Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4128–4132, July 1982
Genetics

Model for the participation of quasi-palindromic DNA sequences in frameshift mutation

(DNA secondary structure/base-substitution mutagenesis/DNA strand switching/cytochrome *c* frameshifts/DNA repair)

LYNN S. RIPLEY

Laboratory of Molecular Genetics, National Institute of Environmental Health Science, Research Triangle Park, North Carolina 27709

Communicated by C. Auerbach, March 29, 1982

ABSTRACT A model is described for the templated production of frameshift and base-substitution mutations mediated through aberrant DNA structures arising as a consequence of quasi-palindromic DNA sequences. Two general mechanisms are considered. One evokes the formation and processing of imperfect DNA secondary structures (hairpins) for the production of mutations. The other evokes a "strand switch" during DNA synthesis which, in a manner unique to quasi-palindromic sequences, may be resolved to produce frameshift or base-substitution mutations, or both. It is the unique combination of symmetrical and asymmetrical elements of the quasi-palindromic sequence itself that provides the basis for both models. Through the mechanisms described, the symmetrical elements permit unusually paired DNA substrates, and the asymmetrical elements permit templated insertions, deletions, and base substitutions. The model predicts a class of mutations—simultaneously frameshifts and base substitutions—whose sequences can be predicted from a local quasi-palindromic sequence. This prediction appears to be met by a significant fraction (more than 15%) of frameshift mutations in the iso-1-cytochrome *c* gene of *Saccharomyces cerevisiae*.

Additions or deletions of small numbers of DNA bases resulting in frameshift mutations represent a sizable proportion of spontaneous mutations, and their frequency is often increased by mutagenic treatments. Studies of frameshift mutational mechanisms have focused upon the DNA sequences in which frameshift mutations arise. In many genetic systems, frameshift mutation has been correlated with DNA sequences comprised of repeated bases (1–4). This observation, first made in the T4 lysozyme gene, led to the suggestion by Streisinger *et al.* that repeated sequences mediated additions or deletions by allowing local misalignments of the complementary strands of DNA (1). Such aberrant DNA intermediates, if formed during replication, recombination, or repair, could then be the precursors of frameshift mutations. These repeated DNA sequences may be the simple reiteration of a single base as observed in the lysozyme and *rII* genes of T4 (ref. 5; J. E. Owen, D. W. Shulz, A. Taylor, and G. R. Smith, personal communication) but also may be more complex repeats. For example, a frameshift hot spot in the *Escherichia coli lacI* gene is a sequence of three tandem C-T-G-G units (4). Frameshift mutations arising at this site were found to be the addition or deletion of one C-T-G-G unit and, thus, were consistent with the prediction of the Streisinger model.

Although the Streisinger model successfully predicts the genetic outcomes of frameshift events in a number of DNA sequences, it fails to explain a sizable class of characterized frameshift mutations that arise within DNA sequences which do not provide the potential misalignments upon which the model is based (ref. 2; J. E. Owen, D. W. Shulz, A. Taylor, and G. R. Smith, personal communication). Thus, alternative models are needed to explain these frameshift mutations. This paper describes the general features of a model for frameshift mutations arising in quasi-palindromic DNA sequences and suggests some alternative mechanisms by which these sequences generate mutations. The model successfully predicts the properties of frameshift mutations in the iso-1-cytochrome *c* gene of *Saccharomyces cerevisiae*, whose properties are inconsistent with the Streisinger model.

Complementarity within quasi-palindromic sequences

Palindromic sequences in double-stranded DNA molecules have the inherent property of self-complementarity within each single strand of the DNA. This complementarity permits such sequences to form uniquely paired DNA structures that are impossible in sequences lacking this complementarity; a classical example is the formation of hairpins or cruciform structures. In quasi-palindromic sequences, where complementarity is imperfect, the formation of the unusual DNA configurations still may be possible, but the noncomplementarity at imperfections in the palindrome provide the potential to generate mutations in a templated and, therefore, predictable manner.

