Adaptive Mutation by Deletions in Small Mononucleotide Repeats

Susan M. Rosenberg,* Simonne Longerich, Pauline Gee, Reuben S. Harris

Adaptive reversion of a +1 frameshift mutation in *Escherichia coli*, which requires homologous recombination functions, is shown here to occur by −1 deletions in regions of small mononucleotide repeats. This pattern makes improbable recombinational mechanisms for adaptive mutation in which blocks of sequences are transferred into the mutating gene, and it supports mechanisms that use DNA polymerase errors. The pattern appears similar to that of mutations found in yeast cells and in hereditary colon cancer cells that are deficient in mismatch repair. These results suggest a recombinational mechanism for adaptive mutation that functions through polymerase errors that persist as a result of a deficiency in post-synthesis mismatch repair.

A new process of mutation has been described that differs in multiple respects from spontaneous mutation in growing cells (*1*). The mutations occur only after exposure to a nonlethal genetic selection (*2–7*), in the apparent absence of cell division (*3–5, 7*), and have been termed adaptive because their formation has been detected in the genes whose function was selected but not in irrelevant genes (*2, 4*) [but see (*7*)]. Adaptive reversion of a *lacZ* frameshift mutation in *E. coli* requires a distinct group of DNA metabolism genes—the genes for homologous recombination, which are not required for growth-dependent mutation (*3, 8*). This requirement implies that a different molecular mechanism generates the adaptive mutations, a mechanism that involves recombination.

Possible recombinational mechanisms for adaptive mutation can be divided into two broad classes (*8*). In the first, templated mutation, blocks of information from preexisting DNA sequences that are similar to those of the mutating gene (templates) are transferred into the gene by recombination (*9–11*). The preexisting templates need be only partially homologous to the mutating gene (*9*). In the second class, nontemplated mutation, recombination is required but the mutant sequences occur de novo (*2, 4, 5, 8, 12, 13*). For example, the mutations could result from errors made either by an RNA polymerase (*2, 12*) or by a DNA polymerase (*8, 13*), with the errors either indirectly (*2, 12, 13*) or directly (*8*) associated with recombination.

These models are distinguishable by the sequences of the reversion mutations that each predicts (*8*). If mutations were templated from somewhat similar sequences, then revertants would usually have the same reversion mutation sequence or a small subset of sequences, because few suitable templates should be available. Also, a distinctive pattern of co-transfer of extraneous nucleotide changes near the reversion mutations could occur if the information-donor template or templates were not perfectly identical to the *lac* gene. In chicken immunoglobulin genes, in which somatic hypermutation occurs by this mechanism, such a pattern has been observed (*14*). If mutant sequences were formed de novo and not templated, then the mutations would be expected to occur anywhere in a region that is capable of restoring the correct reading frame for *lacZ* function. Also, only very rarely would they include any extraneous nucleotide changes. These predictions were distinguished by the sequencing of a 276–base pair (bp) region spanning the *lac*$^-$ +1 frameshift mutation among intragenic *lac*$^+$ adaptive revertants (*15*).

Post-selection reversion mutations of a frameshift mutation (CCC to CCCC) in the *E. coli lacI-lacZ* fusion gene (*3*) were isolated as Lac$^+$ colonies, arising on days 4 through 6 after plating, as previously described (*3, 4, 8*) and purified by streaking on minimal lactose medium (*8*). Figure 1 shows the sequence of a 276-bp region of DNA spanning the frameshift mutation. This region from Lac$^+$ isolates was amplified by polymerase chain reaction (PCR) and sequenced (*16*). Compensatory frameshift mutations capable of restoring function to the *lacZ* gene product would have to occur between the two stop codons boxed in Fig. 1 to avoid creation of a truncated polypeptide chain, stopped in an incorrect reading frame. All of the *lac*$^+$ reversion mutations found are single base deletions in the small mononucleotide repeats highlighted in Fig. 1. Figure 2 shows the distribution of mutations at each site. No extraneous changes were found nearby in any of the adaptive mutants whose *lac* genes were sequenced. The observed spectrum of mutations is characteristic of polymerase errors (*17*).

