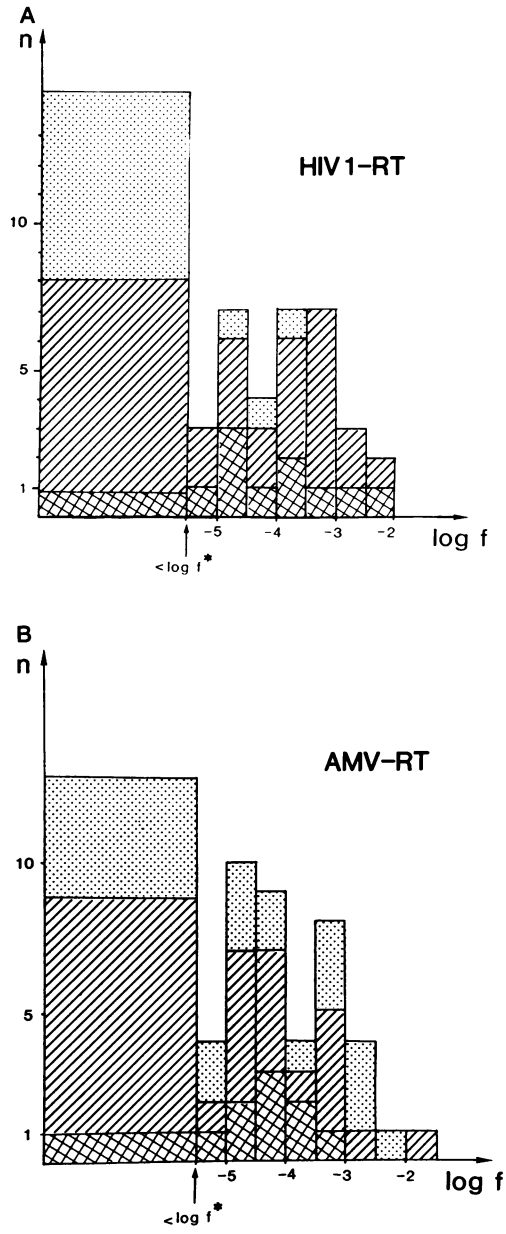

Fig. 6. Histograms of the frequencies of misincorporations assayed with the present *in vitro* test with both HIV-1 RT (**A**) and AMV RT (**B**). Three different classes of mismatches have been considered. Class I (represented here by cross-hatched boxes) corresponds to the 14 mismatches which would have led to the sequence variability observed *in vivo* on the *gag* p25 HIV-1 sequence (cf. Figure 5, upper line). Class II (hatched boxes) includes 56 other mismatches assayed on the same sequence and not found *in vivo*. Class III encompasses 35 misincorporations (stippled boxes) tested on the other DNA templates given in Figure 1. The histogram was constructed by scoring the number of times mismatches occur in a given frequency range (the abscissa is a logarithmic scale). All misincorporations that were not observed are placed in the lower part of the histogram below the corresponding threshold of detection, f*.

For a given enzyme, one scores the number of misincorporations occurring at a frequency higher than a given value, f_x, in class I, or in the overall sample (classes I and II). The corresponding numbers are n_+ and n. The ratio n_+/n is then a measure of the ability of the enzyme to generate preferentially the mismatches observed *in vitro*, in the relevant frequency range ($f > f_x$).