Among the diverse detailed mechanisms by which quasi-palindromic sequences might be imagined to produce mutations through aberrantly templated DNA synthesis are two general classes that may be distinguished on the basis of DNA topology. In one class of mechanisms, the aberrant template is provided within a single strand of DNA through the formation of DNA secondary structures (hairpin loops), thus providing a locally double-stranded structure; this case will be referred to as DNA secondary-structure models. In the second class of mechanisms, the aberrant template is provided by the complementary strand of DNA. This strand is utilized as a template if a strand switch occurs during DNA synthesis. Incorporation of this aberrantly templated DNA into the final product can produce templated mutations as well; this class will be referred to as strand-switching models. Strand-switching models may involve, but do not necessarily require, the formation of hairpin loops.

Secondary structure models

An example of a secondary structure that might form in a quasi-palindromic DNA sequence is shown in Fig. 1. Noncomplementary elements within the quasi-palindrome are found within non-hydrogen-bonded regions of the hairpin loop structure. This configuration of the DNA immediately suggests several mechanisms by which the imperfect complementarity might generate mutations. Such mechanisms are described in Fig. 2 with a schematic representation of the specific secondary structure shown in Fig. 1. Regions A and B in both Figs. 1 and 2

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.

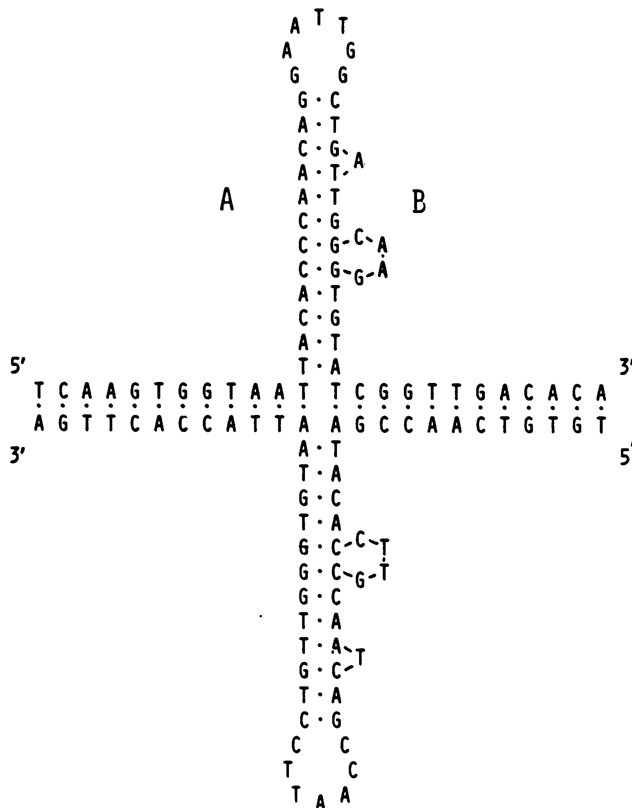

FIG. 1. A quasi-palindromic DNA sequence. This sequence is located in the *rIIB* gene of bacteriophage T4 and includes bases 491 through 550 as reported by Pribnow *et al.* (5). The quasi-palindromic portion is shown in one of its potential hairpin forms and extends from base 503 through base 538. The regions A and B in this figure are topologically the same as regions A and B in the more schematic representation in Fig. 2.

indicate the largely complementary portions of a single strand of the palindromic sequence. However, the A sequence differs from the B sequence by two elements of asymmetry; both elements are extra bases in sequence B. If DNA synthesis in region B is templated by sequence A rather than by the sequence fully complementary to B provided by the opposite strand of DNA, the result (pathway I of Fig. 2) is the deletion of bases in region B that were noncomplementary to sequence A. Alternatively, if DNA synthesis in region A is templated by sequence B rather than by the normal complement to A, addition mutations are produced in sequence A (pathway II of Fig. 2). The added bases in sequence A are complementary to the unpaired bases in sequence B. DNA sequences in which the asymmetry of the palindrome is due to noncomplementary bases rather than to different numbers of bases, template base-substitution mutations by this mechanism. The base substitutions produced are complementary to the formerly noncomplementary bases in the quasi-palindrome. Clearly, quasi-palindromes may contain both unequal numbers of bases and noncomplementary bases and, as will be seen below, permit the concerted production of frameshift and base-substitution mutations.

