Published in final edited form as:

DNA Repair (Amst). 2007 February 4; 6(2): 177–189.

A molecular characterization of spontaneous frameshift mutagenesis within the *trpA* gene of *Escherichia coli*

Aaron Hardin, Christopher F. Villalta¹, Michael Doan, Mouna Jabri, Valliammal Chockalingham, Steven J. White, and Robert G. Fowler*

Department of Biological Sciences, San Jose State University, San Jose, CA 95192, USA

Abstract

Spontaneous frameshift mutations are an important source of genetic variation in all species and cause a large number of genetic disorders in humans. To enhance our understanding of the molecular mechanisms of frameshift mutagenesis, 583 spontaneous Trp⁺ revertants of two *trpA* frameshift alleles in *Escherichia coli* were isolated and DNA sequenced. In order to measure the contribution of methyl-directed mismatch repair to frameshift production, mutational spectra were constructed for both mismatch repair-proficient and repair-defective strains. The molecular origins of practically all of the frameshifts analyzed could be explained by one of six simple models based upon misalignment of the template or nascent DNA strands with or without misincorporation of primer nucleotides during DNA replication. Most frameshifts occurred within mononucleotide runs as has been shown often in previous studies but the location of the 76 frameshift sites was usually outside of runs. Mismatch repair generally was most effective in preventing the occurrence of frameshifts within runs but there was much variation from site to site. Most frameshift sites outside of runs appear to be refractory to mismatch repair although the small number of occurrences at most of these sites make firm conclusions impossible. There was a dense pattern of reversion sites within the *trpA* DNA region where reversion events could occur, suggesting that, in general, most DNA sequences are capable of undergoing spontaneous mutational events during replication that can lead to small deletions and insertions. Many of these errors are likely to occur at low frequencies and be tolerated as events too costly to prevent or repair. These studies also revealed an unpredicted flexibility in the primary amino acid sequence of the *trpA* product, the α subunit of tryptophan synthase.

1. Introduction

As with many other fundamental genetic processes, *Escherichia coli* has been used as a model organism for the study of spontaneous mutagenesis. Many base-pair substitutions are assumed to occur during DNA replication and have been attributed to a failure in one of three steps during DNA replication: base selection, proofreading, and methyl-directed mismatch repair [1,2]. Numerous studies have helped to elucidate many of the specific details of the origins of base-pair substitutions in *E. coli* [3].

Studies on the creation of spontaneous frameshift mutations in *E. coli* are more limited even though frameshifts often lead to more severe phenotypic consequences than base-pair substitutions. The original work of Streisinger and co-workers [4,5] with phage T4 attributed hotspot frameshift events to slippage of either the template strand or the primer strand in

*Corresponding author. Tel.: +1 408-924-4843; fax: +1 408-924-4840, E-mail address: rfowler@email.sjsu.edu (R.G. Fowler).

¹Present address: Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

homopolymer sequences during DNA replication. This model has since been used to explain frameshift hotspots within homopolymers in *E. coli* [6,7] as well as other organisms [8,9].

Many spontaneous frameshifts in *E. coli* don't occur within homopolymeric sequences [10, 11] and alternative models have been developed to account for these events [11–14]. Two mechanisms have been proposed to specifically explain –1 frameshifts. In the first, direct misincorporation, misalignment occurs when the DNA polymerase erroneously creates a misincorporation at the 3' primer end. Next the primer nucleotide realigns by pairing with a complementary template nucleotide one nucleotide down the template strand. Subsequent synthesis creates a single nucleotide deletion [13].

The second model, dNTP-stabilized misalignment, proposes that misalignment of the template strand first occurs. The DNA polymerase then correctly inserts the nucleotide complementary to the next downstream template nucleotide. Subsequent completion of replication may again lead to a –1 frameshift [12,14].

These two models to explain –1 frameshifts as well as the Streisinger model to account for the occurrence of frameshifts in nucleotide runs have been recently reviewed [15]. Most molecular studies of the origins of frameshifts have focused on –1 or, less often, on +1 events. Less is known about how frameshifts of two or more bases arise.

In addition to these “local” models for the origin of frameshifts in nonruns, Ripley and others have shown that some frameshifts may be templated by direct repeats and inverted repeats not located adjacent to the mutant sites [10,11,16,17]. Other authors have suggested the possibility that misalignment, a central component to the above models for the creation of frameshifts, may not always be required. For example, Bichara et al. [18] considered the possibility that deletion frameshifts occurring in (GpC)_n runs may result from the formation of Z DNA within these regions. Such DNA is then processed in some unknown fashion to produce frameshifts.

An important step in determining mechanisms of spontaneous frameshift mutagenesis is the construction of a spectrum of sequenced mutations. In *E. coli* two forward mutational assays have long been used to produce spectra in dividing cells, the N-terminal region of the *lacI* gene where *lacI*^d mutations occur [7] and the *rpsL* gene [19]. Both genes are carried on plasmids. Recently a third forward assay has been utilized, the chromosomal *tonB* gene [20,21].

Forward assays are advantageous in having a large target where mutations can be detected at many sites. However in wild-type backgrounds, apart from hotspots, frameshifts are often a minority among other, more common, types of spontaneous mutations found in such assays. This makes quantifying frameshift events at most individual sites difficult. For example, Schaaper and Dunn [22] found only 18 single-base frameshifts among 414 sequenced *lacI*^d mutations and no frameshift site had more than two occurrences.

In contrast, reversion assays often possess great sensitivity for quantifying particular mutational events but are assumed to have a very limited target of a few nucleotides so that only a small number of specific mutational events can be detected. While this assumption is valid for reversion assays for base-pair substitutions, target size for reversion of frameshift alleles may be larger. The spectrum of spontaneous His⁺ revertant frameshift mutations of the *hisD3052* allele in *Salmonella typhimurium* included full revertants and 182 unique second-site revertants and the reversion target was a minimum of 76 nucleotides [9].