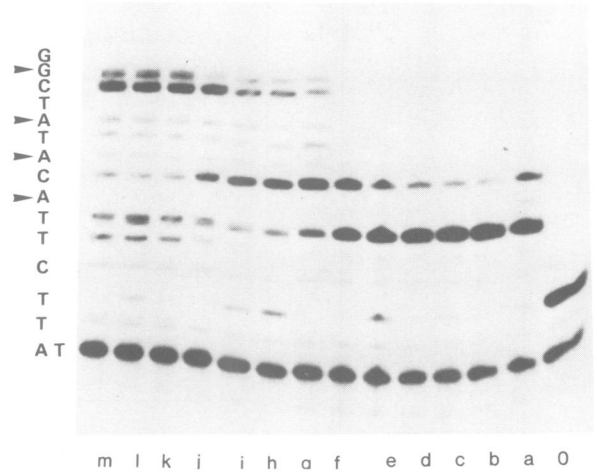

Fig. 7. Gel assay for the incorporation of dATP and dGTP on HIV(−) *gag* template by HIV-1 RT. Synthesis started at position 63. Correct elongation stops after five steps. Addition of dATP alone (100 nM) allows correct elongation until position 2; dGTP promotes correct elongation until position 5. Increasing amounts of dGTP were used (**lanes a−m** indicate respective concentrations of 100 nM, 500 nM, 700 nM, 1 μM, 5 μM, 10 μM, 50 μM, 70 μM, 100 μM, 200 μM, 500 μM, 700 μM, 1 mM. **Lane O** is without dGTP.

Table V. Absence of correlation in the occurrence of T_t:G mismatches, generated by two reverse transcriptases

	$f_y(\times 10^4)$	$f_x(\times 10^4)$	
60 g	<	<	
66 g	1.8	<	Class I
71 g	1.6	7	
79 g	0.16	2.8	
26 g	2	0.03	
28 g	0.22	<	Class II
37 g	<	<	
63 g	0.64	4.1	
12 g	<	–	
13 e	<	0.04	Class III
15 l	<	5	

The first column indicates the position of the template base; in the second column is given the frequency of occurrence of the corresponding T_t:G mismatch created by HIV-1 RT ($\times 10^4$); in the third column the frequency is given when AMV RT is used. A < sign indicates that the corresponding mismatch was not detected. *gag* (g), *env* (e) or *lac* UV5 sequences were used as templates. The overall correlation coefficient, *r*, is equal to 0.155 and is not significant.

The total number of mismatches of class I being 14, and the total number of mismatches examined in the relevant sequence being 70, an absence of selectivity will imply points scattered around $\langle n_+/n \rangle = 0.2$. The maximal possible selectivity would correspond to the 14 mismatches of class I occurring at the highest values of f and would be described by a curve starting at $\langle n_+/n \rangle = 1$ for large values of f and then reaching progressively the average value of 0.2 when all the errors affixed to class II will have been taken into account at their frequency of occurrence. It is clear that for HIV-1 RT, selectivity exists through the whole range of frequencies; some mismatches of class I occur with a particularly high misincorporation frequency ($f > 5 \times$

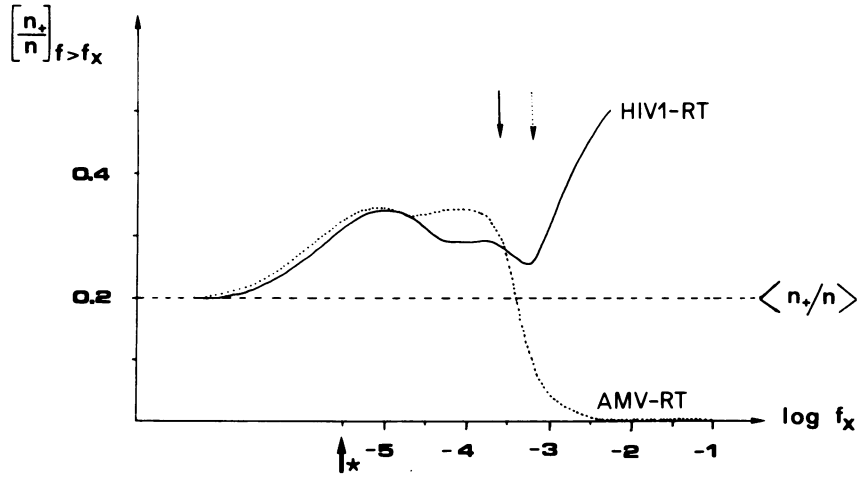

Fig. 8. Ability of two reverse transcriptases to generate at the DNA polymerization step, the substitutions observed *in vivo*. n_+ is the number of mismatches of class I (among 14) occurring at a frequency larger than the one given in the abscissa; n is the corresponding number (among 70) when class II and class I are taken into consideration. n_+/n is plotted against log(f). As explained in the text, the extent by which the curve exceeds 0.2 is a measure of the selectivity of a given enzyme, in the corresponding range of error frequency. HIV-1 RT appears to select mismatches of class I in the whole frequency range unlike AMV RT (compare the full line to the dotted line). The two downward arrows indicate the average error frequencies for the two sets. The upward arrow along the abscissa corresponds to log f*.

