Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 2709–2713, April 1988
Genetics

The *mutY* gene: A mutator locus in *Escherichia coli* that generates G·C → T·A transversions

(*lacZ*/*lacI*/spontaneous mutations/repair)

Y. NGHIEM, M. CABRERA, C. G. CUPPLES, AND J. H. MILLER

Molecular Biology Institute and Department of Biology, University of California, 405 Hilgard Avenue, Los Angeles, CA 90024

Communicated by Jon Beckwith, January 11, 1988 (received for review October 6, 1987)

ABSTRACT **We have used a strain with an altered *lacZ* gene, which reverts to wild type via only certain transversions, to detect transversion-specific mutators in *Escherichia coli*. Detection relied on a papillation technique that uses a combination of β-galactosides to reveal blue Lac$^+$ papillae. One class of mutators is specific for the G·C → T·A transversion as determined by the reversion pattern of a set of *lacZ* mutations and by the distribution of forward nonsense mutations in the *lacI* gene. The locus responsible for the mutator phenotype is designated *mutY* and maps near 64 min on the genetic map of *E. coli*. The *mutY* locus may act in a similar but reciprocal fashion to the previously characterized *mutT* locus, which results in A·T → C·G transversions.**

Mutants with higher than normal rates of spontaneous mutation have facilitated our understanding of mutational pathways. Some of the "mutator" strains have characterized defects in postreplication mismatch repair (1, 2), in specific glycosylases (3), or in the editing function provided by the ε subunit of DNA polymerase III (4). Characterization of additional mutators may reveal pathways of mutagenesis and repair. Toward this end, we have used a highly sensitive screening method to detect mutator strains that revert defined mutations in *lacZ* by a limited number of base substitutions. Here we report the characterization of a mutator locus, *mutY*, which results in the specific generation of G·C → T·A transversions.

MATERIALS AND METHODS

Bacterial Strains. All mutators were selected in strain CC503, which contains an F′*lacproB* episome carrying *lacI378*, *lacZ503*. The chromosome of CC503 is *ara*, Δ(*lacproB*)$_{XIII}$, *rpsL*. The series of strains CC101–106 is similar to CC503 but without the *rpsLA* mutation. Each strain carries a different *lacZ*$^-$ allele (see Table 2). The Hfr strains PK191, KL16, and KL14 were a gift of K. B. Low (Yale University). The point of origin of each Hfr is shown in Fig. 5. Markers used for P1 mapping were carried in strains from the *Escherichia coli* genetic stock center, which were generously supplied by B. Bachmann (Yale University School of Medicine). The relevant mutation in each strain is as follows: *metK110*, CGSC6380 (EWH110); *galP*::Tn*10*, CGSC6902 (JM2071); *nupG511*::Tn*10*, CGSC6568 (SO1023); *metC69*, CGSC4524 (AT2699); Δ(*speC-glc*)*63*, CGSC4969 (PL8-31). Also, strain DF649 from D. Fraenkel was provided as KL472 by K. B. Low. This strain carries a Tn*10* insertion (*zgd*) between *metC* and *metK* (see Fig. 5).

Mutagenesis and Selection of Mutants. Strains were mutagenized with ethyl methanesulfonate as described by Coulondre and Miller (5). After mutagenesis and outgrowth in LB medium overnight, strain CC503 was plated on glucose minimal medium with phenyl β-D-galactoside (P-Gal) and 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal) (see below) to observe mutator colonies. Each mutagenesis was monitored for killing and for the generation of rifampicin-resistant (Rif^r) mutants.

Visualization of Mutators. The medium used to visualize blue papillae contained minimal A salts (6), 0.2% glucose, 0.05% P-Gal (500 μg/ml), and X-Gal (40 μg/ml). Plates were incubated 3–5 days and examined under a low-power stereo microscope to visualize mutator colonies.

Genetic Mapping. All Hfr crosses and P1 transductions were carried out as described (6). Selected markers in Hfr crosses were, for PK191 and KL16, *thyA* (introduced by trimethoprim selection; see ref. 6), and for KL14, *metB*, introduced by an Hfr cross with a Met$^-$ Hfr strain. Except for crosses involving *metC*, all P1 transductions used tetracycline resistance (Tet^r) as the selected marker. Purified colonies were subsequently scored for additional markers.

