J. Mol. Biol. (1986) **189**, 273–284

Mechanisms of Spontaneous Mutagenesis: An Analysis of the Spectrum of Spontaneous Mutation in the *Escherichia coli lacI* Gene

Roel M. Schaaper†, Bryan N. Danforth and Barry W. Glickman‡

Laboratory of Genetics
National Institute of Environmental Health Sciences
Research Triangle Park, NC 27709, U.S.A.

(Received 2 July 1985, and in revised form 20 January 1986)

We have obtained *via* DNA sequence analysis a spectrum of 174 spontaneous mutations occurring in the *lacI* gene of *Escherichia coli*. The spectrum comprised base substitution, frameshift, deletion, duplication and insertion mutations, of which the relative contributions to spontaneous mutation could be estimated. Two thirds of all *lacI* mutations occurred in the frameshift hotspot site. An analysis of the local DNA sequence suggested that the intensity of this hotspot may depend on structural features of the DNA that extend beyond those permitted by the repeated tetramer at this site. Deletions comprised the largest non-hotspot class (37%). They could be divided into two subclasses, depending on whether they included the *lac* operator sequence; the latter was found to be a preferred site for deletion endpoints. Most of the deletions internal to the *lacI* gene were associated with the presence of directly or invertedly repeated sequences capable of accounting for their endpoints. Base substitutions comprised 34% of the non-hotspot events. Unlike the base substitution spectrum obtained *via* nonsense mutations, G·C → A·T transitions do not predominate. A new base substitution hotspot was discovered at position +6 in the *lac* operator; its intensity may reflect specific features of the operator DNA. IS*1* insertion mutations contributed 12% of the non-hotspot mutations and occurred dispersed throughout the gene in both orientations. Since the *lacI* gene is not A+T-rich, the contribution of IS*1* insertion to spontaneous mutation in general might be underestimated. Single-base frameshift mutations were found only infrequently. In general, they did not occur in runs of a common base. Instead, their occurrence seemed based on the "perfection" of direct or inverted repeats in the local DNA sequence. Three (tandem) duplication events were recovered. No repeated sequences were found that might have determined their endpoints.

1. Introduction

The mechanisms by which organisms undergo mutation have been a matter of longstanding interest. It has become clear that cells play an active role in the production of mutations, largely through the processes of DNA replication and repair. Despite this notion, more specific information on the pathways leading to mutation is generally not available. One way to increase our level of understanding is the study of the specificity of mutation. In this approach large collections of mutants are analyzed for their distributions over various mutational classes, such as base substitutions, frameshifts, deletions, etc. Such knowledge, particularly if combined with information on the dependence on the local DNA sequence, will generally provide clues as to the possible origins of mutation.

The *lacI* gene of *Escherichia coli*, which codes for the repressor of the *lac* operon, has been used extensively for the analysis of mutagenic specificity. The development of the *lacI* forward mutation system (Coulondre & Miller, 1977), in which nonsense mutations are analyzed by classical genetic techniques, permits the determination of base substitution mutation at more than 70 sites within the gene. The *lacI* system has revealed information about both spontaneous and induced mutation and can be credited with providing new

† Author to whom all correspondence should be addressed.

‡ Present address: Biology Department, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada.

0022-2836/86/100273-12 $03.00/0

insights into their origins (see, e.g. Coulondre *et al.*, 1978; Duncan & Miller, 1980; Todd & Glickman, 1982; Foster *et al.*, 1983; Miller & Low, 1984; Kunz & Glickman, 1984). This same gene has been the subject of direct analysis by DNA sequencing (Farabaugh *et al.*, 1978; Calos & Miller, 1981). Recently, we have developed a rapid approach to the cloning and sequencing of *lacI* mutants (Schaaper *et al.*, 1985). Using this approach, it now seems practical to record complete mutational spectra by DNA sequence analysis, extending the study of the specificity of mutation to a larger number of mutational classes. In this paper, we analyze a mutational spectrum based on the sequencing of 174 spontaneous *lacI* mutants and consider their potential origins.

2. Materials and Methods

(a) *Bacterial strains*

Spontaneous *lac*I mutations were selected in *E. coli* strain NR3835 (Δ(*pro-lac*), *ara*, *thi*, *trpE*9777, F′*pro-lac*). With the exception of the *trp* mutation, this strain is identical to GM1 (Schmeissner *et al.*, 1977). It carries the I^Q (*lacI*) and L8 (*lacZ*) mutations. All strains utilized for assigning the mutation to specific deletion intervals in the *lacI* gene have been described (Schmeissner *et al.*, 1977) as well as strain S90C (Δ(*pro-lac*), *ara*, *thi*, *strA*), which was used as the host for the F′ *lacI* mutants in recombination experiments with the bacteriophage M13 recovery vector. Strain NR9099 (Δ(*pro-lac*), *ara*, *thi*, *recA*56, F′*pro*$^+$, *lacI*$^-$, *lacZ*$^-$ ΔM15) used for the detection of recombinant phage has been described (Schaaper *et al.*, 1985).

(b) *Bacteriophage*

The M13 recombinant phage mRS81 has been described (Schaaper *et al.*, 1985). It carries the intact *lacI* gene and the portion of the *lacZ* gene required for α-complementation. However, in mRS81 α-complementing ability is abolished by a point mutation.

(c) *Media*

X-gal† minimal plates contained: Vogel–Bonner salts (Vogel & Bonner, 1956); glucose (0·2%); thiamine (5 μg ml^{-1}); tryptophan (50 μg ml^{-1}; X-gal (40 μg ml^{-1}). P-gal plates used for the selection of *lacI*$^-$ mutants contained: Vogel–Bonner salts; phenyl-β-D-galactoside (P-gal) (75 μg ml^{-1}); thiamine (5 μg ml^{-1}); and tryptophan (50 μg ml^{-1}). Luria Broth (LB) contained Difco Bactotryptone (10 g l^{-1}); Difco yeast extract (5 g l^{-1}); and NaCl (10 g l^{-1}).

(d) *Selection of* lacI$^-$ *mutants*

A saturated culture of NR3835 was diluted in LB such that 200 μl added to each well of a 96-well microtiter plate contained fewer than 100 cells. After growth overnight at 37°C, appropriate dilutions were titered on LB plates to determine cell number and on P-gal plates to select *lacI*$^-$ mutants. One *lacI*$^-$ mutant was taken from each culture, gridded and replica-mated (F′ transfer) into strains X^r for subsequent deletion mapping (Schmeissner *et al.*, 1977) and S90C for transfer of the *lacI* gene to phage mRS81 (see below).

(e) *Transfer to M13*lac

F′*lac* × M13*lac* crosses were performed by growing phage mRS81 overnight on strain S90C carrying the F′ *lacI* mutation of interest, in 1 ml LB at 37°C with shaking. The starting inocula were approximately 5×10^7 bacteria and 5×10^6 phages. The next morning the cells were pelleted by centrifugation (4000 revs min^{-1} for 15 min) and samples of the supernatant phage were diluted and plated (10^3 to 10^5 plaques per plate) on X-gal plates with NR9099 as indicator. After 48 h, *lacZ*$^+$-recombinant phages carrying the *lacI*$^-$ gene appear as blue plaques. These recombinant phages were plaque purified on NR9099 and used to prepare single-stranded DNA for DNA sequencing.

(f) *DNA sequence analysis*

Single-stranded DNA was sequenced by the dideoxy chain-termination method of Sanger *et al.* (1977). Oligonucleotide primers complementary to specific positions of the *lacI* and *lacZ* genes (prepared by PL Biochemicals) were used to sequence the specific regions indicated by the deletion mapping of the *lacI* mutations. These primers were 14-mers with their 3′ ends at positions 148, 302, 450, 604, 745, 901 and 1049 of the *lacI* gene and the PL hybridization probe primer (17-mer) starting at position 27 of the *lacZ* mRNA sequence. The resulting DNA sequences were compared with that of the wild-type *lacI* gene (Farabaugh, 1978) and the regulatory region between *lacI* and *lacZ* (Reznikoff & Abelson, 1978).