10^{-3}). In the same frequency range, class I is in fact disfavoured when the assay is performed with AMV RT. In the medium frequency range the two enzymes display about the same propensity in slightly favouring errors belonging to class I. Finally, the two curves tend towards 0.2 as the whole set is exhausted, leaving four mismatches undetected in the case of HIV-1 RT.

The existence of a subset of transitions of class I, which was not found *in vitro* with any of the three enzymes, prompted us to check whether the corresponding errors could rather be attributed to the retrotranscription step occurring on the RNA (+ strand) of the template. Two approaches were used, which will be fully described in a forthcoming publication. Firstly, DNA matrices homologous to the RNA strand were synthesized (cf. Figure 1) and a specific search for G_t:T mismatches occurring at positions 35, 36, 45 and 47 was undertaken. All these misincorporations were found in the enzymatic test; they occurred at a relatively high frequency (contrary to the corresponding C_t:A transitions missing in the previous assay) both with HIV-1 RT and with AMV RT. Secondly when proper RNA templates were synthesized all the missing misincorporations were found (data not shown). Thus full consistency between *in vitro* and *in vivo* results requires taking into account both polymerization reactions catalysed by HIV-1 RT.

Discussion

The elongation assay proposed by Boosalis *et al.* (1987) allows a very good quantitative determination of the overall kinetic parameters for the faithful copy of a given DNA template. Up to now these data have been obtained mainly as average constants measured on various homopolymers. A striking variability was observed for all three reverse transcriptases, particularly when the apparent Michaelis constants are considered. Furthermore no significant correlation is observed between those constants when the enzymes are compared pairwise. In contrast, maximal velocities are more closely related. As a result the stalling patterns observed at low substrate concentrations during the processive elongation of a given primer are much more dependent on the enzyme and on the DNA sequence than at saturating concentrations of dNTPs. This finding could be used for the design of antiviral agents directed against reverse transcriptases. They could be specifically targeted to inactivate the enzyme during a very slow elongation step.

Since the three enzymes under consideration 'see' differently the various DNA templates during faithful incorporation, it was logical to compare their fidelity on the same set of templates. This approach has a methodological advantage, namely to suppress most problems linked to the eventual cross-contamination of the deoxynucleotides. Contrary to expectation, in this particular test which scores only for base substitutions HIV-1 RT did not appear more inaccurate than the two other enzymes tested. The mean frequency of misincorporation was found to be of the order of magnitude (10^{-4}) compatible with *in vivo* estimates (cf. Introduction). However, this particular enzyme has a larger propensity to carry on elongation after the incorporation of a given mismatch. This feature, already noticed by Preston *et al.* (1988), could substantially increase the frequency of misincorporation per completed retrotranscript during *in vivo* replication of an HIV-1 genome. It could also reconcile our results with those obtained through reversion assays, where DNA templates have to be copied full-length to be taken into account and where HIV-1 RT appears more inaccurate than its AMV equivalent (cf. Preston *et al.*, 1988; Weber and Grosse, 1989; Perrino *et al.*, 1989). Here again, the variance of the error spectrum range generated by the three enzymes was extremely large, and the correlation between the frequencies of introduction of a given mismatch at a specific position of the sequence by two given enzymes was very poor. When an improper substrate was used to replace the correct one, it was mainly the change in the apparent Michaelis constant which determined the frequency of misincorporation at a given position, an observation fully consistent with the passive discrimination model put forward by Sloane *et al.* (1988). In this model, polymerases enhance base selection simply because during the initial binding of the substrate (associated with K_m) they differentially

increase the rates of rejection of non-complementary nucleotides as compared to the correct one. In this context, however, because of the extreme specificity of the error spectrum generated by each reverse transcriptase, we predict that each catalytic site would have a different way to position the incoming substrate with respect to the primer template, and therefore a different ability to increase the rate of rejection of the incorrect nucleotide in a sequence specific manner.

