

# Frameshift errors initiated by nucleotide misincorporation

(fidelity/DNA polymerase/DNA synthesis/mutagenesis/deletions)

KATARZYNA BEBENEK AND THOMAS A. KUNKEL\*

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

Communicated by Philip C. Hanawalt, April 16, 1990 (received for review January 22, 1990)

**ABSTRACT** Studies presented here on the fidelity of DNA synthesis *in vitro* support the hypothesis that a classical base-substitution intermediate (i.e., a misincorporated nucleotide) can yield a frameshift mutation. By using a fidelity assay specifically designed to detect minus-one-base errors, nucleotide substrate pool imbalances that have previously been shown to increase the rate of misincorporation are now shown to also increase minus-one-base frameshift error rates. Examination of the specificity of the errors produced in reactions with various dNTP pool imbalances and various DNA templates revealed that template nucleotides were preferentially lost when they had as a 5' neighbor a nucleotide complementary to the dNTP provided in excess. This suggests that when a misincorporated nucleotide is complementary to the next nucleotide in the template, a misaligned intermediate containing a correct terminal base pair can form and be extended by a DNA polymerase, leading to a frameshift mutation. We present evidence that the proposed mechanism may operate *in vivo* and discuss the implications of this model for frameshift mutations induced by DNA damage.

Mutations resulting from the loss or gain of one or more bases are observed *in vivo* in prokaryotes and eukaryotes. Although our knowledge about the mechanisms that govern their production is not extensive, one widely accepted concept, proposed by Streisinger *et al.* (1), is that frameshift mutations in homopolymeric DNA sequences result from slippage of the two strands of DNA. The frameshift mutation frequency should increase with the length of the run; because more misaligned intermediates are possible, these can be stabilized by an increasing number of correct base pairs. Also, the misaligned nucleotide can be moved farther away from the 3'-OH primer terminus and is thus less likely to interfere with subsequent polymerization events (2, 3).

The development of assays to monitor frameshift errors during DNA synthesis *in vitro* (4–6) and the ability to describe mutations at the DNA sequence level have expanded our appreciation of the complexity of frameshift mutagenesis. For example, minus-one-base frameshift mutations at noniterated nucleotide positions comprise a significant proportion of DNA polymerase errors *in vitro* (for review, see ref. 7) and have also been recovered in a number of systems *in vivo* (8–15). We have been interested in how such errors might arise during DNA polymerization, given that it is not obvious how a misaligned intermediate at a noniterated site is stabilized for continued polymerization.

At least three models can be envisioned. One is that a nucleotide assumes a position during polymerization in which it neither instructs incorporation nor interferes with its neighbor's ability to do so. This possibility is supported by structural studies with oligonucleotides demonstrating that an extra base can exist in conformations that do not disrupt hydrogen bonding of adjacent base pairs (refs. 16–21 and, for

review, see ref. 22). As suggested (3, 5, 21) extra bases may be stabilized by interactions with a DNA polymerase.

A second possibility is that a transient misalignment process involving movement and hybridization of the primer to a distant site is followed by limited synthesis to generate the frameshift and then return of the primer, now containing the frameshift error, to its original position. There is substantial experimental support for base-substitution (23, 24) and frameshift (24, 25) mutagenesis by transient misalignment during DNA synthesis *in vitro*. However, frequency and specificity considerations (see below) make it unlikely that all frameshifts are associated with distant template sequences.

Kunkel and Soni (23), therefore, proposed a third possibility (Fig. 1), wherein frameshifts could be initiated by misincorporation of a nucleotide. If this nucleotide is complementary to the next template base, then its relocation to a position one nucleotide ahead could lead to a frameshift intermediate containing a correct base pair with an unpaired nucleotide in the template strand of the template-primer. Just as for classical slippage-initiated frameshifts within homopolymeric runs, the misaligned intermediate would be stabilized by correct base pairing. The difference is that the frameshift error is initiated by misincorporation, not misalignment.