(g) *Enzymes and chemicals*

DNA polymerase I large fragment was obtained from Boehringer-Mannheim Biochemicals, phage T4 polynucleotide kinase from New England Biolabs, deoxy- and dideoxytriphosphates and oligonucleotide primers from PL Biochemicals, [γ-^{32}P]ATP from Amersham and X-gal and P-gal from Bachem Fine Chemicals.

3. Results

(a) *Selection of spontaneous* lacI$^-$ *mutants*

A total of 176 independent *lacI*$^-$ mutants was collected in NR3835 by selecting for growth on P-gal plates as described above. The spontaneous mutation frequency was 2×10^{-6}, similar to that found previously (Coulondre & Miller, 1977; Glickman, 1979). The mutants were then purified and localized in the *lacI* gene by deletion mapping.

(b) *Transfer of* lacI$^-$ *mutations to the M13 cloning vector*

Transfer of *lacI*$^-$ mutations to an M13 vector for DNA sequence analysis was performed as described by Schaaper *et al.* (1985) by the method presented schematically in Figure 1. Phage mRS81 carries an active *lacI* gene but is defective in *lacZ*$_\alpha$ (α-com-

† Abbreviations used: X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; P-gal, phenyl-β-D-galactopyranoside; IPTG, isopropylthiogalactoside.

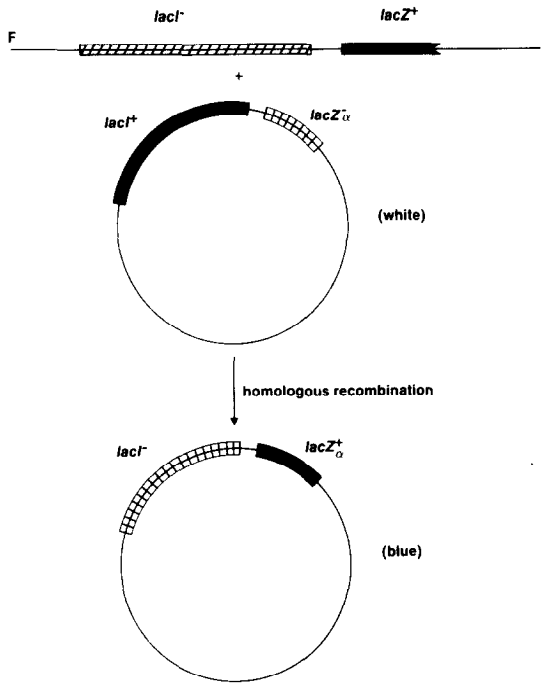

Figure 1. Cloning of *lacI* mutations onto M13 recovery phage mRS81, by recombination *in vivo*. Recombinant phages are detected as blue plaques when plated on NR9099 (*lacI*$^-$, *lacZ*$^-$ ΔM15) in the presence of X-gal but in the absence of an inducer.

plementation). On the *E. coli* host NR9099, which is both *lacI*$^-$ and *lacZ*$_\alpha^-$, it produces colorless plaques on X-gal plates regardless of the presence or absence of the inducer IPTG. When *lacZ*$_\alpha^+$ recombinants of mRS81 are plated on NR9099, plaque color depends on the *lacI* gene carried on the phage and *lacI*$^-$*lacZ*$_\alpha^+$ recombinants can easily be scored as blue plaques in the absence of IPTG against a background of colorless plaques. In other words, *lacI*$^-$ mutants are cloned by recombination *in vivo* and the selection of a double change. This reduces greatly the risk that *lacI*$^-$ mutants will be recovered that are not of the desired origin.

Among the 176 *lacI*$^-$ mutants the frequency of recovery on M13 varied considerably, from approximately 10^{-3} for about 80% of the mutants to less than 10^{-5} for some of the remainder. In two cases, no recombinants could be detected despite repeated efforts. Upon resolution of the DNA sequence (see below), a general trend could be discerned: recombination frequencies were reduced in proportion to the physical extent of the mutations, large deletion and insertion mutations depressing the recombination frequency.

Several control experiments were performed. For example, when progeny of 100 individual crosses with a wild-type host (F′*lacI*$^+$) were plated, not a single blue plaque was observed among a total of 10^7. This indicates that even when recombinants were obtained at low frequencies, there can be little doubt that the *lacI*$^-$ mutations scored were recovered from the F′.

Table 1

Distribution of spontaneous lacI$^-$ *mutants by class*

Class of mutation	Number of occurrences	Frequency × 10^{-7}
(+) T-G-G-C†	95 (54%)	11·0
(−) T-G-G-C†	22 (13%)	2·5
Deletions	22 (13%)	2·5
Base substitutions	20 (11%)	2·2
Insertion elements	7 (4%)	0·8
Single-base frameshifts	5 (3%)	0·7
Duplications	3 (1·7%)	0·3
Unknown‡	2 (1·1%)	0·2
Total	176	20

† Frameshift hotspot at position 620 to 632.
‡ Not recovered onto M13 recovery vector.

(c) *Sequence analysis*

The results of sequencing the 174 cloned mutants are given in Tables 1, 2 and 3. About 67% of the mutations occur at the hotspot site previously reported by Farabaugh *et al.* (1978). The site involves positions 621 to 632 where a triple tandem repeat sequence, T-G-G-C, resides. The mutations involve either the addition of one T-G-G-C sequence (54%) or the loss of one T-G-G-C sequence (12·5%). The fraction of hotspot mutations as well as the ratio of addition to deletion events determined here

Table 2

A. *Base substitutions*

Position	Change	Occurrences	Amino acid change
83	A·T → G·C	1	Thr → Ala
141	T·A → A·T	1	Val → Glu
150	C·G → A·T	1	Ala → Glu
158	G·C → T·A	1	Glu → amber (am7)
183	T·A → A·T	1	Val → Glu
213	T·A → A·T	1	Leu → amber (am11)
222	G·C → A·T	1	Gly → Asp
293	C·G → T·A	1	Gln → ochre (oc13)
419†	C·G → T·A	1	Gln → amber (am15)
582	T·A → G·C	1	Leu → opal
777	C·G → A·T	1	Ala → Glu
845	T·A → G·C	1	Tyr → Asp
885	T·A → G·C	1	Ser → opal
926	A·T → C·G	2	Ser → Arg
+6(*lacO*)	T·A → C·G	5	*Oc* mutation

† G·C → A·T hotspot in spontaneous nonsense spectrum due to deamination of 5-methylcytosine at C-C*A-G-G sequence (Coulondre *et al.*, 1978).

B. *Single-base frameshifts*

Mutant	Position	Alteration	Occurrences	Sequence (wild type)
1-49	52	−T	1	CGA T GTC
4-10	125 to 128	−A	1	G AAAA GC
3-22	286	+C	1	CG C GCCG
3-30	394	−G	1	GCT G ATC
1-20	646	−T	1	CAC T CGC

C. *Deletions*

Class	Mutant	Position	Size (base-pairs)	Repeated bases	Sequence†
I	2-27	91–177	87	8	CCGCGTGG
	1-44	114–127	14	3	ACG
	3-18	146–268	123	8	GCGGCGAT
	2-25	218–262	45	4	ATTG
	1-29	267–282	16	5	TCGCG
	1-11	281–369	89	8	TCTCGCGC
	1-38	313–334	22	7/8	GG$^{T}_{C}$CGTCGA
	1-42	507–524	18	None	—
	4-32	536–616	81	7/8	CGCG$^{A}_{T}$CTG
	2-7	692–770	79	None	—
	2-36	814–1046	233	9	GCGCGTTGG
	1-10	919–1029	111	7	GCAAACC
	3-38	927–1013	87	10/12	AAACCAGC$^{G}_{C}$T$^{G}_{C}$G
	3-16	928–1013	86	10/12	AAACCAGC$^{G}_{C}$T$^{G}_{C}$G
II	2-18	824–(+)8	380	7	GCGGATA
	4-30	900–(+)4	300	None	—
	1-41	916–(+16)	296	None	—
	2-31	954–(+9)	251	None	—
	1-7	1108–(+4)	92	7	GTGAGCG
	4-21	1108–(+4)	92	7	GTGAGCG
	2-28	(−64)–(+7)	71	None	—
	3-10	(−64)–(+12)	76	None	—

D. *Duplications*

Mutant	Duplicated bases	Size (base-pairs)
1-18	362–619	258
1-28	368–824	457
3-25	380–534	155

E. *IS1 Insertion elements*

Mutant	Position†	Orientation‡
3-19	42–50	II
3-24	53–61	I
2-47	148–156	II
3-44	221–229	II
3-2	488–496	II
3-21	496–504	I
1-9	670–678	I

† Nine-base-pair repeat sequence generated upon IS*1* insertion (Calos *et al.*, 1978*b*); only the 3′ end of the IS*1* insertion was sequenced (Johnsrud, 1979).

