© 1996 Oxford University Press

Nucleic Acids Research, 1996, Vol. 24, No. 14 **2627–2631**

Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions

Jean-Pierre Vartanian, Michel Henry and Simon Wain-Hobson*

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

Received May 7, 1996; Revised and Accepted June 5, 1996

ABSTRACT

Very complex mutant libraries of the dihydrofolate reductase (DHFR) gene encoded by the *Escherichia coli* plasmid R67 were created using hypermutagenic PCR with biased deoxynucleotide triphosphate (dNTP) concentrations. Exploiting the particular stability of the G:T mismatch, the DHFR gene could be enriched in A+T by employing biased deoxypyrimidine triphosphate concentrations, i.e. [dTTP] > [dCTP]. A sizeable fraction of hypermutants were functional. A combination of [dTTP] > [dCTP] and [dGTP] > [dATP] biases generated mutations at unexpectedly low frequencies. This could be overcome by the addition of Mn^{2+} cations. Overall mutation frequencies of 10% per amplification (range 4–18% per clone) could be attained. All four transitions and a smaller number of transversions were produced throughout the gene. PCR mutagenesis could be so extensive as to inactivate all amplified versions of the gene.

INTRODUCTION

Although the mutation rates of DNA based organisms vary, they are considerably less than one per genome per cycle. Those of the RNA viruses may approach two to four substitutions per genome per cycle (1). Such rates must represent the upper end of the spectrum compatible with viability as they may be only slightly increased by chemical mutagenesis (2). Higher mutation rates almost certainly result in extinction. However, apart from this obvious restriction there is nothing *per se* to prohibit higher mutation rates *in vitro* or hypermutation restricted to small regions of a genome or gene segment *in vivo* (3). Perhaps the most startling example of this is retroviral G→A hypermutation where hundreds of templated Gs may be copied into As (4–6). This is a particular trait of the lentiviral family of retroviruses, which includes human immunodeficiency virus (HIV), and results from cDNA synthesis in the presence of highly biased [dTTP]/[dCTP] ratios (6).

G→A hypermutation can be reproduced *in vitro* using RNA, biased dNTP concentrations and preferentially the HIV-1 reverse transcriptase (7–9). Referred to as RNA hypermutagenesis, this method delivers elevated mutation and mutant frequencies, ≤0.1 per G per cycle and >0.9 per DHFR gene per cycle respectively. The complexity of the resulting libraries of hypermutated sequences was limited by the monotony of G→A hypermutation. Despite this, iterative hypermutagenesis of a bacterial antibiotic resistance gene, the *Escherichia coli* R67 DHFR, resulted in substitution of up to 23% of amino acids without loss of phenotype (10).

Genes and genomes exhibit G+C- or A+T-rich segments so that it would be useful to have a method capable of enriching any sequence in either. Just as dNTP biases are mutagenic for reverse transcription (7,11) so they are for PCR (12–16), although the magnitude of the bias has to be less to allow reasonably efficient amplification. PCR has the advantage that both strands may be mutated. A [dTTP] > [dCTP] bias would allow enrichment in A and T while a [dGTP] > [dATP] bias would permit the converse. These biases generate $G_{t(template)}$:T and T_t:G mismatches respectively which are the most stable of the 12 possible (17). By combining both deoxypyrimidine and deoxypurine triphosphate biases, it is shown here that PCR can be hypermutagenic to an unprecedented degree.

MATERIALS AND METHODS

The oligonucleotides used for amplification of the R67 DHFR gene have been described (10). PCR reactions were carried out in the following reaction mixture: 10 mM Tris–HCl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 100 pmol of each primer and 5 U *Taq* polymerase (Roche). The dNTP concentrations are described in the tables and legends. Input was ~5 ng plasmid DNA. The cycling parameters were: 50× (95°C, 30 s; 60°C 30 s; 72°C 10 min). Long elongation times were used to favour elongation after mismatches. Vent (Biolabs) and rTth (Roche) DNA polymerases were used at 2 and 2.5 U per reaction. $MnCl_2$ and dNTPs were purchased from Sigma and Pharmacia. PCR products were cloned via *Sac*I and *Bam*HI restriction sites and individual colonies picked, grown up and sequenced as described (10). A few products were cloned into the *Sac*I and *Bam*HI site of M13mp18 RF DNA. Recombinants were sequenced using thermosequenase (USB Amersham).