All media and genetic manipulations, unless otherwise stated, are as described (6).

RESULTS

Selection for Mutators. We have characterized an amber (UAG) mutation in *lacZ* at the position encoding tyrosine-503 in β-galactosidase (C.G.C. and J.H.M., unpublished data), which allows reversion to the full Lac$^+$ phenotype only if tyrosine is inserted at that position in the protein. As Fig. 1 shows, only G·C → T·A and G·C → C·G transversions will result in tyrosine insertion, whether the back mutation occurs at the amber site itself or in a tyrosine tRNA gene (resulting in a nonsense suppressor). To find transversion-specific mutators, we mutagenized CC503, a strain carrying the amber at position 503 on an F′*lacproB* episome, and we screened colonies for higher mutation rates by using indicator medium described below.

Detection of Mutators. Mutation rates can be estimated by exploiting the properties of papillae or microcolonies that grow out from the surface of a colony. Papillation, the formation of papillae, is frequently used with indicator plates such as MacConkey, eosin/methylene blue, or tetrazolium (7). We have developed a papillation test that uses glucose minimal plates containing X-Gal and P-Gal, which is often more sensitive than tests on rich media. On the supplemented minimal medium, constitutive Lac$^+$ revertants will form blue papillae growing out of white colonies, as shown in Fig. 2. The colonies grow until they exhaust the glucose in the medium. Then, only constitutive Lac$^+$ revertants, which can utilize the P-Gal in the medium, can continue to grow. The papillae are stained blue by the X-Gal. Colonies that generate papillae at higher rates than normal are readily detected (Fig. 2). After mutagenesis with ethyl methanesul-

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.

Abbreviations: X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; P-Gal, phenyl β-D-galactoside; Rif^r, rifampicin resistant.

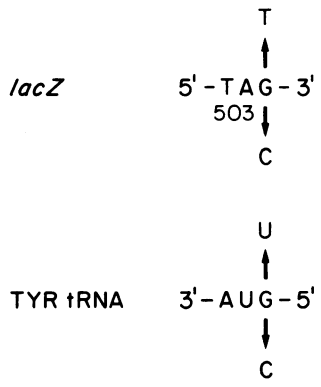

FIG. 1. Specific reversion of *lacZ503*. Effects of an amber mutation corresponding to position 503 in β-galactosidase can be reversed via the G·C → T·A or G·C → C·G transversion, which either recreates a wild-type tyrosine codon (*Upper*) or results in a tyrosine-inserting nonsense suppressor (*Lower*). The suppressor is generated by altering the portion of the tyrosine tRNA gene specifying the tRNA anticodon, as shown here.

fonate blue, we screened ≈50,000 colonies and found 13 strains with increased papillae formation.

Mutagenesis Tests. The 13 strains were retested for papillae formation on several different media and also for the generation of Lac$^+$ and Rif^r revertants in liquid culture. The putative mutators could be placed into four groups based on these tests. Results representative of each group are shown