‡ Orientation I is that of S114 (Calos *et al.*, 1978*b*); orientation II is the reverse insertion.

by DNA sequencing are identical to the values obtained by Farabaugh *et al.* (1978) from mapping and reversion studies.

The remaining 57 mutations were distributed among (larger) deletions (12·5%), base substitutions (11%), insertion events involving IS*1* (4%), single-base frameshifts (3·5%) and duplications (1·7%). These mutants are described in detail in Table 2. The base substitution mutations are further summarized in Table 3. The deletion, insertion and duplication events are presented schematically in Figure 2. Each class of mutations will be further considered in the Discussion.

Table 3

Base substitutions among spontaneous lacI *mutations*

Transitions	Occurrences	Transversions	Occurrences
A·T → G·C	6†	A·T → T·A	3
G·C → A·T	3	A·T → C·G	5
		G·C → T·A	3
		G·C → C·G	0
Total	9		11

† Includes 5 occurrences at position +6 of *lacZ* (*lacO* mutation).

4. Discussion

In this paper we describe the DNA sequence alterations of 174 spontaneous mutations in the *E. coli lacI* gene. We have undertaken this study in the belief that an analysis of the spectrum of spontaneous events in a relatively large target gene will provide a basis for improved understanding of the origin of spontaneous mutations and, by extrapolation, of induced mutations. We have

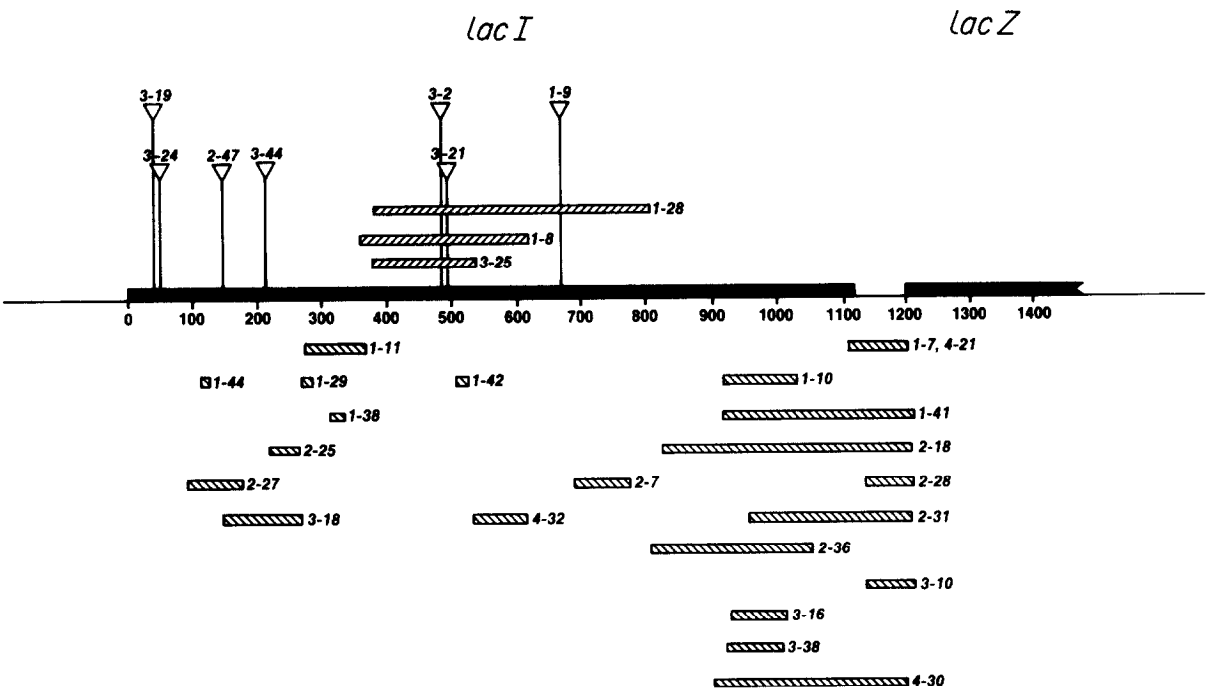

Figure 2. Schematic representation of 22 deletions, 3 duplications and 7 IS*1* insertions recovered among 174 sequenced spontaneous *lacI* mutants. Inverted triangles denote the positions of IS*1* insertions; horizontal bars above the lines denote the positions of the duplications; bars below the line denote the positions of the deletions.

recovered representatives of at least six different classes of mutational events, presumably representing a wide variety of mutational mechanisms. We discuss below each class individually.

(a) *Frameshift hotspot*

As previously reported (Farabaugh *et al.*, 1978), a large fraction of spontaneous *lacI* mutations are the consequence of the gain or loss of one copy of the four-base-pair sequence T-G-G-C, which is tandemly repeated three times at nucleotides 621 to 632. In this study, 67% of the mutations occurred at this site. The ratio of the addition of four base-pairs compared to their loss was 4·3 : 1, similar to that previously determined by genetic analysis (Farabaugh *et al.*, 1978).

Frameshift mutations involving repeated sequences are generally considered to be the consequence of slipped intermediates involving the misalignment of one copy of the repeat on the complement of the other as first proposed by Streisinger *et al.* (1966). Mutations at this hotspot site are thus expected on the basis of this model. However, since mutations at other tandemly repeated sequences (e.g. G-C-C-A-G-C-C-A (107 to 114), G-T-G-G-T-G-G-T-G (308 to 316), C-T-G-C-C-T-G-C (441 to 448), C-G-T-C-T-G-C-G-T-C-T-G (611 to 622) or G-A-A-A-A-G-A-A-A-A (1000 to 1009)) were not recovered, we wondered if additional factors might contribute to the predominance of mutations at the hotspot site. Our analysis of the features of the DNA sequence surrounding the frameshift hotspot leads us to believe that this may indeed be the case.

