ANALYTICAL BIOCHEMISTRY 224, 347–353 (1995)

Direct Random Mutagenesis of Gene-Sized DNA Fragments Using Polymerase Chain Reaction

Michel Fromant, Sylvain Blanquet, and Pierre Plateau
Laboratoire de Biochimie, URA 240 CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France

Received April 19, 1994

The polymerase chain reaction (PCR) can be used to amplify a DNA fragment with the concomitant creation of numerous mutations provided that one dNTP substrate is in excess over the three others. Advantage was taken of this behavior to systematically mutagenize a 291-bp-long DNA fragment and to define the rules relating the frequencies of each possible bp substitution to the set of the dNTP concentrations in the PCR experiment. Sets of parameters governing the rules were determined under various mutagenic conditions including the addition of $MnCl_2$. Finally, validity of the rules was assessed in several mutagenesis experiments showing that a wide range of substitution frequencies including AT → GC and GC → AT transitions as well as AT → TA transversions can be obtained at will. © 1995 Academic Press, Inc.

Although several techniques are available for site-directed mutagenesis of DNA (reviewed in (1)), methods to generate high numbers of random mutations on long DNA fragments have still to be designed. In theory, one efficient approach consists in synthesizing degenerate oligonucleotides by mixing nucleotides at each step. Accordingly, highly mutated synthetic oligonucleotide libraries can be prepared (2). Such a strategy is near from ideal in that it allows (i) to precisely define the randomly mutagenized window, (ii) to mutagenize each nucleotide with a given probability, and (iii) to select the bases introduced at each position. However, this method is limited to DNA fragments containing no more than ~100 residues. Alternative methods are based on the use of chemical mutagens like sodium bisulfite (3), nitrous acid, formic acid, hydrazine (4), or hydroxylamine (5–7). However, these methods do not yield highly mutagenized fragments (5,8) and cannot generate all the possible base substitutions. In turn, random mutations can be enzymatically introduced through misincorporation of nucleotides by polymerases (9–11). In such an approach, the rate of mutations may be increased (i) by using polymerase species lacking proofreading activity (12,13), (ii) by mixing unequal concentrations of the four deoxynucleoside triphosphates (dNTPs) (14–16), and/or (iii) by using PCR[1] to perform successive runs of polymerization and cumulate misincorporations (14,16,17).

In the current study, a procedure displaying several advantages of the above methods is developed. It is based on PCR amplification of DNA by *Taq* DNA polymerase using an excess of one dNTP nucleotide concentration over the others. The procedure was assayed using a 356-bp fragment from the 5'-noncoding region of the *Escherichia coli lysU* gene as template. The influence of various sets of dNTP nucleotide concentrations on the frequencies, the nature, and the locations of the misincorporations was examined through DNA sequencing. From this, rules relating overall mutation rates as well as relative frequencies of the various base pair substitutions to sets of dNTP concentrations could be drawn. The predictive value of these rules was verified experimentally.

MATERIALS AND METHODS

Enzymes and Substrates

Taq DNA polymerase and DNA modification enzymes were purchased from Boehringer (Manheim, Germany), dNTPs were from Pharmacia (Uppsala, Sweden), and [α-^{32}P]dATP (111 TBq/mmol) was from NEN-DuPont (Paris, France). Oligonucleotide primers were synthesized on a Pharmacia gene assembler.

Polymerase Chain Reaction Procedure

The template for PCR was obtained by digesting plasmid pXLysCla3 (18) by *Pvu*II. Amplification of a DNA fragment of 356 bp from this template was achieved by using the following two primers: 5'-AATGAATTCGAT-

[1] Abbreviation used: PCR, polymerase chain reaction.

0003-2697/95 $6.00

CATCGTCGTATTGGCCTTTGC-3' and 5'-GAGGGATCCTCCCGTGTTTCTTGTTCAGACAT-3'. The 5' termini of these primers included a *Bam*HI or an *Eco*RI recognition sequence, respectively.

Unless otherwise stated, the incubation mixture (25 μl) contained 10 mM Tris–HCl (pH 8.7 at 25°C), 50 mM KCl, 5 μg/ml bovine serum albumin, 0.5 μM of each of the two primers, 600 pM of template DNA, and 2 units of *Taq* DNA polymerase. The dNTP concentrations ranged between 0.1 and 10 mM. In some experiments, $MnCl_2$ was present at a final concentration of 0.5 mM. The $MgCl_2$ concentration in the assay was adjusted so that the concentration of added divalent cations ($MgCl_2$ + $MnCl_2$) exceeded the total concentration of dNTPs by either 0.7 mM (conditions of low $MgCl_2$) or 6 mM (conditions of high $MgCl_2$).

