Mutagenesis by Transient Misalignment*

Thomas A. Kunkel and Aruna Soni
From the Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Based upon a consideration of two mutational hot spots produced during DNA synthesis by a eukaryotic DNA repair polymerase, we suggested that certain base substitution errors result not from direct miscoding but from correct coding by a transiently misaligned template-primer (Kunkel, T. A., and Alexander, P. S. (1986) J. Biol. Chem. 261, 160–166). This model, which we called dislocation mutagenesis, has been directly tested. Introducing a single, phenotypically silent G → A base change into the template switches the base substitution specificity at the immediately adjacent hot spot, a T residue, from T → G transitions to T → A transitions. The cumulative change in frequency, represented by the disappearance of the T → G events and the appearance of the T → A events, is >300-fold. These data demonstrate that during DNA synthesis in vitro, a base at one position can code a mutation at another position. This mechanism can operate over greater distances to produce complex mutations as well. We present one example in which a 123-base deletion containing three base changes at one end of the deletion can be precisely explained by transient misalignment. It remains to be established whether mutagenesis by dislocation operates in vivo to produce biologically significant changes in genetic information.

DNA is a structurally dynamic macromolecule capable of assuming a variety of conformations. This presents a significant challenge to an organism during the replication and maintenance of genetic information. The ability of DNA strands to misalign, creating unusual intermediates that can be processed into mutations, provides explanations for a variety of mutations involving the loss or gain of one or more bases (1–11). As early as 1960, Fresco and Alberts (12) suggested that misalignments could in principle also lead to base substitution errors, a suggestion advanced again by Fowler et al. (13) in discussing the base substitution specificity of an *Escherichia coli* mutant strain.

We have been examining molecular mechanisms of mutagenesis during *in vitro* DNA synthesis catalyzed by purified eukaryotic DNA polymerases acting on defined templates (14–19). For this purpose we use a mutagenesis assay which scores a variety of base substitution, frameshift, deletion, and complex errors at numerous positions within the lacZα DNA present in M13mp2 DNA. In the course of studies with rat DNA polymerase-β (pol-β), we observed two mutational hot spots that appeared to be related both by DNA sequence and error specificity. At one site pol-β produced primarily frameshift errors while at the other site this polymerase produced a high frequency of T → G transitions, bordered by a template G neighbor. These data led us to propose a model (14), which we called dislocation mutagenesis, to explain certain base substitution errors based upon a transient template-primer misalignment. In this report we describe results that support this model and then extend it to explain a complex deletion in which a block of genetic information is moved from one position in the genome to a distant position.

**EXPERIMENTAL PROCEDURES**

All materials and procedures have been described in Refs. 14, 18, and 19.

**M13mp2 Mutagenesis Assay**—The assay measures the fidelity of a single round of *in vitro* gap-filling DNA synthesis by a DNA polymerase. A gapped M13mp2 DNA substrate is constructed such that the 390-base single strand gap contains the 250-base mutational target, the lacZα complementation sequence. Correct synthesis to fill the gap produces DNA that, when used to transfect the appropriate *E. coli* host strain, will produce dark blue M13 plaques. Errors during synthesis are scored as lighter blue or colorless plaques. Since the assay measures loss of a gene function that is not essential for phage production, a wide variety of mutations at many different sites can be recovered and scored, including all 12 base substitution errors, frameshifts, deletions, and complex errors.

**RESULTS**

The M13mp2 forward mutation assay was used to describe the spectrum of errors produced by DNA polymerase β during a single round of *in vitro* gap-filling DNA synthesis (14). This enzyme was found to be highly inaccurate, committing a variety of base substitution, one-base frameshift, and deletion errors.