As can be seen in Figure 3(a), the neighboring region of the frameshift hotspot is rich in both palindromic and repeated sequences. Slippage potentiated by the four-base-pair repeated sequence is not the only misalignment that predicts the hotspot event. In fact, the four-base-pair repeats are part of more extensive nine-base-pair repeats (repeats I and II in Fig. 3(a)). The two sequences have a five-base-pair overlap; therefore, the net result of their misalignment is a four-base-pair shift leading to the addition or deletion of a T-G-G-C sequence. As it is likely that the frequency of recovered mutations must to some degree reflect the length of the repeated sequences, these non-base-pair repeats probably represent a major factor contributing to the predominance of mutations at this site, particularly when compared with other tandem repeat sequences in *lacI* (see above) that are four, five or six bases long. However, the misalignment based on the nine-base-pair repeats may not be the entire explanation for the hotspot. Two seven-base-pair repeats (III and IV in Fig. 3(a)) are immediately 5′ to, and partially overlapping with, the nine-base-pair repeats (I and II in Fig. 3(a)). These neighboring repeats jointly form nearly perfect repeats of 19 and 17 base-pairs (denoted as V and VI in Fig. 3(a)). Their misalignment seems to have an even greater potential to direct the gain or loss of the four bases at the hotspot site. Figure 3(b) shows a possible intermediate based on this misalignment. Additional components involving palindromic DNA sequences

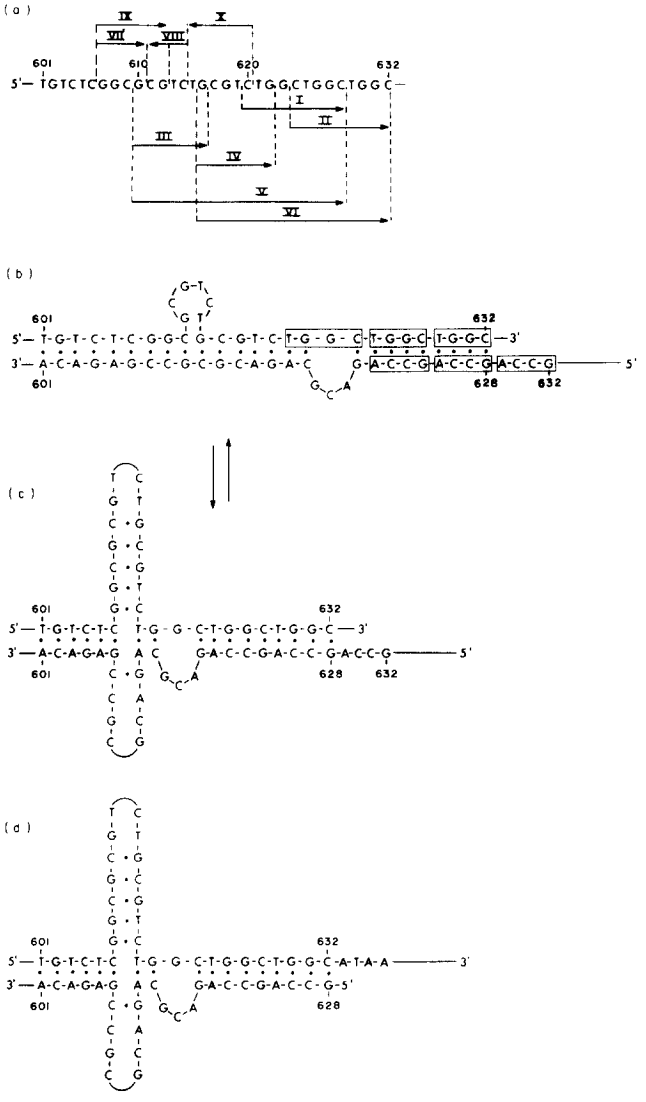

Figure 3. DNA sequence surrounding the *lacI* frameshift hotspot at positions 620 to 631. (a) Regional DNA sequence showing direct and inverted repeats: I and II are overlapping 9-base-pair direct repeats; III and IV are overlapping 7-base-pair repeats; and V and VI are overlapping (19 and 17-base-pair) imperfect repeats composed of I and II, or III and IV. Sequences VII, VIII, IX and X are inverted repeats 4 and 6 base-pairs long, respectively. The structures shown in (b), (c) and (d) are potential intermediates capable of templating the frameshift mutation. Structure (b) is a misalignment mediated by the imperfect repeats V and VI. In structures (c) and (d), the misalignments additionally involve the intra-strand hydrogen bonding permitted by inverted repeats. Structure (c) is an addition intermediate, structure (d) is the corresponding deletion intermediate. The difference between (c) and (d) resides in the strand in which the interruption occurs.

(denoted as VII and VIII, and IX and X in Fig. 3(a)) may co-operate to stabilize certain isomeric forms of the misaligned intermediate (e.g. see Fig. 3(c)). The importance of palindromic components in deletion mutagenesis has been indicated previously (Glickman & Ripley, 1984). Thus, the predominance of mutations at this hotspot may reflect the combined efforts of several structural components, which may co-operate in enhancing both the initial formation and the subsequent stabilization of the structural misalignments that template the mutation.

(b) *Base substitutions*

Base substitutions accounted for 11% of the *lacI*$^-$ mutants and were recovered at a frequency of approximately 2×10^{-7} (Table 1). From this the mutation rate can be calculated to be approximately 10^{-10} mutations per base-pair per round of replication, estimating the number of base-pair sites in the gene that can yield a *lacI*$^-$ mutation roughly as 200 (see below), and taking into account the number of generations the bacterial population had gone through (Drake, 1970). Since five of the mutations occurred at an apparent hotspot site in the operator region of *lacZ* (Table 2A), the actual mutation rate in *lacI* is slightly lower. This estimate agrees well with earlier estimates made by Drake (1969).

The nature of the base substitutions was also examined. The breakdown of the base substitutions in transversion and transition mutations is shown in Table 3. Subtracting the five *lacO*^{c} mutations (A · T → G · C), there is within the *lacI* gene an 11 : 4 bias in favor of transversion events. While the sample size is small, these results differ from those obtained studying the nonsense mutation spectrum. In that case, even after correcting for the spontaneous deamination hotspots, transitions outnumber transversions by severalfold (Coulondre & Miller, 1977; Kunz & Glickman, 1984).

From Table 2A it can be noted that six mutants out of 20 (15 if excluding the *lacO* mutants) were nonsense mutants, a relatively high percentage. In view of the small sample size, the significance of this finding remains to be determined. The potential to detect *lacI* missense mutations *per se* does not seem to be a limiting factor in this respect. We estimate the number of base-pair sites in *lacI* at which missense mutations can be detected to be at least 200. This is based on published data (Miller & Schmeissner, 1979; Miller, 1984), as well as unpublished sequencing data from our laboratory. This number may be contrasted with the 80 base-pair sites at which nonsense mutations can be produced.

The mechanisms responsible for spontaneous base substitutions are potentially severalfold. They may result from the process of DNA replication. In this case their specificity will reflect not only that of misincorporation by the DNA replication complex, but also that of the postreplicative (methylation-instructed) mismatch correction system (Glickman & Radman, 1980). The latter seems to correct transitions more efficiently than transversions (Choy & Fowler, 1985), a preference that may be one reason for the relative predominance of transversion mutations.

Base substitutions may also result from DNA damage. For example, Sargentini & Smith (1981)

noted an increase in levels of spontaneous mutation in strains defective in the *uvr* excision-repair system, implying that excisable damage does occur under spontaneous conditions, and that such damage, if unrepaired, can lead to mutation. It is an interesting question whether such damage-induced mutagenesis, if contributing to spontaneous mutagenesis in repair-proficient strains, would be dependent on the *E. coli* SOS system. A somewhat reduced level of spontaneous mutation has been observed in certain *lexA* or *recA* strains (Kondo *et al.*, 1970; Sargentini & Smith, 1981), consistent with this possibility. The specificity of base substitutions as observed in the present study would also be consistent with such a possibility, since a large majority of SOS-directed mutations seem to be transversions (Schaaper *et al.*, 1983; Miller & Low, 1984).

Base substitution mutations in the operator region of the *lacZ* gene all resulted from an A·T → G·C transition at position +6 of the *lacZ* DNA sequence. This mutation has been shown previously to result in an O^c phenotype (cited by Macquat *et al.*, 1980). The frequency of this base substitution as determined in the present study is 5×10^{-8}. This is at least 50-fold greater than the frequency of base substitution at an average site within the *lacI* gene (10^{-9}) and considerably greater than 5×10^{-9}, the frequency with which mutations at the deamination hotspots were found to occur (Coulondre & Miller, 1977; Glickman, 1979). We note that Cheung *et al.* (1984), in a nuclear magnetic resonance (n.m.r.) study, observed enhanced imino-proton exchange for this particular base-pair compared to others in the region. It seems reasonable to assume that this behavior is somehow related to the origin of the observed mutations at this site. Increased base-pair opening and exchange of the hydrogen bonding protons with water molecules, as implied by the n.m.r. analysis, might lead to enhanced deamination of the adenine residue of the A·T base-pair. The observed mutational specificity is certainly consistent with this possibility, since the resulting hypoxanthine would mispair with cytosine and an A·T → G·C transition would result. Additional or alternative mechanisms, such as reduced mismatch repair due to the particular nature of the *lac* operator or to the protective action of the *lac* repressor, deserve equal consideration.