The reaction mixture was covered with 25 μl light mineral oil (from Sigma) and submitted to 16 cycles of amplification using the Pharmacia Gene ATAK controller: 91°C for 1 min (94°C for 5 min for the first cycle), 51°C for 1 min, and 72°C for 3 min (5 min for the last cycle). At the end of the reaction, the amplified DNA was precipitated with ethanol in the presence of 0.8 M ammonium acetate, in order to remove nucleotides. When indicated, the amount of amplified DNA was estimated by adding [α-^{32}P]dATP (4 GBq/μmol) to the PCR assay mixture. In this case, a 5-μl aliquot of the reaction mixture was withdrawn after the amplification process, precipitated with 5% (w/v) trichloracetic acid, filtered on glass fiber filters, and counted in a liquid scintillation counter (Beckman LS 3801). Another aliquot (6 μl) was used to verify the specificity of the amplification process through electrophoresis on a 5% polyacrylamide gel (300 V, 1.5 h, 15°C) in 50 mM Tris–borate buffer (pH 8.3) containing 1 mM EDTA. After the run, the gel was transferred to Whatman 3MM paper, dried, and autoradiographed.

Cloning and DNA Sequencing of the Amplified Fragments

The amplified DNA was digested by *Bam*HI and *Eco*RI endonucleases and ligated into M13mp19 phage DNA (19) previously digested by the same restriction enzymes. After transformation of *E. coli* strain JM-101TR {(Δ*lac-pro*) *supE thi-1 recA56 srl-300*::Tn*10* [F'(*traD36 proAB lacI*^{q} *lacZ*ΔM15)]} (20) with the ligation mixture, recombinant phages were selected on plates containing 24 μg/ml of 5-bromo-4-chloro-indolyl-β-D-galactopyranoside and 0.4 mM isopropyl-β-D-thiogalactopyranoside. Recombinant phages were purified and their DNA was sequenced by the dideoxynucleotide chain termination method (21), using 1 pmol of M13 universal primer labeled with fluorescein and ~1 μg of single-stranded DNA as template. Sequencing reaction mixtures were electrophoresed on a 6% polyacrylamide gel and analyzed on an A.L.F. DNA sequencer (Pharmacia).

RESULTS AND DISCUSSION

Principle of the Method and Effect of Divalent Cations

The possibility to introduce random-point mutations in a DNA fragment by having the concentration of one dNTP precursor in excess with respect to the concentrations of the others during a PCR amplification was systematically assessed by using a 356-bp DNA template derived from the *E. coli lysU* gene. To ensure a roughly constant amplification yield whatever the concentration of the forcing dNTP in the assay, a high concentration of DNA polymerase (2 U/25 μl) and a long extension time (3 min for a 391-bp amplification product) were chosen. For each dNTP condition assayed, amplified fragments were sequenced after introduction into M13 phages. Routinely, 16 to 40 phage clones were analyzed.

In a first set of experiments, a $MgCl_2$ concentration exceeding by 0.7 mM the sum of the dNTP concentrations was chosen. As shown in Table 1, the addition in excess of a forcing nucleotide strongly influenced the nature of misincorporations during the amplification process. For instance, a high concentration of dTTP mainly resulted in changes toward T or A (much likely through the misincorporation of T in one or the other DNA strand), while a high concentration of dGTP directed changes toward G or C. In all cases, the relative number of transitions was significantly higher than that of transversions. The highest substitution frequency was observed in the case of the AT → GC transition in the presence of high dGTP concentration. This result is in agreement with the already observed tendency of *Taq* DNA polymerase to favor this transition (10,13,22).

Since the manganese ion decreases the fidelity of *Taq* DNA polymerase (14), this metal ion was added at a concentration of 0.5 mM while keeping $MgCl_2$ concentration such that $[MgCl_2]_{total} + [MnCl_2]_{total} = [dNTP]_{total} + 0.7$ mM. Under this condition, the frequencies of all bp substitutions (Table 1), as well as the diversity of the generated mutations, markedly increased. For instance, the AT → TA transversion occurred now at nearly the same rate as the GC → AT transition. Notably, the presence of manganese did not significantly lower the yield of amplification of the 356-bp DNA template (Table 1). In truth, under such conditions, DNA fragments as long as 2000-bp could be efficiently amplified (data not shown).