We previously found that a hyperrecombinagenic *recD* mutant strain is adaptively hypermutable (*8*). The hypermutation in this strain appears to occur by the same mutational pathway as that in *rec*$^+$, because in both strains the hypermutation depends on functional *recA* and *recB* recombination genes (*8*). These dependencies imply that the *recD* mutation does not activate a new

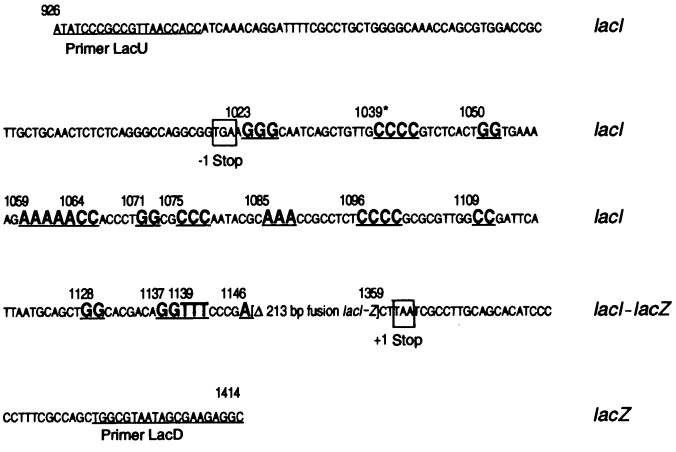

Fig. 1. The DNA sequence of a region of the *lacI-lacZ* fusion gene that was amplified by PCR. The mutant *lac* gene is the one used previously (*3, 4, 8*). The *lacI* gene is fused to *lacZ* at position 1146 so that a frameshift mutation of CCC to CCCC at position 1039 (mutation indicated by an asterisk) is polar on *lacZ*, making the cells Lac$^-$. Intragenic *lac*$^+$ reversion mutations could potentially restore the reading frame by deletion or addition anywhere between the two stop codons (boxed), which occur in incorrect reading frames, and would cause translation stops if a compensatory frameshift mutation were to occur before (first box) or after (second box). The region was amplified by PCR with the LacU and LacD primers indicated (*16*) and then sequenced (*16*). Adaptive revertants each possess only one single base deletion, all occurring at the underlined bases, all but one of which are small mononucleotide repeats. Nucleotide position numbers are those used by Miller (*29*), and the +1 frameshift mutation is not numbered.

S. M. Rosenberg, Department of Biochemistry, University of Alberta Faculty of Medicine, Edmonton, Alberta T6G 2H7, Canada.
S. Longerich and R. S. Harris, Department of Genetics, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.
P. Gee, Division of Biochemistry and Molecular Biology, University of California, Berkeley, CA 94720, USA. Present address: Xenometrics Incorporated, Boulder, CO 80301, USA.
*To whom correspondence should be addressed.

route to the formation of adaptive mutations but more intensively uses the same route that operates in *rec*$^+$ strains. Thus, a correlation exists between the ability to recombine DNA and the ability to generate adaptive mutants: The *recD* cells are hyperrecombinagenic (*18*) and for this reason are hypermutable (*8*). If our argument is correct, then a similar adaptive mutation spectrum would be seen in *recD* as is seen in *rec*$^+$. In *recD*, as in *rec*$^+$, the intragenic adaptive mutations (*15*) are only single base deletions, and nearly all occur in the regions of small mononucleotide repeats (Figs. 1 and 2). Also, as in *rec*$^+$, no extraneous nucleotide changes were found within the region sequenced.

Spontaneous growth-dependent reversion mutations of the same *lac* frameshift mutation do not require recombination functions (*3, 8*) and thus presumably occur by a different mechanism, one that does not require recombination. The idea that the mechanisms are different is supported by the observation of a different and much more heterogeneous mutation spectrum in growing cells (Table 1). Whereas only one sort of adaptive mutation is observed, the growth-dependent mutations include duplications, insertions, and deletions of from 1 to 112 nucleotides, with the smaller deletions not confined to mononucleotide repeats (Table 1). This suggests multiple mechanisms of growth-dependent mutation but a single mechanism of adaptive mutation.

The adaptive mutations that restore Lac$^+$ function to the +1 frameshift mutation in *rec*$^+$ and *recD* strains are all single base deletions, almost all occurring in small mononucleotide repeats, and are not accompanied by nearby extraneous mutations. Mutations such as these are common errors of DNA polymerases (*17*), probably caused by slippage of the newly synthesized strand, which results in mispairing with the template during synthesis (*19*). These results support recombinational models for adaptive mutation that use polymerase error as the mechanism of formation of the mutation. Templated mutation models (*9–11*) are made improbable by these results.