In addition to providing a template, the secondary structure may also provide a substrate that actually promotes metabolic events ultimately yielding frameshift and base-substitution mutations. For example, unpaired regions of DNA hairpins might be particularly prone to endonucleolytic attack. Such nicking, followed by exonucleolytic removal of unpaired bases and then by DNA synthesis or ligation (or both) templated by the other side of the hairpin, can produce mutations as shown in Fig. 2. This series of steps is clearly similar to that of excision repair acting on ordinary double-stranded DNA. It may be particularly attractive to consider the potential contribution to mutations in quasi-palindromic sequences from enzyme systems that remove mismatched bases from double-stranded DNA (6).

There is no requirement that secondary-structure models be limited to repair events. Intermediates of DNA replication also can be imagined to participate. For example, if during DNA synthesis a few bases at the primer terminus located in a quasi-palindromic sequence should dissociate from its normal template and form a short hairpin structure, extension of that structure through a region of imperfect palindromic sequence could generate a mutation. Another instance might be the joining of the Okazaki fragment intermediates of discontinuous DNA synthesis. If the formation of a secondary structure occurred during templated replacement of an RNA primer in a region of imperfect complementarity of the palindrome, mutations would again be expected.

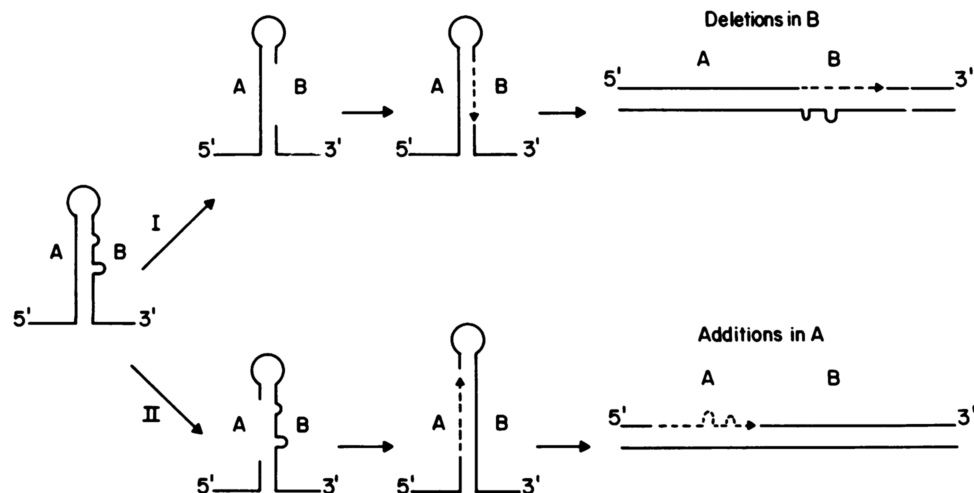

FIG. 2. A general model for producing addition or deletion mutations in quasi-palindromic sequences through the aberrantly templated synthesis of DNA within a secondary structure. The quasi-palindrome is a schematic representation of the specific sequence shown in Fig. 1. Pathway I produces deletions in the B region by the templated incorporation of a sequence complementary to the A region. Pathway II produces additions in the A region by the templated incorporation of a sequence complementary to the B region. ---, Region of DNA synthesis templated from within the hairpin structure.

These secondary-structure models for mutation require the formation of DNA hairpins *in vivo*. The formation of cruciform structures in double-stranded, superhelical DNAs has been detected *in vitro* (7, 8). Furthermore, the formation of hairpins in single-stranded polynucleotides is expected to be energetically favored, at least in the absence of DNA-binding proteins. Both super-helical and partially single-stranded DNAs are expected to form *in vivo* as the result of a variety of DNA metabolic reactions. Secondary structure frequencies at particular quasi-palindromes need not be high, however, because the frequency of mutational events is low. Large differences are anticipated in the ability of quasi-palindromic sequences to form secondary structures.

