Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11787–11791, December 1994
Evolution

Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations

MIGUEL ANGEL MARTINEZ, JEAN-PIERRE VARTANIAN, AND SIMON WAIN-HOBSON

Unité de Rétrovirologie Moléculaire, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France

Communicated by Manfred Eigen, August 22, 1994

ABSTRACT **The finding of G → A hypermutated retroviral genomes in which up to 40% of guanines may be substituted by adenines was proposed to result from the depletion of the intracellular dCTP concentration and suggested a means to hypermutagenize nucleic acids. Using a RNA/reverse transcriptase ratio of ≈1:30, comparable to that within the retroviral replication complex, G → A hypermutants were produced in a simple *in vitro* reaction using highly biased dNTP concentrations—i.e., a low ratio of [dCTP]/[dTTP]. Up to 38% of G residues could be substituted, the proportion being inversely proportional to the concentration of dCTP. As G → A hypermutation resulted from elongation beyond multiple rG·dT mismatches, U → C hypermutants resulting from multiple rU·dG mismatches were sought, and found, during cDNA synthesis using low [dATP] and high [dGTP]. Mixed G → A and U → C hypermutants could also be produced under conditions of low [dCTP] plus low [dATP] and high [dTTP] plus high [dGTP]. Hypermutagenesis should allow jumping through, and subsequent exploration of, sequence space to a greater degree than heretofore and, in conjunction with genetic screening, might be of use in the search of proteins or ribozymes with novel or enhanced properties.**

RNA viruses replicate with an intrinsic replication error some 300 times greater than DNA-based microbes and $\approx 10^6$ times greater than eukaryotic genomes (1). This is the consequence of a total lack of replication proofreading machinery and results in an intrinsic nucleotide substitution error of ≈0.05–1 per genome per cycle (2). Occasionally there is a total breakdown in replication fidelity, giving rise to hypermutated genomes encoding hundreds of monotonously substituted bases (3–10).

To date there are two distinct types of hypermutated RNA viral genomes. A → I hypermutation of measles and vesicular stomatitis viral genomes is thought to result from post-transcriptional enzymatic modification of adenosine to inosine (11). G → A hypermutated genomes have been described for a large number of retroviruses including human immunodeficiency virus type 1 (HIV-1) (5–10) and were hypothesized to arise during reverse transcription as a result of monotonous substitution of dCTP by dTTP due to localized depletion of intracellular dCTP (6, 10).

If retroviral G → A hypermutation could be produced *in vitro* by the manipulation of dNTP pools then it would provide a means to hypermutate nucleic acids and allow the exploration of sequence space to a degree not previously possible. Probably the major obstacle in the construction of complex and varied libraries of mutants is the inability to introduce in a quasi-random manner large numbers of mutations. Contemporary methods rely on enzymatic and chemical methods (12, 13). Invariably one or two mutations per clone may be introduced, thus restricting the complexity of the libraries and the diversity of mutants derived. It is shown here that RNA can be hypermutagenized during cDNA synthesis using the HIV-1 reverse transcriptase (RTase) and highly biased dNTP concentrations.