Because a high concentration of $MgCl_2$ may also contribute to decrease the fidelity of *Taq* DNA polymerase (11), another set of amplifications was performed in the presence of 0.5 mM $MnCl_2$ plus $MgCl_2$, such that $[MgCl_2]_{total} + [MnCl_2]_{total} = [dNTP]_{total} + 6$ mM. Such a condition further increased the mutation rates (Table 1) without significantly changing the distribution of transversions versus transitions. However, under such a high

TABLE 1

Results of Random Mutagenesis by PCR Amplification Using *Taq* DNA Polymerase and One Forcing dNTP Concentration

Forcing dNTP (3.40 mM)	$MgCl_2$ (mM)	$MnCl_2$ (mM)	Number of DNA duplications (n)	Number of bases sequenced	Changes into A	Changes into C	Changes into G	Changes into T	Insertions	Deletions
dATP	4.7	0	8.7	11,640	1	0	1	1	0	0
dCTP	4.7	0	8.9	11,640	0	5	4	0	0	0
dGTP	4.7	0	8.1	11,349	0	15	29	1	0	0
dTTP	4.7	0	8.5	11,349	4	0	3	11	0	0
dATP	4.2	0.5	7.9	11,349	22	1	1	27	0	8
dCTP	4.2	0.5	9.0	11,640	1	21	24	2	0	3
dGTP	4.2	0.5	8.4	11,349	2	84	0	0	0	1
dTTP	4.2	0.5	8.7	11,640	20	1	6	34	0	0
dATP	9.5	0.5	7.0	6984	43	1	0	45	0	3
dCTP	9.5	0.5	8.0	6693	3	22	23	5	0	0
dGTP	9.5	0.5	7.8	6693	0	85	120	3	1	0
dTTP	9.5	0.5	8.1	4656	33	4	6	35	0	2

Note. The concentrations of the nonforcing nucleotides were equal to 0.2 mM. The template DNA was a 356-bp fragment from the *E. coli lysU* gene, added at an initial concentration of 600 pM. The amplified fragments were cloned into phage M13mp19. Sixteen to 40 recombinant clones were sequenced for each PCR condition. Mutations in the DNA regions covered by the two oligonucleotides primers were not taken into account. Consequently, the results only concern a 291-bp sequence containing 102 A, 44 C, 53 G, and 92 T in its wild-type composition. In this set of experiments, the amount of DNA produced by the amplification was estimated from identical experiments performed in parallel and involving [α-^{32}P]dATP. The absence of nonintended amplification products was assessed by polyacrylamide gel electrophoresis analysis.

$MgCl_2$ concentration, template fragments longer than ~400 bp could not be used because of a marked reduction of the yield of the amplification reaction (data not shown).

In all the above experiments, only a few mutations did not relate to the forcing dNTP used. Among them were some deletions and very few insertions (Table 1).

Theory

Assuming that the polymerase follows standard Michaelian kinetics, it can be established that the probability of incorporating a base X instead of a base Y is proportional to the ratio [dXTP]/[dYTP] between the concentrations of the triphosphate precursors in the experiment (23,24).

After the insertion of an incorrect nucleotide, the mismatched terminus must be extended. In theory, the velocity of this process depends on the concentration of the next dNTP substrate (25). However, several observations suggest that this effect may be neglected in the analysis of the substitution frequencies during a PCR amplification: (i) Kwok *et al.* studied the effects of all possible 3′-terminal primer–template mismatches on the yield of DNA amplification by PCR (26) and concluded that T:T (primer/template), T:C, T:G, C:T, C:A, G:T, and A:C mismatches were amplified as efficiently as perfect matches, whatever the concentrations of the added dNTPs (between 6 and 800 μM). A:A, A:G, G:A, and C:C mismatches strongly reduced the overall yield of PCR products. However, this reduction did not depend on the dNTP concentrations. The only mismatch which was sensitive to a variation of dNTP concentration was the G:G. This mismatch corresponds to a bp substitution (GC → CG) which was almost never observed in the current study. Consequently, it seems reasonable to expect that DNA elongation beyond a point of misincorporation will not depend on the set of dNTP concentrations in the PCR assay.