Strand-switching models

In contrast to secondary-structure models in which a hairpin loop provides an aberrant template for DNA synthesis, strand-switch models utilize the displaced DNA strand (complementary to the template strand) as the aberrant template. The internal complementarity of a quasi-palindromic sequence allows the DNA product of a strand switch occurring within the palindrome to be resolved so as to produce many of the same mutations that might be produced through secondary-structure models. The essential features of strand switching are outlined in Fig. 3. DNA synthesis proceeds ordinarily, with displacement of the complementary strand from the template (Fig. 3, structure I). At an arbitrary point in the DNA sequence (between A and B in Fig. 3, structure II), a strand switch occurs, and DNA synthesis continues. A strand switch is topologically relatively simple (9) and has been invoked to explain the highly branched DNA molecules frequently seen after *in vitro* DNA synthesis with DNA polymerases in the absence of accessory proteins (10, 11).

The resolution or repair of branched DNA molecules created by a strand switch might be accomplished in a number of ways, with various consequences. A simple topological resolution is provided by the process of branch migration, which may be viewed as the sequential rotation (rotary diffusion) of the bases from their pairing positions in Fig. 3, structure II, to their pairing positions in Fig. 3, structure III. (This resolution should always be available, because A′ and A are always complementary.) When followed by removal of the hairpin, no mutation results. An alternative resolution is available to palindromic sequences because of their unusual complementarity (Fig. 3, structure IV). In a palindrome, B′ and A′ are complementary (as are B and A). Thus, the elongating DNA strand may diffuse the replication fork forward rather than backward from the point of the strand switch, allowing the incorporation of DNA sequences templated from the wrong strand into the newly synthesized DNA. In the case of a quasi-palindromic sequence, this process would produce mutations at each element of imperfection in the quasi-palindrome.

When a strand switch occurs at the center of symmetry of a palindrome, mutations are produced at palindromic imperfections when resolved to a structure like that of Fig. 3, structure IV. When strand switches are not precisely central, they produce additions or deletions adjacent to the switch point if resolved to structure IV (Fig. 3). In these cases, resolution to structure IV requires bridging a region of noncomplementarity before the replication fork is moved ahead; thus, resolution to structure III (Fig. 3) might be favored. Some palindromes lack a center of symmetry due to noncomplementarity in the central region. In these cases, strand switches should always produce a mutation if resolved to structure IV.