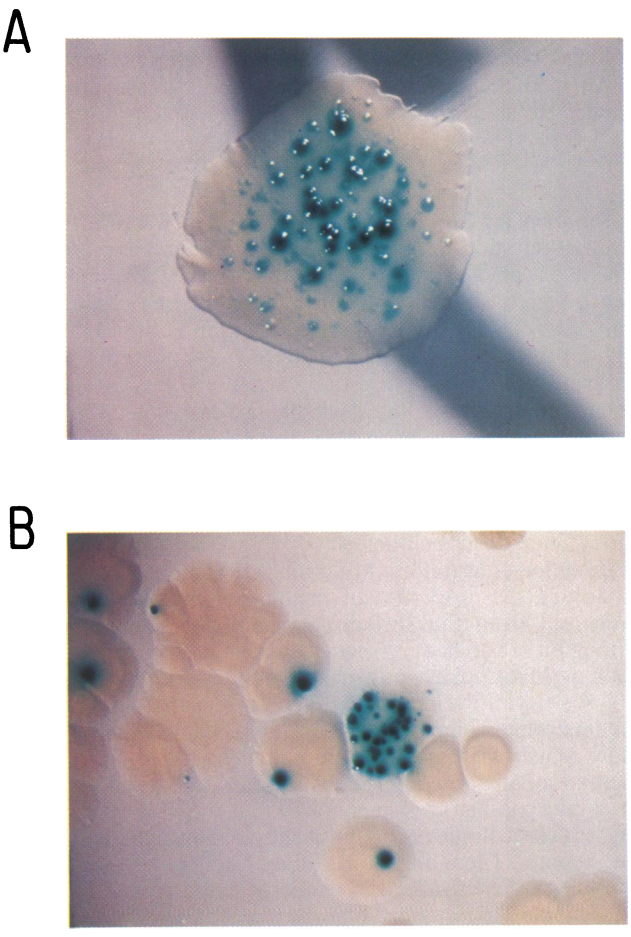

FIG. 2. Papillation tests. (*A*) Reversion to Lac$^+$ results in the formation of blue papillae on medium with X-Gal, P-Gal, and glucose. (*B*) Mutator colonies display more papillae than wild-type colonies.

Table 1. Characteristics of different mutators

Group	Number of mutants	Papillation		Mutants per 10^8 cells	
		X-Gal, P-Gal	Lac Mac	Lac$^+$	Rif^r
I	6	++	++	150–200	150–200
II	1	++	++	500	4000
III	4	+	+	20	5–10
IV	2	+, w	−	5–10	5–10
Control (CC503)				1–2	5–10

The mutators selected from strain CC503 are grouped into four classes based on their papillation response on X-Gal, P-Gal and on lactose MacConkey (Lac Mac) medium and also the number of Lac$^+$ and Rif^r revertants. Papillation is indicated qualitatively. Strong, ++; moderate, +; weak, +, w; no increase over control, −.

in Table 1. The six members of group I and the single member of group II were selected for further study at this point. Initially, we tested the mutators against a series of *lacZ* mutations with alterations at codon 461, which revert to GAG (specifying glutamic acid) via a specific base substitution in each case. Only glutamic acid at position 461 results in the Lac$^+$ phenotype (this system will be described in detail elsewhere; C.G.C. and J.H.M., unpublished data). All members of group I give an identical response in the six strains, and these mutators appear to be specific for the G·C → T·A transversion. Table 2 depicts the results for three representatives of this group. The group II mutator, on the other hand, promotes reversion of five of the six indicator strains at a relatively high rate. Genetic mapping studies (data not shown) indicate that this mutator results from an allele of *mutD*, affecting the ε subunit of DNA polymerase III. We have examined the group I mutators in detail by using the *lacI* system (see below).

Use of *lacI* Nonsense System. To determine the detailed mutagenic specificity of the group I mutators, we used the *lacI* nonsense system (8), which monitors the generation of 90 UAG (amber), UAA (ochre), and UGA mutations at 78 sites in the *lacI* gene. Fig. 3 shows the distribution of nonsense mutations in a wild-type strain. All types of mutations occur, although hot spots for the G·C → A·T transition are seen at 5-methylcytosine residues. The distribution of nonsense mutations in a representative of the group I mutators is markedly different, however, as diagrammed in Fig. 4. It can be seen from both Fig. 4 and Table 3 that only G·C → T·A transversions are stimulated. Of 436 mutations analyzed, 424 result from G·C → T·A transver-

Table 2. Mutational specificity of mutators

Strain donating F′*lacproB*	Base substitution	Lac$^+$ colonies per 10^8 cells				
		Mutator strain				
		Control	Group I			Group II
			a	b	c	
CC101	A·T → C·G	0	0	0	0	611
CC102	G·C → A·T	3	3	4	6	955
CC103	G·C → C·G	0	0	0	0	15
CC104	G·C → T·A	0	159	142	163	1103
CC105	A·T → T·A	0	0	0	0	902
CC106	A·T → G·C	0	0	0	1	527

Six strains (CC101–106) carrying different Z^- alleles on F′*lacproB* episomes. These episomes were donated to different mutator derivatives of CC503 cured of the original episome (Table 1). The number of Lac$^+$ revertants in overnight cultures of each CC503 derivative was monitored. Three strains from group I and one strain from group II (Table 1) were examined. Each value represents the average of several determinations.