In order to obtain as large a spectrum of frameshifts as possible and still be able to calculate frequencies for the more common of these events, we sequenced spontaneous revertants of two chromosomal *E. coli* *trpA* frameshift alleles in wild-type and DNA repair-defective backgrounds. Models are presented to account for these events and we briefly discuss our data

with respect to structure/function relationships in the *trpA* gene product, the α subunit of tryptophan synthase.

2. Materials and Methods

2.1 Bacterial strains

The bacterial strains used in this study are listed in Table 1. P1 transductions were performed using P1*virA*. The *mutL218::Tn10* allele was transferred from KD1093 using the tetracycline resistance conferred by the Tn10 transposon as a selectable marker. Tet^r transductants were tested for mutator activity of the *mutL218* allele by measuring the frequency of nalidixic acid-resistant (Nal^r) mutations. Transductants showing high levels of Nal^r frequencies (1×10^{-6} or higher) were assumed to have acquired the *mutL218* allele.

2.2 Media

LB (Luria-Bertani) broth was 1% tryptone, 0.5% yeast extract and 1% NaCl. Tryptone plates were made of 1% tryptone and 0.5% NaCl solidified with 1.5% agar. Minimal medium (MM) contained Vogel-Bonner salts [25], 0.2% glucose, 1 μ g/ml thiamine and 50 μ g/ml amino acids when required. Mutation to nalidixic acid resistance (Nal^r) was measured on tryptone plates containing 50 μ g/ml of nalidixic acid. Tetracycline-resistant transductants were selected on tryptone plates containing 15 μ g/ml tetracycline. The saline for dilutions and washing was 0.85% NaCl.

2.3 Reversion frequencies

To measure Trp⁺ reversion frequencies, three independent cultures for each *trpA* frameshift strain were started from separate colonies and grown to saturation in 5 ml of LB. Each of these cultures was diluted and about 300 cells were used as inocula to initiate each of four new 5 ml cultures (12 total cultures), which were grown to saturation. The cultures were then washed and concentrated in saline, diluted appropriately and plated on MM without tryptophan to select for Trp⁺ revertants and on tryptone to titer for total cells. The MM selective plates were incubated for 4 days at 37°C and then the Trp⁺ colonies were counted. The tryptone plates were incubated for 24 h and the resulting colonies counted. Reversion frequencies for each strain were calculated by dividing the average number of Trp⁺ revertants by the average cell titer. (In addition to the above cultures, several cultures were grown to saturation and plated on selective medium to secure sufficient numbers of Trp⁺ revertants for DNA sequencing.)

To confirm the presence of the *mutL218* allele in appropriate strains, saturated cultures were plated on selective nalidixic acid plates without diluting to score Nal^r mutations. *mutL* strains normally produce about 100 Nal^r colonies on selective plates while wild-type strains create none or a small number.

2.4 DNA sequencing of Trp⁺ revertants

In order to identify independent Trp⁺ revertants, every Trp⁺ colony was taken from the selective plates and restreaked on the same medium. Because multiple revertants were recovered from the same culture for sequencing, it is possible that they may not all represent independent events. The purifications were then incubated for 48 h. A small number of “revertants” failed to grow when restreaked and were discarded.

Direct DNA sequencing was used to deduce the nucleotide changes associated with the two *trpA* frameshift alleles and their revertants. We used the procedure of Fix [26] with slight modifications. For each *trpA* mutant or revertant strain, a single colony was completely resuspended in 30 μ l of sterile water. The sample was placed in a boiling water bath for 2 min and frozen at -70°C. For PCR, oligonucleotide primers specific for the 5' and 3' sides (5'-

GATCCGGGCATTGAGCAGTCA -3' and 3'- CATGTTGGCTACTTTTCGCCGC -5', respectively) of the *trpA* frameshift alleles and their revertants within the *trpA* gene were used [27]. The 717 base-pair product was cleaned and sent out for sequencing (Gene Gateway, LLC, Hayward, CA). All samples were sequenced at least twice (once for each strand) with the same primers used for the PCR reactions.

2.5 Software

Alignments of sequences were performed using Vector NTI. “etandem” (Emboss) was used to locate tandem repeats and “einverted” (Emboss) was used to find inverted repeats within the target regions for reversion of the *trpA9813* and *trpA540* alleles. Direct repeats between 2 and 50 nucleotides were searched for as were inverted repeats between 3 and 100 nucleotides. All repeats that were greater than 10 nucleotides apart were excluded.

3. Results

3.1 Nucleotide sequences of the *trpA* frameshift alleles

The *trpA21* allele and the *trpA540* allele, UV-induced and ICR191-induced respectively, were originally analyzed by Berger et al [28] who deduced that both mutations resulted from single or double base deletions and that the mutant sites were very close to each other. Later it was shown that *trpA21* resulted from a single base deletion [29]. The UV-induced *trpA9813* allele was first studied by Brammar et al [30] who concluded that it must result from the loss of an A from a tyrosine codon. Sargentini and Smith [23], using the wild-type *trpA* sequence [27] and the analysis of Siegel and Vaccaro [31], concluded that *trpA21* results from loss of a single C or G in codons 216 or 217 of the *trpA* gene and *trpA9813* results from the loss of an A in codon 175.

Our direct sequencing confirms the deduced *trpA9813* change and reveals that both *trpA21* and *trpA540* result from the loss of a C from a run of four C's in codons 216 and 217 (Fig. 1). Because *trpA21* and *trpA540* have identical mutations, *trpA21* was not used in subsequent reversion analysis.