This hypothesis was suggested by several observations in studies of the fidelity of DNA synthesis by DNA polymerases. The error rate for minus-one-base errors at noniterated template positions is surprisingly high, being similar to base-substitution error rates (3, 6), and these errors are proofread as effectively as are base-substitution errors (26). Minus-one-base errors at noniterated template positions are mostly the loss of a template purine that has as a 5'-nearest neighbor a template pyrimidine (3, 5, 6, 26), and DNA polymerases frequently misinsert dAMP and dGMP opposite template purines (4, 6, 26–28), generating purine-purine mispairs that are poor substrates for further incorporation by DNA polymerases (26, 29–31). Finally, using a DNA substrate that can form either a terminal mispair or the proposed one-base misalignment, exonuclease-deficient DNA polymerases will extend from the misaligned template-primer with various efficiencies (6, 26, 28).

We present here a direct test of the hypothesis that minus-one-base frameshifts can be initiated by nucleotide misincorporation during an ongoing polymerization reaction. We have chosen to test the model first with the Klenow fragment of DNA polymerase I (Klenow polymerase) because more is known about this polymerase than about any other.

## MATERIALS AND METHODS

**Materials.** A mutant derivative of bacteriophage M13mp2 was used that is missing two of the four thymidines at positions 70–73 of the *lacZ*  $\alpha$ -complementation coding sequence. The D355A, E357A mutant form of Klenow poly-

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.

\*To whom reprint requests should be addressed.

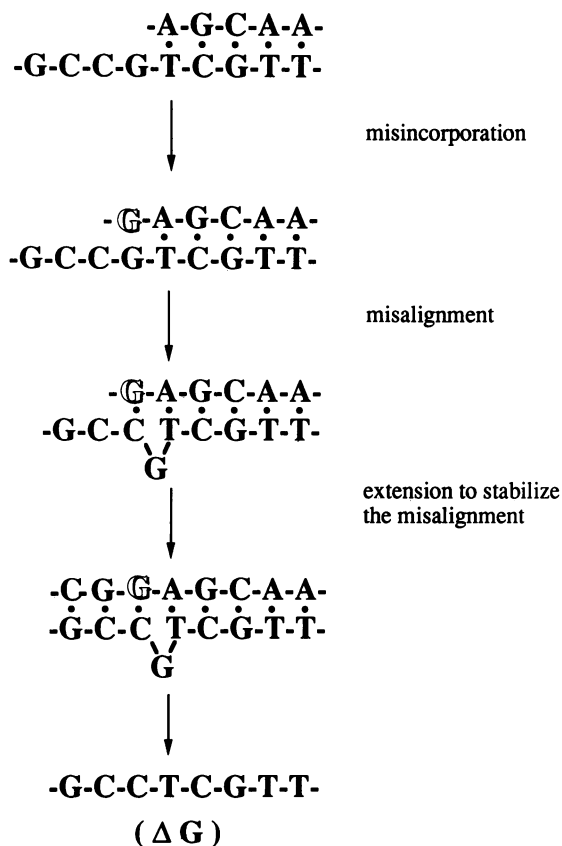


FIG. 1. Model for minus-one-base frameshifts resulting from nucleotide misincorporation.

merase was a kind gift from Catherine M. Joyce (Yale University). This polymerase lacks a proofreading exonuclease activity (32) and, therefore, cannot excise misinserted nucleotides (26). The two amino acid substitutions introduced into the small domain to inactivate the exonuclease do not affect the structure, specific activity, or selectivity (26, 32) of the DNA polymerase, whose active site is the large domain (33). The sources of other materials have been described (4, 34, 35).

**DNA Polymerase Reactions.** Polymerase reaction mixtures (100  $\mu$ l) contained 20 mM Hepes (pH 7.8), 2 mM dithiothreitol, 10 mM  $MgCl_2$ , 600 ng of gapped DNA, and all four dNTPs, each at 1 mM (or, for the biased-pool reaction mixtures, one dNTP at 1 mM and the other three dNTPs, each at 50  $\mu$ M), and 1 unit of exonuclease-deficient Klenow polymerase. The mixtures were incubated for 10 min (equal pools) or 20 min (biased pools) and terminated by adding EDTA to a final concentration of 15 mM.

**Other Procedures** Site-directed mutagenesis, preparation of substrates, analysis of reaction products, transfections, and sequence analysis were performed as described (4, 34–36).