**The Dislocation Model**—The dislocation model for base substitution errors was suggested by considering two mutational hot spots in the spectrum of errors produced by pol-β. These two hot spots were related in both their template DNA sequences and their error specificities. At the template sequence 5'-C-G-T-T-T-T-A-C-3', a high frequency of T frameshifts was observed. At a lower (but still high) frequency, T → G transversions were observed at the 5'-most T in the run (position 70), having as its 5'-nearest neighbor a template G residue. The second hot spot 33 bases downstream contained a quite similar template sequence, 5'-C-G-T-T-A-C-3'. Here both T frameshifts and T → G transversions at the 5'-most T in the run (position 103) were again observed, but the relative frequencies of the two errors were reversed. At this shorter (two-base) T run, the frameshift frequency was much lower and the T → G events predominated.

The specificity of the base substitutions, exclusively T → G transversions, was unexpected. This mutation involves a template T-dCMP mispair. The simplest way to produce this intermediate is by direct misinsertion by pol-β of dCMP opposite the template T residue (Fig. 1A, pathway B, for Base

---

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

† The abbreviation used is: pol-β, polymerase-β.
FIG. 1. Mutational pathways and intermediates at the hot spot. Only one of the two hot spots is shown. The underlined base is position 102, with position 1 being the first transcribed base of the lacZα gene. Using our previous convention (14), the open triangle indicates a minus-one-base frameshift. Since the exact base lost from the run is not known, the frameshift is described by the 5'-most position in the run, in this instance position 103.

substitution). This possibility seems reasonable since pol-β is error prone for a variety of misinsertions throughout the lacZα mutational target, precedents exist for large and specific site-to-site differences in mutational phenomena, and the precise rules governing misinsertion fidelity, particularly with eukaryotic enzymes, have not yet been defined. Once misinsertion has occurred, pol-β, which cannot remove the mistake since it lacks associated proofreading exonuclease activity (20), could process the mispaired intermediate through the base substitution pathway (Fig. 1A, pathway BB). Alternatively, because the misinserted nucleotide (dCMP) is complementary to the next template base (a G), a rearrangement could occur...
(pathway BF) to produce a −T frameshift.

Considering the chemistry and structure of mismatches (21–23) and measurements of misincorporation of pyrimidines opposite pyrimidines with prokaryotic DNA polymerases (24, 25), such direct misinsertion might be expected to occur rarely, while the other two base substitutions that are known to be detectable at these positions (16) should occur at equal or higher frequencies. However, no T → C transitions (via a more favorable T-dCMP intermediate) or T → A transversions (via a T-dTMP mismatch) were observed. This unique base substitution specificity, the tendency of pol-β to produce minus-one-base framematches within runs of a common base (14) by a slippage mechanism (17) and the fact that errors occurred at the 5′-most T in a run of T residues followed by a template G residue, led to the model shown in Fig. 1A (pathway F, for Frameshift). In this instance, correct synthesis proceeds until a misalignment occurs, resulting in an extrahelical template T. Continued correct incorporation of additional nucleotides from the misaligned template-primer (pathway FF) fixes the extrahelical T and results in the −T event. However, if a single correct incorporation of dCMP occurs followed by realignment of the extrahelical T prior to the next incorporation event (pathway FB), a T-dCMP mismatch is created at the 5′-most T in the run. Continued correct incorporation from this mispair results in the T → A base substitution. This base substitution is not due to miscoding by pol-β but rather to repetitive use of the template G, first from a misaligned intermediate and then from a misaligned primer terminus.

Oligonucleotide-directed Mutagenesis of the Hot Spots—The dislocation model invokes multiple and complex interactions to explain an unexpectedly frequent event. The specificity and symmetry observations for the hot spots do not reveal whether misalignment or misinsertion is the initiating event for errors at these sites. To directly distinguish between these alternatives, we performed oligonucleotide-directed mutagenesis (26) to change a G to an A at both 5′-neighbors to the positions where the T → G errors were observed. If the initiating event is misinsertion of dCMP, then the expected results with this new template would continue to be T → G transversions (Fig. 1B, pathway B then BB) and either a double change (−T and A → G, following pathway BF) or a decrease in −T frequency, since the realigned intermediate contains both an A-C mismatch and an extrahelical T. Alternatively, if the initiating event for the base substitution errors is misalignment, then the base substitution specificity should switch from T → G to T → A transversions.

