JOURNAL OF BACTERIOLOGY, July 1986, p. 130–137
0021-9193/86/070130-08$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 167, No. 1

Characterization of Mutational Specificity within the *lacI* Gene for a *mutD5* Mutator Strain of *Escherichia coli* Defective in 3′ → 5′ Exonuclease (Proofreading) Activity

ROBERT G. FOWLER,†* ROEL M. SCHAAPER, AND BARRY W. GLICKMAN‡

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

Received 24 October 1985/Accepted 3 March 1986

The *mutD* (*dnaQ*) gene of *Escherichia coli* codes for the epsilon subunit of the DNA polymerase III holoenzyme which is involved in 3′ → 5′ exonuclease proofreading activity. We determined the mutational specificity of the mutator allele, *mutD5*, in the *lacI* gene of *E. coli*. The *mutD5* mutation preferentially produces single base substitutions as judged from the enhanced fraction of *lacI* nonsense mutations and the spectrum of sequenced dominant *lacI* (*lacI*^{d}) and constitutive *lacO* (*lacO*^{c}) mutations which were predominantly (69/71) single nucleotide substitutions. The distribution of amber *lacI* and sequenced *lacI*^{d} mutations revealed that transitions occur more frequently than transversions. A · T→G · C and G · C→A · T transitions were equally frequent and, with one major exception, evenly distributed among numerous sites. Among the transversions, A · T→T · A events were the most common, A · T → C · G substitutions were rare, and G · C→C · G changes were not detected. Transversions were unequally distributed among a limited number of sites with obvious hotspots. All 11 sequenced transversions had a consensus neighboring sequence of 5′-C-C-(mutated G or A)-C-3′. Although no large deletions or complex mutational events were recovered, sequencing revealed that *mutD5* induced single nucleotide deletions within consecutive G · C sequences. An extraordinary A · T → G · C transition hotspot occurred at nucleotide position +6 in the *lac* operator region; the *mutD5* mutation frequency of this single base pair was calculated to be 1.2×10^{-3}.

The high fidelity of DNA replication in *Escherichia coli* is under genetic control. One gene whose product is essential for accurate DNA replication is *mutD* (*dnaQ*), which is located at about 5 min on the *E. coli* genetic map and is closely linked to the structural gene for RNase H, *rnh* (10, 22, 23). Two well-characterized mutant alleles of this gene, the dominant *mutD5* and the recessive *dnaQ49*, are strong mutators and raise spontaneous mutation frequencies to 10^3 to 10^5 times wild-type levels (10, 23, 29). The *mutD5* allele is conditional, with its phenotype dependent upon growth conditions (10). In rich medium such as L-broth or in minimal medium supplemented with thymidine, *mutD5* strains display extremely high mutation frequencies, whereas in minimal medium without thymidine, mutation frequencies are much lower (50 to 100 times wild-type frequencies). While the nature of this physiological conditionality remains incompletely understood, it has been shown that the thymidine effector which enhances the mutagenic activity of *mutD5* strains must be phosphorylated (14).

The *mutD* gene product is a 25 to 28-kilodalton protein (9, 22, 27) which is also produced by *mutD5* mutator strains (9, 29). Since the dominant *mutD5* allele exhibits negative complementation in *mutD5*/*mut*$^+$ merodiploids, it has been suggested that the *mutD* protein is part of a multimeric structure which could be at the replication fork (9). Recently it has been shown that *mutD5* and *dnaQ49* mutator strains have reduced levels of the DNA polymerase III 3′ → 5′ exonuclease (proofreading) activity (11, 13) and that the *mutD* gene product is the epsilon subunit of DNA polymerase III (11, 44). The *mutD* protein, the epsilon subunit, catalyzes the error-removing 3′ → 5′ exonuclease activity (43).

One approach toward discerning the role of the 3′ → 5′ exonuclease editing activity in mutation avoidance is to characterize the mutational spectra in strains deficient in the proofreading activity, i.e., *mutD5* strains. Knowledge of the exact molecular nature and precise location of a large number of mutations caused by the lack of normal 3′ → 5′ exonuclease activity should provide insights into the proofreading process. Earlier work with the *trpA* reversion system (50) had demonstrated that *mutD5* strains were generalized mutators since all possible base-pair substitution and frameshift mutations were induced (17). However, the *trpA* reversion system provides only a limited number of sites at which mutational events can be observed, and complex mutational events cannot be detected at all. A more complete characterization of the *mutD5* mutational spectrum is needed for a fuller understanding of the role of the editing activity of the DNA polymerase III 3′ → 5′ exonuclease in replication fidelity. In this report we present the mutational specificity of the *mutD5* allele derived from two distinct approaches. A mutational spectrum of amber *lacI* mutations was produced based upon genetic analysis as described by Coulondre and Miller (5, 6) by which base substitutions at 36 sites within the gene can be scored. In addition, a mutational spectrum was produced by the DNA sequencing of *lacI*^{d} (dominant *lacI*) mutations. This approach made use of a recently developed method for the rapid cloning and sequencing of *lacI* mutants (41).

* Corresponding author.
† Present address; Department of Biological Sciences, San Jose State University, San Jose, CA 95192.
‡ Present address: Biology Department, York University, Toronto, Ontario M3J 1P3, Canada.

TABLE 1. Influence of *mutD5* on spontaneous *lacI* mutation frequencies

Strain	*lacI* mutations		Amber mutations[a]		Ochre mutations		Total nonsense mutations	
	Frequency (10^{-6})	*mutD5*/*mut*$^+$	Frequency (%) (10^{-8})	*mutD5*/*mut*$^+$	Frequency (%) (10^{-8})	*mutD5*/*mut*$^+$	Frequency (%) (10^{-8})	*mutD5*/*mut*$^+$
NR8093	7,900	2,800	67,000 (8.5)	22,000	46,000 (5.8)	21,000	113,000 (14.3)	21,000
NR3835	2.8		3.1 (1.1)		2.2 (0.8)		5.3 (1.9)	

[a] A total of 4,491 independent *lacI* mutants were screened for nonsense mutations in strain NR8093. The nonsense mutation data for strain NR3835 (*mut*$^+$) are taken from a previous study (25) in which a total of 61,487 *lacI* mutants were screened.