Unlike the *E.coli* chromosomal counterpart, the R67 DHFR gene is resistant to trimethoprim ($trim^R$). As the pTrc99A (Stratagene) cloning vector confers resistance to ampicillin ($ampi^R$) the ratio of the number of colonies on trimethoprim plus ampicillin and ampicillin only plates yields the proportion of functional genes post-PCR. The plating efficiencies of wild-type

* To whom correspondence should be addressed

DHFR construct on trimethoprim and ampicillin plates were comparable. Greater than 90% of ampi^R colonies had DHFR inserts.

RESULTS

DNA hypermutagenesis

Given the modified amplification protocol PCR conditions were first optimized for primer, magnesium, *Taq* DNA polymerase concentrations and number of cycles. As usual there was a strong Mg^{2+} dependence for all the thermostable polymerases used, the 2.5–5 mM range proving satisfactory. Particularly with large dNTP biases 30 cycles of PCR yielded relatively little product. Fifty cycles allowed adequate recovery for all but one reaction involving a 1000-fold [dTTP]/[dCTP] bias. In this case a further 25 cycles with equimolar dNTPs were performed as a chase. The efficiency of a standard amplification with equimolar 50 µM dNTPs was not affected by the addition of 1 mM ATP indicating that any increase in the ionic strength resulting from the addition of millimolar triphosphate did not alter PCR yields (data not shown).

Table 1 gives viable mutant frequencies following DNA hypermutagenesis with increasing [dTTP] > [dCTP] biases. The inverse relationship between the proportion of trim^R colonies as a function of the total (i.e. ampi^R) with increasing bias reflects the extent of DNA hypermutation. The overall mutation frequency for the entire amplification was inversely proportional to the dNTP bias and attained values as high as 2.9×10^{-2} substitutions per base per reaction for the ampi^R clones (Table 2). A collection of hypermutated trim^R sequences is given in Figure 1A. Up to five amino acid substitutions per functional clone (6.5%) were obtained which were generally well distributed throughout the sequence. Among the most hypermutated ampi^R clones up to 15 (6.5%) nucleotides and 11 (14%) amino acids respectively were replaced (not shown). The vast majority of substitutions were GC→AT transitions, as predicted from G_t:T mispairing on both strands due to the [dTTP] > [dCTP] bias. A small proportion (6%) of transversions were noted, uniquely A→T and T→A, to be expected from what is known about the ability of *Taq* DNA polymerase to elongate after mismatches (18,19).

Use of the thermostable rTth DNA polymerase, with reverse transcriptase activity, upon PCR with biased dNTP pools did not differ significantly from *Taq* DNA polymerase as judged by the ratio of trim^R and ampi^R clones and was not pursued. As a control DNA hypermutagenesis was performed using the Vent DNA polymerase which has a 3′-exonuclease activity. As evidenced by sequencing 20 trim^R clones, Vent completely protected DNA amplification against base misincorporation with a 500-fold [dTTP]/[dCTP] bias (data not shown).

Table 1. Inverse relationship between functional R67 DHFR mutants and dNTP pool biases

dNTP/µM C	T	A	G	R67 DHFR trim^R/ampi^R	
1000	1000	50	50	1100/1160	95%
10	1000	50	50	189/463	41%
3	1000	50	50	221/1373	16%
1	1000	50	50	42/420	10%

The [dTTP] > [dCTP] bias would favour GC→AT transitions. Plating of cloned PCR products on ampicillin (ampi^R) yields the total number of recombinants, while plating on trimethoprim and ampicillin (trim^R) yields the number of functional recombinants.