3.2 Trp⁺ reversion frequencies of two *trpA* frameshift alleles in wild-type and *mutL218* strains

Trp⁺ reversion frequencies for the *trpA9813* and *trpA540* frameshift alleles are shown in Fig. 1. The wild-type frequencies for both are similar to those found in an earlier study [31]. The *mutL218* allele causes defective methyl-directed mismatch repair that results in reduced genomic mutation avoidance [7,24]. Our data indicate that the Trp⁺ reversion frequencies of *trpA9813* and *trpA540* are slightly enhanced (2–5-fold) by the presence of *mutL218* (Fig. 1). The values for *trpA9813* and *trpA540* in a *mutL218* background are similar to reversion frequencies found previously for the same *trpA* alleles in other strains defective for mismatch repair [31].

3.3 DNA sequencing of Trp⁺ revertants

For the *trpA9813* allele, Trp⁺ revertants were DNA sequenced to discern the molecular nature of the reversion events. The results are shown in Table 2 for both genetic backgrounds.

19 different frameshift mutations that reverted *trpA9813* (Table 2) were located within the region from nucleotide position 516 to 536 (Fig. 1). [Frameshift mutation is used here and elsewhere to refer to any loss or gain of a small number of nucleotides.] About half of these 19 mutations involve nucleotide deletions. Only four of the frameshift events resulted from the addition of a nucleotide to a run of two or more bases and two of these were insertions of a nucleotide different from the one that composed the run.

For the wild-type spectrum, four sites were warm spots and accounted for approximately 83% of all the mutations (Table 2). An approximate range of reversion frequencies for each of the warm spots can be calculated as between 2.4 and 4.0×10^{-9} (reversion frequency of *trpA9813* [Fig. 1] $\times .15-.25$). For comparison, in the forward *lacI* mutation assay the frequency of -1 events at the warmest sites was approximately 6.2×10^{-9} [22]. (+1 frameshifts are not detected in this system.) Because of the low level of occurrence for the remaining revertant classes, it is not possible to calculate precise individual reversion frequencies, but all appear to be less than 1×10^{-9} .

For the *mutL218* strain one frameshift, R2, resulting from the insertion of a T into a run of four T's dominated the spectrum, accounting for 78% of all the revertants. No *trpA9813* revertants were recovered that fully restored the wild-type sequence in either background.

For the *trpA540* allele, 57 different reversion events occurred from nucleotide position 603 to 660 (Fig. 1) and are shown in Table 3. (No changes in the *trpA* DNA sequence were found for 5 *trpA540* revertants and these "revertants" are not included in Table 4. They may be suppressor mutations.) 39 of the frameshift events resulted from insertions with the remainder involving deletions. 14 of the 57 frameshift events are +1 insertions in runs. Four of these +1 insertions involve the addition of a nucleotide that differs from the polymer of the run.

For the *trpA540* wild-type spectrum, R1 revertants, resulting from a C insertion in a run of two C's, were the most common mutation and account for approximately 38% of all the *trpA540* revertants. The reversion frequency for this class is calculated to be 1.1×10^{-8} . The remaining 42 revertant classes occur at lower frequencies. Of these, only the R5 class has a sufficient number of occurrences to make a reversion frequency calculation (2.5×10^{-9}). None of the revertants sequenced in the wild-type strain were true revertants that restored the wild-type *trpA* sequence even though this change involves a simple insertion of a C in a run of three C's (Fig. 1).

The *mutL218* spectrum was also dominated by the R1 revertant class that accounts for approximately 50% of all revertants. In contrast to the wild-type spectrum, revertants that restore the wild-type *trpA* sequence were recovered in the *mutL218* background as a significant proportion (13%) of the spectrum.

3.4 Mechanisms of frameshift mutagenesis

Most of the revertants for both *trpA* alleles can be explained by one of six models involving misalignment during DNA replication (Table 4; Fig. 2). The first model (Fig. 2a) is the original Streisinger model [4] with slippage occurring in the primer strand and resulting in +1 insertions within a polynucleotide run. The second model assumes that misincorporation first occurs with the 3' primer nucleotide followed by misalignment of the nascent strand with the primer now correctly paired with the adjacent template nucleotide (Fig. 2b). This model accounts for several of the +1, +4, and +7 insertions.

The third model (Fig. 2c), originally described by Kunkel [12], is similar but assumes the misalignment occurs with the template strand to affect correct pairing. This is postulated to have resulted in some of the -2 , -5 , and -8 deletions. The -2 events can also be explained by the dNTP-stabilized misalignment model (not shown). Here misalignment occurs first and then the complementary nucleotide is inserted opposite the template base two positions downstream [12,14]. This mechanism would seem unlikely to account for the -5 and -8 deletions.

The fourth model (Fig. 2d) is similar to the second model except that correct incorporation of the primer nucleotide precedes misalignment of the primer strand. This model could account for some of the +4 and +7 insertions. The fifth model (Fig. 2e) is identical to the third model

except that correct incorporation precedes misalignment of the template strand and could explain some of the -2 and -5 events and all of the -11 deletions.

The sixth model involves misincorporation of the primer nucleotide followed by correct incorporation of the next nucleotide and the misalignment of the primer strand (Fig. 2f). This model accounts for the unusual +1 insertions that were surprisingly common in the frameshift spectra. These were insertions of nucleotides that differed from the nucleotide of the homopolymer run into which they were inserted.

Four revertants do not appear to be explained by any of the above models. Models for these revertants are shown in Fig. 3. Two revertants, R18 of *trpA9813* and R17 of *trpA540*, can be accounted for by assuming misincorporation first occurs with the 3' primer nucleotide (shown for R18 in Fig. 3a). This mispairing is not corrected, the polymerase proceeds and misincorporation occurs a second time. Then misalignment of the nascent strand occurs so that the 3' primer nucleotide is correctly paired, resulting in a +1 insertion.