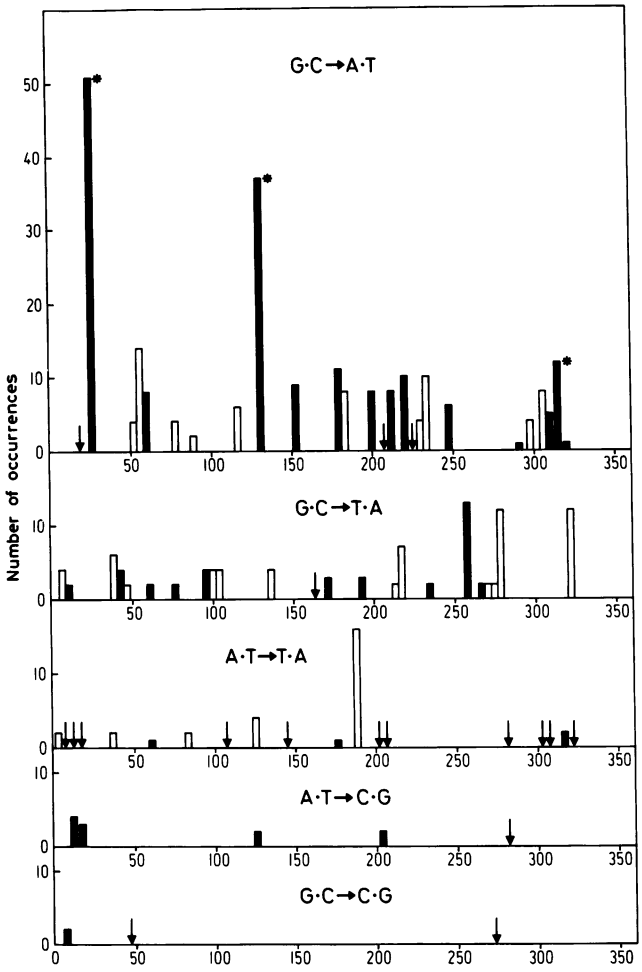

FIG. 3. Spontaneous amber (solid bars) and ochre (open bars) mutations occurring in a wild-type (*mutY*$^+$) strain. The height of each bar represents the number of independent occurrences in a collection of 306 nonsense mutations. (The ochre bar heights have been normalized to account for a smaller sample size.) Arrows indicate the position of nonsense sites at which there were zero occurrences in this collection. Asterisks indicate 5-methylcytosine residues. The position of sites in the *lacI* gene is indicated on the horizontal axis by the number of the corresponding amino acid in the *lac* repressor. (Redrawn from ref. 8.)

sions. We consider these data further in the *Discussion*. A second member of group I gives similar results (data not shown).

At five sites we can monitor both G·C → T·A and G·C → C·G transversions at the same base pair; the former yields TAA in all five cases, and the latter yields TAG in the case of TAC tyrosine codons (third position) and TGA in the case of TCA serine codons (second position). As Table 4 shows, G·C → T·A transversions occurred a total of 49 times at these five sites, compared with zero occurrences of G·C → C·G transversions at the same sites.

Table 3. The distribution of nonsense mutations in a *mutY* strain

Substitution	Available sites	Sites found	Total occurrences	% of total occurrences
G·C → A·T	30	9	11	2.5
G·C → T·A	25	25	424	97.2
A·T → T·A	21	1	1	0.2
A·T → C·G	6	0	0	0
G·C → C·G	5	0	0	0
Total	87	35	436	