This purely enzymatic approach throws some light on the functional problem of the variability of the various viral genomes: in our assay, the mismatches predicted to occur from the analysis of viral HIV-1 isolates are indeed introduced with a greater frequency by HIV-1 RT than by the two other enzymes tested. Some striking findings illustrate the validity of the *in vitro* approach. For example the only T_t:T mismatch created by the homologous enzyme is the one that can account for the sole A to T transversion observed by Goodenow *et al.* in this region of the HIV-1 genome; see also two specific misincorporations predicted at positions 24 (G_t:T) and 66 (T_t:G) which occur only at a reasonable frequency when the homologous system is used. However, this trend towards selectivity is only qualitative and some mismatches, predicted from the *in vivo* results are not observed when the DNA polymerization of the minus strand only is considered. Yet, this particular subset of changes, observed *in vivo* in sequencing the plus strand of the proviral DNA, was fully accounted for when HIV-1 RT directed polymerization was followed on a complementary template mimicking the RNA plus strand of the virus. Thus mistakes occurring at both the reverse transcription level and at the DNA polymerization level have to be taken into consideration, if one wants to account for the *in situ* variability of the HIV-1 genome.

The next step is obviously to understand how the sequence of the template introduces such large fluctuations in the values of K_m and f for a given enzyme, and why each enzyme has its own way of reading these various sequences. We are presently trying to establish those rules, using a systematic approach in the design of adequate templates. Several empirical rules can already be suggested from the present study. Firstly the part of the template located downstream of the elongation step considered does not play a major role in specifying the kinetic constants V_{max}, V'_{max}, K_M and K'_M (positions 26 *gag*, 71 *gag* and 15 *lac* correspond to T followed by the same sequence ATPy; they show a broad variability in all the parameters tested). Secondly a single base located immediately upstream of the nucleotide that is incorporated does not unequivocally determine its kinetic constants for any enzyme tested. In particular, our results could generally not be accounted for by a misalignment model of mutagenesis proposed by Kunkel and Alexander (1986). Thirdly we can make reasonable predictions by taking into consideration the 3′ doublet upstream of the relevant base of the template. These correlations sometimes fail, when the upstream part of the primer template can adopt some unusual DNA structure (for example, an A_t:C mismatch occurring at the 5′ end of a TTA template sequence arises with a low frequency except at position 23 *gag* which is preceded by a stretch of 3T). For correct incorporations, HIV RT incorporates a T more easily when the template A is preceded by a purine than when it follows a pyrimidine (lower K_m, higher V_{max}). By contrast, AMV RT always incorporates C more easily.

As far as mismatches are concerned, one can already understand why HIV-1 RT shows a higher propensity than AMV RT to create A_t:G mismatches in our sample. For HIV-1 RT this mismatch occurs when the template adenine residue is preceded by the doublets AT, TT, CT, or AC and is prevented when the 3′ sequence is NG, GT, TC or GC. The rules are different for AMV RT: the mismatch is facilitated by the doublets AT, TT, GT and AG and disfavoured by AC, TC, GC, GG or CT sequences. The larger error frequency found in the first case may thus be mainly due to the composition of our sample (where CTA sequences are overrepresented) rather than to an intrinsically larger inability of HIV-1 RT to avoid this particular type of mismatch.

Similar observations can be made on the 3′ template sequences favoring each type of transition for either AMV RT or HIV-1 RT. A significant correlation between error frequencies occurs only in the case of A_t:C mismatch which are at least five times more frequently generated by HIV-1 RT. In both cases, GG, AT and TT favour the appearance of this mismatch while NC has never been shown to permit it. It is interesting to analyse in parallel the patterns observed for T_t:G mismatches, since in a cycle of replication U_t:G mispairings occurring during reverse transcription of the + strand and A_t:C mismatches occurring during DNA polymerization yield the same final U → C transition (or vice versa to the same final A → G change). As shown in Table IV, the frequencies of occurrence of these mismatches are absolutely not correlated and indeed the 3′ doublets that promote them are different (GA, CA, TA, TC for HIV-1 RT, CA, GA and AG for AMV RT). When the other pair of linked transitions, C_t:A and G_t:T are examined, similar deductions can be made. C_t:A transversions were found to occur more frequently when AMV RT was used; in this case only the doublets GA and AG prevented the appearance of this mismatch. HIV-1 RT, on the contrary, did not introduce this mismatch in the six cases where a purine preceded the relevant cytosine in the template sequence. For the G_t:T transversions these empirical rules are not yet established because of their poor occurrence in our present sample.