The conditions that promote strand switching are certainly

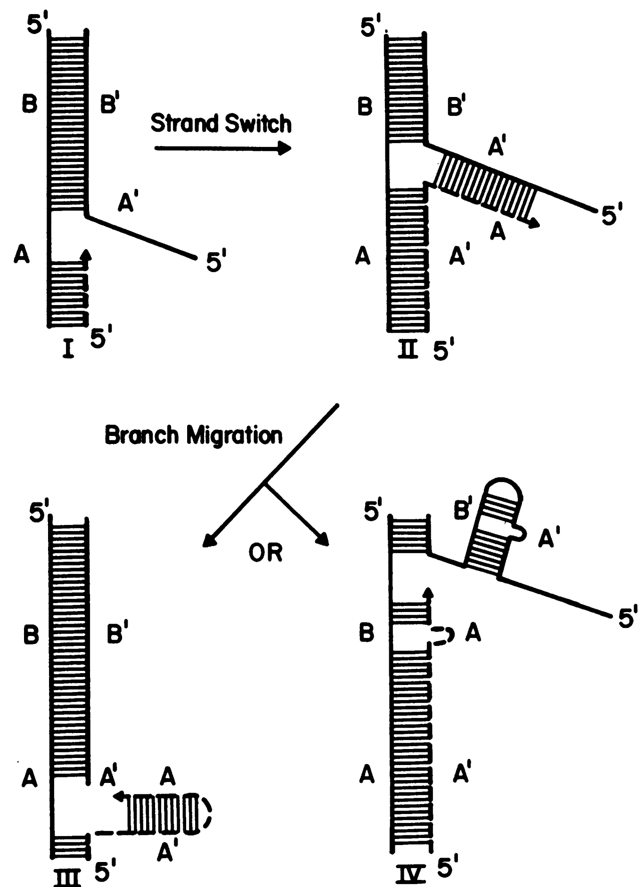

FIG. 3. A model for the production of mutations by a strand switch during DNA synthesis in a quasi-palindromic DNA sequence. Dashed lines represent newly synthesized DNA. The DNA sequences A and B are complementary to A′ and B′, respectively. As DNA synthesis proceeds displacing the complementary strand (structure I), a strand switch may occur leading to incorrect templating of the elongating DNA sequence from the displaced complementary DNA strand (structure II). This structure is subject to isomerization (structure III) brought about by branch migration. Isomerization is permitted because of the complementarity of A to A′ and B to B′. However, if A and B represent complementary symmetry elements of a palindrome, an additional mode of isomerization (structure IV) becomes possible, in which the replication fork is advanced because A and B are exactly complementary. Should this mode of resolution occur in a quasi-palindromic sequence, the result would be mutations at the imperfections of the palindrome. In the example illustrated in structure IV, an addition mutation would be produced in the B region of the DNA due to a few additional bases occurring in the A sequence for which no complementary bases exist in the B sequence.

not well understood. It is intriguing to note, however, that strand switching has been observed *in vivo* in association with a palindromic sequence (12). The strand-switch products were isolated as trapped replicative intermediates of a plasmid, pBR345, when grown in cells treated with chloramphenicol. Characterization of the intermediates demonstrated that they had arisen through a strand switch located in the center of symmetry of a palindromic DNA sequence provided by a pair of head-to-head *lac* operator sequences.

Yeast frameshift mutations and quasi-palindromic DNA sequences

The attractiveness of these models for the participation of quasi-palindromic DNA sequences in templated frameshift mutagenesis would be enhanced if it were to explain heretofore mys-

terious types of frameshift mutations. In fact, the properties of an important class of frameshift mutations characterized in the yeast iso-1-cytochrome *c* gene are consistent with the mutagenic potential of quasi-palindromic sequences.

Many NH_2-terminal *cyc1* frameshift mutations have been isolated as suppressors of other frameshift mutations in the same region (2). Among 36 amino acid sequences determined in those studies, at least 8 clearly could not have resulted from the simple addition or deletion of DNA bases as templated through Streisinger-type mispaired intermediates. The explanation for each of these eight amino acid sequences requires a complex combination of frameshift and base-substitution mutations. Because three of the eight sequences had been isolated more than once, the explanation cannot be merely random insertion of nucleotides into a DNA gap. Furthermore, examples of these eight complex mutations were isolated 16 times among 105 isolates, thus representing a significant fraction of all frameshift mutations detected in this system. Each of these eight types of mutation can be explained on the basis of quasi-palindromic DNA sequences.

The most frequently isolated protein sequence, type 14 from *cyc1-239* (2), was found six times: once in a spontaneously arising revertant, once after α-promoted mutation, and four times after treatment with di(2-chloroethyl)methylamine. Fig. 4 demonstrates how a potential DNA secondary structure predicts the observed protein sequence. The DNA sequence of *cyc1-239* is given first; it is from this sequence that the revertant arose. The protein sequence of the revertant is given next, followed by a DNA sequence that would produce the protein sequence of the revertant and that contains the least number of changes from the DNA sequence of *cyc1-239*. That change is consistent with the deletion of the underlined C-G-G sequence of *cyc1-239* and the synthesis of A-C-C-T in its place, templated from within the secondary structure. (The sequence A-T-C-T is also consistent with the protein sequence but is not the sequence predicted from the quasi-palindrome.)