MATERIALS AND METHODS

Strains and media. Unless otherwise stated, the *E. coli* suppressor and deletion strains, as well as the media and techniques for the *lacI* system, were the same as those previously described (5, 6). The *mutD5* strain, KH1214 (F$^-$*zaf-13*::Tn*10 mutD5*), was kindly supplied by D. Horner and E. Cox (8). The *mutD5* allele was transferred from KH1214 to NR8036[F$^-$Δ(*pro-lac*) *ara thi trpE9777*] (20) by P1 transduction (32). Selection for *mutD5* was indirect via selection for tetracycline resistance (carried by Tn*10*). The presence of the closely linked *mutD5* allele in recipient cells was subsequently established by the high frequency of nalidixic acid-resistant mutations produced by *mutD5* mutator strains. Strain NR8093 [F′ *pro-lac* Δ(*pro-lac*) *ara thi trpE9777 mutD5* Tn*10*], used for the selection of *lacI* mutants (see below), was then constructed by F′ transfer from NR3835 [F′ *pro-lac* Δ(*pro-lac*) *ara thi trpE9777 mut*$^+$]. Selection was for *pro*$^+$ Tet^r conjugants, and one such isolate with strong mutator activity, NR8093, was used in this study. The F′ *lac* contained in this strain carries the *lacI*^{q} and the *lacZ* L8 promoter mutants commonly used in *lacI* mutational studies (5). The construction of the recombinant M13 phage vector, mRS81 (carrying *lacI*$^+$ and a point mutation which prevents alpha-complementation [*lacZ*$_\alpha$]), and the alpha-complementing host NR9099 [Δ(*pro-lac*) *ara thi recA56*/F′ (*pro-lacIZΔM15*)] used for the DNA sequencing of *lacI* mutations has been described in detail (41). The *lacI* mutants were selected and scored on minimal medium supplemented with the noninducing sugar phenyl-β-D-galactoside as the sole carbon source (32).

***lacI* mutant selection.** To select spontaneous *lacI* mutants, cells of strain NR8093 (F′ *lac*$^+$/*mutD5*) in exponential phase were inoculated into wells of 96-well microtiter plates containing 0.2 ml of L-broth (approximately 10 cells per well) and incubated at 37°C overnight as previously described (25). Samples of 10 μl were withdrawn from each well, plated on phenyl-β-D-galactoside medium, and incubated at 37°C for 60 h. Only one mutant per culture was selected to insure the independent origin of the *lacI* mutants used to create the mutational spectra.

***lacI* mutation frequencies.** The *lacI*$^+$ to *lacI* mutation frequencies for strains NR8093 (*mutD5*) and NR3835 (*mut*$^+$) were determined by inoculating six 5-ml L-broth cultures for each strain with a small number (200 to 800) of cells. The cultures were grown at 37°C to saturation (24 h). The cells were then plated on tryptone medium (32) to titrate for total cells and on phenyl-β-D-galactoside medium to determine the number of *lacI* mutants. Mutation frequencies were calculated by dividing the number of *lacI* mutants by the number of total cells, and an average value over six cultures was determined for each strain.

Identification and mapping of nonsense mutations. Nonsense mutations were scored and identified as previously described (5, 6).

Transfer of *lacI* mutations to M13 *lac*. A procedure has been recently developed to efficiently transfer *lacI* mutations from F′ *lac* onto a single-stranded M13 phage vector (41), and this method was used to prepare the *mutD5*-induced *lacI* mutations for DNA sequencing. Only *lacI*^{q} mutations were sequenced. These mutations were identified by their *trans* dominance to the *lacI*$^+$ allele when crossed by replica mating into strain CSH52 [Δ(*lac-pro*) *ara rpsL thi recA* (φ80 d*lac*$^+$)] (32). Concurrently, the *lacI* mutants were crossed into strain S90C [Δ(*pro-lac*) *ara thi rpsL*] (41) to permit recovery of *lacI*^{q} mutations on the recombinant M13 phage mRS81. This vector carries the *lacI* gene as well as the adjacent alpha-complementation region of the *lacZ* gene. The *lacZ* gene fragment contains a single point mutation which destroys its alpha-complementing ability. During the infection cycle, homologous recombination takes place (usually at a frequency of 10^{-3} to 10^{-4}) between this *lacI*$^+$*Z*$_\alpha$ phage and the resident F′ *lacIZ*$^+$ carrying the *lacI* mutation of interest, which can result in the transfer of both the *lacI* and *lacZ*$^+$ genes from the F′ to the M13 vector, producing an M13 *lacIZ*$^+_\alpha$ recombinant. These recombinants can be identified by subsequent plating on alpha-complementation host NR9099 [Δ(*pro-lac*) *recA56*/F′ *lacIZΔM15*] (41). The presence of the recombinant *lacIZ*$^+_\alpha$ phage in the *lacIZΔM15* host permits alpha-complementation in the absence of the inducer isopropyl-β-D-thiogalactose (blue plaques when plated on medium containing X-gal). The parental *lacI*$^+$*Z* phage and other recombinant forms produce white plaques under these conditions. The selection for the double change from *lacI*$^+$*Z*$_\alpha$ to *lacIZ*$^+_\alpha$ practically eliminates the possibility of selecting a *lacI* phage which originated by spontaneous mutation in the phage rather than by recombinational exchange.

DNA sequence analysis. Purification of the M13 *lacIZ*$^+$ plaques and preparation of single-stranded M13 DNA were performed as previously described (31). DNA sequencing was carried out by the dideoxy chain termination method (40) on single-stranded M13 DNA preparations. Three synthetic oligonucleotides complementary to specific positions along the *lacI* and *lacZ* gene sequences (prepared by P-L Biochemicals) were used as primers: two 14-mers with their 3′ ends at positions 148 and 302 of the *lacI* gene, and the hybridization probe primer (17-mer; P-L Biochemicals) starting at position 27 of the *lacZ* sequence. The resulting sequences were compared with the wild-type *lacI* sequence (15) or the *lacZ* regulatory region (38).