In order to determine which base substitution errors at the hot spots (positions T70 and T103) were detectable as light blue or colorless α-complementation plaques when their neighbors (positions 69 and 102) were either G or A, oligonucleotide-directed mutagenesis experiments were performed to establish the phenotypes of all codons of interest. Fortunately, both G → A changes are phenotypically silent, i.e. the resulting M13mp2 plaques are dark blue on X-gal plates, indistinguishable from true wild-type plaques. The altered template can therefore be used in the standard M13mp2 forward mutation assay (14) in which mutants are identified as light blue or colorless plaques. With wild-type DNA (i.e. G69 and G102), all three possible base substitutions were constructed at both sites, and all were found to result in scorable color phenotypes (all light blue). With the pseudo-wild-type derivative (A69A102), all three substitutions were again constructed at both sites. With this template, three of the six possible changes were found to be detectable (also light blue phenotypes). These are T → G and T → A at position 103 and T → G at position 70.

Mutational Specificity of Pol-β at the Hot Spots—To test the dislocation model, gapped templates were prepared (19), using either the original wild-type M13mp2 DNA (to confirm our previous observations performed with different preparations of DNA and pol-β) or M13mp2A69A102 DNA. Pol-β gap-filling synthesis reactions were performed and an aliquot of each reaction was analyzed (14), confirming that the 390-base gap was filled (data not shown). A second aliquot of the products was used to transfect competent cells to score light blue and colorless mutants (18, 19). The overall frequency of mutants produced by pol-β was similar for the two templates (Table I), being only slightly lower with the A69A102 derivative, and was similar to previous measurements of pol-β fidelity (14).

In order to determine mutational specificity at the hot spots, the sequence of the DNA from each of 494 independent and randomly chosen mutants was determined (Table I and Fig. 2). With wild-type DNA the results were essentially as previously described, demonstrating the reproducibility of the original observations. Thus, −T framematches and primarily T → G transversions were produced at high frequency at the two hot spots and the ratio of these two errors differed at the two sites. However, the substitution of A for G at position 102 had a marked effect on base substitution specificity at position 103. Only one T → G transversion was observed with the new template while T → A transversions were now produced at a high frequency. Comparing the two templates at the base substitution hot spot where both errors could have been scored with both templates, the frequency of T → G events decreased more than 30-fold, while the frequency of T → A events increased more than 10-fold. The cumulative effect is thus >300-fold and precisely as predicted by the dislocation model, following pathways F and FB. These frequencies, plus the absence of T → G mutants and −T103, A → G102 double mutants in the pseudowild-type spectrum, suggest that base substitutions are produced at position 103 at least 10–30-fold more frequently by dislocation than by direct miscoding.

TABLE I

| Nucleotide at positions 69 and 102 | G | A |
|-----------------------------------|---|---|
| Number of determinations          | 4 | 249 |
| Mutation frequency ± S.D.*        | 470 ± 81 | 400 ± 92 |
| Mutants sequenced                 | 259 | 233 |

| Mutants | M.F. * | Mutants | M.F. * |
|---------|--------|---------|--------|
| T → G70 | 7      | 13 (12) | 1      |
| −T70    | 77     | 140 (200) | 23     | 35     |
| T → A103 | 0    | <1.9 (<4) | Not scorable |
| −T103   | 4      | 7.4 (8)  | 1      |

* × 10^−4.
position 69 demonstrates that, as expected, simple direct miscoding can contribute to the base substitution spectrum at these sites. The decrease in T→G frequency at position 70 (Table I) also is predicted by the dislocation model. Unfortunately, T→A events are not scorable at this site. The appearance of several T→C transitions at the hot spots and a T→G transversion at position 70 when an A is present at position 69 demonstrates that, as expected, simple direct miscoding can contribute to the base substitution spectrum at these sites.