Table 2. DNA hypermutation frequencies

dNTP/µM C	T	A	G	Mn^{2+} mM	Colonies sequenced	No. mut.[a]	Ti/Tv[b]	N→A,T[c]	N→G,C[d]	Mutation frequency[e]
1	1000	50	50	–	37 trim^R	126	125/1	121	5	1.5×10^{-2}
1	1000	50	50	–	24 ampi^R	162	157/5	157	5	2.9×10^{-2}
3	1000	50	50	–	20 trim^R	24	22/2	24	0	5.2×10^{-3}
3	1000	50	50	–	20 ampi^R	29	27/2	28	1	6.3×10^{-3}
10	1000	50	50	–	20 trim^R	15	12/3	12	3	3.2×10^{-3}
10	1000	50	50	–	22 ampi^R	21	16/5	19	2	4.1×10^{-3}
5	1000	5	1000	–	18 trim^R	3	3/0	3	0	7.2×10^{-4}
5	1000	5	1000	–	18 ampi^R	0	0/0	0	0	$\leq 10^{-4}$
30	1000	30	1000	–	18 ampi^R	4	4/0	1	3	10^{-3}
30	1000	30	1000	0.5	34 ampi^R	755	521/234	256	499	10^{-1}
100	1000	100	1000	0.5	18 trim^R	19	16/3	8	11	4.5×10^{-3}
100	1000	100	1000	0.5	24ampiR	41	33/8	11	30	7.4×10^{-3}

On the left are the experimental conditions in terms of dNTP and manganese cation concentrations. On the right the analysis of mutants.

[a]Total number of mutations scored.

[b]Ti/Tv, number of transitions/number of transversions.

[c]Number of substitutions from non-A→A and non-T→T combined.

[d]Number of substitutions from non-G→G and non-C→C combined.

[e]The average mutation frequency was calculated as the number of substitutions divided by product of the number of clones sequenced and the number of bases between the PCR primers (231 bp).

A

```
     1        10        20        30        40        50        60        70      78    no. subs.
                                                                                         aa   na
R67  MERSSNEVSNPVAGNFVFPSDATFGMGDRVRKKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALERIN
  1  ......................M.......................Y.........................T.....      3    3
  2  ................I.............................................S........V......      3    4
  3  ..........L.....I.......................................I.....................      3    4
  4  ......K.............N...........N..V..........................................      4    4
  5  ..Q................L....SI....................................................      4    6
  6  ........N.............M............V..........Y...............................      4    8
  7  ...........I.......LN...............T.......................................C.      5    5
  8  ..........LI..................................Y.............T.S...............      5    8
```

B

```
     1        10        20        30        40        50        60        70      78    no. subs.
                                                                                         aa   na
R67  MERSSNEVSNPVAGNFVFPSDATFGMGDRVRKKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALERIN
  1  ........R......................E.....................R......T....E..........F.      6   10
  2  ......Q................S.....................C........H....V....P.*.........V.      8   11
  3  ........N.....D.....................................-....G......P.W.*....P.CD.      9   18
  4  ....C..A..........S.............................A...GDC..........A...........D      9   16
  5  .............S.SI...G..S..........P...........Y.....----..PG......R...........     10   15
  6  .........D........L..........G........V..A......*...GR...............RI.....T.     11   16
  7  ..Q............S.....V.S.............R....T...R...A..........L.L.............I     11   16
  8  .G..G........T..........I.......................G....T..I........ER..F......V.     11   18
  9  ...............LAS.F.....T.................A..Y.....QA...................P..T.     11   18
 10  ........T.......IL.....-.K..P...E.....R...D.....S..-............P...........T.     11   21
 11  ..Q.G....K....S......T..............................G......G.........L...RG.TD     12   14
 12  .....D.A........A................P....L...............CT.....R....R...AV....T.     12   19
 13  ....R..A.S.AT....Y..V..L...............R..........................R...GT......     12   19
 14  ...............P.S....AS.....L...P.........ER....S.......V.........T...T..G...     13   18
 15  ....N......A.S....L...................Y...I...R...A......K......P..V..D...G.T.     14   21
 16  ....N.G.......SS............C........R....I....AI....T..........L..V..I...G..G     15   18
 17  ..............I.....G.-L....C..E.........V.....A........I..G.....ALV.....P..VS     15   20
 18  .....SG..D...D...I.......L.....RR........VI.....I...V.HV..........R.....T.....     16   20
 19  ...I....ND....SS-...G.V..TEN...L...T......A.....K..-......................G.V.     16   21
 20  .K.......S..........G..........R...........R.*Y.S....D..AK...L.P..R......P...D     16   23
 21  .......ACD.A..Y.........................R....*..Q...G...LK...L...P..S......GV.     16   25
 22  .D.I...AG..............P..E..A........*.*........S.-...T...G.....IRV..A......H     17   20
 23  ...N.Y..T......-....G.A.......G.......-C....R..I..A...F...P....CP.*.......D.T.     17   22
 24  ...G...AG.....D.......-....G...R....T..D......R.S..S.C...G.G.R..T...........V.     17   23
 25  ..*..S...K....K....L...L...G..........H.RT..RC...S..............P.L.......G..Y     17   23
 26  .Q............S..L.P...S........E....R....FR.....S.....T.........PT..ITV.D....     17   26
 27  ...........A..C..............AH.E..S.........-.A..L.GS.......R..P.*A.....P..VD     17   28
 28  ......GDC.SA...L.S....V.S....A.R............G.......G........R.....V.....PK..D     18   30
 29  ...G.D..G.S...S.........D...C....Y..T...R.....G..*A.......P..R..P.....A..P...Y     19   25
 30  .....K.....A..E.A..T..AL....H..G.TD........R*.R.........A......S..R..S......V.     19   29
 31  ..G.....G....S...L...TAL..EV.A.R.....R*....-..R.....K...A.LG...............C.D     20   29
 32  ...R.S..G..........L.....ERG...ER..T.R..R...R.........H.A....R....L.C-.R....FD     21   38
 33  .R.TGG...D.L.......PV..PA....A....D..HSR.........S.....T.....R....H.....V...TD     22   34
 34  .G....GA.I..-VS..L..G..L.ARV.A.GNT.....RT...R.R...SS..F.....T.....*..SD..P..VS     30   41
```