We therefore tentatively propose that the differential effects analysed here in the correct and incorrect incorporations of deoxynucleotides reflect a different ability for the three reverse transcriptases to stack correctly the incoming substrate on the two preceding base pairs of primer-template hybrid. The broad variability in the kinetic constants that have been measured reflects the heterogeneity in the kinetic constants which should be attributed to at least 64 combinations. The present *in vitro* assay is rather imperfect since the nucleoprotein structures of the physiological templates are not reassembled, and since we do not take into account the pattern of errors due to the host RNA polymerase. However, it allows us to conclude that genomic variability of retroviruses is a property depending not only on the nucleic acid sequence but also on the reverse transcriptase used, and that both polymerization steps have to be taken into account to explain the sequence dependence of the changes recorded *in vivo*.

Materials and methods

HIV-1 RT was purified from an overproducing strain (gift of Dr S.H.Hughes, Frederick, USA) (Hizi *et al.*, 1988) using a published procedure (cf. Hansen

et al., 1987). AMV RT was purchased from Pharmacia and MoMLV RT from BRL. The deoxynucleotide primers and templates were synthesized using β-cyanoethylphosphoramidite supports with an LKB Gene Assembler Synthesizer (Pharmacia). The dNTP substrates, FPLC pure, were purchased from Pharmacia; their purity was checked by FPLC analysis (cf. below). [γ-^{32}P]ATP (10 Ci/mol, 10 mCi/ml) was purchased from Amersham and T4 polynucleotide kinase from BRL.

Primer labelling

The 5′ termini of DNA primers were labelled with ^{32}P in a 50 μl reaction mixture containing 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT (dithiothreitol), 0.7 mM [γ-^{32}P]ATP, 0.2 mM primer oligonucleotide and two units of polynucleotide kinase. The solution was incubated at 37°C for 1.5 h and the reaction terminated by addition of 10 mM EDTA followed by heating at 90°C for 20 min. The DNA was then purified by preparative electrophoresis on 20% polyacrylamide gels containing 8 M urea, the radioactive band being cut out, eluted by overnight shaking in TE buffer (10 mM Tris−HCl, 1 mM EDTA, pH 7.6) followed by three phenol extractions and two precipitations in 100% ethanol. The primer was washed once with 70% ethanol, dried in a Savant speed-vac concentrator and resuspended in doubly distilled H_2O.

Template purification and primer annealing

The template DNA was prepared and purified as described above. During the preparative electrophoresis step a small amount of radioactive fragment was co-electrophoresed in order to determine the position of the correct band. Hybridization with the primer was performed at a constant amount of primer (0.4 nM) and increasing amounts of template, in the buffers used for the assay (see below), for 20 min. The amount of active hybrid was determined by the enzymatic assay of reverse transcriptases to see whether or not all the radioactive hybridized primer could be displaced into its normal products, in the presence of high concentrations of RT (10 nM) and dNTPs at 37°C for 10 min. As a result, the primer/template ratio was always adjusted to 0.4 nM primer for 10 nM template (ratio 1:25).

Enzymatic assays

Slightly different reaction buffers were used for each reverse transcriptase; for AMV RT, 50 mM Tris−HCl, 5 mM $MgCl_2$, 40 mM KCl, 1 mM DTT, pH 8.3; for MoMLV RT, 50 mM Tris−HCl, 3 mM $MgCl_2$, 75 mM KCl, 1 mM DTT, pH 7.5; for HIV-1 RT, 50 mM Tris−HCl, 6 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, pH 7.8 (we checked that these changes do not appreciably affect the values of the kinetic constants). A given amount of the specified reverse transcriptase was incubated with the hybrid for 10 min at 37°C. The dNTPs were then added at various concentrations usually for 5 min. The reaction was stopped by the addition of 10 mM EDTA in 90% formamide and heating for 15 min at 90°C. Firstly titration experiments were performed at a constant amount of primer/template and increasing amounts of enzyme, under initial velocity conditions (cf. Boosalis *et al.*, 1987). Secondly a concentration of enzyme corresponding to half the equivalence point was used. At each substrate concentration, initial velocities could then be computed in terms of the percentage of the displacable primer that is converted into product per unit time.