(ii) Eckert and Kunkel followed the consequence of a 1000-fold increase in the four dNTP concentrations added at equimolar concentrations during a PCR experiment. An increase in the dNTP concentrations from 1 μM to 1 mM only enhanced the substitution frequencies by a factor of 2. This result was obtained under $MgCl_2$ conditions of low fidelity (10 mM) (11).

(iii) An influence of the concentration of the next dNTP substrate on the misincorporation frequency at one position would result in a correlation between the substitution frequency at this position and the nature of the following template base. Such a correlation could never be observed in our experiments. For instance, when using 3.4 mM dGTP in the presence of both 0.5 mM $MnCl_2$ and the smaller concentration of $MgCl_2$, the A → G transition frequency was found to be nearly constant whatever the nucleotide following the A. Measured frequency values were 27, 23, 26, and 23‰ for a following G, A, T, or C base, respectively.

In contrast, the misincorporation probability depends on the number of DNA duplications (n) during the PCR (27). When the mutation rate is low (i.e., when the probability of reversing an acquired mutation remains negligible), the misincorporation probability can be shown to be directly proportional to the n value (28,29). Hence, let us consider the AT → GC transition. The probability

of incorporating a G instead of an A is proportional to n[dGTP]/[dATP] and that of incorporating a C instead of a T is proportional to n[dCTP]/[dTTP]. If the values of these probabilities are small, then the probability of substituting an AT base pair by a GC ($p_{AT\rightarrow GC}$) will be the sum of the two above probabilities:

$$p_{AT\rightarrow GC} = n\left(k_{AG}\frac{[dGTP]}{[dATP]} + k_{TC}\frac{[dCTP]}{[dTTP]}\right). \quad [1]$$

By analogy, the probabilities of the five other bp substitutions are described by the following combinations of k_{XY} constants:

$$p_{AT\rightarrow CG} = n\left(k_{AC}\frac{[dCTP]}{[dATP]} + k_{TG}\frac{[dGTP]}{[dTTP]}\right); \quad [2]$$

$$p_{AT\rightarrow TA} = n\left(k_{AT}\frac{[dTTP]}{[dATP]} + k_{TA}\frac{[dATP]}{[dTTP]}\right); \quad [3]$$

$$p_{GC\rightarrow AT} = n\left(k_{GA}\frac{[dATP]}{[dGTP]} + k_{CT}\frac{[dTTP]}{[dCTP]}\right); \quad [4]$$

$$p_{GC\rightarrow TA} = n\left(k_{GT}\frac{[dTTP]}{[dGTP]} + k_{CA}\frac{[dATP]}{[dCTP]}\right); \quad [5]$$

and

$$p_{GC\rightarrow CG} = n\left(k_{GC}\frac{[dCTP]}{[dGTP]} + k_{CG}\frac{[dGTP]}{[dCTP]}\right). \quad [6]$$

Experimental Adjustment of Parameters

To evaluate the k_{XY} contants, two more assumptions were made. (i) Since, in this study, at most 35% of each of the initially added nucleotides was consumed during amplification (Table 1), we admitted that the free nucleotide concentrations remained constant in the assay. (ii) We observed that the distribution of mutated bases along the DNA template did not significantly differ from a binomial distribution, i.e., from the theoretical distribution in the absence of any context dependence. In particular, hot spots of misincorporation in the DNA fragment under study could not be evidenced (see, for example, in Fig. 1, the distribution of the mutations resulting from the addition of an excess of 3.4 mM dGTP). Consequently, each k_{XY} value could be assumed to be independent of the sequence context of the mutations. Under such assumptions, a set of k_{XY} constants could be derived from experimental sequencing data by least-square linear regression analysis (Table 2). From these constants and Eqs. [1] to [6], misincorporation frequencies during a PCR experiment can be predicted as a function of the dNTP concentrations.