RESULTS

***lacI* mutagenesis in a *mutD5* strain.** We used strains NR8093 (*mutD5*) and NR3835 (*mut*$^+$) for a comparison of their *lacI* mutation frequencies and mutational spectra. *mutD5* caused an increase in the *lacI* mutation frequency of over 2,000-fold compared with the *mut*$^+$ strain (Table 1). Moreover, the percentage of nonsense mutations among the

mutD5 lacI mutations (Table 1) increased seven-to-eightfold compared to the *mut*$^+$ control. The low percentage of nonsense mutations among spontaneous *lacI* mutations in the *mut*$^+$ strain is the consequence of most *lacI* mutations being the result of frameshift and deletion events (16). Nonsense mutations were enhanced over 20,000-fold by *mutD5* (Table 1). Since amber (TAG) and ochre (TAA) nonsense codons result from base-pair substitutions, it can be concluded that base-pair substitutions are readily created by *mutD5* mutator activity. It must be noted, however, that whereas the increase in proportion of nonsense mutations in the *mutD5* strain may indicate a preference for base-pair substitution mutations, this does not preclude the possibility that this potent mutator also causes other classes of mutational events.

mutD5-induced amber mutational spectrum of the *lacI* gene. We have used the *lacI* system (5) to determine the mutational specificity of *mutD5*. In this system 36 sites have been identified where a single base substitution results in an amber codon. All four transversions can be detected as well as G · C → A · T transitions. Only the A · T → G · C transition event cannot be scored. This is because an A · T → G · C event cannot convert a sense codon into an amber codon (TAG).

The distribution of *lacI*(Am) mutations induced by *mutD5* is shown in Table 2. Even though more transversion sites (22) are available than transition sites (14), the majority (69%) of *mutD5*-induced base-pair substitutions are G · C → A · T transitions. They are relatively evenly distributed over the 14 transition sites, being absent only at the A34 site (Table 2). Because of the very high *mutD5 lacI* mutation frequency, 20,000-fold above background, it can be assumed that most, if not all, of the *lacI* mutations recovered actually result from the *mutD5* mutator activity. Indeed, the spectrum is quite different from that seen in *mut*$^+$ strains. In the *mut*$^+$ spectrum about 60% of the amber transitions are produced at three hotspots, A6, A15, and A34, which result from the spontaneous deamination of 5-methylcytosine to thymine (7). The *mutD5* transition distribution is characterized by the absence of highly predominant hotspots, although not all sites are equally recovered; more transition occurrences are found at A15 than any other single site.

In contrast to the *mutD5*-induced transitions, the transversions are distributed unequally over the 22 available sites. The A18 and A36 sites, where A · T → T · A transversions are scored, contributed 68% (23/34) to the transversion population (Table 2); one-half (17/34) of the total transversions occurred at the A36 site. Transversions were not recovered at one-half (11/22) of the available sites, and 7 other sites had only 1 occurrence. Most transversions (31/34) resulted in the insertion of an A · T base pair at the mutated site. Only 9% (3/34) involved the insertion of a new G · C base pair. No G · C → C · G transversions were recovered at the three available sites.

Specificity of *mutD5*-induced mutagenesis as determined by DNA sequencing. To characterize more fully the mutational spectrum of *mutD5*, we cloned and sequenced a number of *lacI* mutants. The methods employed have been recently described (41). To facilitate this study, we restricted our analysis to the *lacI*^{d} mutations. This approach limits the analysis to the proximal portion of the gene (first 200 nucleotides) but still allows all classes of mutations to be recovered (R. M. Schaaper, B. N. Danforth, and B. W. Glickman, J. Mol. Biol., in press).

The *lacI*^{d} mutations were identified by their dominance in *lacI*^{d}/*lacI*$^+$ merodiploids (32). In addition to *lacI*^{d} mutations, this test selects for *lacO*^{c} mutations. Among the mutants recovered in the *mutD5* strain, 40% (76/188) were scored as positive. This compares with 11% (20/176) as observed previously in a *mut*$^+$ strain (Schaaper et al., in press). Sequencing of these 76 mutations revealed that only 44 resulted from nucleotide changes within the proximal *lacI* region, i.e., they were true *lacI*^{d} mutations. Of the remaining 32 mutations, 27 were found to be located in the *lac* operator region (*lacO*^{c} mutations). The nucleotide changes responsible for the remaining five mutants were not determined. They may therefore represent *lacI*^{d} mutations at sites elsewhere within the *lacI* gene (35). The results of the sequencing are presented in Table 3. Of the 44 *lacI*^{d} mutations and the 27 *lacO*^{c} mutations, 69 resulted from single base-pair substitutions (Table 3). The two exceptions were *lacI*^{d} frameshift mutations resulting from single G · C base-pair deletions in a G · C doublet (nucleotides 148 to 149; Fig. 1) and in a G · C triplet (nucleotides 90 to 92; Fig. 1). No large

TABLE 2. Spontaneous mutations in a *mutD5* strain: distribution of *lacI* amber mutations[a]

Base substitution	Site	Sequence[b] (5′ → 3′)	No. of independent occurrences at each site
G · C → A · T	A5	A-T-C-A-G-A	7
	A6	A-C-C-A-G-G	8
	A9	A-A-C-A-G-T	4
	A15	A-C-C-A-G-G	15[c]
	A16	A-C-C-A-G-A	2
	A19	A-C-C-A-G-C	8[c]
	A21	G-C-T-G-G-C	2
	A23	T-T-C-A-G-C	2
	A24	A-C-T-G-G-A	5
	A26	A-T-C-A-G-A	10
	A31	A-A-C-A-G-G	2
	A33	C-T-C-A-G-G	2
	A34	G-C-C-A-G-G	0
	A35	A-T-C-A-G-C	8[c]
G · C → T · A	A2	C-A-G-A-G-T	0
	A7	C-G-G-A-G-C	0
	A10	A-G-T-C-G-T	1
	A12	C-G-T-C-G-C	2
	A13	T-G-T-C-G-A	1
	A17	T-G-G-A-G-C	0
	A20	T-C-T-C-G-G	1
	A25	A-T-G-A-G-G	1
	A27	C-C-G-A-G-T	1
	A28	T-C-T-C-G-G	1
A · T → T · A	A11	C-G-T-T-G-C	0
	A18	C-A-T-T-G-G	6[c]
	A32	G-C-T-T-G-C	0
	A36	T-G-T-T-G-C	17[c]
A · T → C · G	A3	G-T-A-T-G-C	0
	A4	T-T-A-T-C-A	0
	A14	C-T-A-T-C-C	1
	A22	A-T-A-T-C-T	2
	A30	T-T-A-T-A-T	0
G · C → C · G	A1	A-T-A-C-G-A	0
	A8	T-T-A-C-A-T	0
	A29	A-T-A-C-G-A	0

[a] Total transitions, 75 (68.8%); total transversions, 34 (31.2%). Only G · C → A · T transitions can be detected in the *lacI* nonsense spectrum.

[b] Nucleotide sequence surrounding each nucleotide (underlined) that mutates to produce an amber codon.

[c] One or more isolates at this site were confirmed by DNA sequencing.