Gel electrophoresis, autoradiography and densitometry

The DNA samples in formamide were run on a 20% polyacrylamide gel containing 8 M urea. The gels were overlayed with a Kodak XAR-5 film in the presence of an intensifying screen and kept at −80°C for times varying between 5 and 14 h. Exposure was adjusted to obtain bands of which the intensity was in the range of a linear response of the film, as previously evaluated by multiple expositions of a standard gel. Autoradiographs were scanned on a Biorad 620 video densitometer (which gave a linear response up to a total absorbance equal to 2.0). Band intensities were computed using a Shimadzu integrator.

Curve fitting procedure

Estimates and errors on both the apparent Michaelis constant and the maximal velocity for the incorporation of a given deoxynucleotide were obtained from a plot of the initial velocity against the substrate concentration, using a non-linear regression analysis initially described by Wilkinson (1961), implemented as an 'Enzfitter' program on an IBM-PC computer. From these values, obtained for correct and incorrect nucleotides, the frequencies of misincorporations, f, were calculated as the ratio $\frac{V'_{max}}{K'_m}$ over $\frac{V_{max}}{K_m}$ where the dash quantities refer to the incorrect nucleotides, and the others to the correct nucleotide.

Statistical analysis

Data obtained with the various enzymes at homologous positions (apparent Michaelis constants, maximal velocities, frequencies of misincorporation) were compared two by two by correlation analysis. The approach is similar to methods used for objective comparison of two sets of sequences (cf. for example Bougueleret *et al.*, 1988; Rogerson, 1989). The Pearson-product moment correlation coefficient was used as a measure of the similarity of a given variable when observed on two sets of data. In cases where the correlation appeared significant, $P < 0.05$, a simple linear regression analysis was then performed.

Computation of f*

Frequencies of misincorporations are not equally detected for all the positions considered, the limit of detection f*, being strongly dependent on the kinetic constants for the incorporation of the correct nucleotide. The larger the value of $\frac{V_{max}}{K_m}$, the easier it is to pick up misincorporations occurring at a low frequency. The lowest detectable value of f can be determined as follows. An incorporation of 2.5% in 10 min is our present threshold of detection; only rates of incorporation larger than 2.5×10^{-3} min^{-1} are then considered as positive. The maximal substrate concentration used here was 250 μM. These two conditions impose $\frac{V_i}{S} \geq 10^{-8}$ nM^{-1} min^{-1} for detection to occur. For weak mismatches, $\frac{V_i}{S}$ is a fair estimate of $\frac{V'_{max}}{K'_m}$. Hence, to be detected, f must be larger than f* given by $f^* = 10^{-8} (K_m/V_{max})$.

Nucleotide purity

The ability of the enzyme to elongate beyond a position for which the complementary substrate was not provided can be attributed either to a misincorporation or to the presence of a contaminating deoxynucleotide triphosphate in the cocktail mixture. Purity of nucleotides was first analysed with a FPLC system equipped with a PepRPC HR5/5™ (Pharmacia) column. Elution was performed with a 0%−50% CH_3CN gradient buffered with 50 mM triethylammonium acetate, pH 6.5. A second analysis was based on the observation that several 'incorrect' elongations were present only when using one out of three or two out of the three RTs. The negative control given by the RTs which in the same conditions did not elongate the primer excluded in those cases a contamination problem.

More specifically we could define positions in our sequences where no misincorporation was observed with a given enzyme and where the kinetic constants for the correct substrate are very good. For example at position 78 of the *gag* sequence, the kinetic constants for incorporation of dTTP by HIV-1 RT are: $V_{max} = 0.2$ min^{-1} and $K_m = 10$ nM. As the minimal detectable velocity of incorporation is $V_i = 2.5 \times 10^{-3}$ nM^{-1} min^{-1} (see above), putting $V_i = 0.2$ min^{-1} $\frac{S}{10\text{ nM}}$, implies that one could in principle detect a concentration of dTTP contaminants larger than 0.12 nM by following an eventual incorporation at this template position with HIV-1 RT. Reconstitution experiments could be performed by deliberately contaminating other dNTP solutions (used at 100 μM or 1 mM) with dTTP. They show that this level of detection should be raised to 2.5 nM to take into account the inhibition of the rate of dTTP incorporation due to the presence of the other nucleotide. Similar experiments were conducted at other positions on the sequence in order to investigate any contamination of any source for the dNTP used. The comparative studies performed here with the various enzymes ruled out any significant contamination in all cases reported.