The R8 and R13 -8 deletion revertants of *trpA9813* occur at the same site and likely have similar origins although they are not identical. We assume that the short three nucleotide inverted repeat allows a hairpin loop to be formed. This may occur during replication or any other time when this region becomes denatured. A uncharacterized nuclease activity then removes the single-stranded region (8 nucleotides) of the hairpin. For R8 this removal is precise, while for R13 it is imprecise with the A in the stem that is adjacent to the loop being removed while the C at the other end of the loop is retained.

The four revertants in this final group are not easily explained by previously described mechanisms of frameshift mutagenesis including templating as by quasipalindromes [16] and the above models presented (Fig. 3a, b) are admittedly not satisfying. Each of these events only occurred once which is consistent with a complex origin.

3.5 Amino acid changes in the tryptophan synthase α chain of Trp⁺ revertants

The amino acid changes associated with revertants of the *trpA9813* and *trpA540* alleles are shown in Fig. 4. The *trpA* gene encodes the α subunit of the tetrameric ($\alpha\beta\beta\alpha$) tryptophan synthase complex that catalyzes the final two steps in tryptophan biosynthesis [32]. The α subunit catalyzes the conversion of indole 3-glycerol phosphate (IGP) to indole and glyceraldehyde-3-phosphate (G3P), the α reaction, and the β subunit catalyzes the production of tryptophan from indole and L-serine, the β reaction. The physiologically important overall reaction catalyzed by the $\alpha\beta\beta\alpha$ tetramer is: $\text{IGP} + \text{L-serine} \rightarrow \text{L-tryptophan} + \text{G3P} + \text{H}_2\text{O}$ ($\alpha\beta$ -reaction).

A series of early genetic studies revealed that only a small number of sites in the 268 amino acid α subunit were capable of being mutated to produce an inactive protein [33–35]. Most of these substitutions were later found to be located in the active center of the α subunit [36]. Five of these sites (positions 175, 177, 179, 211 and 213) are within the two regions affected by reversion of the *trpA9813* and *trpA540* alleles. These two regions also flank loop αL6 ($\alpha\text{Arg 179} - \alpha\text{Leu193}$), which participates in substrate binding and transmitting the effects of such binding to the β subunit [37,38].

Reversion of *trpA9813* involves amino acids from position 172 to 179 (Fig. 4). Tyrosine 175 is assumed to be critical for the proper geometry of the active site since a central role in catalysis or substrate binding was ruled out by site-directed mutagenesis experiments [39]. Substitution of tyrosine 175 with phenylalanine retained some activity of the α and $\alpha\beta$ reactions [39]. Some *trpA9813* revertants have retained the mutant phenylalanine 175 substitution (not shown in Fig. 4) and their prototrophy is consistent with the previous *in vitro* activity measurements. The

substitution leucine at position 175 indicates that at least one additional amino acid is able to function at this position.

Several *trpA9813* revertants have multiple substitutions involving position 175 and other sites and some have deletions or gains of amino acids indicating the tolerance to change this region has in spite of its importance in enzymatic activity. This is exemplified by the loss of arginine 179 which was previously shown to be involved in transmission of substrate –induced conformation changes between the two subunits [40].

The amino acid changes associated with reversion of *trpA540* involve amino acid positions 201 to 223 (Fig. 4). Within this region the 211 and 213 sites have previously been identified as possessing important roles in α subunit function [41]. Glycine 211 has been shown to be involved in maintaining the geometry of the active site [39]. All of the *trpA540* revertants have glycine 211; however +1 DNA changes before the 211 codon do not change the codon meaning. The glycine 213 site probably also has some role in facilitating the structural integrity of the active site since a valine at this position reduces enzymatic activity if a valine is also at position 211 [35]. A single valine at either position results in nearly full activity. Some of the *trpA540* revertants have tryptophan at position 213 showing that this substitution allows some activity.

Many of the *trpA540* revertants have multiple substitutions (up to 13 amino acids) as well as amino acid insertions and deletions. Again this is in a region of the α subunit known to be in or near the active site but seemingly having a large degree of flexibility in primary structure while still retaining some catalytic activity.

4. Discussion

4.1 Structure/function relationships in the α subunit

The amino acid changes associated with the Trp⁺ revertants recovered in this study reveal unexpected flexibility in the structure of the α subunit of tryptophan synthase. The tyrosine 175, leucine 177, arginine 179, and glycine 213 sites, previously demonstrated to be critical positions near or at the active site [39,40], can be altered by substitution with retention of at least partial enzymatic activity. Perhaps more unexpected are the losses and gains of single and multiple amino acids that permit partial function. For example, losses of up to three amino acids in revertants of *trpA9813* and up to four amino acids in revertants of *trpA540* are tolerated.

We compared the positions altered to all available *trpA* sequences in GenBank from other *E. coli* strains. Only two amino acid changes found among our revertants appear as apparently neutral substitutions among 109 laboratory, commensal, and pathogenic *E. coli* strains, a serine at position 208 and a phenylalanine at position 215. In some cases the limited substitutions among the diverse strains may reflect the limited possibility of a mutational event giving rise to a neutral substitution rather than selection against an unfit substitution. For example, the substitution of arginine for the wild-type alanine at position 216 (revertant R1 of *trpA540*) appears to allow near normal levels of enzyme activity, at least as measured by growth rate in minimal medium (R.G. Fowler, unpublished data), but would require a rare tandem base-pair substitution to occur spontaneously in a wild-type strain.

4.2 Frameshift mutagenesis in a wild-type background

Spontaneous frameshifts has long been assumed to occur during DNA replication in dividing *E. coli* cells [42]. We are suggesting that those recovered here mostly result from local misalignments occurring at or near the replication fork. This requires that the replisome accommodate such alterations during the normal replication cycle. DNA polymerase III

hologenzyme (HE) is highly processive and simultaneously replicates both the leading and lagging chromosomal strands (reviewed by McHenry [43]).