C

```
     1        10        20        30        40        50        60        70      78
R67  MERSSNEVSNPVAGNFVFPSDATFGMGDRVRKKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALERIN

      Q*TNIKAGDSIVDDLILLLNVMLELENHMHERPRTV**D*NIE**YI**AS*R*TIKPGTYLDPI*V*LIVVPKCTI
      GQIRSGRTSLATVSSAI PVIASIIRGPACRNTNVTSHSRTDRRHRAISI KTHVLGFVVLSSLAPTCSFTTRGHVD
      KGNGDQICIKL SYI S FGTVILK ECGGMEYDS RYR VSAACG DQS GAFIAVL  R NRERFRFGRIFDGFY
      D R C  DR    KP Y N P PST V L L  S  CRC F G   K L  VSC        CI LA RD    YCG
      R   Y  N     I    T    AA     G  P   L  A     G    QD          T H   A     YS
          K        E         DE            V        S     C            W         DH
                   C
```

Figure 1. Collections of R67 DHFR hypermutants. Amino acid sequences were aligned to the reference R67 sequence using the one letter code, only differences being noted (10). A dot indicates sequence identity, a hyphen (-) represents codons harbouring nucleotide deletions. An asterisk (*) defines an in phase stop codon. To the left is the clone designation, to the right the number of amino acid (aa) and nucleic acid (na) differences with respect to the wild-type DHFR sequence. (**A**) A selection of trim^R hypermutants derived from the single bias [dTTP] = 1 mM, [dCTP] = 1 µM reaction (Tables 1 and 2). (**B**) Thirty-four clones derived from hypermutation involving two dNTP biases ([dTTP] = [dGTP] = 1 mM, [dCTP] = [dATP] = 30 µM) with 0.5 mM manganese cations (Table 2). All clones were trim^S. Among all data sets no two sequences were identical. Clone 33 and 34 encoded single nucleotide insertions at codons 13 and 31 (not shown). (**C**) Summary of all amino acid substitutions from all data sets. On average 3.7 substitutions per residue (range 1–7) were identified from a trivial number of clones. The single letter amino acid code is: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Double dNTP biases and manganese ions

The particularities of the G_t:T mismatch ensured A+T enrichment of the R67 DHFR gene. Alternatively a [dGTP] > [dATP] bias would have generated T_t:G mismatches with resulting G+C enrichment. Yet if all four base transitions could be generated during a single reaction the resulting mutant libraries would be among the most complex possible accessing an even greater proportion of sequence space. This is in principle possible if both a [dTTP] > [dCTP] and [dGTP] > [dATP] bias were used during PCR. However, no product whatsoever was obtained with a 1000- or 300-fold biases in both ratios. Only with <200-fold biases was

→	A	T	C	G
A	-	0.21	0.05	0.98
T	0.16	-	1	0.06
C	0.02	0.28	-	0.06
G	0.31	0.04	0.08	-

Figure 2. Base substitution matrix for the 755 mutation data set normalized to the most frequent substitution, T→C. Substitution frequencies (e.g. G→A/755) given with respect to the (+) DNA strand were normalized to its base composition (52 T, 60 C, 65 G and 54 A excluding the ATG initiator codon derived from the forward PCR primer).

this possible. Sequencing of the trim^R hypermutated products yielded unexpectedly low mutation frequencies (Table 2).