TABLE 3. Spontaneous mutation in a *mutD5* strain: distribution of sequenced *lacI*^{d} base substitution mutations

Base substitution	Nucleotide position[a] in *lacI* gene	Sequence[b] (5′ → 3′)	No. of independent occurrences	Amino acid changes
A · T → G · C	41	TAACG	1	Thr → Ala
	54	TGTCG	2	Val → Ala
	72	TGTCT	1	Val → Ala
	83	AGACC	3	Thr → Ala
	89	TTTCC	2	Ser → Pro
	117	CGTTT	3	Val → Ala
	119	TTTCT	1	Ser → Pro
	168	TTACA	1[c]	Tyr → Cys
	171	CATTC	2	Ile → Thr
	−3 (operator)	TGTGG	1[d]	O^c mutation
	+6 (operator)	TGTGA	26	O^c mutations
G · C → A · T	56	TCGCA	1	Ala → Thr
	80	ATCAG	3	Gln → amber (A5)
	84	GACCG	1	Thr → Ile
	93	CCGCG	2	Arg → His
	129	AACGC	1	Thr → Met
	186	GGCAC	1	Ala → Val
	191	AACAA	4	Gln → ochre (O10)
	198	GGCGG	1	Ala → Val
	206	AACAG	1	Gln → amber (A9)
	+5 (operator)	TTGTG	1	O^c mutation
A · T → T · A	96	CGTGG	1	Val → Glu
	141	AGTGG	3	Val → Glu
	183	CGTGG	4	Val → Glu
G · C → T · A	93	CCGCG	2	Arg → Leu
	198	GGCGG	1	Ala → Glu

[a] The nucleotide positions are as described by Farabaugh (15); the numbering for *lacZ* regulatory nucleotides is as described by Reznikoff and Abelson (38), where +1 is the first nucleotide transcribed.

[b] Nucleotide sequence surrounding each nucleotide (underlined) that mutates to produce a *lacI*^{d} or *lacO*^{c} mutation.

[c] This change is part of a double mutation, with the second event being an A · T → G · C substitution occurring at position 171. Single substitutions at either position 168 or 171 result in *lacI*^{d} mutations (Schaaper, unpublished data). For the purpose of this table, each substitution is considered an independent event.

[d] This change is part of a double mutation, with the second event being a G · C → A · T transition at the +5 position in the *lac* operator. While the G · C → A · T substitution at the +5 position leads to the *LacO*^{c} phenotype (28), the phenotype of the A · T → G · C transition at position −3 is unknown.

deletions or complex mutational events were recovered, although two mutants contained two single base substitutions in close proximity. Since 26 of the 27 *lacO*^{c} mutations were A · T → G · C transitions at a single nucleotide position (+6 in *lacO*), this site must be considered a unique hotspot. Hence, the *lacO*^{c} mutations are not included in the analysis of *mutD5*-induced *lacI*^{d} mutations in Table 4. *mutD5*-induced *lacI*^{d} base-pair substitution mutations result from A · T → G · C and G · C → A · T transitions as well as from A · T → T · A and G · C → T · A transversions (Table 3). No A · T → C · G or G · C → C · G transversions were detected. As with the *lacI* amber spectrum among the *lacI* mutations, transitions (70%) occurred more frequently than transversions (25%). The A · T → G · C and G · C → A · T transitions occurred with equal frequency. However, among the transversions, A · T → T · A substitutions predominated (8/11). Again, reminiscent of the distribution of *lacI* amber mutations, the transitions as observed within the proximal portion of the *lacI* gene were distributed over many different sites (Table 3; Fig. 1). Thus, with the obvious exception of the +6 site in the *lac* operator region, hotspots for transition events do not appear to exist. In contrast, but similar to what was observed in the nonsense spectrum, transversions do not appear to be randomly distributed. Transversions occur at just five sites, even though numerous sites exist where transversions give rise to the LacI^d phenotype (33; Schaaper et al., in press).

We examined possible site specificity of *mutD5*-induced mutagenesis by attempting to correlate transition and transversion events with neighboring nucleotide sequences. At the top of Fig. 2 are displayed the four types of base-pair substitutions observed. Below each transition or transversion are given the 16 possible combinations of neighboring nucleotides. We have normalized all events to the strand containing the purine base, although this does not imply that the initial mutational event took place in this strand. For A · T → G · C transitions, the base 5′ to the mutated adenine is usually a purine (15/16), the base on the 3′ side is usually a pyrimidine (13/16), and 75% (12/16) of these transitions have both a 5′ purine and a 3′ pyrimidine flanking the mutated adenine. For G · C → A · T transitions, the base 5′ to the mutated guanine is usually a pyrimidine (14/15), the base on the 3′ side is usually a pyrimidine (12/15), and 73% (11/15) of these transitions have both a 5′ and a 3′ pyrimidine flanking the mutated guanine. These specificities are not the result of any bias in the available target site. A total of 17 sites are currently known within *lacI* where A · T → G · C transitions lead to the LacI^d phenotype (33; Schaaper, et al., in press; R. M. Schaaper, B. N. Danforth, and B. W. Glickman, unpublished data); 59% (10/17) of these sites have a purine on the 5′ side of the mutated adenine, 47% (8/17) have a 3′ pyrimidine, and 35% (6/17) have both a 5′ purine and a 3′ pyrimidine. A total of 28 sites have been found where G · C → A · T transitions lead to the LacI^d phenotype (33; Schaaper et al., in press; Schaaper et al., unpublished data); 43% (12/28) of these have a pyrimidine on the