In a number of these eight examples, the amino acid sequence alone does not strongly predict the particular mutated DNA sequence because many codons are consistent with the observed amino acids. On the other hand, this model for frameshift mutagenesis, based on the templating properties of quasi-palindromic sequences, makes very strong predictions of DNA sequence. Therefore, a determination of the DNA sequences of these yeast *cyc1* frameshift mutations will permit a useful evaluation of the role of quasi-palindromic sequences in the production of these particular frameshift mutations. It may be that these DNA sequences will provide some clues about the participation of secondary-structure mechanisms versus strand-switching mechanisms because, in the case of quasi-palindromic sequences having an asymmetric central sequence, the two mechanisms can generate different mutational outcomes.

Frameshift models and DNA metabolism

The Streisinger model and the quasi-palindromic DNA sequence model for frameshift mutagenesis predict DNA sequences prone to frameshifting and suggest potential DNA intermediates, but the models do not predict the precise enzymology involved in the production of the mutations. Because both models evoke templated DNA synthesis as an intermediate, the DNA polymerase itself may be expected to influence the frequency or the nature of the frameshifts produced. This appears to be the case from our recent studies of the effects of mutant T4 DNA polymerases on frameshift mutations in the *rII* genes (14). I have found that the frequency of frameshift mutations at different sites may be influenced by mutant polymerases in profoundly different ways. For instance, one polymerase allele decreased the frequency of frameshifts at some sites by as much as a factor of 10, while increasing their frequency at other sites by as much as a factor of 40. It is particularly intriguing that the increased mutation frequencies were observed in the DNA region containing the quasi-palindromic

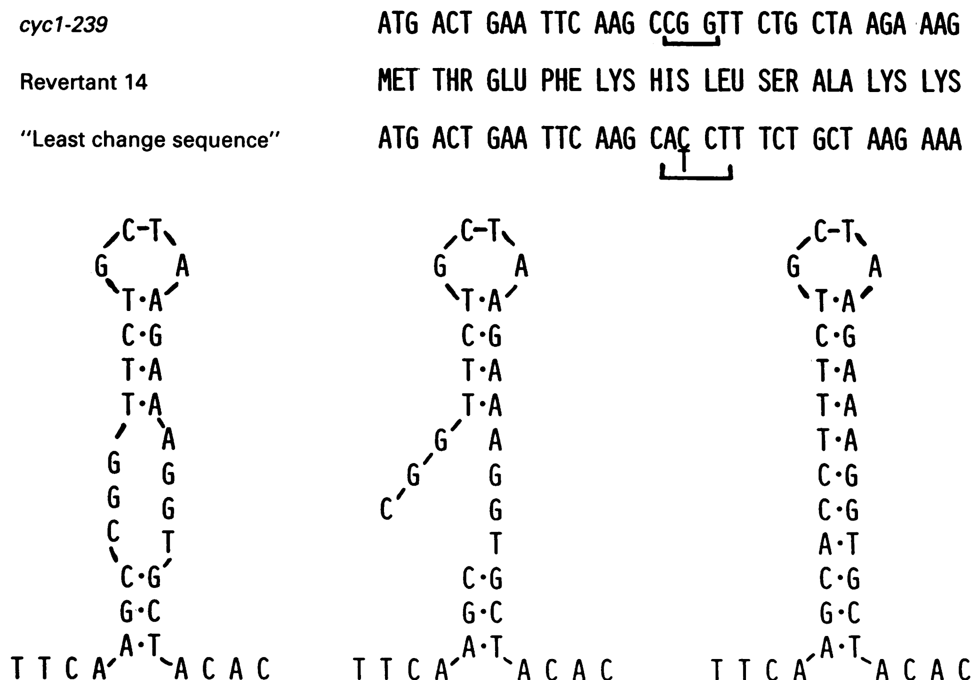

FIG. 4. The quasi-palindromic sequence predicting a frameshift revertant of *cyc1-239*. The DNA sequence of *cyc1-239* is given beginning at the ATG initiation codon (2, 13). The protein sequence (type 14) isolated 6 times among 37 revertants of *cyc1-239* is given below. A DNA sequence showing the fewest number of changes in sequence from that of *cyc1-239* and producing the amino acid sequence of the revertant is shown below the protein sequence. Deletion of C-G-G followed by the insertion of A-C-C-T will generate the revertant. A potential secondary structure within the *cyc1-239* sequence might template this change as shown.