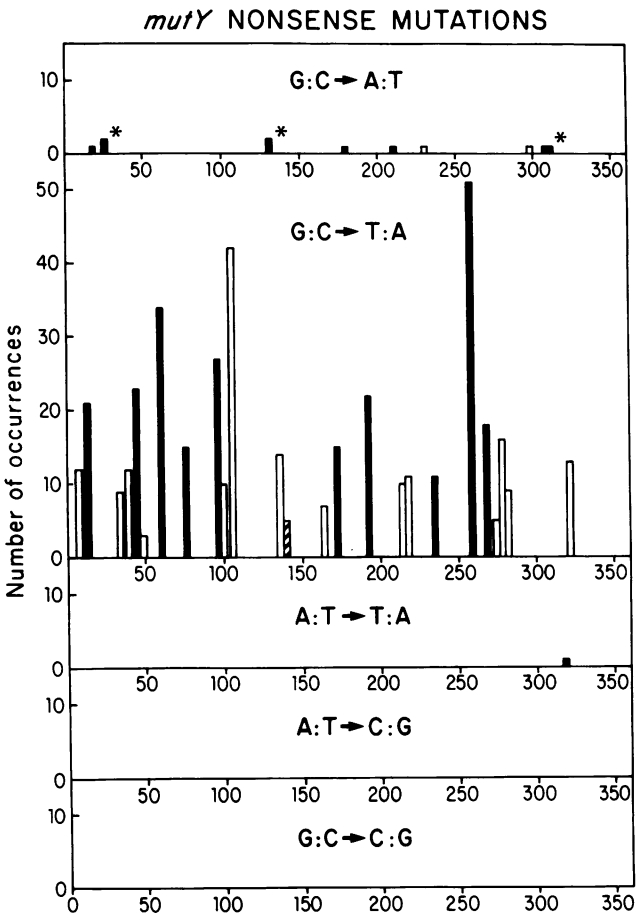

FIG. 4. Distribution of nonsense mutations in *lacI* in a *mutY* strain. Position and number of occurrences of 436 amber, ochre, and UGA mutations are shown. As in Fig. 3, solid bars represent amber mutations and open bars represent ochre mutations. Here, the single UGA site detected is indicated by a striped bar. All bar heights represent the exact number of mutations detected. Each mutation is of independent origin. The mutations have been analyzed as described (9). For further details, see legend to Fig. 3.

Genetic Mapping. We used a series of Hfr strains to determine the approximate position of the locus or loci responsible for the mutator phenotype described above. We selected for markers donated by each respective Hfr and then scored for the mutator character by papillation (see Fig. 2). All of these crosses indicated that the mutator locus mapped between 61 and 66 min on the *E. coli* genetic map (ref. 10; see Fig. 5 *Upper*). P1 cotransduction established

Table 4. Mutations occurring at the three TAC tyrosine and two TCA serine codons in the *lacI* gene

Site	Coding position	Independent occurrences TAC → TAA (G·C → T·A)	TAC → TAG (G·C → C·G)
A1,O3	Tyr-7	12	0
A8,O8	Tyr-47	3	0
A29,O30	Tyr-273	5	0
		TCA → TAA	TCA → TGA
O31/U9	Ser-280	16	0
O36/U10	Ser-322	13	0
	Total	49	0

Both amber (*A*) and ochre (*O*) mutations can occur at the TAC codons and both ochre and UGA (*U*) mutations can occur at the TCA codons, allowing two different transversions to be monitored at each site (see text).