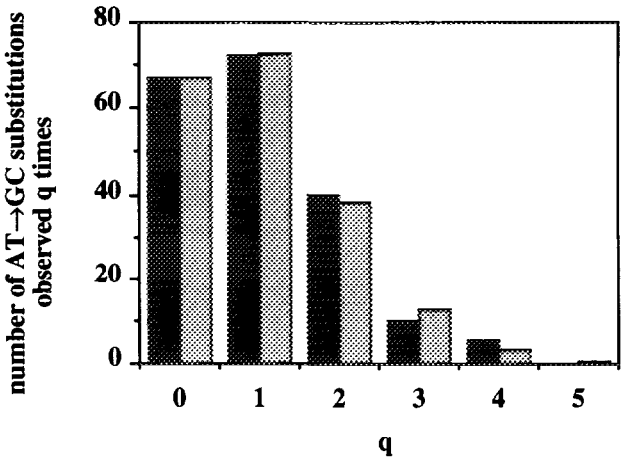

FIG. 1. Distribution of AT → GC substitutions after PCR amplification in the presence of 9.5 mM $MgCl_2$, 0.5 mM $MnCl_2$, 3.4 mM dGTP, and 0.2 mM of the other three dNTPs. The template DNA carried 194 AT base pairs susceptible to be mutagenized. Upon sequencing DNA from 23 randomly picked mutants, a total of 202 AT → GC substitutions were observed. The number of given AT bp which were substituted q times is plotted as a function of q. The experimental results (dark bars) are compared with the theoretical binomial distribution corresponding to the same substitution frequency (light bars).

Predictive Value of the Method

According to the above equations, a linear relationship between the frequency of one given bp substitution and the concentration of one forcing nucleotide in the PCR reaction must occur. This relationship could be validated by experiments performed at various dGTP

TABLE 2

k_{XY} Values According to Various Divalent Cation Conditions

	Value of the k_{XY} constants (multiplied by 10^6)		
	Low $MgCl_2$ No $MnCl_2$	Low $MgCl_2$, 0.5 mM $MnCl_2$	High $MgCl_2$, 0.5 mM $MnCl_2$
k_{AG}	42	160	330
k_{TC}	4.3	22	33
k_{AC}	0.8	5.8	15
k_{TG}	0.9	1.5	8.8
k_{AT}	3.4	23	100
k_{TA}	0.8	34	78
k_{GA}	0.6	19	57
k_{CT}	19	37	69
k_{GT}	<0.5	4.7	12
k_{CA}	<0.5	7.4	20
k_{GC}	<0.5	<0.5	<0.5
k_{CG}	1.9	5.5	15

Note. The k_{XY} constants were derived from the fitting in of sequencing data to Eqs. [1] to [6] by a least-square linear regression analysis. The $MgCl_2$ concentration was adjusted so that the total concentration of divalent cations exceeded the sum of the dNTP concentrations by either 0.7 mM (low $MgCl_2$) or 6 mM (high $MgCl_2$).

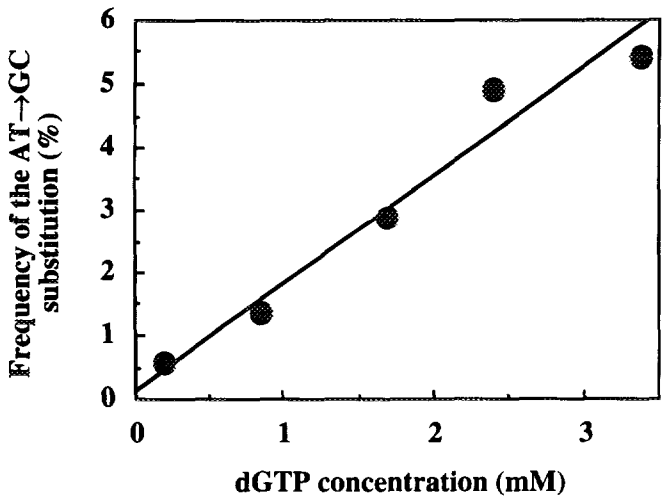

FIG. 2. Variation of the frequency of the AT → GC substitution as a function of the dGTP concentration in the assay. The concentration of the other nucleotides was equal to 0.2 mM. Amplification was carried out as described under Materials and Methods, in the presence of 0.5 mM $MnCl_2$ and high $MgCl_2$ concentration ($[MgCl_2]_{total}$ + $[MnCl_2]_{total}$ = $[dNTP]_{total}$ + 6 mM). For each PCR condition, the actual number of DNA duplications (n) was obtained from additional experiments performed with [α-^{32}P]dATP. From one experiment to another, there were slight differences in the measured n values. Consequently, the presented substitution frequencies were all normalized for n = 10, assuming that the substitution frequency value varied proportionally to n. At least 5820 bp were sequenced for each PCR condition assayed.

concentrations (Fig. 2), which confirmed the expected linear dependence ($p = p_0 + a$[dGTP]).