MATERIALS AND METHODS

DNA from M13mp18 recombinant containing the *env* V1–V2 regions of HIV-1 B40 (6) was PCR-amplified with oligonucleotide 1 (5′-GCGAAGCTTAATTTAATACGACTCACTATAGGGACAAAGCCTAAAGCCATGTGTA) and oligonucleotide 2 (5′-GCGGAATTCTAATGTATGGGAATTGGCTCAA). Oligonucleotide 1 contains the T7 RNA polymerase promoter sequence (underlined) allowing production of plus strand transcripts of the *env* V1–V2 region. The 342-bp DNA fragment, with unique *Hin*dIII and *Eco*RI restriction sites (boldface), was digested and ligated into a pBluescript SK (+) vector. The resulting plasmid was digested with *Hin*dIII and *Eco*RI, and the fragment was purified from a 2% agarose gel. Insert DNA was used as template for *in vitro* transcription using T7 RNA polymerase. Reaction conditions were 40 mM Tris·HCl (pH 8), 30 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 50 μg of RNase/DNase-free bovine serum albumin per ml, 500 μM each NTP, 100 ng of template, 0.3 unit of RNase inhibitor per μl (Pharmacia), and 2 units of T7 RNA polymerase per μl (Pharmacia) in a final volume of 100 μl. After incubation for 1 hr at 37°C, the DNA template was digested with 0.075 unit of RNase-free DNase I per μl (Pharmacia) for 30 min at 37°C. RNA was phenol extracted and ethanol precipitated. The reverse transcription reaction buffer was 50 mM Hepes (pH 7), 15 mM magnesium aspartate, 10 mM dithiothreitol, 55 mM potassium acetate, 15 mM NaCl, and varying dNTP concentrations (see legend to Fig. 1). Two picomoles of oligonucleotide 3 (5′-GCGTCTAGAAGTATCATTATCTATTGGTA, complementary to positions 224–243 of the 342-bp plus strand DNA fragment) was annealed to 0.5 pmol of the template RNA in 50 μl of the reaction by first heating to 65°C for 1 min followed by incubation at 37°C for 1 min, after which 0.3 unit of RNase inhibitor per μl and 15 pmol (6.25 units) of HIV-1 RTase (Boehringer Mannheim) were added. The reaction was incubated at 37°C for 3 hr.

To recover sufficient material for subsequent cloning, cDNA was amplified by PCR with oligonucleotide 3 and oligonucleotide 4 (5′-GCGGTCGACCAAAGCCTAAAGCCATGTGTA), producing a 231-bp DNA fragment with restriction sites at its ends, *Xba* I and *Sal* I, respectively (boldface). Optimized PCR conditions were 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris·HCl (pH 8.3), 200 μM each dNTP, 20 μl of the reverse transcription reaction, and 2.5 units of *Taq* DNA polymerase (Cetus) in a final volume of 100 μl. Annealing, extension, and denaturation cycling parameters were 37°C (30 s), 72°C (30 s), and 95°C (30 s) for 2 cycles and 55°C (30 s), 72°C (30 s), and 95°C (30 s) for 10 cycles. Southern blot analysis was performed on 5% of the PCR-amplified material electrophoresed through a 2% agarose gel and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV-1, human immunodeficiency virus type 1; RTase, reverse transcriptase.

blotted into a nylon membrane by capillary transfer. The filter was hybridized with a ^{32}P 5′-labeled probe 5′-TGTAAAA(T/C)(T/C)AACCCCAC(T/C)C, complementary to positions 53–71 of the *env* V1–V2 342-bp minus strand. The specific 231-bp PCR fragments were purified from a 2% agarose gel, digested with *Xba* I and *Sal* I, and ligated into a pBluescript SK (+) vector. Clones were dideoxy-sequenced using the *Taq* dye primer cycle sequencing kit (Applied Biosystems), with resolution of the products on an Applied Biosystems 373A sequencer.

RESULTS

cDNA Synthesis Is Restricted by Biased dNTP Concentrations. DNA corresponding to the V1 and V2 hypervariable regions of the HIV-1 gp120 sequence was cloned downstream of the T7 DNA-dependent RNA polymerase promoter and used as a substrate for *in vitro* transcription. DNase-treated RNA was used in a 3-hr *in vitro* HIV-1 RTase reaction using a DNA oligonucleotide as primer and biased dNTP substrate concentrations. The ratios of RNA to HIV-1 RTase were comparable to those within the replication complex, with a 30-fold molar excess of RTase (14). Biased pyrimidine concentrations of the order of 10 nM–1 μM dCTP and 440 μM dTTP were used; the concentrations of the remaining dNTPs were comparable to free intracellular dNTP concentrations (i.e., 10 μM dCTP, 20 μM dGTP, 44 μM dTTP, and 40 μM dATP) for T-lymphoblastoid cell lines (15). To recover sufficient material for subsequent cloning, cDNA was amplified in 12 cycles of PCR.