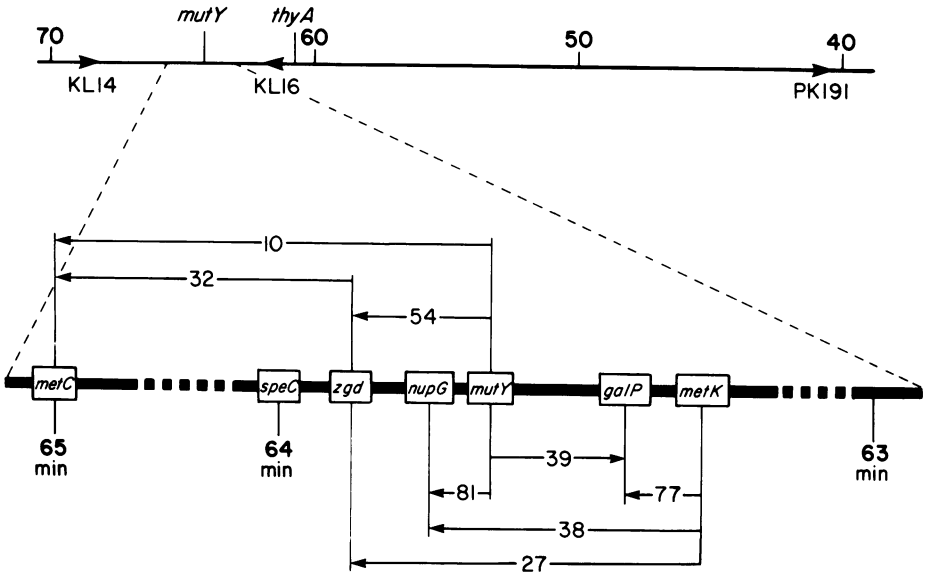

FIG. 5. Genetic mapping of *mutY*. (*Upper*) Point of origin of Hfr strains used to position *mutY*. (*Lower*) P1 cotransduction frequencies are given as percentage cotransduction with the arrow pointing to the selected marker. See *Materials and Methods* for a description of each mutation used. The cotransduction frequencies represent the average of several experiments. Most ordering was achieved by three-factor crosses. [The relative position of *speC* (10) is inferred from additional experiments; data not shown.] The map is not drawn to exact physical scale.

that the locus, which we designate *mutY*, maps very near *nupG*, between *metC* and *metK*, at ≈64 min. Fig. 5 (*Lower*) depicts the map location together with relevant surrounding loci.

DISCUSSION

We have developed a system that is capable of detecting mutators specific for certain base substitutions (C.G.C. and J.H.M., unpublished data). This system relies on the use of specific alleles of *lacZ* (Fig. 1) and also on a sensitive visualization assay involving histochemically stained papillae (Fig. 2). Here we report the characterization of a mutator locus, *mutY*, which maps at ≈64 min on the *E. coli* genetic map.

Strains that are *mutY* stimulate only G·C → T·A transversions (at least to the limits of our detection systems). This can be seen not only by examining the *lacZ* reversion system (Table 2) but also by using the *lacI* forward nonsense system (Fig. 4 and Table 3). The *lacI* system monitors 90-base substitutions at 78 different nonsense sites. Nonsense mutations constitute close to 30% of all *lacI* mutations generated by *mutY*. From 436 nonsense mutations, all but 12 were at one of the 25 nonsense sites arising from the G·C → T·A transversion. Moreover, a system capable of detecting frameshifts and deletions (23) indicates that *mutY* does not induce either of these types of mutations (data not shown).

mutY is highly specific compared to most mutators. The only other mutator that stimulates a specific transversion is *mutT*, which generates only A·T → C·G transversions (11, 12, 24). Both G·C → T·A and A·T → T·A transversions are stimulated by the SOS mutator effect (13). Among the G·C → T·A transversions in *lacI*, *mutY* displays a different site specificity than the SOS system (13). Strains defective in the mismatch repair system (*mutH, mutL, mutS, uvrD*, and *dam*) preferentially stimulate transitions and some frameshifts (9, 12, 14, 15), those lacking the editing function of DNA polymerase III (*mutD*) stimulate all types of base changes, although transitions predominate (12, 16) and cells lacking uracil DNA-glycosylase (*ung*) generate G·C → A·T transitions (17). It is not clear at this point whether an uncharacterized mutator locus, *mutB*, in *Salmonella typhimurium*, which maps between 62 and 72 min (18), is the counterpart of the *mutY* locus described here.

Several possible explanations for the specificity of *mutY* readily come to mind. The *mutY* locus might encode a glycosylase that excises spontaneously damaged guanine residues. Altered guanines might fail to pair properly and translesion synthesis could result in the preferential insertion of an adenine residue across from the damaged base, as appears to occur for a number of other lesions (19–21). However, the specificity appears far too tight to accommodate this type of mechanism. A second hypothesis is based on the presumed action of *mutT*, the one other transversion-specific mutator. Fresco and Topal (22) have proposed that replication errors that result in transversions might occur by the incorporation of the *syn* isomer of guanine or adenine across from either of the tautomers imino-adenine or *enol*-imino-guanine. The *mutT* locus might control a protein that normally prevents the incorporation of *syn*-guanine across from imino-adenine (12, 22), thus explaining its unidirectional action and predicting the existence of a similar protein that prevents the incorporation of *syn*-adenine (12). The *mutY* locus would be an excellent candidate to encode such a protein. (It is interesting to note that the levels of increased spontaneous mutations displayed by *mutT* and *mutY* are on the same order of magnitude.) One problem with this idea in its simplest form is that *mutT* might be expected to also cause G·C → C·G transversions if it were simply preventing the incorporation of *syn*-guanine (12), and *mutY* would be expected to also cause A·T → T·A transversions if it were just blocking the incorporation of *syn*-adenine. Perhaps additional complexities exist for this putative system. An exciting prospect would be to clone and sequence the *mutY* gene and compare it with the sequence of the *mutT* gene. Enzymes with functions as closely related as those postulated above might have recognizable homology.