DNA sequence shown in Fig. 1. We are currently determining the sequence of these frameshift mutations in an attempt to evaluate the potential of this quasi-palindrome to produce frameshifts and to understand the influence of different mutant DNA polymerase alleles upon the process.

This discussion of quasi-palindromic DNA sequences and their role in frameshift and base-substitution mutation has focused upon direct-templating models, but there may be other mechanisms by which these sequences can affect the frequency or nature of mutational events. For example, quasi-palindromic sequences may influence the metabolic events involved in UV-induced mutation (15). UV-induced base-substitution hotspots in the *E. coli lacI* gene are all found in quasi-palindromic sequences clustered in a small fraction of the gene (<100 base pairs of DNA). Furthermore, one of these hotspots appeared only in an excision-defective *E. coli* strain. Quasi-palindromic sequences also might produce strong effects upon recombination events. Palindromic sequences have been hypothesized to play a special role in initiating recombination (16), and palindromic recombinational intermediates might well be precursors to mutation by extension of the models presented here. The potential involvement of DNA secondary structures in large addition and deletion mutations and base-substitution mutation will be addressed elsewhere.

Note Added in Proof. The DNA sequence of a type 14 revertant of *cyc1-239* has been determined and shown to be consistent with the model (Fig. 4) (J. Ernst and F. Sherman, personal communication).

I thank B. W. Glickman, whose unquenchable enthusiasm for a role for DNA secondary structures in mutation contributed substantially to this paper and J. W. Drake for his critical reading of the manuscript.

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. Stewart, J. W. & Sherman, F. (1974) in *Molecular and Environmental Aspects of Mutagenesis*, eds. Prakash, L., Sherman, F., Miller, M. W., Lawrence, C. W., & Taber, H. W. (Charles C. Thomas, Springfield), pp. 102–127.
3. Isono, K. & Yourno, J. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1612–1617.
4. Farabaugh, P. J., Schmeissner, U., Hofer, M., & Miller, J. H. (1978) *J. Mol. Biol.* **126**, 847–863.
5. Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., Brosius, J., Dull, T. J. & Noller, H. F. (1981) *J. Mol. Biol.* **149**, 337–376.
6. Glickman, B. W. & Radman, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1063–1067.
7. Lilley, D. M. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6468–6472.
8. Panayotatos, N. & Wells, R. D. (1981) *Nature (London)* **289**, 466–470.
9. Kornberg, A. (1980) *DNA Replication* (Freeman, San Francisco), pp. 153–154.
10. Inman, R. B., Schildkraut, C. L. & Kornberg, A. (1965) *J. Mol. Biol.* **11**, 285–292.
11. Sinha, N. K., Morris, C. F. & Alberts, B. M. (1980) *J. Biol. Chem.* **255**, 4290–4303.
12. Bolivar, F., Betlach, M. C., Heyneker, H. L., Shine, J., Rodriguez, R. L. & Boyer, H. W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5265–5269.
13. Smith, M., Leung, D. W., Gillman, S., Astell, C. R., Montgomery, D. L. & Hall, B. D. (1979) *Cell* **16**, 753–761.
14. Ripley, L. S. & Shoemaker, N. B. (1982) in *Molecular and Cellular Mechanisms of Mutagenesis*, eds. Lemontt, J. F. & Generoso, W. M. (Plenum, New York), pp. 161–178.
15. Todd, P. & Glickman, B. W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4123–4127.
16. Wagner, R. E. & Radman, M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3619–3622.