(c) *Single-base frameshifts*

Five independent frameshift mutations were recovered in this study. As can be seen in Table 2, four of the frameshifts involved a single-base deletion and one a single-base addition.

The DNA sequence information provides details of two aspects of mutation: identification of the sequence change at the site of the mutation, and the larger DNA context of the site. This information provides an opportunity of examining the mechanisms by which these frameshift mutations may have occurred. The proposed mechanisms share the feature that mutation proceeds through misaligned or "slipped" intermediates. Such intermediates were first proposed by Streisinger *et al.* (1966), who suggested that frameshift mutations were the consequence of slippage during strand extension at sites of repetitive sequences. This model was later supported by several instances in which the DNA sequence of frameshift sites could be resolved (Okada *et al.*, 1972; Pribnow *et al.*, 1981) and by the sequencing of the *lacI* frameshift hotspot (Farabaugh *et al.*, 1978). The Streisinger model has been extended to slippage schemes involving non-adjacent repeats. For example, the reversion of the T4 *rIIB* frameshift mutation *FC47* is thought to be templated by a 16-base-pair repeat 256 base-pairs downstream from the mutated site (Pribnow *et al.*, 1981; Ripley & Glickman, 1983*b*). More recent studies have suggested that not only direct repeats, but also inverted repeats, i.e. palindromic sequences, may provide opportunities for misaligned intermediates (Ripley, 1982; Ripley & Glickman, 1983*a*).

An initial examination of the single-base-pair frameshifts isolated in this study revealed that only one could have arisen as a consequence of misalignments involving adjacent repeated sequences: mutation 4-10 (Table 2), which resulted from the loss of a single A in a run of four A residues. Since four of the frameshifts could not be easily explained by mechanisms involving misalignments on locally repeated sequences, we undertook a computer search for direct and/or inverted repeats capable of templating the mutational events. The search was limited to within 100 base-pairs of the site of the mutation. All five sites were subjected to this analysis and, for each of them, nearby sequences were found that were capable of accounting for the specificity of the recovered mutations. The analysis was performed as reported by Golding & Glickman (1985). The potential misalignments capable of directing these frameshift mutations are as follows.

Mutation 1-49 (loss of a T) could have been templated by a four-base-pair direct repeat not adjacent to the mutated site, but located nine base-pairs in the 3′ direction (Fig. 4(a)).

Mutation 3-22 results in the addition of a C. This mutation could have been templated by a local inverted repeat as shown in Figure 4(b). The mutation leads to the perfection of an extensive palindrome. A G-A-T-C *dam* methylation sequence is located at the base of the stem.

Mutation 4-10, which involves the loss of an A in a run of four A residues, could have been directed by a nearby inverted repeat (Fig. 4(c). As predicted for mutations directed by inverted repeats, it results in the perfection of the stem of the potential palindrome. We have no way of knowing whether the mutation was the result of a misalignment involving the repetition of the four A residues or occurred as the consequence of the misalignment

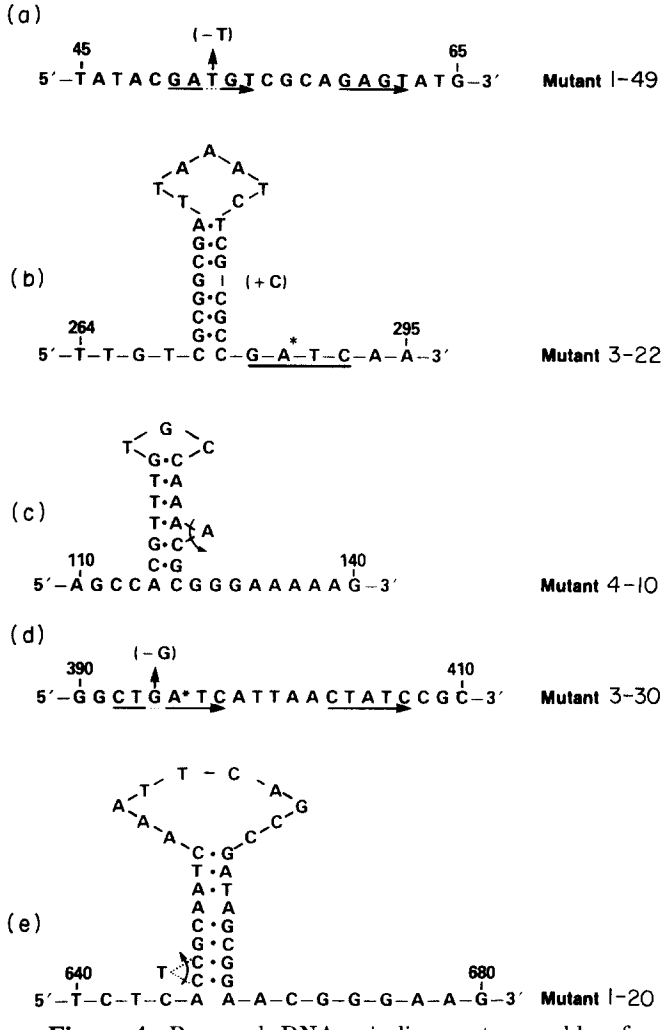

Figure 4. Proposed DNA misalignments capable of directing the formation of the spontaneous single-base frameshift mutations. (a) Loss of a T templated by a misalignment mediated by a nearby direct repeat. (b) Insertion of a C templated by an inverted repeat. The asterisk denotes a 6-methyladenine residue at the *dam* methylation sequence. (c) Loss of an A templated by an inverted repeat. (d) Loss of a G templated by a direct repeat. The asterisk denotes a 6-methyladenine residue. (e) Loss of a T directed by an inverted repeat.

made possible by the inverted repeat. Interestingly, evidence for the possible significance *in vivo* of the proposed structure was obtained from the DNA sequence analysis of this mutant. On sequencing gels a strong compression phenomenon was observed at the mutant location (but not at the wild-type location), presumably attributable to the formation of the proposed hairpin even under the denaturing conditions of the urea/polyacrylamide gel. Only at higher temperature did the anomaly disappear (data not shown).

Mutation 3-30 involves the loss of a G·C base-pair. This sequence alteration could have been templated by a direct repeat located 12 base-pairs 3′ to the mutated sequence (Fig. 4(d)). We note that this frameshift involves the loss of a G·C base-pair contained within a *dam* methylation sequence.

Mutation 1-20 involves the loss of a T and could not have been templated by a local direct repeat. However, a nearby inverted repeat could have directed the mutation (Fig. 4(e)). The eight-base-pair inverted repeat is separated further from the mutational site than in the previous examples and includes one mismatched base-pair.

From the sequenced examples of frameshift mutations we can draw several conclusions. All of the single-base frameshifts recovered in the *lacI* gene of *E. coli* can be modeled as sequence-directed i.e. templated by a nearby sequence. This sequence can be either a direct or an inverted repeat and need not be directly adjacent to the site of the mutation. In addition, we note that two out of the five mutations are at or near a G-A-T-C *dam* methylation sequence. We looked for an association of mutations with these sites because G-A-T-C sequences might be preferred sites for nick and/or gap formation (Gomez-Eichelmann & Lark, 1977; Lu *et al.*, 1984). Frameshift mutations are expected to occur more frequently in the neighborhood of such sites. For example, in bacteriophage T4, the frequency of frameshift events at different runs of A residues was related to the proximity to the dinucleotide A-C, where T4 Okazaki fragments are known to be initiated (Ripley & Shoemaker, 1983; Ripley & Glickman, 1983*b*). In the present case, two out of five frameshifts were at or immediately adjacent to a G-A-T-C site, although only three such sites are present in the approximately 1100-base-pair target. Although these data provide some evidence, a larger number of mutations is required to delineate the possible role of G-A-T-C sites in frameshift mutagenesis.