An *in vitro* study has shown that HE is capable of creating -1 frameshifts at a high frequency [44]. Data from that study supported the conclusion that these events were caused by a misincorporation/misalignment model of frameshift mutagenesis [13] and it was proposed that the initial misincorporation caused HE to stall. If proofreading didn't occur, the primer nucleotide usually paired with the next downstream template nucleotide if it was complementary, producing a -1 frameshift.

While the above study showed that HE *in vitro* caused one kind of small frameshift, the *in vivo* situation is more complex. HE may need to dissociate from replicating DNA to accommodate the formation, via misalignment, of some or all of the mutations that we observed. There are several observations suggesting that the processive HE does dissociate regularly during replication: (1) Many DNA polymerases tend to stall following misincorporation [45] and probably also at lesions and some forms of secondary structure. One consequence of such lagging has been suggested to be dissociation of the polymerase [32, 46]. (2) Other *E. coli* DNA polymerases gain access to the growing point during the replication cycle. Overexpression of Pol IV, one of the Y-family error-prone polymerases [47], leads to a mutator effect, showing that Pol IV has access to the site of polymerization and replaces HE [48]. Another recent study demonstrated that Pol II also gains access to the replication point, possibly in a role as proofreader for previous HE-induced misincorporations that caused its dissociation [49]. However it is unlikely that Pol V, another of the Y family of error-prone polymerases, affects reversion of *trpA9813* or *trpA540* since wild-type reversion frequencies are not altered in *umuC* strains, which are defective in this polymerase activity [50]. (3) There are differences in accuracy of replication between the leading and lagging strands with the latter more accurate during normal replication [46]. Perhaps HE can more readily dissociate after misincorporation on the lagging strand because dissociation is a normal part of lagging strand replication.

4.3 The role of mismatch repair in the formation of frameshifts

The *mutL218* strains generate spectra of frameshifts that have resulted from replication errors without subsequent methyl-directed mismatch repair. (However there may be *mutL*-independent repair or processing that affects the fixation of frameshifts after replication that is not abolished in the *mutL* background.)

While the overall reversion frequency of *trpA9813* is only modestly enhanced (4.6x) by the presence of *mutL218*, the spectrum of reversion events is dramatically altered. The R2 revertant class, a T insertion in a run of 4 T's, is very common (78% of all occurrences). The frequency of this single mutation can be calculated to be 5.7×10^{-8} ($7.4 \times 10^{-8} \times .77$), which represents about a 15x increase in the frequency of this event when mismatch repair is abolished. Three classes of revertants appear in the *mutL* spectrum that were not recovered in a wild-type background (Table 4). One of these, R18, a -2 deletion, occurred 9 times and represented about 8% of all *mutL218* revertants.

For the *trpA9813* revertant classes in the wild-type background other than R2, there is little evidence of enhancement by *mutL218*. For example, R3 revertants occur with a frequency of about 2.4×10^{-9} in the wild-type background and with a frequency of 3.2×10^{-9} in the *mutL218* strain, only a 1.3x enhancement. The R4 revertants occur with a frequency of 3.1×10^{-9} in the wild-type strain and with a frequency of 5.1×10^{-9} in the *mutL218* background, a 1.7x increase. The other revertant classes found in the wild-type spectrum show little or no *mutL*-induced increases, although for most classes the numbers of occurrences are too low to make credible comparisons.

The R1 class of *trpA9813* actually shows a decrease (5.8x) in frequency in the *mutL218* spectrum, although the single occurrence of this event in the mismatch-deficient background makes the exact level of decrease tentative. Several other reversion events that appear in the wild-type spectra of both *trpA* frameshift alleles are not found in the *mutL218* strains. If the decrease in *mutL218* strains for one or more frameshift events is real, it would imply that functional mismatch repair can paradoxically create some frameshift events. Such a possibility has been previously considered in yeast [51]. One explanation for the reduction of R1 revertants and possibly others in the *mutL* background is that mismatch repair is being inappropriately directed to the wrong DNA strand in wild-type cells.

For the *trpA540* allele there was only a small (2.1x) enhancement of reversion frequencies in the *mutL* background. Two reversion events are clearly enhanced. The R1 +1C insertions, the most frequent mutation in the wild-type distribution, is still the most common, showing a 2.8x increase with *mutL218*. The R44 revertant event, a +1 C insertion that restores the wild-type *trpA* sequence (Fig 1), occurs 17 times (13%) in the *mutL218* spectrum (Table 3), but is not found in the wild-type distribution. Other reversion events, also not found in the wild-type background, occur a small number of times with *mutL218*, but the small numbers preclude definitive conclusions.

Three sites, R2 for *trpA9813* and R1 and R44 for *trpA540*, show significant frequency enhancements in the *mutL218* background and result from +1 insertions within runs. Most of the previous studies that have specifically addressed the role of mismatch repair in frameshift mutagenesis have observed increased frameshift frequencies in runs or simple repeated sequences with mismatch repair-defective backgrounds. Both in *E. coli* with chromosomal, phage or plasmid targets and in *Saccharomyces cerevisiae* with chromosomal insertion targets, lack of mismatch repair leads to greatly elevated levels of frameshift mutagenesis [52–56]. Some homopolymer runs are more mutable than others and the most common frameshifts are generally –1 deletions [7,22,51,56], a class of frameshifts that can't be recovered in this study. For example in the *lacI* gene of *E. coli*, a –1 frameshift is enhanced at least 400x in a *mutL* strain [7].

The R1 and R44 *trpA540* revertant classes suggest a complex relationship involving the length of a nucleotide run, mismatch repair, and the final appearance of frameshifts in wild-type cells. R1 mutations are more frequent than R44 changes in both mismatch repair-proficient and mismatch-deficient backgrounds indicating the short run of two Cs (R1) is more prone to suffer premutational misalignments than the nearby run of three Cs (R44). Mismatch repair corrects the premutational perturbances at both sites, since both events are enhanced in the *mutL218* background, but does this more efficiently at the R44 site. Recently it has been shown that variation exists for the efficiency of mismatch repair at different sites within the yeast genome [57] and apparently the same situation occurs in *E. coli*.