Transition metal ions such as manganese (Mn^{2+}) and cobalt (Co^{2+}) may decrease the fidelity of DNA synthesis including PCR (12,20,21). Addition of $MnCl_2$ to a final concentration of 0.5 mM in a reaction with both [dTTP]/[dCTP] = [dGTP]/[dATP] = 1000 µM/30 µM overcame the enhanced fidelity noted above. The overall base mutation frequency could be increased from ~10^{-3} to ~10^{-1} per site per amplification (Table 2). In fact, the PCR was so error prone that no trim^R colonies (0 trim^R/600 ampi^R) were identified. A collection of 34 clones is given in Figure 1B, mutants starting with a minimum of 10 substitutions (4%) per clone. The maximum number was 41 (18%) per clone. The proportion of transversions (31%) was greatly enhanced by the addition of Mn^{2+} and was accompanied by a few deletions and even fewer single base insertions (Fig. 1B). There was no correlation between the proportion of synonymous (s) to non-synonymous (ns) base substitutions within this or any other data sets (not shown).

Figure 1C collates amino acid replacements from all the data sets and indicates that hypermutagenic PCR may introduce between one and seven (mean 3.7) different amino acids per residue. The large 755 mutation data set resulting from manganese mutagenesis was analyzed for substitution biases. The mutation matrix, normalized for base composition effects, showed almost perfect strand symmetry (i.e. G→C ≈ C→G, etc.) (Fig. 2). However, there was a bias for AT→GC transitions which perhaps may be attributable to subtle differences between G_t:T and T_t:G mismatches in the *Taq* DNA polymerization site. Once again, A→T and T→A were the most frequent transversions.

The proportions of adjacent double and triple substitutions were as expected given the mutation frequency of isolated changes. The distribution of substitutions per site differed significantly from that expected from a binomial distribution (not shown). A χ^2 analysis of the dinucleotide context showed that there were a few substitution preferences, notably an excess of AT→GC transitions as well as A→T and T→A transversions in the context of CpA/TpG, and a dearth of the same transitions in the TpA dinucleotide. Only a one significant bias (GC→AT in GpC) was seen in reactions with [dTTP] > [dCTP] biases.

DISCUSSION

Balanced DNA precursor concentrations are clearly crucial to the fidelity of cellular DNA or retroviral cDNA synthesis *in vivo* and *in vitro* (22–25). The same is true of PCR, the present findings reproducing and extending earlier work (12,13,16). The nature of the dNTP bias generally produced the substitution expected from G:T mispairing once again highlighting the importance of this most stable of base mismatches to hypermutation (7,8). Perhaps surprisingly, the fidelity of amplification was enhanced many fold when both deoxypyrimidine and deoxypurine triphosphate biases were used (Table 2). This might result from the fact that although G:T mismatches are being forced so were G:G and T:T mispairs. From what is known of *Taq* DNA polymerase elongation beyond mismatches, G:G represents one of the most substantial blocks to elongation and consequently amplification (18,19). By contrast T:T mismatches pose fewer problems. The addition of Mn^{2+} ions, known to be mutagenic for DNA synthesis by a variety of mechanisms including modification of the relative K_ms of mismatches and matches (20,21), overcame this problem. The 100-fold enhanced overall mutation frequency was indeed so great that no trim^R clones could be derived.

With a double dNTP bias and manganese ions there was an excess of transitions towards G+C which was not strand-specific (Fig. 2). Clearly this could be countered by increasing [dTTP] or decreasing [dGTP] in the reaction. There was evidence that the distribution of mutations was not completely random. However, significant deviations from the expected values were noted for only a few substitutions.