The mutation frequencies at both hot spots are lower when A rather than G is present at positions 69 and 102. The decrease is 4-fold for −T70 mutations, 5-fold for −T103 mutations, and 2.3-fold for dislocation base substitutions at position 103. The reason(s) for these decreases is not known.

Complex Deletion Mutagenesis by Transient Misalignment—The steps involved in the dislocation process, misalignment, limited incorporation, realignment, and continued incorporation, are not limited to base substitution errors. The original pol−β mutant collection contained a number of deletions (14). One particular deletion was recovered five times and at a frequency >200-fold above the spontaneous background frequency. It was the loss of 317 bases of the 390-base single-stranded gap (Fig. 3A). This included one of two 5-base direct repeat sequences, having the template sequence 5’-C-C-C-G-C-3’ (positions 166–170 in the lacZα coding sequence) and all 312 intervening bases.

The simplest model (Fig. 3B) to explain this mutation involves the following steps: 1) synthesis by pol−β of nine nucleotides starting with the provided 3’-OH primer and proceeding through the first C-C-C-G-C template repeat, 2) disruption of (at least) the five G-C primer-terminal base pairs, 3) rearrangement of the DNA with reformation of five hydrogen-bonded G-C base pairs involving the newly made DNA and the second direct repeat 317 bases downstream. This intermediate may be further stabilized by three G-C base pairs that can potentially form a stem within the resulting loop of the heteroduplex intermediate. Through this step, this mechanism is formally equivalent to forming a frameshift intermediate (pathway F, Fig. 1), but it involves more nucleotides, greater distances, and a larger misaligned heteroduplex. 4) Continued synthesis from the intermediate (equivalent to pathway FF in Fig. 1) produces the heteroduplex which upon transfection and expression of the minus strand yields the 317-base deletion.

In the collection of mutants generated by rat, pol−β was a complex mutant in which 123 bases were deleted, and three base changes were present at one end of the deletion (Fig. 3A). This complex deletion can be precisely explained by a transient misalignment mechanism (Fig. 3C). The model involves four blocks of DNA sequences positioned at three different locations spread over the entire 390-base single-stranded gap. The initial events (steps 1 and 2 in Fig. 3C) are as just described for a simple direct repeat mechanism, but in this case involving two 5’-G-C-C-C-G-3’ sequences 370 bases apart. The first of these repeats is almost the same repeat as for the 317-base deletion, except that one additional base has been incorporated; the repeat is thus at positions 165–169. The resulting intermediate contains a five-G-C base pair primer stem and a 375-base loop, in this case without a stem as just described for a simple direct repeat mechanism, but in this case involving two 5’-G-C-C-C-G-3’ sequences 370 bases apart. The first of these repeats is almost the same repeat as for the 317-base deletion, except that one additional base has been incorporated; the repeat is thus at positions 165–169. The resulting intermediate contains a five-G-C base pair primer stem and a 375-base loop, in this case without a stem involved in this model (Fig. 3B) to explain this mutation.

The 9-fold decrease in T→G frequency at position 70 (Table I) also is predicted by the dislocation model. Unfortunately, T→A events are not scorable at this site. The appearance of several T→C transitions at the hot spots and a T→G transversion at position 70 when an A is present at position 69 demonstrates that, as expected, simple direct miscoding can contribute to the base substitution spectrum at these sites.

The mutation frequencies at both hot spots are lower when A rather than G is present at positions 69 and 102. The decrease is 4-fold for −T70 mutations, 5-fold for −T103
FIG. 3. Deletions by pol-β. A, the top line represents the wild-type sequence with the end points of the 390-base gap indicated by the misaligned intermediate, then realignment, followed by continued synthesis from the second heteroduplex intermediate. The result is that a block of 11 nucleotides has been moved to a position 283 nucleotides distant.