As can be seen from Fig. 1, the efficiency of cDNA synthesis decreased as the bias in the pyrimidine (lanes 2–5) or purine (lanes 6–8) concentrations increased. Very low dNTP concentrations also reduced DNA synthesis; thus there was less product when the concentration of two dNTPs was each 30 μM (lane 9 with respect to lane 4 or lane 7) and even less when the concentration of the four was 30 μM each (lane 10). The yield of cDNA was 3- to 5-fold more when equimolar dNTPs at 200 μM were used as opposed to biased intracellular dNTP concentrations (15), indicating that *in vivo* reverse transcription was occurring under suboptimal conditions with respect to the dNTP substrates (data not shown and refs. 16–18).

G → A Hypermutation Correlates with Low dCTP Substrate Concentration. All PCR-amplified cDNAs in Fig. 1 were cloned into pBluescript SK (+) and 334 recombinants were sequenced, some of which are shown here (Fig. 2 and Table 1). Fifteen G → A hypermutated clones derived from cDNA synthesis using 10 nM dCTP and 440 μM dTTP are shown in Fig. 2*A*. The average number of G → A substitutions per clone was 4.6, with a range of 1–12 (Table 1). The frequency of G → A substitutions, $f_{G \to A}$, when normalized to the number of G residues in the target sequence, was 1.2×10^{-1}—i.e., on average every eighth G residue was substituted by A. $f_{G \to A}$ was inversely proportional to the logarithm of the dCTP concentration, the relationship being $f_{G \to A} = 0.17 - 0.06\log[\text{dCTP}]$, R = 0.97, for the four data points (10 nM, 30 nM, 100 nM, and 1 μM dCTP). There was no significant preference for G → A substitutions with respect to any dinucleotide context—i.e., GpN or NpG in any one sample. While there appeared to be a few hot and cold spots, there was insufficient data to define the underlying sequence motifs. Concomitant with increasing RTase infidelity was the proportion of deletions and insertions accompanying cDNA synthesis (Table 1).

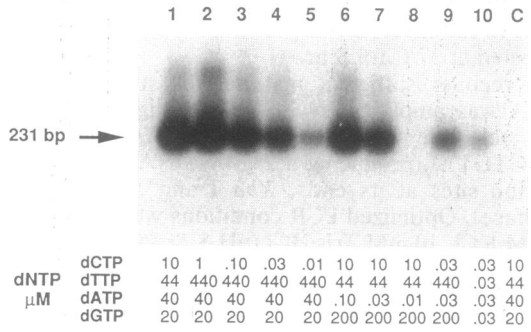

FIG. 1. Southern blot analysis of the PCR-amplified HIV-1 RTase reactions. Lane 1, RTase reaction performed with dNTP concentrations comparable to the free intracellular dNTP concentrations in T-lymphoblast cell lines (10)—i.e., 10 μM dCTP, 44 μM dTTP, 40 μM dATP, and 20 μM dGTP. Lane C, PCR control without HIV-1 RTase showing that amplification of residual DNA after DNase treatment of the input RNA was negligible.

The vast majority of substitutions were those expected—i.e., G → A—given the highly biased [dCTP]/[dTTP] ratio. However, a small proportion (8.6% of the total) of other substitutions was also noted. These are unlikely to be due to *Taq* polymerase as the error rate on this locus is low, having been estimated to be of the order of 1/3000 bases sequenced after 30 cycles (19). Indeed, the number of PCR cycles, 12, was deliberately kept low in order to limit *Taq* error.