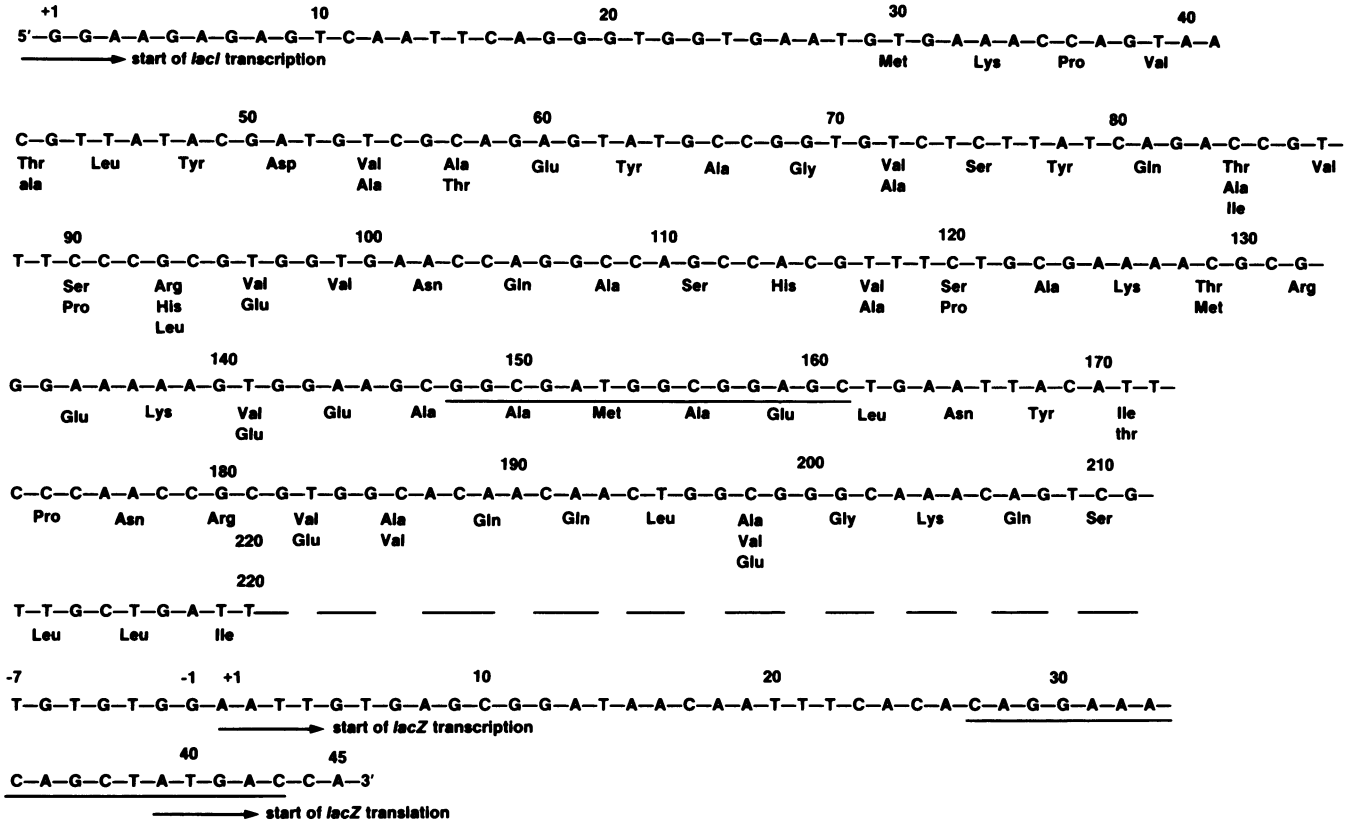

FIG. 1. Wild-type DNA sequence for the proximal region (nucleotide +1 to nucleotide 220) of the *lacI* gene (15), along with the *lac* operator sequence and the start of the *lacZ* gene (38). All *lacI*^{d} mutations sequenced were located within the first 220 nucleotides of the *lacI* gene. Only one strand of the DNA duplex is shown along with the starts for transcription. Beneath the *lacI* DNA sequence are displayed the corresponding amino acids in the *lacI* repressor. Under the wild-type amino acids are given the replacements created by the missense *lacI*^{d} mutations found in this study. The nucleotide sequences underlined within the *lacI* gene are complementary to the primers used to sequence the *lacI*^{d} and *lacO*^{c} mutations.

5′ side of the mutated guanine, 57% (16/28) have a 3′ pyrimidine, and 43% (12/28) have both a 5′ and a 3′ pyrimidine.

An even greater degree of neighboring base specificity was revealed by an analysis of the 11 transversion events; the mutated adenine of A · T → T · A transversions and the mutated guanine of G · C → T · A transversions were flanked on both sides by cytosines (Fig. 2). The specificity of neighboring bases seems to include the two immediate flanking 5′ bases (Fig. 1, Table 3); the mutated adenine of A · T → T · A substitutions is preceded by two cytosines, as is the mutated guanine of the G · C → T · A events. Among the sites with the remaining 30 possible combinations of neighboring bases where we did not find *mutD5*-induced transversions (Fig. 2), *lacI*^{d} mutations can be scored for at least 29 such locations (33; Schaaper et al., in press; Schaaper et al., unpublished data).

TABLE 4. Summary of the distribution of sequenced *lacI*^{d} mutations recovered in a *mutD5* strain[a]

Determination	No. of occurrences	% of total *lacI*^{d} mutations
Total *lacI*^{d} mutations sequenced	44	
Total base-pair substitutions	42	95.5
Transitions		
A · T → G · C	16	36.4
G · C → A · T	15	34.1
Transversions		
A · T → T · A	8	18.2
G · C → T · A	3	6.8
A · T → C · G	0	0
G · C → C · G	0	0
Frameshifts		
Single nucleotide deletions	2	4.5
Single nucleotide insertions	0	0
Large deletions and complex mutations	0	0

[a] The *lacO*^{c} mutations induced by *mutD5* are not included in this summary. Number of sites where *lacI*^{d} mutations were recovered = 25.

DISCUSSION

We have characterized the mutational specificity of spontaneous mutation in a strain carrying the *mutD5* mutation. Our finding that *mutD5* greatly enhanced the frequency of *lacI* nonsense mutations and created single base-pair substitutions in 69 of 71 sequenced *lacI*^{d} and *lacO*^{c} mutations demonstrates that *mutD5* preferentially causes single base substitutions, including both transitions and transversions. However, the mutational specificity of *mutD5* is probably broader than appears here, since the sampling of a modest number of mutations in a forward mutation system revealed only the most frequent events. A previous study, using as a measure the reversion of specific alleles (17), has demonstrated that *mutD5* can stimulate the yield of G · C → C · G