(d) *Deletions*

After hotspot frameshift mutations, deletions constituted the most frequent spontaneous event in the *lacI* gene. They represent 37% of all the non-hotspot mutations. While the total number of potential deletions is very large, of the 22 independently isolated deletions sequenced in this study, two were identical. In addition, one of the 22 mutants (mutant 3-18 in Table 2C) is identical to a deletion recovered in a previous study (Farabaugh *et al.*, 1978). The recovery of mutants more than once among the relatively small sample sizes in both these studies indicates that, while the potential for deletion formation seems rich and varied, the events are certainly not random.

An examination of the deletion mutations isolated in this study suggests that they may fall into two distinct classes. The majority (14/22) lie fully within the *lacI* gene, tend to be relatively short and are usually (12/14) flanked by repeated sequences. These mutations very much resemble those recovered in an earlier study by Farabaugh *et al.* (1978). We have called these deletions class I events (see Table 2). Class II deletions (Table 2C) include a group of eight that are unique in that they include portions of the regulatory region of

lacZ as shown in Figure 2. These deletions, regardless of whether or not they disrupt the *lacI* gene sufficiently to result in a non-functional repressor protein, include the *lac* operator, and the resulting mutants are therefore operator-constitutive (O^c). They differ from the *lacI–Z* fusions described by Müller-Hill & Kania (1974) in that they terminate before the translational start signal of the *lacZ* gene. These deletions therefore should be primarily regarded as *lacI–Z* "messenger" fusions from which *lacI* and *lacZ* are translated independently, although in the case of mutations 1-41, 2-18 and 2-31, a fused protein is also likely to result, since the two transcripts are in-frame without interposing nonsense codons. Class II deletions are longer than class I deletions: their average length is 195 base-pairs and, while none of the class I deletions (average length 78 base-pairs) extends over 250 base-pairs, full half of the class II events comprised more than 250 base-pairs. Furthermore, unlike class I deletions, the majority of class II deletions (5/8) lack repeated sequences at their endpoints. The two classes appear to be quite distinct and may reflect different mutational mechanisms.

The mechanism of deletion formation has been the subject of much discussion in the recent literature. In many cases, short stretches of repeated bases are found at the deletion endpoints (Farabaugh *et al.*, 1978; Albertini *et al.*, 1982). This observation has lent support to the hypothesis that deletion mutation occurs *via* intermediates in which one copy of the repeat is misaligned on the complement of the other. The deletion then comprises one copy of the repeats and all the sequence between them. Palindromic structures, made possible by the presence of inverted repeats, may also play a role in a deletion formation, either by shortening the distance between otherwise greatly separated sequences (Albertini *et al.*, 1982), or by exactly defining the endpoints of a deletion (Glickman & Ripley, 1984). Other models for deletion (and addition) formation are based on recombination events between directly repeated sequences, mediated by enzymes that recognize such homologies. It has been shown (Albertini *et al.*, 1982) that the frequency of larger deletions (comprising several hundreds of base-pairs) is reduced in a *recA*$^-$ strain, whereas that of the *lacI* frameshift hotspot is unchanged. As stated by these authors, the *recA* dependence of deletion formation is consistent with both recombination and slipped-mispairing models, since in the latter case RecA protein may be needed for the formation and/or stabilization of the misaligned intermediate.

An analysis of the *lacI* deletion mutants recovered in this study revealed the potential interaction of both repeated and palindromic sequences in determining the deletion endpoints. For example, in each of the 12 cases of class I deletions that are flanked by direct repeats, the presumed misaligned intermediates have the added potential for the formation of some form of stem–loop structure based on the palindromic components of the local DNA sequence. This is expected to contribute significantly to their stability. Two examples drawn from mutants isolated in this study are illustrated in Figure 5(a) and (b).

The relative roles of inverted and direct repeats in the formation of deletion remains an interesting question and one that is difficult to answer. However, a comparison of deletions 3-38 and 3-16 provides perhaps the most compelling evidence of the involvement of these structures in the production of deletion mutations. These two deletions differ by a single G at position 927 (Table 2D). Deletion 3-38 (87 base-pairs) is flanked by the six-base-pair repeat A-A-A-C-C-A. Hence, this deletion may be viewed as the consequence of a misaligned intermediate mediated by a direct repeat. The deletion of 3-16 (86 base-pairs), however, is one base out of phase relative to these direct repeats and cannot be simply explained by their misalignment. We therefore undertook a more detailed analysis of the potential of the local DNA sequence to provide structural intermediates.

The misalignments that can be drawn based on the six-base-pair repeat A-A-A-C-C-A are characterized by a lack of stabilization *via* palindromic components. This contrasts with what is generally observed in this group of deletions and raises some doubts as to the full explanation for mutant 3-38, even though this deletion is exactly flanked by this sequence. The sequence is part of a larger, slightly imperfect, 12-base-pair repeat (see Fig. 5(c)). A search for stable misalignments along these repeat units led to intermediate 5D in which some palindromic stabilization is attained. In this intermediate, a single C is unpaired in the stem, and a G·G mismatch is present in the misalignments mediated by the near-repeats. This structure, if formed, is likely to isomerize to intermediate 5E by extrusion of the extra base from the stem into the misaligned repeats, reducing the number of unpaired bases to a single G. This intermediate (5E) seems well suited to explain the occurrence of both deletions 3-38 and 3-16: the 87-base-pair deletion 3-38 on the basis of the increased stabilization of the misalignment of the two DNA strands (noting that 87-base-pair slippage has occurred between them) and the 86-base-pair deletion 3-16 by further enzymatic processing of the intermediate (noting that the stem–loop structure now contains 86 bases and is lacking guanine residue 927, which is exactly the residue retained in mutant 3-16 compared to 3-38). This pair of deletions thus provides a particularly interesting example of both repeated and inverted sequences in deletion mutagenesis.

(e) *Analysis of class II events*

The eight class II deletions all have one endpoint in a narrow stretch of DNA defining the *lac* operator. In two cases, 2-28 and 3-10, the other endpoint also falls in a palindromic sequence, the