## RESULTS

The fidelity assay for detecting minus-one-base frameshifts used a double-stranded M13mp2 DNA substrate with a 361-nucleotide single-stranded gap. The template sequence in the gap contained a mutation in the *lacZ*  $\alpha$ -complementation coding sequence, the deletion of two template nucleotides from a TTTT run at positions 70–73 (where position 1 is the first transcribed nucleotide). This altered the reading frame and created a downstream UGA termination codon. The resulting plaque phenotype was colorless. Gap-filling DNA

synthesis reactions were performed using the exonuclease-deficient Klenow polymerase, samples of the reaction mixtures were analyzed by agarose gel electrophoresis to assure complete gap-filling (as was achieved for all reactions described here), and the remaining products were used to transfect competent host cells to score minus-one-base frameshifts as blue revertants. The loss of a single nucleotide at any site between position 46 and position 85 restored the reading frame, yielding a blue plaque phenotype. DNAs from independent mutants were sequenced to define error specificity.

**Effect of dNTP Substrate Imbalances on Error Specificity.** To test the model, we reasoned that forcing misincorporations by a DNA polymerase lacking a proofreading exonuclease activity using an excess of one dNTP should yield a high frequency of loss of those template nucleotides whose 5'-nearest neighbor is complementary to the incorrect nucleotide provided in excess. Since previous observations (3–6, 26–31) suggested that misincorporation of purine nucleotides could be responsible for some minus-one-base errors, reactions were performed with a 20-fold excess of either dGTP or dATP.

These pool imbalances did not dramatically affect the overall average frameshift fidelity, since the reversion frequencies (blue/total plaques) obtained upon transfection were similar to each other and to that obtained from reactions performed at equal substrate concentrations (see Table 1). However, DNA sequence analysis of independent mutants generated in each biased substrate reaction demonstrated substantial differences in frameshift error specificity (Table 1 and Fig. 2A). Thirteen of 18 mutants from reaction mixtures with excess dGTP resulted from loss of a template nucleotide whose 5' neighbor was a cytidine. The resulting error rate per nucleotide polymerized for loss of template bases having a neighboring cytidine was at least 16-fold higher than for those having a thymidine neighbor (Table 1). Conversely, 14 of 19 minus-one-base deletions with excess dATP had as a 5' neighbor a template thymidine. (This assumes that the guanosine at position 62 was lost; we cannot be sure of this in this two-base run.) Thus, when excess dATP was present, the

Table 1. Minus-one-base error rates at template sites with cytidine and thymidine neighbors

Template site		Error rate ( $\times 10^6$ )	
5'-Neighboring base	Base lost	Excess dGTP	Excess dATP
Results with initial template			
C	X (9 sites)	21	16
T	X (11 sites)	$\leq 1.3$	47
C	G-66	150	37
T	G-62	$\leq 14$	410
Results with altered template			
C	X (9 sites)	31	5.1
T	X (11 sites)	9.1	44
T	G-66	$\leq 20$	$\leq 23$
C	G-62	240	$\leq 23$

Revertant frequencies for the initial template were  $360 \times 10^{-6}$  with equal dNTP pools,  $160 \times 10^{-6}$  with a 20-fold excess of dGTP, and  $420 \times 10^{-6}$  with a 20-fold excess of dATP. The corresponding revertant frequencies for the altered template were  $400 \times 10^{-6}$ ,  $240 \times 10^{-6}$ , and  $330 \times 10^{-6}$ . X indicates the template nucleotide lost. G-66 and G-62 refer to the nucleotide and position of the observed mutational hot spots for minus-one-base frameshifts in the target sequences (the target sequences are shown in Fig. 2). Error rates are expressed per nucleotide polymerized and were calculated by multiplying the overall minus-one-base frameshift revertant frequency by the proportion of mutants belonging to that class (from Fig. 2), dividing by 0.6 to correct for the probability of expressing an error in the newly synthesized strand (34), and then dividing by the number of detectable sites for that class of errors.