U → C Hypermutation Correlates with Low dATP Substrate Concentration. The multiple G → A transitions occurred via rG·dT mismatches, which are thermodynamically the most stable of all base mismatches (20, 21). The penchant of the HIV-1 RTase to efficiently elongate beyond G·T mismatches has already been noted (22–25). Consequently, U → C hypermutation, via rU·dG mismatches, might result from a bias in the [dATP]/[dGTP] ratio. As can be seen, cDNA synthesis was less efficient under biased purine concentrations (Fig. 1, lanes 6–8). Subsequent sequencing of recombinants derived from these cDNAs proved that U → C hypermutation was indeed possible (Fig. 2*B*). The frequency, $f_{U \to C}$, of base substitution was inversely proportional to the dATP concentration (Table 1). Mixed hypermutants could be derived from cDNA synthesis under both biased pyrimidine and purine concentrations (i.e., 30 nM dCTP, 30 nM dATP, 440 μM dTTP, and 200 μM dGTP, Fig. 2*C* and Table 1). G → A were more frequent than U → C substitutions, although the frequency of both was approximately half the value determined for cDNA synthesis in the presence of either a pyrimidine or purine bias. Under low, yet equimolar, dNTP substrate concentrations, cDNA synthesis, albeit inefficient (Fig. 1, lane 10), was not conducive to hypermutation (Table 1), emphasizing the importance of the bias in pyrimidine and purine concentrations.

DISCUSSION

These *in vitro* data clearly demonstrate that G → A hypermutation occurs during cDNA synthesis in the presence of biased or unbalanced dNTP precursor concentrations. The data are consistent with the hypothesis that G → A hypermutation *in vivo* results from local fluctuations in the intracellular dCTP concentration (6, 10). RTase is capable of both RNA-dependent and DNA-dependent DNA synthesis. G → A hypermutation is also likely to occur during DNA-dependent DNA synthesis for two reasons. (*i*) In the polymerization site of the HIV-1 RTase, double-stranded DNA, rather than being in the B form, is in the A conformation typical of RNA·DNA hybrids, thus attenuating any local stereochemical differences between RNA·DNA and DNA·DNA templates/primer structures during reverse transcription (26). (*ii*) In the original study (6) of HIV-1 hypermutants a single case of a C → T hypermutated genome (i.e., with respect to the viral plus strand) was noted and postulated