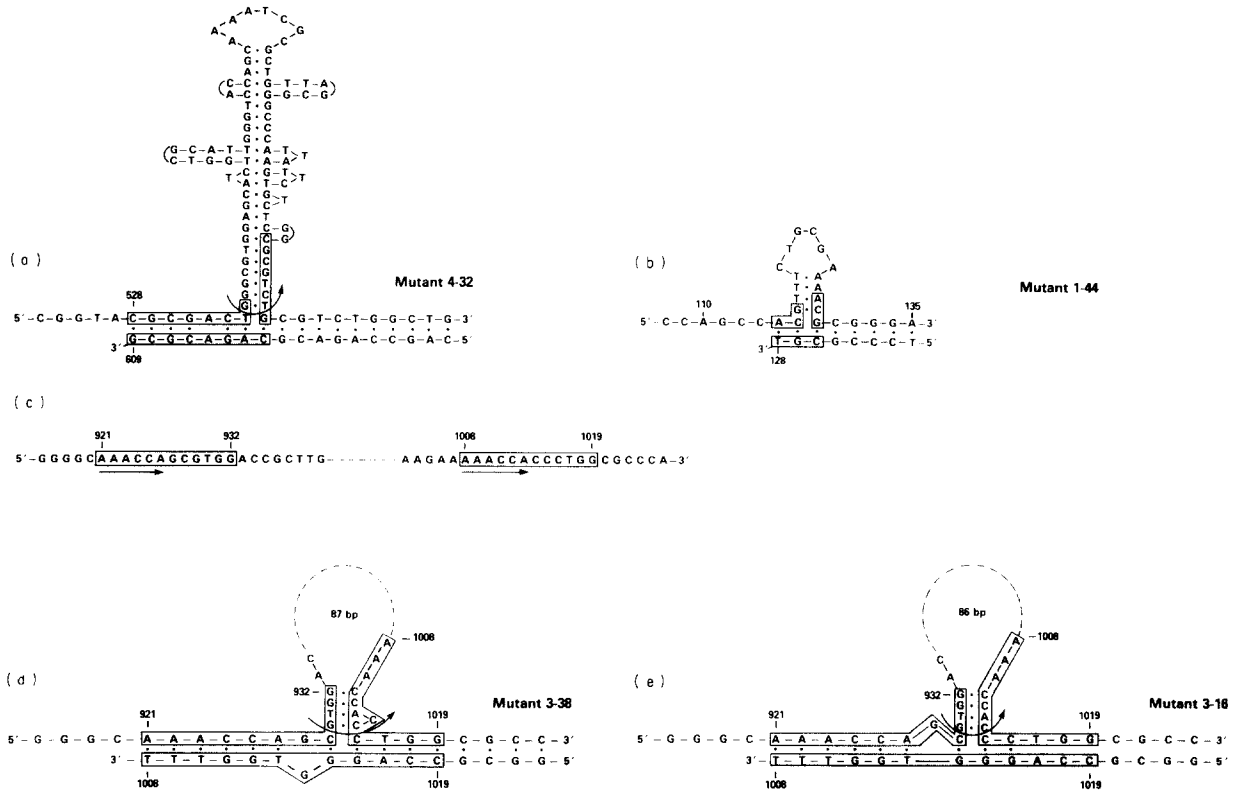

Figure 5. Structural intermediates of spontaneous deletions. (a) Mutant 4-32 is an 81-base-pair deletion mediated by both a direct repeat (7/8) and inverted repeat. (b) Mutant 1-44 is a 14-base-pair deletion mediated by both direct and inverted repeats. (c) The DNA sequences that define the endpoints of deletions 3-38 and 3-16 based on a 6-base-pair direct repeat (arrow) or imperfect 12-base-pair repeat (boxed sequence). Hyphens have been omitted for clarity. (d) Mutant 3-38 is an 87-base-pair deletion having both repeated and inverted DNA repeats at its endpoint. A single-base-pair mismatch is present at the stem of the secondary structure. Further isomerization of this structure yields (e), a structure that can account for the endpoints of the 86-base-pair deletion 3-16.

CAP site (Fig. 6). Of the remaining, three are flanked by repeated sequences and three have no direct repeats at their endpoints (Table 2). In the three cases lacking direct repeats, the breakage point in the 5′ end of the deletion occurred at the sequence CAG. We are not aware of any metabolic process that might preferentially involve this sequence. Thus, with the possible exception of the three deletions flanked by repeated sequences, the origin of class II deletions is obscure. Their origin

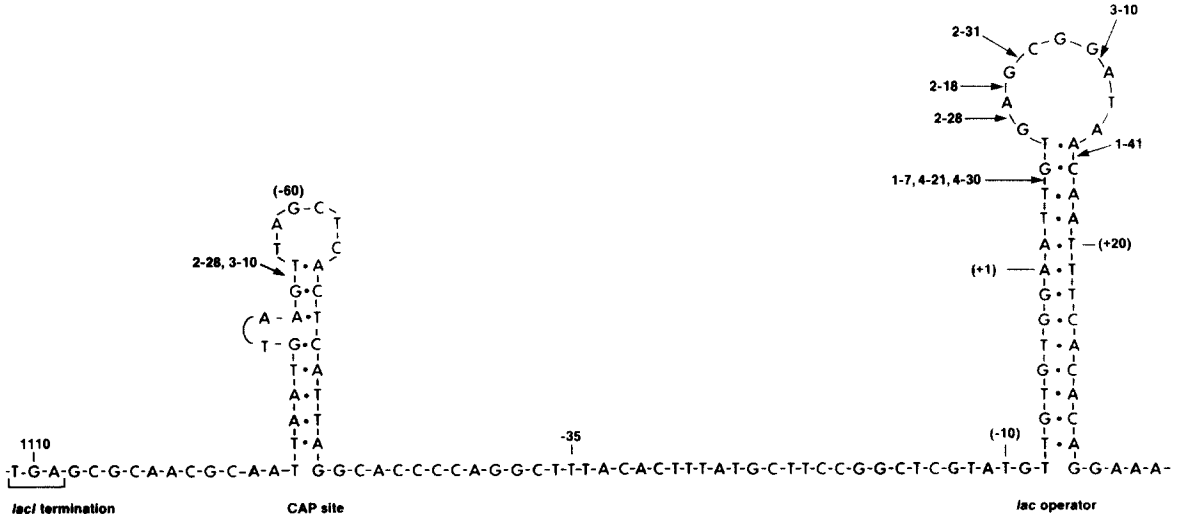

Figure 6. DNA sequence of the *lacZ* regulatory region with possible secondary structures at the CAP binding site and the *lac* operator. These hypothetical structures appear to be the preferred endpoints for class II deletions as described in the text. Arrows indicate deletion endpoints. This sequence contains a G instead of A at position −66 due to the L8 mutation present on F′*lac* in NR3835.

might reflect some aspect of DNA metabolism specific to this regulatory region, but the nature of these activities is unknown.

(f) *Duplications*

Three tandem duplications were recovered, of sizes 155, 258 and 457 base-pairs (Table 2D). They represent about 5% of the non-hotspot mutations obtained in this study. While only a small fraction of the mutations recovered in this study are duplications, evidence of their importance in evolution is considerable (Ohno, 1972; Hood *et al.*, 1978). The origin of duplication mutations is not well understood. None of the duplications recovered in this study was flanked by either repeated or inverted sequences capable of directing their formation. This was also the case for a duplication of 88 base-pairs previously recovered in the *lacI* gene by Calos *et al.* (1978*a*). We propose that the mechanism of addition formation is slightly different from that delineated for deletions. Whereas at any repeat sequence, reciprocal addition and deletion intermediates can be envisaged, two factors are likely to weigh against duplications when compared with deletions: a higher energy cost associated with the formation of an addition intermediate (since a larger stretch of DNA has to be melted out), and subsequent correction processes that may preferentially incise the loop-out strand (Abastado *et al.*, 1984), thereby fixing deletions and destroying addition mutations. We therefore envisage addition intermediates, while still being directed by repeated and inverted sequences, as being more complex in nature. In particular, they might contain loop structures in both DNA strands, negating the possible bias against additions that are due to correction mechanisms. We have discovered the potential for extensive secondary structure formation within each of the duplication sequences. How such structures might function in juxtaposing the exact endpoints of the observed duplications is difficult to show, however, and the answer to this problem requires a larger data base than is available at present.

(g) *Transposable elements*

Transposable elements are normal constituents of most bacterial genomes and of many extrachromosomal elements. They can alter the organization and expression of prokaryotic genes at frequencies similar to, and often greater than, other classes of spontaneous mutation (for a review, see Kleckner, 1981). Thus, among our collection of spontaneous *lacI*$^-$ mutations, representatives of this class of events were expected, and seven were observed. In each case the insertion was of IS*1*, a 768-base-pair element characterized by short inverted terminal repeats and the production of a direct repeat of a nine-base-pair segment at the site of insertion (Johnsrud, 1979; Calos & Miller, 1980). The *E. coli* chromosome typically contains between 8 and 14 copies of IS*1* at different sites (Saedler & Heiss, 1973; Chow, 1977). The seven IS*1* insertions occurred at seven different sites in the *lacI* gene (Table 2). Two insertions of IS*1* into *lacI* reported earlier by Calos *et al.* (1978*b*) occurred at sites identical to our mutants 3-19 and 3-24. However, in each case the insertion had the reverse orientation to that found here.