DISCUSSION

The results presented here demonstrate that a template sequence at one position can code for what is observed as new genetic information, i.e., a mutation, at an entirely different position. The dislocation mechanism has implications for a variety of mutational phenomena. The data presented here provide a novel alternative to direct misinsertion to explain a subset of base substitution errors. In addition, just as these results demonstrate that base substitutions can be initiated by misalignments, frameshifts could be initiated by misinsertions, by following pathway B and then BF in Fig. 1A. This could explain frameshifts at nonrun or short run sequences. The interplay between mispaired and misaligned intermediates depicted in Fig. 1 provides one framework for considering multiple mutational endpoints induced by a single DNA adduct, such as frameshifts resulting from miscoding lesions. Furthermore, transient misalignment provides one possible mechanism by which various barriers to DNA synthesis, such as DNA adducts or palindromic sequences, might be bypassed, albeit mutagenically. Each of these possibilities can be examined with existing mutational assays.

Dislocation is only one of several mechanisms for producing errors during DNA synthesis in vitro. Thus far, base substitution mutagenesis by dislocation has been clearly demonstrated at only one template site with one set of reaction conditions and the simplest eukaryotic DNA polymerase known. It does not operate at detectable frequency with more complex polymerases, including the human KB cell DNA polymerase α-primase complex or DNA polymerase γ (15), which contains a proofreading exonuclease (19). Nevertheless, the mechanism was suggested by considering two hot spots that contained 35% of the mutants in the collection (14). While we do not yet know why this site is so prone to dislocation mutagenesis, the site and enzyme specificity demonstrate that both the primary DNA sequence and protein interactions with the template-primer stem are obviously important.

A number of approaches are available to examine the parameters that influence this process, including further engineering of the DNA sequence at the hot spots and measuring polymerase kinetic constants for the intermediates shown in Fig. 1.

Whether the dislocation mechanism operates to generate mutants in vivo remains to be determined. The idea that distant sequences can template mutations during DNA replication, repair, or recombination has been proposed to explain a variety of mutations generated in vivo. For example, the processing of imperfect palindromes (6, 7) is one mechanism that could explain some complex mutations in the E. coli lacZ (30), T4 rII-B (31), and yeast iso-1-cytochrome c (2, 32) genes. The data in Fig. 3 suggest that dislocation during DNA synthesis provides an additional mechanism to explain the origin of certain other mutations recovered in these same genetic systems (2, 11, 32–35). This mechanism could have biological significance for the evolution of mammalian multigene families such as the human interferon genes (36) or for systems requiring rapid generation of diversity such as the arrows. The second line describes the position of the 317-base deletion (dashed line), and the third line describes the position of the complex deletion, with asterisks to indicate the three unexpected bases. B, the direct repeat model for the 317-base deletion. C, the transient misalignment model for the complex deletion.
immunoglobulin genes (37). We have, in fact, already reported that synthesis by terminal deoxynucleotidyltransferase, either alone or in combination with pol-β, produces, through imperfect, misaligned intermediates, complex errors that are similar to sequences in rearranged immunoglobulin genes (38, 39). It will also be interesting to see if any of the mutants presently being generated and recovered from eukaryotic systems in vivo and analyzed by DNA sequencing (9, 10 and, for review, see Ref. 40) are consistent with this mechanism. Even more importantly, it should be possible in the future to directly test this model in vivo.

Acknowledgments—We thank Jan Drake and Roel Schaaper for critical evaluation of the manuscript.