Recombinational models have been proposed in which polymerase error during normal replicative synthesis (*13*) and during unusual DNA synthesis events (*2, 4*) is indirectly associated with recombination. In a model with direct association of synthesis and recombination, we have suggested that repair synthesis associated with recombination could be error-prone (*8*). This mutagenic recombination model is made tenable by observations of mutation that is associated with normal homologous recombination (*20*) or sex (or both) (*21*) in bacteria, yeast, and other fungi. In this model, hyperrecombination would yield hypermutation because of the increased opportunity for faulty synthesis. On the basis of our previous results, we argued that recombination is probably increased because of the presence of double-strand DNA breaks in adaptively mutating cells (*8*). Here we modify this view to include a second parameter—decreased post-synthesis mismatch repair—that would produce the distinctive mutation spectrum observed and would further increase mutation.

The different mutation spectra for growth-dependent and adaptive mutations could be caused if the polymerase generating the adaptive mutations were different from the normal replicative polymerase and

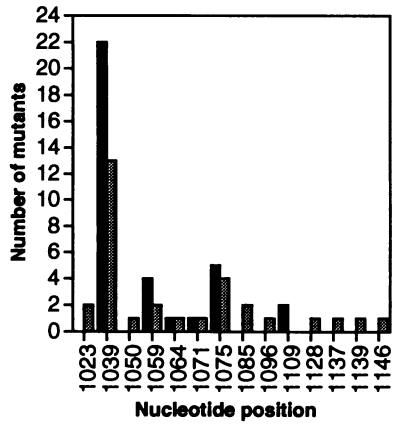

Fig. 2. Positions of adaptive reversion, single base deletion mutations. Nucleotide positions are shown in Fig. 1. Black bars indicate *rec*$^+$; gray bars indicate *recD*.

Table 1. Spectrum of growth-dependent mutations (*30*).

Nucleotide position	Strain	
	rec$^+$	*recD*
955		−73
968†		−92
972		−110
979†	−110	−112*
1014†		−13
1016†	−2 with −1 at 1022*	
1018		−24*
1019†		−1
1021		−4
1022		−1; −7‡, −19
1023	−1	−1
1030†		(−1) × 2
1035†	−1	−1; −43
1039	(−1) × 7	(−1) × 13
1042†		−4‡; CG to (CG)2
1043†	(TC)2 to (TC)3	−1
1044†		−1
1045†		−1
1048†		CT to (CT)2
1058†		−1
1059	−1	(−1) × 2; +2
1061		−13
1064		−1
1067	−1	
1071		−1
1073†		−4
1075	(−1) × 2	(−1) × 5
1078		(−1) × 2
1085		+2; −1
1088		−1
1091	+5	
1092		[(CT)2 to (CT)3] × 2
1095†		−1
1096		−1; +2
1099	+8 (not mononucleotide); [(CG)3 to (CG)4] × 2	
1113	−1	
1117		Duplication of 26 (1093–1117) beginning after 1117
1128		−1
1134†		−1
1137		(−1) × 3
1361	Duplication of 67 (1082–1361) beginning after 1361	

*These deletions do not restore the correct reading frame but do disrupt or delete the first stop codon indicated (Fig. 1) and so are presumably accompanied by a different frameshifting mutation upstream of the sequenced region. †Not part of a mononucleotide repeat. ‡Including bases not in repeat.

made different mistakes more frequently. However, a second, perhaps additional, explanation appears likely in view of the prevalence of the adaptive mutations in simple repeats. Template-slippage mutations in simple repeats are characteristic of yeast cells (*22*) and of hereditary colon cancer cells (*23–25*) that are deficient for post-synthesis mismatch repair. Thus, the hypermutability caused by apparent polymerase errors in *E. coli* cells exposed to selection could also indicate decreased mismatch repair (*26*). Down regulation of mismatch repair during adaptive mutation has been suggested (*6, 9, 27*) but not in a context of recombination-dependent, nontemplated adaptive mutations. In both the cancer cells and in the bacteria exposed to nonlethal selection, the ability to mutate adaptively confers the ability to grow and divide. Mechanistic as well as formal similarities in the two processes may exist (*28*), which raises the possibility of bacterial model systems for mutagenesis in cancer.