A comparison of RNA and DNA hypermutagenesis is telling (7,8). The HIV-1 reverse transcriptase error rate per pass is clearly greater than *Taq* DNA polymerase. Among the hundreds of RNA molecules hypermutated *in vitro* by the HIV-1 reverse transcriptase, up to 32% of G targets were substituted for one clone with a best mean of 11%, all in a single cycle of cDNA synthesis (7). However, given the monotony (e.g. G→A) of RNA hypermutagenesis these numbers translates into best and average overall mutation frequencies of ~7 and 3% respectively. To date, DNA hypermutagenesis has produced up to 18% base substitution per clone with a best mean of 10% involving copying of both strands.

Despite the intrinsic properties of the HIV-1 RT the advantages of DNA hypermutagenesis by PCR are manifold. First, the complexity of the mutant libraries are incomparably greater providing access to even larger fraction of sequence space. Secondly, the procedure is faster being reduced to a single reaction. Thirdly, as the PCR step is mutagenic there is in principal no need to clone before undertaking a second cycle of DNA hypermutagenesis. However, the power of DNA hypermutagenesis is now so great that iteration without some sort of phenotypic selection is probably unwise because the information threshold can be crossed. In addition, preliminary work suggests that primer dimers and deleted molecules may be preferentially amplified upon cycling without phenotypic selection or purification of the DNA band. The conditions can surely be refined to purge the present GC→AT bias.

The extent of mutation described above, as well as the complexity of the mutant libraries, exceeds that generated by any biological method to date. A recent paper described hypermutagenic PCR using modified dCTP and dGTP substrates (26). The best and average mutation frequencies described here

(0.18 and 0.1 per base per reaction) are highly comparable with those reported, notably 0.19 and 0.1 per base per reaction. The modified bases generally produced AT→GC transitions and a small percentage (<10%) of transversions. The present protocol used standard bases, generates at high frequencies all four transitions and, given the presence of manganese cations, approximately one third transversions. Clearly there is considerable flexibility and choice in the production of hypermutants which could be tailored to the desires or needs of the experimentalist.

Although DNA hypermutagenesis allows huge leaps through sequence space, viable hypermutants are to be had. The diversity currently accessible is so great that any screening procedure will explore only a minute fraction of the sequence space accessed. The simplicity and efficiency of DNA hypermutagenesis transfers the burden of work in protein evolution *in vitro* onto analytical procedures. The potential of the method is such that, after iterative DNA hypermutagenesis, the historical information content of a sequence might be annihilated, defying recognition.

The choice of the small DHFR gene was particularly propitious. PCR product yield decreases with dNTP pool bias and is further reduced upon addition of Mn^{2+} cations. This can be alleviated to some extent by a chase PCR with equimolar dNTPs. Alternatively cycling the product from an agarose gel purified band should allow one to extensively hypermutate larger genes. Yet as the probability of introducing deleterious mutations increases with target DNA length, inevitably hypermutation of such genes might not prove as informative, unless some form of biological selection is used. A further reservation concerns the nature of the transversions observed. That A→T and T→A transitions were the most common may be attributed to the ability of *Taq* DNA polymerase to elongate after T:T mismatches (18,19). Inversely, the dearth of a number transitions correlates well with the relative inefficiency of the enzyme to elongate after A_t:G, G_t:A, G_t:G and C_t:C mismatches. Thus the mutation spectrum is shaped to some extent by *Taq* DNA polymerase. It is possible that different thermostable enzymes might show subtle differences. Alternatively, modifications to the reaction mix might be introduced in an attempt to alleviate such preferences.

DNA hypermutation accelerates what may occur under more physiological circumstances over much longer time periods. Indeed there is a wealth of experimental data associating dNTP pool biases, mutation and cancer (22,23,25). The consequences of an intracellular [dTTP] > [dCTP] bias are particularly intriguing. Among eukaryotic cells the intracellular dNTP concentrations are invariably [dATP] ≥ [dTTP] > [dCTP] ≥ [dGTP] or, in other words, [dTTP] > [dCTP] and [dATP] > [dGTP] (25). Given the particular properties of the G:T mismatch any increase in the deoxypyrimidine triphosphate bias would help enrich the sequence in A+T. The potential mutagenic effects resulting from fluctuations in the deoxypurine triphosphate bias would have to be even more substantial as they would need to invert the natural [dATP] > [dGTP] bias (25). From this it might be surmised that any exacerbation of the natural [dTTP] > [dCTP] bias should have more long term impact on the genome. In this context it is interesting to note that among vertebrate cells non-coding segments are generally A+T rich.