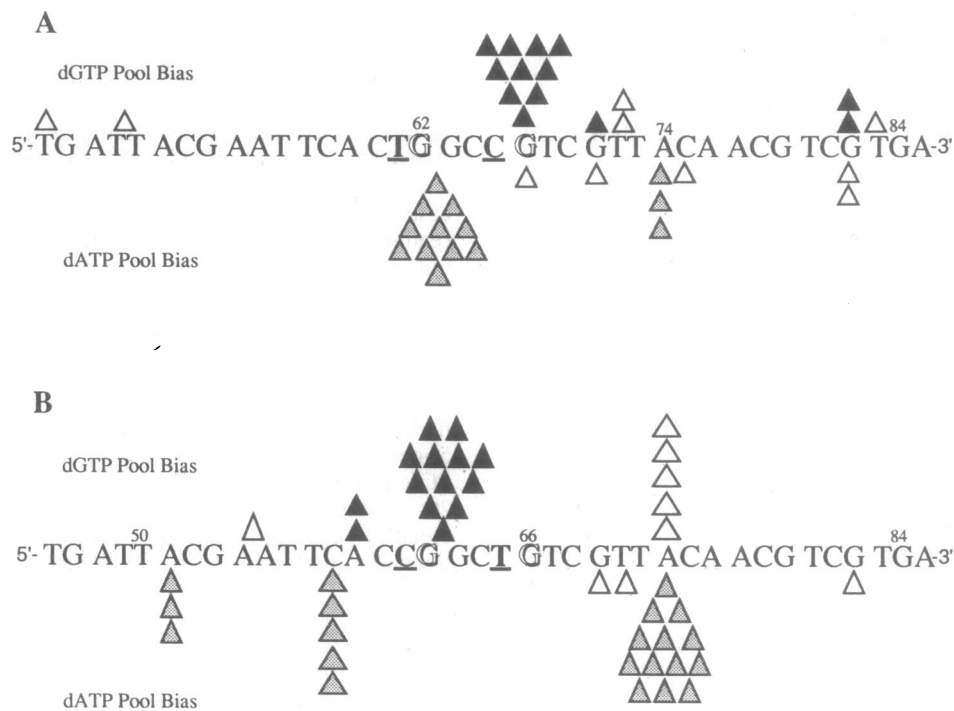


FIG. 2. Spectra of frameshifts from reactions with excess dGTP and dATP. The lines of DNA sequence represent the initial (A) and the altered (B) mutational targets. The nucleotides shown are from position 46 to position 85, where +1 is the first transcribed base of the *lacZ*  $\alpha$ -complementation gene in M13mp2. Although two thymidines have been deleted (nucleotides 72 and 73), the original numbering of the bases has been retained for convenience. DNA synthesis across these templates is from right to left. Each triangle represents an independent mutant whose DNA sequence was determined. Triangles directly above or below a nucleotide indicate that that nucleotide was deleted. For mutants missing one of two consecutive identical nucleotides, the triangle is centered between the two. The underlined nucleotides are the positions that were changed by site-directed mutagenesis, and the outlined nucleotides are the hot spots for minus-one-base errors in the initial template. (A) Mutants generated in reactions with the initial template. (B) Mutants generated in reactions with the template altered by site-directed mutagenesis. Solid triangles indicate frameshifts at template sites followed by a cytidine, from reactions with the dGTP pool bias. Stippled triangles indicate frameshifts at template sites followed by a thymidine, from reactions with the dATP pool bias. Open triangles represent frameshifts at all other sites.

error rate for loss of template nucleotides with a neighboring thymidine was higher than for loss of those template nucleotides with a neighboring cytidine (Table 1). Both specificities were as predicted by the model in Fig. 1, which suggests that it is complementarity between the forced misincorporation and the 5' template neighbor that determines whether a template nucleotide is lost.

**Effect of Neighboring Nucleotide Changes on Error Specificity.** Frameshift errors are not randomly distributed in the target (Fig. 2A). For example, with excess dGTP, the error rate is 7-fold higher at position 66 than the average for all nine sites in the target that have a 5'-neighboring cytidine (Table 1). Similarly, with excess dATP, the error rate at position 62–63 is 9-fold greater than the average error rate for all sites that have a 5'-neighboring thymidine. We exploited this site specificity to further test the model by changing the template thymidine at position 61 to a cytidine and the cytidine at position 65 to a thymidine. We then repeated the analysis (Fig. 2B and Table 1, altered template).