To further establish the validity of the method, we attempted to obtain a large diversity of mutations by performing two runs of PCR amplification starting from a single sample and using successively different nucleotide conditions. The first run was in the presence of forcing dGTP (0.2 mM dATP, 0.2 mM dCTP, 0.85 mM dGTP, and 0.2 mM dTTP). These conditions were expected to produce mainly AT → GC substitutions. After the first run, the sample was diluted 220-fold, a factor which corresponded to the amplification factor (2^n) in this PCR run. Then, a second PCR was performed using a high dATP/dCTP ratio (3.4 mM dATP, 0.1 mM dCTP) in order to maximize the generation of GC → TA substitutions. The concentrations of the other nucleotides (0.21 mM dGTP, 0.27 mM dTTP) were adjusted from the k_{XY} constants in Table 2 with the aim of finally producing mutation frequencies nearly identical for the three AT → GC, AT → TA, and GC → AT substitutions. The experiment was in the presence of 0.5 mM $MnCl_2$ and high $MgCl_2$ concentration ($[MgCl_2]_{total}$ + $[MnCl_2]_{total}$ = $[dNTP]_{total}$ + 6 mM). At the end of the experiment, amplified fragments were cloned into phage M13 mp19, and 29 amplified fragments of 291 bp each were sequenced. AT → GC, AT → TA, GC → AT, and GC → TA substitutions were found at frequencies of 12.3, 12.4, 8.5, and 6.4‰, respectively. The frequencies of the AT → CG and GC → CG substitutions (1 and 0.7‰, respectively) were much lower. A few deletions, amounting to 18, were also observed. From calculations using the k_{XY} values in Table 2 and an n value of 7.8 as measured in each PCR run, AT → GC, AT → TA, GC → AT, and GC → TA substitutions in this experiment were expected at frequencies of 11.4, 9.1, 9.3, and 5.6‰, respectively. AT → CG and GC → CG substitutions were expected at frequencies of 0.5 and 0.7‰, respectively. The satisfying agreement between the experiment and the theory gives credit to the data in Table 2 to predict the frequencies of the substitutions which can be obtained in a PCR experiment.

Choice of the Nucleotide Concentrations

The k_{AC}, k_{TG}, k_{GC}, k_{CG}, k_{GT}, and k_{CA} values being small, the AT → CG, GC → CG, and GC → TA transversions are more difficult to generate than the other substitutions. Noteworthy, the low values of the k_{CA}, k_{AC}, k_{GC}, and k_{CG} constants agree with the above remark that *Taq* DNA polymerase poorly elongates downstream of A:G, G:A, C:C, and G:G mispairings (26,30). However, through an adequate choice of the dNTP concentrations, a wide range of values can in theory be obtained for the relative occurrences of the three most frequent substitutions (AT → GC, AT → TA, and GC → AT). For instance, Table 3 gives sets of nucleotide concentrations ensuring at a same time (i) substitution of GC and AT bp with the same probability and (ii) equiprobability for the AT → GC and AT → TA substitutions.

By increasing the concentration(s) of the forcing nucleotide(s), high mutation probabilities should theoretically be obtained. However, at the highest concentrations assayed (>4 mM), we noted a decrease in the yield of the amplified product, especially when $MgCl_2$ concentration was high (data not shown).

Choice and Estimation of the Number of Duplications

As outlined above, the mutation frequency is a function of the actual number of DNA duplications (n). Under our experimental conditions, the programming of 16 cycles of amplification with 600 pM of a 356-bp-long DNA template (i.e., about 9 × 10^9 double-stranded DNA molecules in 25 μl) resulted in an n value of roughly 8. Noteworthy, various parameters including concentration of DNA polymerase, initial copy number of the template, annealing temperature, and extension time may affect the value of n and, hence, the mutagenesis rate (reviewed by Eckert and Kunkel (27)). Since the number of duplications may also slightly depend on the sequence and/or the length of the DNA template, it is recommendable to measure the n value when searching for precise substitution

TABLE 3

Calculated Probabilities of bp Substitutions (‰) for Various Sets of Nucleotide Concentrations (mM)