4.4 Some conclusions about spontaneous frameshift mutagenesis

Two short DNA regions provided the targets for spontaneous frameshift mutagenesis in this study. For reversion of *trpA9813*, 16 reversion sites were found in wild-type cells while for *trpA540*, 43 sites were revealed. Additional sites were found in the mismatch-deficient strains. Since most sites have only one occurrence, it may be predicted that many more additional reversion sites exist that are not revealed by the limited number of revertants sequenced. In addition most frameshifts in these regions would not be expected to lead to Trp⁺ revertant phenotypes. It may be concluded that frameshift sites within these regions are dense and numerous.

The combined revertants of both frameshift alleles indicate that most of the sequenced mutations were +1 insertions within simple runs while most sites for reversion occurred outside

runs. For the wild-type strains there were a total of 59 reversion sites with 14 of these sites being in homopolymer runs. (This includes runs where the inserted nucleotide differs from that of the run.) A total of 152 mutations occurred at these 14 sites. In contrast, the remaining 45 sites that didn't involve +1 insertions within runs accumulated only a total of 183 mutations, indicating that, on average, simple repeats were the most frameshift mutable sites.

For the *mutL218* strains this trend is more severe. A total of 12 sites in runs produced 192 +1 insertions and only 55 mutations occurred in the 20 sites outside of runs. While there are large variations among individual sites, clearly frameshift mutagenesis is more likely to occur within simple runs than at most other sites. These data also suggest that among mutations that we can recover as revertants, mismatch repair is generally more effective in reducing +1 insertions within runs than frameshift events at sites elsewhere.

If the regions analyzed in this study are representative of the *E. coli* genome, spontaneous frameshift mutagenesis may be described as follows. Single nucleotide insertions (and presumably deletions) within simple runs are likely to be the most common type of frameshift mutation in wild-type cells and these events are often minimized by effective mismatch repair [this study; 7]. However the majority of frameshift sites are likely to be outside of runs and are so numerous as to suggest that almost any sequence can be a site susceptible to some type of misalignment error during DNA replication. A similar situation likely occurs in the *S. typhimurium* genome based upon an analysis of the *his* region [9].

As with base-pair substitutions, frameshifts probably occur at any nucleotide position although at frequencies that vary greatly from site to site. Most sites suffer spontaneous frameshifts at very low frequencies and many of these events seem to be unaffected by mismatch repair. Possibly the value in correcting such infrequent lesions is not worth the cost of possessing a mismatch repair system that would be sensitive enough to fix them.

Acknowledgements

We acknowledge the financial support of grant 1 R15 GM070563-01 from NIH.

References

1. Echols H, Goodman MF. Fidelity mechanisms in DNA replication. *Ann Rev Biochem* 1991;60:477–511. [PubMed: 1883202]
2. Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J Biol Chem* 1993;268:23762–23765. [PubMed: 8226906]
3. Maki H. Origins of spontaneous mutations: specificity and directionality of base substitutions, frameshifts and sequence-substitution mutagenesis. *Ann Rev Genet* 2002;36:279–303. [PubMed: 12429694]
4. Streisinger G, Okada Y, Emrich J, Newton A, Tsugita E, Terzhagi E, Inouye M. Frameshift mutations and the genetic code. *Cold Spring Harbor Symp Quant Biol* 1966;31:77–84. [PubMed: 5237214]
5. Streisinger G, Owen JE. Mechanisms of spontaneous and induced frameshift mutations in bacteriophage T4. *Genetics* 1985;109:633–659. [PubMed: 3988038]
6. Farabaugh PJ, Schmeissner U, Hoffer M, Miller JH. Genetic studies on the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J Mol Biol* 1978;126:847–863. [PubMed: 370408]
7. Schaaper RM, Dunn RL. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: The nature of *in vivo* DNA replication errors. *Proc Natl Acad Sci USA* 1987;84:6220–6224. [PubMed: 3306672]
8. Strand M, Prolla T, Liskay RM, Petes T. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 1993;365:274–276. [PubMed: 8371783]