The insertion of IS*1* is apparently not random (Kuhn *et al.*, 1979) but demonstrates what has been termed regional rather than site specificity (Kleckner, 1981). The recovery of IS*1* at seven different sites in the *lacI* gene would confirm the relaxed aspect of integration specificity. G·C base-pairs were found at one and often both ends of the nine-base-pair duplication (11 out of 14), a feature of IS*1*-mediated insertion noted previously by Galas *et al.* (1980). It has been reported that insertions of IS*1* and also of transposon Tn*9*, which carries IS*1* as its flanking elements, occurs preferentially in A+T-rich regions (Galas *et al.*, 1980). The sites at which IS*1* insertions were recovered in the *lacI* gene were not particularly rich in A·T base-pairs. In fact, the *lacI* gene is G+C-rich (56%) and has virtually no locally A+T-rich regions. The seven IS*1* insertions recovered in this study represent 12% of the non-hotspot mutations. These insertions are therefore not rare events. In addition, if IS*1* insertion generally demonstrates a significant preference for A+T-rich regions, the absence of such sequences in the *lacI* gene has as a consequence the fact that our estimate for the contribution of this element to spontaneous mutagenesis in general might be an underestimate.

(h) *Concluding remarks*

In this paper we have reported the DNA sequence changes of a large collection of spontaneous mutations in the *lacI* gene of *E. coli*, and demonstrated the feasibility of our approach to studying the specificity of mutation by direct DNA sequence analysis, using a newly developed system for recovering episomal *lacI*$^-$ mutations on an M13 vector. This method provides easy access to the gene for sequence analysis. The study has demonstrated the complexity of spontaneous mutation: at least six different classes of mutation were discovered, which presumably represent a variety of mutational mechanisms. Since the analysis was non-selective as to the type of mutation and comprised a large sequence target, the data may be regarded as an indicator of the relative importance of the various classes of spontaneous mutation. The large frameshift hotspot in *lacI* should perhaps not be included in such a comparison, although mutational hotspots, as most visible indication of the non-randomness of mutation, will be likely components of any comprehensive mutational study. Of the non-hotspot mutations, deletions formed the largest contributor, closely followed by base substitutions. Most efforts directed at the question of how organisms would lower their mutation rates have

focused on the accuracy of DNA replication. Our results seem to support the notion that organisms would not lower their rates of replication errors below levels dictated by other, independent sources of mutation such as perhaps deletion formation. The next most frequent mutational events are IS*1* insertions, single-base frameshifts and duplications, which are lesser events but, if combined, contributed approximately 30% to the non-hotspot events. In further studies we will place greater emphasis on specific subgroups of mutations and their possible origins. The use of bacterial mutant strains defective in repair and/or replication pathways will be an important aspect of such studies aimed at unraveling the contributing mechanisms in spontaneous mutation.

We thank J. Drake and T. Kunkel of this Institute for their critical reading and helpful comments on this manuscript.

References

Abastado, J. P., Cami, B., Dinh, T. H., Igolen, J. & Kourilsky, P. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 5792–5796.

Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. (1982). *Cell*, **29**, 319–328.

Calos, M. P. & Miller, J. H. (1980). *Cell*, **20**, 579–595.

Calos, M. P. & Miller, J. H. (1981). *J. Mol. Biol.* **153**, 39–66.

Calos, M. P., Galas, D. & Miller, J. H. (1978*a*). *J. Mol. Biol.* **126**, 865–869.

Calos, M. P., Johnsrud, L. & Miller, J. H. (1978*b*). *Cell*, **13**, 411–418.

Cheung, S., Arndt, K. & Lu, P. (1984). *Proc. Nat Acad. Sci., U.S.A.* **81**, 3665–3669.

Chow, L. T. (1977). In *DNA Insertion Elements, Plasmids and Episomes* (Bukhari, A. I., Shapiro, J. A. & Adhya, S. L., eds), pp. 73–80, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Choy, H. E. & Fowler, R. G. (1985). *Mutat. Res.* **142**, 93–97.

Coulondre, C. & Miller, J. H. (1977). *J. Mol. Biol.* **177**, 577–606.

Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. (1978). *Nature (London)*, **274**, 775–780.

Drake, J. W. (1969). *Nature (London)*, **221**, 1132.

Drake, J. W. (1970). *The Molecular Basis of Mutation*, Holden-Day, San Francisco.

Duncan, B. K. & Miller, J. H. (1980). *Nature (London)*, **287**, 560–561.

Farabaugh, P. J. (1978). *Nature (London)*, **274**, 765–769.

Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. (1978). *J. Mol. Biol.* **126**, 847–863.

Foster, P. L., Eisenstadt, E. & Miller, J. H. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 2695–2698.

Galas, D. J., Calos, M. P. & Miller, J. H. (1980). *J. Mol. Biol.* **144**, 19–41.

Glickman, B. W. (1979). *Mutat. Res.* **61**, 153–162.

Glickman, B. W. & Radman, M. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 1063–1067.

Glickman, B. W. & Ripley, L. S. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 512–516.

Golding, G. B. & Glickman, B. W. (1986). *Can. J. Genet.* In the press.

Gomez-Eichelman, M. C. & Lark, K. G. (1977). *J. Mol. Biol.* **117**, 621–635.

Hood, J. M., Fowler, A. B. & Zabin, I. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 113–116.

Jonsrud, L. (1979). *Mol. Gen. Genet.* **169**, 213–218.

Kleckner, N. (1981). *Annu. Rev. Genet.* **15**, 391–404.

Kondo, S., Ichikawa, H., Iwo, K. & Kato, T. *Genetics*, **66**, 187–217.

Kuhn, S., Starlinger, P. & Fritz, H. J. (1979). *Mol. Gen. Genet.* **167**, 235–242.

Kunz, B. A. & Glickman, B. W. (1984). *Genetics*, **106**, 347–364.

Lu, A. L., Welsh, K., Clark, S., Su, S. S. & Modrich, P. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 589–596.

Macquat, L., Thornton, K. & Reznikoff, W. (1980). *J. Mol. Biol.* **139**, 537–549.

Miller, J. H. (1984). *J. Mol. Biol.* **180**, 205–212.

Miller, J. H. & Low, K. B. (1984). *Cell*, **37**, 675–682.

Miller, J. H. & Schmeissner, U. (1979). *J. Mol. Biol.* **131**, 223–248.

Müller-Hill, B. & Kania, J. (1974). *Nature (London)*, **249**, 561–563.

Ohno, S. (1972). *Evolution by Gene Duplication*, Springer-Verlag, New York.

Okada, Y., Streisinger, G., Owen, J., Newton, J., Tsugita, A. & Inouye, M. (1972). *Nature (London)*, **263**, 338–341.

Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., Brosius, J., Dull, T. J. & Noller, M. F. (1981). *J. Mol. Biol.* **149**, 337–376.

Reznikoff, W. S. & Abelson, J. N. (1978). In *The Operon*. (Miller, J. H. & Reznikoff, W. S., eds), pp. 221–243, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Ripley, L. S. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 4128–4132.

Ripley, L. S. & Glickman, B. W. (1983*a*). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 851–861.

Ripley, L. S. & Glickman, B. W. (1983*b*). In *Cellular Responses to DNA Damage* (Friedberg, E., ed.), pp. 524–540, Alan Liss, New York.

Ripley, L. S. & Shoemaker, N. B. (1983). *Genetics*, **103**, 353–366.

Saedler, H. & Heiss, B. (1973). *Mol. Gen. Genet.* **122**, 267–277.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Sargentini, N. J. & Smith, K. C. (1981). *Carcinogenesis*, **2**, 863–872.

Schaaper, R. M., Kunkel, T. A. & Loeb, L. A. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 487–491.

Schaaper, R. M., Danforth, B. N. & Glickman, B. W. (1985). *Gene*, **39**, 181–189.

Schmeissner, U., Ganem, D. & Miller, J. H. (1977). *J. Mol. Biol.* **109**, 303–326.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzhagi, E. & Inouye, M. ((1966). *Cold Spring Harbor Symp. Quant. Biol.* **33**, 77–84.

Todd, P. A. & Glickman, B. W. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 4123–4127.

Vogel, H. J. & Bonner, D. M. (1956). *J. Biol. Chem.* **218**, 97–106.

Edited by M. E. Gottesman