[dATP]	[dCTP]	[dGTP]	[dTTP]	$p_{overall}$	$p_{AT\to GC}$	$p_{AT\to CG}$	$p_{AT\to TA}$	$p_{GC\to AT}$	$p_{GC\to TA}$	$p_{GC\to CG}$
Low $MgCl_2$										
0.56	0.90	0.20	1.40	1.5	0.7	0.1	0.7	1.1	0.4	<0.1
0.35	0.40	0.20	1.35	2	1.0	<0.1	1.0	1.6	0.4	<0.1
0.20	0.20	0.18	1.26	3	1.5	<0.1	1.5	2.6	0.4	<0.1
0.22	0.20	0.27	1.82	4	2.0	<0.1	2.0	3.6	0.4	<0.1
0.22	0.20	0.34	2.36	5	2.5	<0.1	2.5	4.6	0.4	<0.1
0.23	0.20	0.42	2.90	6	3.0	<0.1	3.0	5.5	0.4	0.1
0.23	0.20	0.57	4.00	8	4.0	<0.1	4.0	7.5	0.4	0.1
0.12	0.10	0.36	2.50	10	5.0	<0.1	5.0	9.4	0.4	0.2
0.12	0.10	0.55	3.85	15	7.5	<0.1	7.5	14.3	0.4	0.3
High $MgCl_2$										
0.51	0.20	1.15	3.76	15	7.5	<0.1	7.5	13.2	0.9	0.9
0.39	0.15	1.17	3.85	20	10.0	<0.1	10.0	17.9	0.9	1.2
0.26	0.10	1.20	3.94	30	15.0	<0.1	15.0	27.3	0.9	1.8

Note. Sets of nucleotide concentrations were chosen to ensure (i) substitutions of AT and GC bp with the same probability and (ii) equiprobability of the AT → GC and AT → TA substitutions. For each nucleotide condition, the probabilities of the three possible substitutions at a given bp ($p_{XY\to X'Y'}$) are indicated. In addition, the corresponding overall probability of substituting this bp ($p_{overall} = p_{AT\to GC} + p_{AT\to CG} + p_{AT\to TA}$ or $p_{GC\to AT} + p_{GC\to TA} + p_{GC\to CG}$) is given. The data were calculated using Eqs. [1] to [6], as defined under Results, and the k_{XY} values of Table 2 corresponding to the presence of 0.5 mM $MnCl_2$ and of either low $MgCl_2$ ($[MgCl_2]_{total} + [MgCl_2]_{total} = [dNTP]_{total} + 0.7$ mM) or high $MgCl_2$ ($[MgCl_2]_{total} + [MnCl_2]_{total} = [dNTP]_{total} + 6$ mM) concentration. The substitution probabilities ($p_{XY\to X'Y'}$) are for a number of DNA duplications (n) equal to 10. For an n value different from 10, the substitution probabilities in the table should be multiplied by $n/10$.

frequencies. This can be rapidly done by comparing the amount of the amplified fragment with known amounts of standard DNA on an agarose gel. The accuracy of this measurement is sufficient since the n value is proportional to the logarithm of the amount of amplified DNA and, therefore, quite insensitive to an imprecision in the estimation of this amount. To obtain small mutation frequencies, it may be useful to decrease the value of n by reducing the number of PCR cycles and, consequently, to increase the initial concentration of the DNA fragment to be mutagenized. In turn, the mutation probabilities can be enhanced by decreasing the initial concentration of the DNA fragment or by performing successive runs of amplification with appropriate dilutions of the sample between the runs.

Concluding Remarks

(i) The values of error frequency during PCR elongation already reported in the literature vary from one case to another by factors sometimes greater than 10 (29). This renders difficult a comparison with the set of the k_{XY} values determined in this study. In fact, several causes may explain such a scattering of the published error frequencies: (a) various reaction buffers were used and, as already noted (11,12), some of them may be more mutagenic than the others. (b) Different relative proportions of AT base pairs in the template may introduce biases, caused by the fact that the AT → GC transition is, by far, the most frequent bp substitution under standard PCR conditions. Consequently, the overall mutation frequency obtained from an AT-rich template will be higher than that from a GC-rich fragment. (c) The value to be taken into account, when substitution frequencies are compared, is the number (n) of actual DNA duplications in the experiment, not the number of programmed PCR cycles. Most often, the reports in the literature cannot be usefully compared because of the unavailability of the n values.

(ii) Hot spots of *in vitro* misincorporation (10) may interfere with the predictions made from the k_{XY} values. However, as already mentioned, with the DNA template we used, we failed to detect any dependence of the substitution frequencies on the sequence context. Cadwell and Joyce reached the same conclusion using the gene of *Tetrahymena* ribozyme as template (16). Therefore, the occurrence of hot spots of misincorporation may remain exceptional.