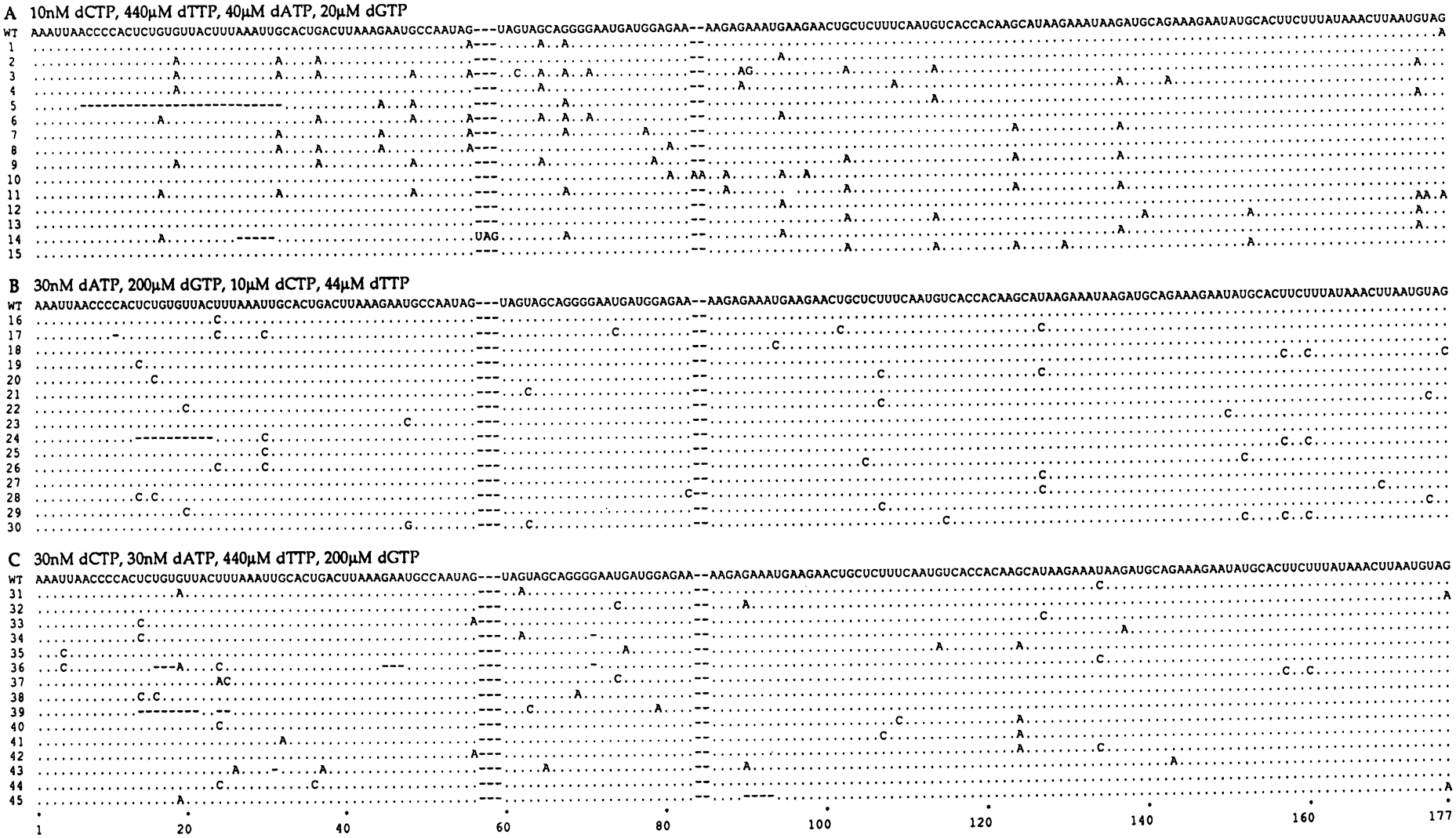

FIG. 2. Spectrum of 45 hypermutants resulting from cDNA synthesis with HIV-1 RTase and highly biased dNTP substrate conditions. The DNA sequences are aligned to that of the input (WT) sequence. Only sequences differences are noted, a dot indicating sequence identity. Hyphens indicate gaps introduced to maximize sequence homology. Sequences 1–15 were derived from reverse transcription using 10 nM dCTP, 440 μM dTTP, 40 μM dATP, and 20 μM dGTP (Fig. 1, lane 5); sequences 16–30 were from the reaction using 10 μM dCTP, 44 μM dTTP, 30 nM dATP, and 200 μM dGTP (Fig. 1, lane 7); sequences 31–45 were from the reaction using 30 nM dCTP, 440 μM dTTP, 30 nM dATP, and 200 μM dGTP (Fig. 1, lane 9).

Table 1. Biased dNTP concentrations give rise to hypermutated cDNA sequences

dNTP, μM				Plaques, no. analyzed			G → A subs	$f_{G \to A}$*	Deletions or insertions	Other point mutations
C	T	A	G	Total	WT	Mutated				
10	44	40	20	47†	46	1	0	0	1 0	0†
1	440	40	20	90†	78	12	12	4.1×10^{-3}	0 0	1†
0.1	440	40	20	31	10	21	38	3.8×10^{-2}	3 0	10
0.03	440	40	20	27	2	25	70	8.1×10^{-2}	3 4	4
0.01	440	40	20	38	6	32	146	1.2×10^{-1}	3 2	10
							U → C	$f_{U \to C}$*		
10	44	0.1	200	18	16	2	2	2.5×10^{-3}	0 0	0
10	44	0.03	200	33	6	27	58	4.0×10^{-2}	3 0	8
							G → A U → C	$f_{G \to A}$ $f_{U \to C}$		
0.03	440	0.03	200	27	3	24	40 24	4.6×10^{-2} 2.0×10^{-2}	9 1	9
0.03	0.03	0.03	0.03	22	19	3	1 0	1.5×10^{-3}	0 0	2

WT, wild type; subs, substitutions.