A·T → G·C	G·C → A·T	A·T → T·A	G·C → T·A
5′ A<u>A</u>A 3′	5′ A<u>G</u>A 3′	5′ A<u>A</u>A 3′	5′ A<u>G</u>A 3′
A<u>A</u>T (2)	A<u>G</u>T	A<u>A</u>T	A<u>G</u>T
A<u>A</u>G	A<u>G</u>G	A<u>A</u>G	A<u>G</u>G
A<u>A</u>C (4)	A<u>G</u>C	A<u>A</u>C	A<u>G</u>C
T<u>A</u>A	T<u>G</u>A (3)	T<u>A</u>A	T<u>G</u>A
T<u>A</u>T	T<u>G</u>T (5)	T<u>A</u>T	T<u>G</u>T
T<u>A</u>G	T<u>G</u>G	T<u>A</u>G	T<u>G</u>G
T<u>A</u>C (1)	T<u>G</u>C (1)	T<u>A</u>C	T<u>G</u>C
G<u>A</u>A (3)	G<u>G</u>A	G<u>A</u>A	G<u>G</u>A
G<u>A</u>T	G<u>G</u>T (1)	G<u>A</u>T	G<u>G</u>T
G<u>A</u>G	G<u>G</u>G	G<u>A</u>G	G<u>G</u>G
G<u>A</u>C (6)	G<u>G</u>C	G<u>A</u>C	G<u>G</u>C
C<u>A</u>A	C<u>G</u>A	C<u>A</u>A	C<u>G</u>A
C<u>A</u>T	C<u>G</u>T (1)	C<u>A</u>T	C<u>G</u>T
C<u>A</u>G	C<u>G</u>G	C<u>A</u>G	C<u>G</u>G
C<u>A</u>C	C<u>G</u>C (4)	C<u>A</u>C (8)	C<u>G</u>C (3)

FIG. 2. Effect of neighboring sequences on *mutD5*-induced *lacI*^{d} mutations. At the top are the four types of base-pair substitutions observed. Below each transition and transversion are the 16 possible orders for neighboring nucleotides on each side of the mutated nucleotide (underlined). The number of independent occurrences for each site with that particular configuration of neighboring bases is given in parentheses. For comparisons, the sequences shown are those of the DNA strand which contains the purine of the affected base pair. The initial mutational event may have occurred in either this strand or its complement.

transversions and addition frameshift mutations, neither of which were detected in this study.

The combined distribution of both the amber and sequenced *lacI*^{d} mutations indicates that the *mutD5* allele produces mutational events at a large number of sites throughout the *lacI* gene. Transitions occur within *lacI* at many sites and seem to be distributed approximately evenly. (The exception, a hotspot at the +6 position of *lacO*, is discussed below.) The opposite appears to be true when transversion events are considered. The 11 sequenced transversions all have the same neighboring sequence, 5′-C-C-(mutated G or A)-C-3′. Similarly, the nonsense transversion mutations occur preferentially at certain sites (Table 2).

The *mutD*-encoded protein is the epsilon subunit of DNA polymerase III, which provides the 3′ → 5′ exonuclease or proofreading activity of this enzyme (11, 43, 44). This suggests that the base-substitution mutations recovered may be the result of mistakes occurring at the replication fork that have gone uncorrected i.e., incorporation errors. However, the postreplicative mismatch repair system (21, 37) is expected to correct some of the mispaired bases (4) that have resulted from misincorporation so that the *mutD5* mutational spectrum is not a direct indicator of incorporation errors. That the *mutD5* phenotype was due to defective proofreading was first proposed by Topal and Fresco (47). According to their model, purine-pyrimidine mispairings, which give rise to transitions, should occur much more often than purine-purine mispairings, which give rise to transversion events. Transversions were predicted not to arise through pyrimidine-pyrimidine mispairings since such intermediates were considered energetically unfavorable as compared with the corresponding purine-purine mismatches (47). As in an earlier case (17), and consistent with the Topal-Fresco model, our analysis of the mutational specificity of *mutD5* indicates that, indeed, transitions do occur more frequently than transversions. Moreover, the frequencies of A · T → G · C and G · C → A · T transitions are equal, a result also predicted by the Topal-Fresco model (47). However, on the average, transitions occur only about three times as frequently as do transversions, and the frequencies at some transversion sites are as high as or higher than those at transition sites.

Although *mutD5*-induced transitions are widely distributed, there may be some site specificity determined by neighboring nucleotides. Mutated adenines are usually flanked by a 5′ purine and a 3′ pyrimidine, whereas mutated guanines are often flanked by 5′ and 3′ pyrimidines. If these limited data are significant, nearest-neighbor interactions may play an important role in the *mutD5* spontaneous transition spectrum. Base stacking interactions between the incoming nucleotide and the primer terminus have been shown to be involved in in vitro incorporation with T4 DNA polymerase (46) and could be critical here. Fresco and co-workers (18) have concluded that base stacking interactions in both DNA strands at transition sites influence the stability of purine-pyrimidine mispairings.

Although transitions occur more frequently than transversions in the *mutD5* spectrum, a number of sites do exist where the frequency of transversion events is as high as or even higher than mutation frequencies at transition sites. Such transversion sites are not random; rather, they appear to depend upon neighboring base sequences. Possibly nearest-neighbor interactions strongly influence the frequencies of transversion mispairings at these sites. It is not likely that *mutD5* is capable of creating transversions only at 5′-C-C-(A/G)-C-3′ sites. A previous study demonstrated that *mutD5* modestly enhanced mutation frequencies at all transversion sites that could be scored (17), and our amber spectrum includes transversions outside of 5′-C-C-(A/G)-C-3′ sites. We assume that *mutD5* causes transversions at additional sites which result in *lacI*^{d} mutations but at frequencies too low for these events to be included within our sequenced sample.

The possibility that *mutD5* mutator activity reflects a reduction in the proofreading activity of DNA polymerase III leads to interesting speculation. Current models for inducible, error-prone DNA repair suggest that the "bypass" of lesions that otherwise block DNA replication is facilitated by the suppression of proofreading (see, e.g., reference 37). If this assumption is correct, then the specificity of mutations occurring in the presence of the *mutD5* allele might be expected to approximate that observed after the induction of error-prone repair. The limited data available on the specificity of inducible, error-prone repair indicate that the insertion of an adenine residue across from such lesions is preferred (24, 34, 42). Hence, a reduction in proofreading might be expected to lead to an increase in mutations to A · T base pairs. While this specificity is not seen in the case of transitions (Table 3), it is clearly evident among transversions (Tables 2 and 3).

The *mutD5* allele also enhances the frequency of frameshift mutations (17, 48). In this study, only two frameshift mutations were identified. In both cases they involved the loss of a G · C base pair in a run of two or more G · C base pairs. Similar results were obtained by Lorenzetti et al. (26), who sequenced eight *mutD5*-induced frameshifts within a 91-nucleotide insertion derived from the *tet* gene of pBR322. In each case the frameshift involved the loss of a G · C base pair located within a run of G · Cs. The mechanism for frameshift mutagenesis originally proposed by Streisinger et

al. (45) predicts that such mutations will occur in runs. The frameshift mutations recovered in the *mutD5* genetic background are at least consistent with this model. While misaligned templates created by slippage of repeated sequences are thought to be the underlying principle, the molecular details of the origin of the misalignments and of their fixation into mutations remain largely a matter of conjecture. The influence of DNA replication on frameshift mutagenesis has also been demonstrated for the DNA polymerase of bacteriophage T4, where gene *43* mutator and antimutator alleles have been shown to influence the frequency of frameshift events in repeated A · T sequences (39).