9. De Marini DM, Sheldon ML, Abu-Shakra A, Szakmary A, Levine JG. Spectra of spontaneous frameshift mutations at the *hisD3052* allele of *Salmonella typhimurium* in four DNA repair backgrounds. *Genetics* 1998;149:17–36. [PubMed: 9584083]
10. Schaaper RM, Danforth BN, Glickman BW. Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J Mol Biol* 1986;189:273–284. [PubMed: 3018259]
11. Ripley LS. Frameshift mutation: determinants of specificity. *Ann Rev Genet* 1990;24:189–213. [PubMed: 2088167]
12. Kunkel TA. Misalignment-mediated DNA synthesis errors. *Biochemistry* 1990;29:8003–8011.
13. Kunkel TA, Soni A. Mutagenesis by transient misalignment. *J Biol Chem* 1988;263:14784–14789. [PubMed: 3049589]
14. Bloom LB, Chen X, Fygenon DK, Turner J, O'Donnell M, Goodman MF. Fidelity of *Escherichia coli* DNA polymerase III holoenzyme. The effect of beta, gamma complex processivity protein and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies. *J Biol Chem* 1997;272:27919–27930. [PubMed: 9346941]
15. Garcia-Diaz M, Kunkel TA. Mechanism of a genetic glissando: structural biology of indel mutations. *TBS* 2006;31:206–214.
16. Ripley LS. Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proc Natl Acad Sci USA* 1982;79:4128–4132. [PubMed: 7051004]
17. Ripley LS, Glickman BW. The unique self-complementarity of palindromic sequences provides DNA structural intermediates of mutation. *Cold Spring Harbor Symp Quant Biol* 1983;47:851–861. [PubMed: 6345080]
18. Bichara M, Shumacher S, Fuchs RPP. Genetic instability within monotonous runs of CpG sequences in *Escherichia coli*. *Genetics* 1995;140:897–907. [PubMed: 7672590]
19. Mo JM, Maki H, Sekiguchi M. Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. *J Mol Biol* 1991;222:925–936. [PubMed: 1762158]
20. Nagata Y, Mashimo K, Kawata M, Yamamoto K. The roles of Klenow processing and flap processing activities of DNA polymerase I in chromosome instability in *Escherichia coli* K12 strains. *Genetics* 2002;160:13–23. [PubMed: 11805041]
21. Tago Y, Imai M, Ihara M, Atofuji H, Nagata Y, Yamamoto K. *Escherichia coli* mutator *polA* is defective in base mismatch correction: the nature of *in vivo* DNA replication errors. *J Mol Biol* 2005;351:299–308. [PubMed: 16005896]
22. Schaaper RM, Dunn RL. Spontaneous mutations in the *Escherichia coli lacI* gene. *Genetics* 1991;129:317–326. [PubMed: 1660424]
23. Sargentini NJ, Smith KC. Ionizing and ultraviolet radiation-induced reversion of sequenced frameshift mutations in *Escherichia coli*: a new role for *umuDC* suggested by delayed photoreactivation. *Mutat Res* 1987;179:55–63. [PubMed: 3037365]
24. Choy HE, Fowler RG. The specificity of base-pair substitution induced by the *mutL* and *mutS* mutators in *E. coli*. *Mutat Res* 1985;142:93–97. [PubMed: 3883143]
25. Vogel HJ, Bonner DM. Acetylornithinase of *Escherichia coli*: partial purification and some properties. *J Biol Chem* 1956;218:97–106. [PubMed: 13278318]
26. Fix D. N-ethyl-N-nitrosurea-induced mutagenesis in *Escherichia coli*: multiple roles for the UmuC protein. *Mutat Res* 1993;294:127–138. [PubMed: 7687005]
27. Nichols BP, Yanofsky C. Nucleotide sequences of *trpA* of *Salmonella typhimurium* and *Escherichia coli*: An evolutionary comparison. *Proc Natl Acad Sci USA* 1979;76:5244–5248. [PubMed: 388433]
28. Berger H, Brammar WJ, Yanofsky C. Spontaneous and ICR191-A induced frameshift mutations in the A gene of *Escherichia coli* tryptophan synthetase. *J Bacteriol* 1968;96:1672–1679. [PubMed: 4882018]
29. Brammar WJ. Independent expression of the A gene of the tryptophan operon of *Escherichia coli* during tryptophan starvation. *J Gen Microbiol* 1973;76:395–405. [PubMed: 4579131]
30. Brammar WJ, Berger H, Yanofsky C. Altered amino acid sequences produced by reversion of frameshift mutants of tryptophan synthetase A gene of *Escherichia coli*. *Proc Natl Acad USA* 1967;58:1499–1506.

31. Siegel EC, Vaccaro KK. The reversion of *trpA* frameshift mutations in *mut*, *polA*, *lig*, and *dnaE* mutant strains of *Escherichia coli*. *Mutat Res* 1978;50:9–17.
32. Fijalkowska IJ, Schaaper RM. Effects of *Escherichia coli dnaE* antimutator alleles in a proofreading-deficient *mutD5* strain. *J Bacteriol* 1995;177:5979–5986. [PubMed: 7592352]
33. Helinski DR, Yanofsky C. A genetic and biochemical analysis of second-site reversion. *J Biol Chem* 1963;238:1043–1048. [PubMed: 13953567]
34. Allen MK, Yanofsky C. A biochemical and genetic study of reversion with the A- gene A- protein system of *Escherichia coli* tryptophan synthetase. *Genetics* 1963;72:1065–1083. [PubMed: 14050802]
35. Yanofsky C, Horn V, Thorpe D. Protein structure relationships revealed by mutational analysis. *Science* 1964;146:1593–1594. [PubMed: 14224506]
36. Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. *J Biol Chem* 1988;263:17857–17871. [PubMed: 3053720]
37. Miles EW. The tryptophan synthase $\alpha_2\beta_2$ complex. *J Biol Chem* 1991;266:10715–10718. [PubMed: 1904055]
38. Kulik V, Weyand M, Seidel R, Niks D, Arac D, Dunn MF, Schlichting I. On the role of α Thr183 in the allosteric regulation and catalytic mechanism of tryptophan synthase. *J Mol Biol* 2002;324:677–690. [PubMed: 12460570]
39. Nagata S, Hyde CC, Miles EW. The α subunit of tryptophan synthase. Evidence that aspartic acid 60 is a catalytic residue and that the double alteration of residues 175 and 211 in a second-site revertant restores the proper geometry of the substrate binding site. *J Biol Chem* 1989;264:6288–6296. [PubMed: 2649498]
40. Kawasaki H, Bauerle R, Zon G, Ahmed SA, Miles EW. Site-specific mutagenesis of the α subunit of tryptophan synthase from *Salmonella typhimurium*. Changing arginine 179 to leucine alters the reciprocal transmission of substrate-induced conformational changes between the α and β_2 subunits. *J Biol Chem* 1987;262:10678–10683. [PubMed: 3112150]
41. Murgola EJ, Yanofsky C. Structural interactions between amino acid residues at positions 22 and 211 in the tryptophan synthetase α chain of *Escherichia coli*. *J Mol Biol* 1974;86:775–784. [PubMed: 4610146]
42. Newton A, Masys D, Leonardi E, Wygal D. Association of induced frameshift mutagenesis and DNA replication in *Escherichia coli*. *Nat New Biol* 1972;236:19–22.
43. McHenry CS. Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences. *Mol Microbiol* 2003;49:1157–1165. [PubMed: 12940977]
44. Pham PT, Olson NW, McHenry CS, Schaaper RM. Mismatch extension by *Escherichia coli* DNA polymerase III holoenzyme. *J Biol Chem* 1999;274:3705–3710. [PubMed: 9920922]
45. Joyce CM, Sun XC, Grindley NDF. Reactions at the polymerase active site that contribute to the fidelity of *Escherichia coli* DNA polymerase I (Klenow fragment). *J Biol Chem* 1992;267:24485–24500. [PubMed: 1447195]
46. Fijalkowska IJ, Jonczyk P, Maliszewska-Tkaczyk M, Bialeskorska M, Schaaper RM. Unequal fidelity of leading and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* 1998;95:10020–10025. [PubMed: 9707593]
47. Goodman MF. Error-prone repair polymerases in prokaryotes and eukaryotes. *Ann Rev Biochem* 2002;70:17–50. [PubMed: 12045089]
48. Kuban W, Banach-Orlowska M, Bialoskorska M, Lipowska A, Schaaper RM, Jonczyk P, Fijalkowska IJ. Mutator phenotype resulting from Pol IV overproduction in *Escherichia coli*: preferential mutagenesis on the lagging strand. *J Bacteriol* 2005;187:6862–6866. [PubMed: 16166552]
49. Banach-Orlowska M, Fijalkowska IJ, Schaaper RM, Jonczyk P. DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in *Escherichia coli*. *Mol Microbiol* 2005;58:61–70. [PubMed: 16164549]
50. Bhamre S, Gadea BB, Koyama CA, White SJ, Fowler RG. An anaerobic *recA*-, *umuC*- dependent pathway of spontaneous base-pair substitution mutagenesis in *Escherichia coli*. *Mutat Res* 2001;473:229–247. [PubMed: 11166040]