*Frequency (f) of G → A or U → C substitutions was calculated as follows: number of G → A substitutions/(32 Gs in target sequence × number of clones analyzed) or number of U → C substitutions/(44 Ts in the target sequence × number of clones analyzed).

†Hypermutants were identified by A-tracking only. Consequently, only G → A and A → non-A transitions could be observed. As the majority of HIV-1 RTase substitutions are purine–purine transitions, the number of other substitutions will always represent a minority. All other recombinants were sequenced. As can be seen from lane 8 in Fig. 1, cDNA synthesis using 10 nM dATP was highly inefficient. Only a single hypermutant encoding 6 U → C transitions was identified.

to represent an example of G → A hypermutation during DNA plus strand synthesis.

The degree of infidelity noted here, $f \approx 0.1$, was surprisingly elevated given *in vitro* data where base misincorporation frequencies of the order of 10^{-4} per site have been reported (22–25). While it is difficult to compare these very different systems employing different buffers, these latter studies invariably used [template/primer hybrid] >> [RTase] such that the K_m and V_{max} parameters could be determined. The ratios of template to RTase used here are comparable to those within the retroviral replication complex (14)—i.e., ≈30-fold molar excess of RTase. It is possible that the combination of excess RTase and long elongation times results in extensive recycling between the template/nascent strand hybrid and RTase, allowing elongation beyond the multiple rG·dT or rU·dC mismatches.

One may ask why U → C hypermutation of HIV genomes has not yet been reported *in vivo*. Low [dCTP]/high [dTTP] exaggerates an existing bias, whereas low [dATP]/high [dGTP] is the converse of the natural steady-state concentrations. It could simply be that a biased and inverted [dATP]/[dGTP] ratio is a more infrequent event than a highly biased [dCTP]/[dTTP] ratio. Alternatively, very low dATP concentrations (<100 nM) conducive to U → C hypermutation might be incompatible with host cell survival.

By exploiting the ability of HIV-1 RTase to elongate beyond G·T(U) mismatches it is possible to generate all six forms of hypermutants. If the target sequence was cloned in a pBluescript vector and RNA made separately from both strands using the T3 and T7 RNA polymerases, reverse transcription of RNA in the presence of biased pyrimidine or purine concentrations would allow hypermutation of both strands: G → A and U → C hypermutants from the sense strand and apparent C → U and A → G hypermutants being derived from the complementary strand. Thus all four types of monotonous hypermutants (i.e., G → A, A → G, U → C, and C → U), as well as two mixed forms (i.e., G → A plus U → C; C → U plus A → G), could be generated from any sequence. Rapid quenching of the reaction with excess dNTPs resulting in hypermutagenesis of the 5′ or 3′ end of a template could be envisioned, as could sequential cycles of the same, or even different, types of hypermutation.

There is a substantial body of data demonstrating that DNA polymerization in the presence of unbalanced dNTP concentrations results in mutation (27–32). Hypermutagenesis extends these findings and allows the construction of libraries of immense and variable complexity with up to 25–30% of some bases substituted. The mutants at the upper end of the spectrum represent jumps through sequence space of unknown potential. By combining hypermutagenesis with bacterial, yeast-based screening systems, it should be possible to select for novel phenotypes, such as altered substrate dependence, drug resistance, or improved thermal stability. A glance at *Escherichia coli* genetics suggests an embarrassing wealth of opportunities. Together they allow exploration of sequence space to a far greater degree than hitherto possible.

These findings may have some implications for evolution. It is generally agreed that RNA genomes preceded DNA genomes (33, 34). In the early genesis of dNTP metabolism, pool imbalances might have been commonplace. Reverse transcription under such conditions could have provided a rich source of genetic diversity among which a few sequences might have survived the rigors of selection.

M.A.M. was supported by a European Union Fellowship. This work was supported by grants from the Institut Pasteur, l'Agence Nationale de Recherches sur le SIDA (ANRS), and the European Union.

1. Drake, J. W. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4171–4175.
2. Holland, J. J., de la Torre, J. C. & Steinhauer, D. A. (1992) *Curr. Top. Microbiol. Immunol.* **176**, 1–20.