A more straightforward model would involve a postreplication repair system specific for correcting mismatches leading to G·C → T·A transversions. Evidence for a repair pathway that corrects G·A mispairs to G·C base pairs in a reaction that is independent of *mutH, -L, -S* gene products has been observed in *E. coli* extracts (S.-S. Su, R. S. Lahue, G. Au, and P. Modrich, personal communication). There is

as yet no evidence for such a pathway operating *in vivo*. A tantalizing possibility is that *mutY* affects one of the components of this putative pathway.

Of course, other mechanisms, such as effects on precursor pools, might be involved in the specific transversions seen with *mutY*.

We thank Drs. B. Bachmann and K. B. Low for bacterial strains, and Drs. P. Modrich, P. Foster, E. Eisenstadt, E. C. Cox, and J. R. Fresco for helpful discussions. This work was supported by a grant from the National Institutes of Health (GM32184) to J.H.M. C.G.C. was the recipient of a fellowship from the Natural Sciences and Engineering Research Council of Canada.

1. Radman, M., Wagner, R. E., Glickman, B. W. & Meselson, M. (1980) in *Progress in Environmental Mutagenesis*, ed. Alacevic, M. (Elsevier/North-Holland, Amsterdam), pp. 121–130.
2. Lu, A.-L., Welsh, K., Clark, S., Su, S.-S. & Modrich, P. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 589–596.
3. Duncan, B. K., Rockstroh, P. A. & Warner, H. R. (1978) *J. Bacteriol.* **134**, 1039–1045.
4. Scheuermann, R., Tann, S., Burgers, P. M. J., Lu, C. & Echols, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7085–7089.
5. Coulondre, C. & Miller, J. H. (1977) *J. Mol. Biol.* **117**, 525–567.
6. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
7. Konrad, E. B. (1978) *J. Bacteriol.* **133**, 1197–1202.
8. Coulondre, C. & Miller, J. H. (1977) *J. Mol. Biol.* **117**, 577–606.
9. Leong, P.-M., Hsia, H. C. & Miller, J. H. (1986) *J. Bacteriol.* **168**, 412–416.
10. Bachmann, B. J. (1983) *Microbiol. Rev.* **47**, 180–230.
11. Cox, E. C. (1973) *Genetics* **73**, Suppl., 67–80.
12. Cox, E. C. (1976) *Annu. Rev. Genet.* **10**, 135–156.
13. Miller, J. H. & Low, K. B. (1984) *Cell* **37**, 675–682.
14. Choy, H. E. & Fowler, R. G. (1985) *Mutat. Res.* **142**, 93–97.
15. Glickman, B. W. (1979) *Mutat. Res.* **61**, 153–162.
16. Fowler, R. G., Degnen, G. E. & Cox, E. C. (1974) *Mol. Gen. Genet.* **133**, 179–191.
17. Duncan, B. & Miller, J. H. (1980) *Nature (London)* **287**, 560–561.
18. Shanabruch, W. G., Behlau, I. & Walker, G. C. (1981) *J. Bacteriol.* **147**, 827–835.
19. Loeb, L. A. & Preston, B. D. (1986) *Annu. Rev. Genet.* **20**, 201–230.
20. Eisenstadt, E., Warren, A. J., Porter, J., Atkins, D. & Miller, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1945–1949.
21. Foster, P. L., Eisenstadt, E. & Miller, J. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2695–2698.
22. Fresco, J. R. & Topal, M. D. (1976) *Nature (London)* **263**, 285–293.
23. Miller, J. H. (1985) *J. Mol. Biol.* **182**, 45–68.
24. Yanofsky, C., Cox, E. C. & Horn, V. (1966) *Proc. Natl. Acad. Sci. USA* **55**, 274–281.