The present results suggest that the 3′ → 5′ nuclease activity of DNA polymerase III has a role in removing the misaligned bases that are precursors for frameshift mutations. Although no firm conclusions can be drawn from two observed frameshift mutations, a comparison with *lacI* frameshift frequencies in a *mut*$^+$ strain (Schaaper et al., in press) shows that the effect may be several thousand-fold. Similar data were reported for other *mutD5* and *mut*$^+$ strains based upon reversion analysis of *trpA* frameshift mutations (17).

The most frequently recovered mutation in this study was an A · T → G · C transition at the +6 position of *lacO*. The frequency of mutation at this hotspot was found to be approximately 10^{-3}. This translates into an error rate of approximately 6×10^{-5} per nucleotide replicated (method of Drake [12]) and is extraordinarily high compared to the frequently cited value rate of 10^{-9} to 10^{-12} per nucleotide replicated for spontaneous mutation in *mut*$^+$ strains (12). Most notably, it occurs in a strain that presumably has a functional mismatch repair system (21, 37), although it cannot be excluded that, in the *mutD5* background, the mismatch repair system is so overloaded that it ceases to make a significant contribution to the reduction in error rates. Mutations at this site are also frequently recovered in a *mut*$^+$ spectrum, where they account for about 3% of all spontaneous *lacI* mutations (Schaaper et al., in press). The enhanced production of these transitions by the *mutD5* allele strongly suggests that these errors are being made during DNA replication (unless the *mutD* protein also functions away from the replication fork). This makes their origin through a replication-independent event, e.g., deamination, unlikely. This latter possibility was put forward (Schaaper et al., in press) in view of the atypical nuclear magnetic resonance data for the AT base pair at this site (3).

The *lac* repressor functions as a tetramer consisting of four identical subunits (19) that each contain 360 amino acids (1, 2). Models have been proposed that predict specific regions of contact between DNA sequences of the *lac* operator and amino acid sequences of the *lac* repressor (30, 36, 49). Recently Miller (33) compiled a list of *lacI*^{d} mutations that cause the repressor to fail to bind to the *lac* operator. Several of the *lacI*^{d} mutations generated in this study result in amino acid substitutions that have not yet been described and may provide information on *lac* repressor-operator interactions. The ease and specificity with which *mutD5* produces base-pair substitutions throughout the *lacI* gene should allow an examination of the effects of amino acid substitutions on various repressor functions.

In summary, the potent mutator allele *mutD5* preferentially produces single base substitutions, including both transitions and transversions. Transitions occur more frequently than transversions and are distributed among numerous sites. Transversions are restricted to a limited number of sites, are strongly influenced by neighboring sequences, and have a specificity similar to that predicted for inducible error-prone repair. Frameshift mutations are also induced by *mutD5* in runs of G · C base pairs. Most likely these mutations result from replication errors associated with the loss of the 3′ → 5′ proofreading ability of DNA polymerase III.

ACKNOWLEDGMENTS

We thank Ronnie Dunn and Don Halderman for expert technical assistance and D. Horner and E. Cox for supplying the *mutD5* allele used in this study. R.G.F. expresses his appreciation to the staff of the National Institute of Environmental Health Sciences, Research Triangle Park, N.C., for their hospitality during a sabbatical leave spent there when most of this work was conducted.

LITERATURE CITED

1. **Beyreuther, K.** 1978. Chemical structure and functional organization of *lac* repressor from *Escherichia coli*, p. 123–154. *In* J. H. Miller and W. S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Beyreuther, K.** 1978. Revised sequence for the *lac* repressor. Nature (London) **274**:767.
3. **Cheung, S., K. Arndt, and P. Lu.** 1984. Correlation of *lac* operator DNA imino proton exchange kinetics with its function. Proc. Natl. Acad. Sci. USA **81**:3665–3669.
4. **Choy, H. E., and R. G. Fowler.** 1985. The specificity of base-pair substitution induced by the *mutL* and *mutS* mutators in *E. coli*. Mutat. Res. **142**:93–97.
5. **Coulondre, C., and J. H. Miller.** 1977. Genetic studies of the *lac* repressor. III. Additional correlation of mutational sites with specific amino acid residues. J. Mol. Biol. **117**:525–575.
6. **Coulondre, C., and J. H. Miller.** 1977. Genetic studies of the *lac* repressor. IV. Mutagenic specificity in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **117**:577–606.
7. **Coulondre, C., J. H. Miller, P. J. Farabaugh, and W. Gilbert.** 1978. Molecular basis of base substitution hotspots in the *lacI* gene of *E. coli*. Nature (London) **274**:775–780.
8. **Cox, E. C., and D. L. Horner.** 1982. Dominant mutators in *Escherichia coli*. Genetics **100**:7–18.
9. **Cox, E. C., and D. L. Horner.** 1983. Structure and coding properties of a dominant *Escherichia coli* mutator gene, *mutD*. Proc. Natl. Acad. Sci. USA **80**:2295–2299.
10. **Degnen, G. E., and E. C. Cox.** 1974. A conditional mutator gene in *Escherichia coli*: isolation, mapping, and effector studies. J. Bacteriol. **117**:477–487.
11. **DiFrancesco, R., S. K. Bhatnagar, A. Brown, and M. J. Bessman.** 1984. The interaction of DNA polymerase III and the product of the *Escherichia coli* mutator gene, *mutD*. J. Biol. Chem. **259**:5567–5573.
12. **Drake, J. W.** 1969. Comparative rates of spontaneous mutation. Nature (London) **221**:1132.
13. **Echols, H., C. Lu, and P. M. J. Burgers.** 1983. Mutator strains of *Escherichia coli*, *mutD* and *dnaQ*, with defective exonucleolytic editing by DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **80**:2189–2192.
14. **Erlich, H. A., and E. C. Cox.** 1980. Interaction of an *Escherichia coli* mutator gene with a deoxyribonucleotide effector. Mol. Gen. Genet. **178**:703–708.
15. **Farabaugh, P. J.** 1978. Sequence of the *lacI* gene. Nature (London) **274**:765–769.
16. **Farabaugh, P. J., U. Schmeissner, M. Hofer, and J. H. Miller.** 1978. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **126**:847–857.
17. **Fowler, R. G., G. E. Degnen, and E. C. Cox.** 1974. Mutational specificity of a conditional *Escherichia coli* mutator, *mutD5*. Mol. Gen. Genet. **133**:179–191.
18. **Fresco, J. R., S. Broitman, and A.-E. Lane.** 1980. Base mispairing and nearest-neighbor effects in transition mutations. ICN-UCLA Symp. Mol. Cell. Biol. **19**:753–768.
19. **Gilbert, W., and B. Muller-Hill.** 1966. Isolation of the *lac* repressor. Proc. Natl. Acad. Sci. USA **56**:1891–1898.
20. **Glickman, B. W.** 1979. Spontaneous mutagenesis in *Escherichia*

coli strains lacking 6-methyladenine residues in their DNA. An altered mutational spectrum in *dam*$^-$ mutants. Mutat. Res. **61**:153–162.