The predicted result with excess dGTP is that the one-base-deletion error frequency at position 66 should decrease, since when dGMP is misincorporated, it can no longer pair with the neighboring template nucleotide, now a thymidine. This prediction was fulfilled, since no errors were detected in this situation (Fig. 2B), representing a  $\geq 7.5$ -fold decrease in error rate (Table 1, compare  $150 \times 10^{-6}$  to  $\leq 20 \times 10^{-6}$ ). A second prediction was that, because position 62–63 is now followed by a template cytidine, this site may yield a high error rate with excess dGTP. Indeed, the dGTP-driven error rate increased at least 15-fold (Table 1, compare  $240 \times 10^{-6}$  to  $\leq 14 \times 10^{-6}$ ).

The predicted result with excess dATP is that the one-base-deletion error rate at position 62–63 should decrease, since if dAMP is misincorporated, it can no longer pair with the neighboring template nucleotide, now a cytidine. Consistent with this logic, no dATP-driven errors were observed (Fig. 2B), reflecting at least an 18 times decrease in error rate (Table 1, compare  $410 \times 10^{-6}$  to  $\leq 23 \times 10^{-6}$ ). When followed by a template thymidine, position 66 was not a hot spot with excess dATP. This neither supports nor contradicts the model.

**Antimutator Effect with Excess dCTP.** The above results suggest that substrate imbalances that favor misincorporation of a nucleotide increase the frameshift error rate at certain template sites. If this also occurs during DNA synthesis with equimolar concentrations of dNTP substrates, then conditions that reduce misincorporation at a template site should decrease the frequency of loss of that nucleotide. To test this, we first analyzed error specificity for reactions performed with equimolar concentrations of all four substrates to ascertain if there was preferential loss of particular template nucleotides in the 38-base target. Template purines were deleted in 17 of 20 mutants (Fig. 3), with 10 mutants having lost a guanosine. We, therefore, focused on template guanosine, where the model predicts that a reaction mixture containing excess dCTP should be antimutagenic for minus-guanosine frameshifts. As predicted by the model, excess dCTP was 5-fold antimutagenic for deletions of template guanosine. (Because the overall frequencies were similar for the two reactions, the quantitative effect can be seen by directly comparing the solid triangles in Fig. 3.) This suggests that misincorporations initiate frameshifts even at equimolar substrate concentrations. In the same reaction, excess dCTP



In principle, the model in Fig. 1 is not limited to minus-one events at noniterated sites. Minus-one-base errors within runs might be initiated by misincorporation, as well as plus-one-base errors and frameshifts involving loss or gain of more than one nucleotide. The results presented here do not exclude the involvement of other processes in producing frameshifts. For example, it is still theoretically possible that frameshifts even at noniterated sites might be initiated by misalignment. This might occur if a template nucleotide assumes a position during polymerization in which it neither instructs incorporation nor interferes with its neighbors ability to do so.

The model may be relevant to frameshift mutagenesis induced by DNA damage. Some of the minus-one-base errors resulting from depurination of DNA (35) may have resulted from misincorporation opposite abasic sites. In some circumstances these are known to be difficult to extend (38). In attempting to explain certain frameshift mutations induced by the carcinogen *N*-acetyl-2-aminofluorene, it has been suggested (39) that, after insertion of cytidine opposite a damaged guanosine, relocation of the cytidine to a 5'-neighboring template guanosine could create a misaligned but properly hydrogen-bonded terminal base pair that could be extended to generate the frameshift mutation. It will be interesting to examine the relationship between the degree of difficulty in extension from a particular template-primer (whether mispaired or damaged) and the probability that, in the appropriate sequence context, relocation to a new position will provide a misaligned but more favorable substrate for extension. The choice made between a mispaired and misaligned terminus will be influenced by the proteins and the nature of the mispair or damage.

We thank John W. Drake and Roel M. Schaaper for critical evaluation of the manuscript and Catherine M. Joyce for her kind gift of exonuclease-deficient Klenow polymerase.

1. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77-84.
2. Streisinger, G. & Owen, J. E. (1985) *Genetics* **109**, 633-659.
3. Kunkel, T. A. (1986) *J. Biol. Chem.* **261**, 13581-13587.
4. Kunkel, T. A. (1985) *J. Biol. Chem.* **260**, 5787-5796.
5. de Boer, J. G. & Ripley, L. S. (1988) *Genetics* **118**, 181-191.
6. Kunkel, T. A., Hamatake, R. K., Motto-Fox, J., Fitzgerald, M. P. & Sugino, A. (1989) *Mol. Cell. Biol.* **9**, 4447-4458.
7. Kunkel, T. A. & Bebenek, K. (1988) *Biochim. Biophys. Acta* **951**, 1-15.
8. Schaaper, R. M., Danforth, B. N. & Glickman, B. W. (1986) *J. Mol. Biol.* **186**, 273-284.
9. Nalbantoglu, J., Hartley, D., Phear, G., Tear, G. & Meuth, M. (1986) *EMBO J.* **5**, 1199-1204.
10. Ripley, L. S., Clark, A. & de Boer, J. G. (1986) *J. Mol. Biol.* **191**, 601-613.
11. Ashman, C. & Davidson, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3354-3358.
12. Giroux, C. N., Mis, J. R. A., Pierce, M. K., Kohalmi, S. E. & Kunz, B. A. (1988) *Mol. Cell. Biol.* **8**, 978-981.
13. Lee, G. S.-F., Savage, E. A., Ritzel, R. G. & von Borstel, R. C. (1988) *Mol. Gen. Genet.* **214**, 396-404.
14. Ikehata, H., Akagi, T., Kimura, H., Akasaka, S. & Kato, T. (1989) *Mol. Gen. Genet.* **219**, 349-358.
15. Phear, G., Armstrong, W. & Meuth, M. (1989) *J. Mol. Biol.* **209**, 577-582.
16. Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K. & Breslauer, K. J. (1982) *Biochemistry* **21**, 445-451.
17. Hare, D., Shapiro, L. & Patel, D. J. (1986) *Biochemistry* **25**, 7456-7464.
18. Roy, S., Sklenar, V., Appella, E. & Cohen, J. S. (1987) *Biopolymers* **26**, 2041-2052.
19. Woodson, S. A. & Crothers, D. M. (1988) *Biochemistry* **27**, 3130-3141.
20. Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E. & Sussman, J. L. (1988) *Nature (London)* **334**, 82-84.
21. Miller, M., Harrison, R. W., Wlodawer, A., Appella, E. & Sussman, J. L. (1988) *Nature (London)* **334**, 85-86.
22. Patel, D. J., Shapiro, L. & Hare, D. Q. (1987) *Rev. Biophys.* **20**, 35-112.
23. Kunkel, T. A. & Soni, A. (1988) *J. Biol. Chem.* **263**, 14784-14789.
24. Boosalis, M. S., Mosbaugh, D. W., Hamatake, R., Sugino, A., Kunkel, T. A. & Goodman, M. F. (1989) *J. Biol. Chem.* **264**, 11360-11366.
25. Papanicolaou, C. & Ripley, L. S. (1989) *J. Mol. Biol.* **207**, 335-353.
26. Bebenek, K., Joyce, C. M., Fitzgerald, M. P. & Kunkel, T. A. (1990) *J. Biol. Chem.*, in press.
27. Mendelman, L. V., Boosalis, M. S., Petruska, J. & Goodman, M. F. (1989) *J. Biol. Chem.* **264**, 14415-14423.
28. Roberts, J. D., Preston, B. D., Johnson, L. A., Soni, A., Loeb, L. A. & Kunkel, T. A. (1989) *Mol. Cell. Biol.* **9**, 469-476.
29. Perrino, F. W. & Loeb, L. A. (1989) *J. Biol. Chem.* **264**, 2898-2905.
30. Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265**, 2338-2346.
31. Joyce, C. M. (1989) *J. Biol. Chem.* **264**, 10858-10866.
32. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M. & Steitz, T. A. (1988) *Science* **240**, 199-201.
33. Freemont, P. S., Ollis, D. L., Steitz, T. A. & Joyce, C. M. (1986) *Proteins* **1**, 66-73.
34. Kunkel, T. A. & Soni, A. (1988) *J. Biol. Chem.* **263**, 4450-4459.
35. Kunkel, T. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1494-1498.
36. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154**, 362-387.
37. Roberts, J. D. & Kunkel, T. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7064-7068.
38. Bonner, C. A., Randall, S. K., Rayssiguier, C., Radman, M., Eritja, R., Kaplan, B. E., McEntee, K. & Goodman, M. F. (1988) *J. Biol. Chem.* **263**, 18946-18952.
39. Schaaper, R. M., Koffel-Schwartz, N. & Fuchs, R. P. P. (1990) *Carcinogenesis*, in press.