REFERENCES
1. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1968) Cold Spring Harbor Symp. Quant. Biol. 33, 77-84
2. Stewart, J. W., and Sherman, F. (1974) in Molecular and Environmental Aspects of Mutagenesis (Prakash, L., Sherman, F., Miller, M. W., Lawrence, C. W., and Taber, H. W. eds.) pp. 102-127, Charles C. Thomas, Springfield
3. Farabaugh, P. J., Schmeissner, U., Hofer, M., and Miller, J. H. (1978) J. Mol. Biol. 126, 847-863
4. Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., Napoli, C., Brosius, J., Dull, T. J., and Noller, H. F. (1981) J. Mol. Biol. 149, 337-376
5. Albertini, A. M., Hofer, M., Calos, M. P., and Miller, J. H. (1982) Cell 29, 319-328
6. Riplcy, L. S. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4128-4132
7. Glickman, B. W., and Riplcy, L. S. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 512-516
8. Streisinger, G., and Owen J. (1985) Genetics 109, 633-659
9. Nalbantoglu, J., Hartley, D., Phier, G., Tear, G., and Mouth, M. (1986) EMBO J. 5, 1199-1204
10. Ashman, C., and Davidson, R. L. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 3554-3558
11. Shinedler, S., Singer, B. S., Gayle, M., Pribnow, D., Jarvis, E., Edgar, B., and Gold, L. (1987) J. Mol. Biol. 195, 471-480
12. Fresco, J. R., and Alberts, B. M. (1960) Proc. Natl. Acad. Sci. U. S. A. 46, 311-321
13. Fowler, R. G., Degnem, G. E., and Cox, E. C. (1974) Mol. Genet. 133, 179-191
14. Kunkel, T. A. (1985) J. Biol. Chem. 260, 5787-5796
15. Kunkel, T. A. (1985) J. Biol. Chem. 260, 12866-12874
16. Kunkel, T. A., and Alexander, P. S. (1986) J. Biol. Chem. 261, 160-166.
17. Kunkel, T. A. (1986) J. Biol. Chem. 261, 13581-13587
18. Kunkel, T. A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1494-1498
19. Kunkel, T. A., and Soni, A. (1988) J. Biol. Chem. 263, 4450-4459
20. Stalker, D. M., Mosbaugh, D. W., and Meyer, R. R. (1976) Biochemistry 15, 3114-3121
21. Watson, J. D., and Crick, F. H. C. (1953) Nature 171, 964-967
22. Watson, J. D., and Crick, F. H. C. (1953) Cold Spring Harbor Symp. Quant. Biol. 18, 123-131
23. Topal, M. D., and Fresco, J. R. (1976) Nature 263, 285-289
24. Fersht, A. R., and Knill-Jones, J. W. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4251-4255
25. Sinha, N. K., and Haimes, M. D. (1981) J. Biol. Chem. 256, 10671-10683
26. Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 488-492
27. Mosbaugh, D. W., and Linn, S. (1983) J. Biol. Chem. 258, 108-118
28. Soltysk, A., Siedlecki, J. A., Pietrzykowska, I., and Zmudzka, B. (1981) FEBS Lett. 125, 227-230
29. Wang, T. S.-F., and Korn, D. (1980) Biochemistry 19, 1782-1790
30. Todd, F., and Glickman, B. W. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4123-4127
31. deBoer, J. G., and Ripley, L. S. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 5528-5531
32. Das, G., Stewart, J. W., and Sherman, F. (1986) Mutation Res. 163, 233-245
33. Coulondre, C., and Miller, J. H. (1977) J. Mol. Biol. 117, 525-575
34. Ripley, L. S., Clark, A., and deBoer, J. G. (1986) J. Mol. Biol. 191, 601-613
35. Schaaper, R. M., Danfor, B. N., and Glickman, B. W. (1986) J. Mol. Biol. 189, 273-284
36. Golding, G. B., and Glickman, B. W. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 8577-8581
37. Golding, G. B., Gearhart, P. J., and Glickman, B. W. (1987) Genetics 115, 169-176
38. Kunkel, T. A., Gopinathan, K. P., and Soni, A. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1867-1871
39. Snow, E. T., Kunkel, T. A., and Loeb, L. A. (1987) Mutation Res. 180, 137-146
40. Moore, M. M., DeMarini, D. M., deSerres, F. J., and Tindall, K. T. (1987) Banbury Report 28: Mammalian Cell Mutagenesis, CSH Publications, CSH, New York