3. O'Hara, P. J., Nichol, S. T., Horodoyski, F. M. & Holland, J. J. (1984) *Cell* **36,** 915–924.
4. Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V. & Billeter, M. A. (1988) *Cell* **55,** 255–265.
5. Pathak, V. K. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6024–6028.
6. Vartanian, J. P., Meyerhans, A., Åsjö, B. & Wain-Hobson, S. (1991) *J. Virol.* **65,** 1779–1788.
7. Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. (1991) *Virology* **185,** 217–228.
8. Gao, F., Yue, F., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M. & Hahn, B. H. (1992) *Nature (London)* **359,** 495–499.
9. Perry, S., Flaherty, T., Kelly, M. T., Clabough, M. J., Tronick, D. L., Coggins, S. R., Whetter, L., Lengel, C. R. & Fuller, F. (1992) *J. Virol.* **66,** 4085–4097.
10. Vartanian, J. P., Meyerhans, A., Sala, M. & Wain-Hobson, S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3092–3096.
11. Bass, B. L., Weintraub, H., Cattaneo, R. & Billeter, M. A. (1989) *Cell* **56,** 331.
12. Lehtovaara, P. M., Koivula, A. K., Bamford, J. & Knowles, J. K. C. (1988) *Protein Eng.* **2,** 63–68.
13. Pjura, P., Matsumura, M., Baase, W. A. & Matthews, B. W. (1993) *Protein Sci.* **2,** 2217–2225.
14. Layne, S. P., Merges, M. J., Dembo, M., Spouge, J. L., Conley, S. R., Moore, J. P., Raina, J. L., Renz, H., Gelderblom, H. R. & Nara, P. L. (1992) *Virology* **189,** 695–714.
15. Dahbo, Y. & Eriksson, S. (1985) *Eur. J. Biochem.* **150,** 429–434.
16. Meyerhans, A., Vartanian, J. P., Hultgren, C., Plikat, U., Karlsson, A., Wang, L., Eriksson, S. & Wain-Hobson, S. (1994) *J. Virol.* **68,** 535–540.
17. Gao, W.-Y., Cara, A., Gallo, R. C. & Lori, F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8925–8928.
18. O'Brien, W. A., Namazi, A., Kalhor, H., Mao, S.-H., Zack, J. A. & Chen, I. S. Y. (1994) *J. Virol.* **68,** 1258–1263.
19. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. & Wain-Hobson, S. (1989) *J. AIDS* **2,** 344–352.
20. Ikuta, S., Takagi, K., Wallace, R. B. & Itakura, K. (1987) *Nucleic Acids Res.* **15,** 797–811.
21. Modrich, P. (1987) *Annu. Rev. Biochem.* **56,** 435–466.
22. Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265,** 2338–2346.
23. Bebenek, K., Roberts, J. D. & Kunkel, T. A. (1992) *J. Biol. Chem.* **267,** 3589–3596.
24. Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) *Science* **242,** 1168–1171.
25. Ricchetti, M. & Buc, H. (1990) *EMBO J.* **9,** 1583–1593.
26. Jacobo-Molina, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6320–6324.
27. Kunz, B. A. & Kohalmi, S. E. (1991) *Annu. Rev. Genet.* **25,** 339–359.
28. Mathews, C. K. & Ji, J. (1992) *BioEssays* **14,** 295–301.
29. Kunkel, T. A. (1992) *BioEssays* **14,** 303–308.
30. Meuth, M. (1989) *Exp. Cell Res.* **181,** 305–316.
31. Phear, G., Nalbantoglu, J. & Meuth, M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4450–4454.
32. Slabaugh, M. B., Howell, M. L., Wang, Y. & Mathews, C. K. (1991) *J. Virol.* **65,** 2290–2298.
33. Eigen, M. (1971) *Naturwissenschaften* **58,** 65–523.
34. Miller, S. L. & Orgel, L. E. (1974) *The Origins of Life on Earth* (Prentice Hall, Englewood Cliffs, NJ).