REFERENCES AND NOTES

1. S. E. Luria and M. Delbrück, *Genetics* **28**, 491 (1943).
2. J. Cairns, J. Overbaugh, S. Miller, *Nature* **335**, 142 (1988).
3. J. Cairns and P. L. Foster, *Genetics* **128**, 695 (1991).
4. P. L. Foster and J. Cairns, *ibid.* **131**, 783 (1992).
5. P. L. Foster, *Annu. Rev. Microbiol.* **47**, 467 (1993).
6. R. Jayaraman, *J. Genet.* **71**, 23 (1992).
7. B. G. Hall, *Genetics* **125**, 5 (1990).
8. R. S. Harris, S. Longerich, S. M. Rosenberg, *Science* **264**, 258 (1994).
9. P. J. Hastings and S. M. Rosenberg, in *Encyclopedia of Immunology*, I. M. Roitt and P. J. Delves, Eds. (Saunders, London, 1992), pp. 602–605.
10. D. S. Thaler, J. R. Roth, L. Hirschbein, in *The Bacterial Chromosome*, K. Drlica and M. Riley, Eds. (American Society for Microbiology, Washington, DC, 1990) pp. 445–456; D. S. Thaler, *Trends Ecol. Evol.* **9**, 108 (1994).
11. A. Grafen, *Nature* **336**, 525 (1988).
12. E. J. Steele, *Somatic Selection and Adaptive Evolution* (Univ. of Chicago Press, Chicago, 1979); H. S. Rothenfluh and E. J. Steele, *Today's Life Science* **5**, 8 (1993).
13. J. Roth and F. W. Stahl, personal communication; F. W. Stahl, *Genetics* **132**, 865 (1992).
14. N. Maizels, *Cell* **48**, 359 (1991).
15. The frameshift mutation resides in a *lac* operon that is present on an F′ episome in tetracycline-sensitive cells with the chromosomal *lac* region deleted. We identified intragenic reversion mutations by mating the episome into a *lac*-deleted, tetracycline-resistant strain and assaying tetracycline-resistant conjugants for β-galactosidase activity on X-Gal plates. Out of 109 adaptive revertants, none was found to be extragenic. An additional 10 *lac*$^+$ isolates were sequenced directly, without mating, and all but one (a *recD* isolate) were also found, by sequencing, to be intragenic mutations.
16. Synthetic oligonucleotide primers corresponding to the sequence indicated as primer LacU and complementary to the sequence indicated as primer LacD (Fig. 1) (synthesized on an ABI Model 392 Synthesizer, Applied Biosystems, Foster City, CA) were used in asymmetric PCR to amplify single-strand DNA from the 276-bp region for sequencing. The PCR solution consisted of 50 mM KCl, 10 mM tris-HCl (pH 8.6), 2.5 mM MgCl, bovine serum albumin (150 μg/ml), 400 μM deoxynucleotide triphosphates, one primer at 1 μM and the other primer at 0.01 μM, and 1.25 units of Taq polymerase in a total volume of 50 μl, 2 μl of which consisted of bacterial cells suspended in water. These were cycled 40 times at 94°C for 1 min, at 50°C for 1 min, and at 72°C for 1 min in a RoboCycler 40 (Stratagene). The PCR products were purified in G50 Spin Columns (Boehringer Mannheim, Mannheim, Germany) according to the manufacturer's instructions, primers were removed by isopropanol precipitation followed by a 70% ethanol wash, and products were then sequenced either with the Sequenase 2.0 kit (U.S. Biochemicals) according to the manufacturer's instructions or in an ABI Model 373 Sequencer (Applied Biosystems) with the use of dye-terminator nucleotides and with the primer used in the PCR at 0.01 μM used to prime the sequencing reactions.
17. L. S. Ripley, *Annu. Rev. Genet.* **24**, 189 (1990).
18. D. P. Biek and S. N. Cohen, *J. Bacteriol.* **167**, 594 (1986); S. K. Amundsen, A. F. Taylor, A. M. Chaudhury, G. R. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5558 (1986); D. S. Thaler *et al.*, *Genome* **31** (1), 53 (1989); S. M. Rosenberg and P. J. Hastings, *Biochimie* **73**, 385 (1991).
19. G. Streisinger *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77 (1966).
20. C. McGill, B. Schafer, J. N. Strathern, personal communication; M. S. Esposito and C. V. Bruschi, *Curr. Genet.* **23**, 430 (1993).
21. M. Demerec, *Proc. Natl. Acad. Sci. U.S.A.* **48**, 1696 (1962); G. E. Magni and R. C. von Borstel, *Genetics* **47**, 1097 (1962); A. Paszewski and S. Surzycki, *Nature* **204**, 809 (1964).
22. M. Strand, T. A. Prolla, R. M. Liskay, T. D. Petes, *Nature* **365**, 274 (1993).
23. L. A. Aaltonen *et al.*, *Science* **260**, 812 (1993); S. N. Thibodeau, G. Bren, D. Schaid, *ibid.*, p. 816.
24. Y. A. Ionov, M. A. Peinado, S. Malkbosyan, D. Shibata, M. Perucho, *Nature* **363**, 558 (1993).
25. R. Fishel *et al.*, *Cell* **75**, 1027 (1993); F. S. Leach *et al.*, *ibid.*, p. 1215; R. Parsons *et al.*, *ibid.*, p. 1227.
26. Repeat instability in colon cancers has been shown to occur in mono- (*24*) and di-nucleotide repeats (*23*), as deletions only (*24*), and as deletions and insertions (*23*). Our selection for reversion of a +1 frameshift mutation could have revealed +2 insertions in di-nucleotide repeats as well as the −1 deletions in mononucleotide repeats, but the former were not observed among adaptive revertants. This would be expected if deletions were a more common template-slippage error in *E. coli* than were insertions or could be due to the presence of more mononucleotide repeats than di-nucleotide repeats in the region (Fig. 1). In the human cancer cells and yeast cells that are mismatch repair–defective, only deletions were observed in some cell lines (*22, 24*). Thus, the comparison of those systems with ours seems reasonable.
27. F. W. Stahl, *Nature* **355**, 112 (1989).
28. B. S. Strauss, *Cancer Res.* **52**, 249 (1992).
29. J. H. Miller, *A Short Course in Bacterial Genetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992).
30. Growth-dependent Lac$^+$ revertants of the +1 frameshift strain were isolated as described (*8*) (with only one isolate obtained from each culture to avoid duplication in jackpots) and then sequenced (*16*). For mutations greater than 1 bp, position numbers indicate the first nucleotide affected. For mutations that occurred in more than one isolate, the number of isolates is indicated after ×: "× 6" indicates a mutation found in six separate isolates. One *rec*$^+$ isolate has no visible sequence change and so presumably carries an extragenic mutation. It is possible that the few −1 deletions occurring in regions of small mononucleotide repeats are early-appearing adaptive mutations.