Acknowledgements

We thank S.Wain-Hobson and F.Barré-Sinoussi, who introduced us to the subject, for many helpful suggestions, P.Roux and L.Bracco for their assistance, F.Tekaia for his guidance in the statistical analysis of data, and O.Delpech for typing the manuscript. This work was supported by grants from the M.R.E.S. (no. 87C0396), the I.N.S.E.R.M. (no. 88-1003), the A.N.R.S. and the F.R.M. M.R. is a recipient of a fellowship from the E.E.C.

Footnote

A preliminary account of this work was given at the EMBO Workshop 'The Molecular Biology of Retroid Viruses and Elements', Flumersberg-CH, 3−7 April 1989; see comment by Hull and Will (1989) *Trends Genet.*, **5**, 357−359.

References

Battula,N. and Loeb,L.A. (1974) *J. Biol. Chem.*, **249**, 4086–4093.

Bebenek,K., Abbotts,J., Roberts,J.D., Wilson,S.H. and Kunkel,T.A. (1989) *J. Biol. Chem.*, **264**, 16948–16956.

Boosalis,M.S., Petruska,J. and Goodman,M.F. (1987) *J. Biol. Chem.*, **262**, 14689–14696.

Bougueleret,L., Tekaia,F., Sauvaget,I. and Claverie,J.-M. (1988) *Nucleic Acids Res.*, **16**, 1729–1738.

Coffin,J.M. (1986) *Cell*, **46**, 1–4.

Coffin,J.M., Tsichlis,P.V., Barker,C.S. and Voynow,S. (1980) *Ann. NY Acad. Sci.*, **354**, 410–425.

Dougherty,J.P. and Temin,H.M. (1988) *J. Virol.*, **62**, 2817–2827.

Fersht,A. (1977) In *Enzyme Structure and Mechanism*. W.H.Freeman and Co., Reading and San Francisco.

Gojobori,T. and Yokoyama,S. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4198–4201.

Goodenow,M., Huet,T., Saurin,W., Kwok,S., Sninsky,J. and Wain-Hobson,S. (1989) *J. Acquired Immune Deficiency Syndrome*, **2**, 344–352.

Hansen,S., Schulze,T. and Moelling,K. (1987) *J. Biol. Chem.*, **262**, 12393–12396.

Hizi,A., McGill,C. and Hughes,S. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1218–1222.

Kunkel,T.A. and Alexander,S. (1986) *J. Biol. Chem.*, **261**, 160–166.

Leider,J.M., Palese,P. and Smith,F.I. (1988) *J. Virol.*, **62**, 3084–3091.

Mendelman,L.V., Boosalis,M.S., Petruska,J. and Goodman,M.F. (1989) *J. Biol. Chem.*, **264**, 14415–14423.

Myers,G., Rabson,A.B., Josephs,S.F., Smith,T.F. and Wong-Staal,F. (1988) In *Human Retroviruses and AIDS*. Los Alamos Natl. Lab., Los Alamos, Vol. I, pp. 1–195.

Perrino,F.W., Preston,B.D., Sandell,L.L. and Loeb,L.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8343–8347.

Preston,B.D., Poresz,B.J. and Loeb,L.A. (1988) *Science*, **242**, 1168–1171.

Roberts,J.D., Bebenek,K. and Kunkel,T.A. (1988) *Science*, **242**, 1171–1173.

Rogerson,A.C. (1989) *Nucleic Acids Res.*, **17**, 5547–5563.

Sloane,D.L., Goodman,M.F. and Echols,H. (1988) *Nucleic Acids Res.*, **16**, 6465–6475.

Steinhauer,D.A. and Holland,J.J. (1987) *Annu. Rev. Microbiol*, **41**, 409–433.

Weber,J. and Grosse,F. (1989) *Nucleic Acids Res.*, **17**, 1379–1393.

Wilkinson,G.N. (1961) *Biochem. J.*, **80**, 324–332.

Received on December 11, 1989; revised on February 7, 1990