It is salutary to realize that DNA synthesis can be so error prone. It might be supposed that during the evolution of primitive DNA based replicons and before highly integrated dNTP metabolism, biased dNTP concentrations alone, or in conjunction with dilute solutions of some transition metal ions, might have contributed to the genesis of DNA sequence diversity upon which natural selection could work.

ACKNOWLEDGEMENTS

We would like to thank Drs Fredj Tekaia and Christophe Terzian for statistical analyses. This work was supported by grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA.

REFERENCES

1 Drake, J. W. (1993) *Proc. Natl Acad. Sci. USA*, **90,** 4171–4175.
2 Holland, J. J., Domingo, E., de la Torre, J. C. and Steinhauer, D. A. (1990) *J. Virol.*, **64,** 3960–3962.
3 Rueda, P., Garcia-Barreno, B. and Melero, J. A. (1994) *Virology*, **198,** 653–662.
4 Pathak, V. K. and Temin, H. M. (1990) *Proc. Natl Acad. Sci. USA*, **87,** 6019–6023.
5 Vartanian, J. P., Meyerhans, A., Åsjö, B. and Wain-Hobson, S. (1991) *J. Virol.*, **65,** 1779–1788.
6 Vartanian, J. P., Meyerhans, A., Sala, M. and Wain-Hobson, S. (1994) *Proc. Natl Acad. Sci. USA*, **91,** 3092–3096.
7 Martinez, M. A., Vartanian, J. P. and Wain-Hobson, S. (1994) *Proc. Natl Acad. Sci. USA*, **91,** 11787–11791.
8 Martinez, M. A., Sala, M., Vartanian, J. P. and Wain-Hobson, S. (1995) *Nucleic Acids Res.*, **14,** 2573–2578.
9 Pezo, V., Martinez, M. A. and Wain-Hobson, S. (1996) *Nucleic Acids Res.*, **24,** 253–256.
10 Martinez, M. A., Pezo, V., Marliere, P. and Wain-Hobson, S. (1996) *EMBO J.*, **15,** 1203–1210.
11 Battula, N. and Loeb, L. A. (1974) *J. Biol. Chem.*, **249,** 4086–4093.
12 Leung, D., Chen, E. and Goeddel, D. (1989) *Technique*, **1,** 11–15.
13 Caldwell, R. C. and Joyce, G. F. (1992) *PCR Meth. Appl.*, **2,** 28–33.
14 Chen, K. and Arnold, F. H. (1993) *Proc. Natl Acad. Sci. USA*, **90,** 5618–5622.
15 Rellos, P. and Scopes, R. K. (1994) *Prot. Eng. Purif.*, **5,** 270–277.
16 Fromant, M., Blanquet, S. and Plateau, P. (1995) *Anal. Biochem.*, **224,** 347–353.
17 Modrich, P. (1987) *Annu. Rev. Biochem.*, **56,** 435–466.
18 Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. (1990) *Nucleic Acids Res.*, **18,** 999–1005.
19 Huang, M. M., Arnhein, N. and Goodman, M. F. (1992) *Nucleic Acids Res.*, **20,** 4567–4573.
20 Beckman, R. A., Mildvan, A. S. and Loeb, L. A. (1985) *Biochemistry*, **24,** 5810–5817.
21 Goodman, M. F., Keener, S., Guidotti, S. and Branscomb, E. W. (1983) *J. Biol. Chem.*, **258,** 3469–3475.
22 Phear, G., Nalbantoglu, J. and Meuth, M. (1987) *Proc. Natl Acad. Sci. USA*, **84,** 4450–4454.
23 Meuth, M. (1989) *Expt. Cell. Res.*, **181,** 305–316.
24 Kunz, B. and Kohalmi, S. (1991) *Annu. Rev. Genet.*, **25,** 339–359.
25 Mathews, C. K. and Ji, J. (1992) *BioEssays*, **14,** 295–301.
26 Zaccolo, M., Williams, D. M., Brown, D. M. and Gherardi, E. (1996) *J. Mol. Biol.*, **255,** 589–603.