31. We are indebted to B. Malcolm for sharing his lab, for discussions, and for facilitating this work, and to J. Elliott for enzymes and advice. We thank R. Kolodner and L. Reha-Krantz for discussions; J. Cairns, P. Foster, J. Haber, P. Hastings, and J. Stone for comments on the manuscript; and C. Thulin for technical assistance. Supported by grants from the Natural Sciences and Engineering Research Council (NSERC), Canada; the Medical Research Council, Canada; and the Alberta Heritage Foundation for Medical Research. Also supported by an NSERC graduate studentship to R.S.H. and a Walter H. Johns Graduate Fellowship to R.S.H. S.M.R. is an Alberta Heritage Medical Scholar.

31 March 1994; accepted 19 May 1994

Adaptive Reversion of a Frameshift Mutation in *Escherichia coli* by Simple Base Deletions in Homopolymeric Runs

Patricia L. Foster* and Jeffrey M. Trimarchi

Spontaneous mutations are thought to occur primarily in growing cells. However, spontaneous mutations also arise in nutritionally deprived cells, and in some cases this process appears to be adaptive. Here it is reported that when a Lac$^-$ strain of *Escherichia coli* is under selection for lactose use, the spectrum of Lac$^+$ mutations that arises is different, and simpler, than that arising without selection. Mutations appearing during selection were mainly one-base deletions in runs of iterated bases. Similar mutations occurring in repetitive DNA elements are associated with a variety of human hereditary diseases and are increased in cells that cannot correct heteroduplex DNA.

The mechanisms by which spontaneous mutations arise in growing cells have been the subject of much research. Two causes of spontaneous mutation that are often cited are intrinsic polymerase errors and endogenous DNA lesions. Indeed, many of the mutations that arise spontaneously in bacteria and in their viruses are the same types of errors as those made by DNA polymerases replicating damaged and undamaged templates in vitro (*1, 2*). However, it is unclear if such mechanisms can account for the mutations that arise among cells that

Department of Environmental Health, Boston University School of Public Health, Boston University School of Medicine, Boston, MA 02118, USA.

*To whom correspondence should be addressed.