21. **Glickman, B. W., and M. Radman.** 1980. *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc. Natl. Acad. Sci. USA **77**:1063–1067.
22. **Horiuchi, T., H. Maki, M. Muruyama, and M. Sekiguchi.** 1981. Identification of the *dnaQ* gene product and location of the structural gene for RNase H of *Escherichia coli* by cloning of the genes. Proc. Natl. Acad. Sci. USA **78**:3770–3774.
23. **Horiuchi, T., H. Maki, and M. Sekiguchi.** 1978. A new conditional lethal mutator (*dnaQ49*) in *Escherichia coli* K12. Mol. Gen. Genet. **163**:277–283.
24. **Kunkel, T. A.** 1984. Mutational specificity of depurination. Proc. Natl. Acad. Sci. USA **81**:1494–1498.
25. **Kunz, B. A., and B. W. Glickman.** 1983. The infidelity of conjugal DNA transfer in *Escherichia coli*. Genetics **105**: 489–500.
26. **Lorenzetti, R., G. Cesareni, and R. Cortese.** 1983. Frameshift mutations induced by an *Escherichia coli* strain carrying a mutator gene, *mutD5*. Mol. Gen. Genet. **192**:515–516.
27. **Maki, H., T. Horiuchi, and M. Sekiguchi.** 1983. Structure and expression of the *dnaQ* mutator and the RNase H genes of *Escherichia coli*: overlap of the promoter regions. Proc. Natl. Acad. Sci. USA **80**:7137–7141.
28. **Maquat, L. E., K. Thornton, and W. S. Reznikoff.** 1980. *lac* promoter mutations located downstream from the transcription start site. J. Mol. Biol. **139**:537–549.
29. **Maruyama, M., T. Horiuchi, H. Maki, and M. Sekiguchi.** 1983. A dominant (*mutD5*) and a recessive (*dnaQ49*) mutator of *Escherichia coli*. J. Mol. Biol. **167**:757–771.
30. **Matthews, B. W., D. H. Ohlendorf, W. F. Anderson, and Y. Takeda.** 1982. Structure of the DNA-binding region of *lac* repressor inferred from its homology with *cro* repressor. Proc. Natl. Acad. Sci. USA **79**:1428–1432.
31. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
32. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Miller, J. H.** 1984. Genetic studies of the *lac* repressor. XII. Amino acid replacements in the DNA binding domain of the *Escherichia coli lac* repressor. J. Mol. Biol. **180**:205–212.
34. **Miller, J. H., and K. B. Low.** 1984. Specificity of mutagenesis resulting from the induction of the SOS system in the absence of mutagenic treatment. Cell **37**:675–682.
35. **Miwa, J., and J. R. Sadler.** 1977. Characterization of i^{-d} repressor mutations of the lactose operon. J. Mol. Biol. **117**:843–868.
36. **Ogata, R., and W. Gilbert.** 1978. An amino-terminal fragment of *lac* repressor binds specifically to *lac* operator. Proc. Natl. Acad. Sci. USA **75**:5851–5854.
37. **Radman, M., G. Villani, S. Boiteux, A. R. Kinsella, B. W. Glickman, and S. Spaderi.** 1979. Replication fidelity: mechanisms of mutation avoidance and mutation fixation. Cold Spring Harbor Symp. Quant. Biol. **43**:937–946.
38. **Reznikoff, W. S., and J. N. Abelson.** 1978. The *lac* promoter, p. 221–243. *In* J. H. Miller and W. S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Ripley, L. S., and N. B. Shoemaker.** 1983. A major role for bacteriophage T4 DNA polymerase in frameshift mutagenesis. Genetics **103**:353–366.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
41. **Schaaper, R. M., B. N. Danforth, and B. W. Glickman.** 1985. Rapid repeated cloning of mutant *lac* repressor genes. Gene **39**:181–189.
42. **Schaaper, R. M., T. A. Kunkel, and L. Loeb.** 1983. Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc. Natl. Acad. Sci. USA **80**:487–491.
43. **Schuermann, R. H., and H. Echols.** 1984. A separate editing exonuclease for DNA replicating: The ϵ subunit of *Escherichia coli* DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. USA **81**:7747–7751.
44. **Scheuermann, R., S. Tam, P. M. J. Burgers, C. Lu, and H. Echols.** 1983. Identification of the ϵ-subunit of *Escherichia coli* DNA polymerase III holoenzyme as the *dnaQ* gene product: a fidelity subunit for DNA replication. Proc. Natl. Acad. Sci. USA **80**:7085–7089.
45. **Streisinger, G., Y. Ikada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye.** 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**:77–84.
46. **Topal, M. D., S. R. DiGuiseppi, and N. K. Sinha.** 1980. Molecular basis for substitution mutations. Effect of primer terminal and template residues on nucleotide selection by phage T4 DNA polymerase *in vitro*. J. Biol. Chem. **255**:11717–11724.
47. **Topal, M. D., and J. R. Fresco.** 1976. Complementary base pairing and the origin of substitution mutations. Nature (London) **263**:285–289.
48. **Vacarro, K. K., and E. C. Siegel.** 1977. The frameshift mutability of *polA1* and *recA1* derivatives of mutator strains of *Escherichia coli*. Mutat. Res. **42**:443–446.
49. **Weber, I. T., O. B. McKay, and T. A. Steitz.** 1982. Two helix DNA binding motif of CAP found in *lac* repressor and *gal* repressor. Nucleic Acids Res. **10**:5085–5102.
50. **Yanofsky, C., J. Ito, and V. D. Horn.** 1966. Amino acid replacements and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**:151–162.