(iii) Instead of introducing mutations, our approach may also be used to improve the fidelity of the PCR amplification process. Indeed, because the AT → GC and GC → AT transitions are more frequent than the other substitutions under standard PCR conditions (low concentration of free $MgCl_2$ and absence of $MnCl_2$), the global accuracy of a PCR amplification experiment may be improved by using dGTP and dTTP at lower concentrations than dATP and dCTP. From the data in Table 2, it may be predicted that in the presence of 1 mM dATP, 1 mM dCTP, 0.2 mM dGTP, and 0.5 mM dTTP, instead of 0.2 mM for the four triphosphates, the accu-

racy can be increased twice, i.e., the number of misincorporations during amplification can be lowered by a factor of two.

ACKNOWLEDGMENT

We gratefully acknowledge F. Dardel for fruitful advice.

REFERENCES

1. Zoller, M. J. (1991) *Curr. Opin. Biotechnol.* **2,** 526–531.
2. Hutchison, C. A., Nordeen, S. K., Vogt, K., and Edgell, M. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 710–714.
3. Shortle, D., and Botstein, D. (1983) *Methods Enzymol.* **100,** 457–468.
4. Myers, R. M., Lerman, L. S., and Maniatis, T. (1985) *Science* **229,** 242–247.
5. Busby, S., Irani, M., and de Crombrugghe, B. (1982) *J. Mol. Biol.* **154,** 197–209.
6. Grompe, M., Muzny, D. M., and Caskey, C. T. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5888–5892.
7. Montandon, A. J., Green, P. M., Giannelli, F., and Bentley, D. R. (1989) *Nucleic Acids Res.* **17,** 3347–3358.
8. Kadonaga, J. T., and Knowles, J. R. (1985) *Nucleic Acids Res.* **13,** 1733–1745.
9. Lehtovaara, P. M., Koivula, A. K., Bamford, J., and Knowles, J. K. C. (1988) *Protein Eng.* **2,** 63–68.
10. Tindall, K. R., and Kunkel, T. A. (1988) *Biochemistry* **27,** 6008–6013.
11. Eckert, K. A., and Kunkel, T. A. (1990) *Nucleic Acids Res.* **18,** 3739–3744.
12. Bloch, W. (1991) *Biochemistry* **30,** 2735–2747.
13. Keohavong, P., and Thilly, W. G. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9253–9257.
14. Leung, D. W., Chen, E., and Goeddel, D. V. (1989) *Technique* **1,** 11–15.
15. Liao, X., and Wise, J. A. (1990) *Gene* **88,** 107–111.
16. Cadwell, R. C., and Joyce, G. F. (1992) *PCR Methods Appl.* **2,** 28–33.
17. Zhou, Y., Zhang, X., and Ebright, R. H. (1991) *Nucleic Acids Res.* **19,** 6052.
18. Lévêque, F., Plateau, P., Dessen, P., and Blanquet, S. (1990) *Nucleic Acids Res.* **18,** 305–312.
19. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33,** 103–119.
20. Hirel, P. H., Lévêque, F., Mellot, P., Dardel, F., Panvert, M., Mechulam, Y., and Fayat, G. (1988) *Biochimie* **70,** 773–782.
21. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Dunning, A. M., Talmud, P., and Humphries, S. E. (1988) *Nucleic Acids Res.* **16,** 10393.
23. Echols, H. (1991) *Annu. Rev. Biochem.* **60,** 477–511.
24. Fersht, A. (1985) Enzyme Structure and Mechanism, pp. 112, Freeman, New York.
25. Mendelman, L. V., Boosalis, M. S., Petruska, J., and Goodman, M. F. (1989) *J. Biol. Chem.* **264,** 14415–14423.
26. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J. J. (1990) *Nucleic Acids Res.* **18,** 999–1005.
27. Eckert, K. A., and Kunkel, T. A. (1992) *in* PCR: A Practical Approach (McPherson, M. J., Quirke, P., and Taylor, G. R., Eds.), pp. 225–244, Oxford Univ. Press, New York.
28. Krawczak, M., Reiss, J., Schmidtke, J., and Rösler, U. (1989) *Nucleic Acids Res.* **17,** 2197–2201.
29. Eckert, K. A., and Kunkel, T. A. (1991) *PCR Methods Appl.* **1,** 17–24.
30. Huang, M. M., Arnheim, N., and Goodman, M. F. (1992) *Nucleic Acids Res.* **20,** 4567–4573.