51. Greene CN, Jinks-Robertson S. Spontaneous frameshift mutations in *Saccharomyces cerevisiae*: Accumulation during DNA replication and removal by proofreading and mismatch repair activities. *Genetics* 2001;159:65–75. [PubMed: 11560887]
52. Levinson G, Gutman GA. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* 1987;15:5323–5338. [PubMed: 3299269]
53. Strauss BS, Sagher D, Acharya S. Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. *Nucleic Acids Res* 1997;25:806–813. [PubMed: 9064658]
54. Tran HT, Keen JD, Krickler M, Resnick MA, Gordenin DA. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol Cell Biol* 1997;17:2859–2865. [PubMed: 9111358]
55. Harfe BD, Jinks-Robertson S. Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae*. *Genetics* 2000;156:571–578. [PubMed: 11014807]
56. Gragg H, Harfe BD, Jinks-Robertson S. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2000;22:8756–8762. [PubMed: 12446792]
57. Hawk J, Stefanovic L, Boyer RC, Petes TD, Farber RA. Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. *Proc Natl Acad Sci USA* 2005;102:8639–8643. [PubMed: 15932942]

Fig. 1.

The nucleotide and amino acid changes associated with the *trpA9813*, *trpA540*, and *trpA21* frameshift mutations and reversion frequencies of these alleles are shown below. The pertinent portions of the *trpA* gene and its product the alpha subunit of tryptophan synthase are given. The numbers for the gene and amino acid sequence correspond to those used by Nichols and Yanofsky [27]. The nucleotide deleted in each frameshift is boxed. For the *trpA540* and *trpA21* alleles, the single C deleted can be any in a run of four.

Fig. 2.

Proposed mechanisms (listed in Table 4) for the occurrence of representative frameshift revertants: (a) Simple slippage of the nascent strand in mononucleotide runs; (b) Misincorporation of a primer nucleotide followed by misalignment of the nascent strand; (c) Misincorporation of the primer nucleotide followed by misalignment of the template strand; (d) Misalignment of the nascent strand with no misincorporation of a primer nucleotide; (e) Misalignment of the template strand with no misincorporation of a primer nucleotide; and (f) Misincorporation of a primer nucleotide followed by correct incorporation of the next nucleotide followed by misalignment of the nascent strand. For most frameshift events, except for 2f, either DNA strand could be the primer strand. In these cases the choice of primer and

template strands for the models was based upon the following: (1) A transversion mispairing is assumed to cause the DNA polymerase complex to stall longer at the mismatch than a transition mismatch, providing a greater opportunity for misalignment to occur [32] and (2) AT-richer sequences in newly synthesized DNA provide greater instability leading to misalignment. (See text for further detail on these models.)

Fig. 3.

Possible mechanisms (called 'complex' in Table 4) to account for frameshift revertants not explained by the models of Fig. 2. (a) Misincorporation of the primer nucleotide followed by a second misincorporation of the next primer nucleotide and finally misalignment of the primer strand (shown here for *trpA9813* revertant R18). (b) Creation of a hairpin loop followed by either by precise excision of the single-stranded region (*trpA9813* R8) or imprecise excision (*trpA9813* R13). As shown here the formation and excision of the hairpin loop is occurring during replication but could occur anytime this region is single-stranded.

trpA9813

Position	172	173	174	175	176	177	178	179
Wild-type	G	Y	T	Y	L	L	S	R
<i>trpA9813</i>	G	Y	T	F	C	C	H	E
Revertant Changes	-	L	H/N/Y/-	L/+/-	L/M/-	L/M	S/Q/-	R/-

trpA540

Position	201	202	203	204	205	206	207	208	209	210	211	212
Wild-type	K	E	Y	N	A	A	P	P	L	Q	G	F
<i>trpA540</i>	K	E	Y	N	A	A	P	P	L	Q	G	F
Revertant Changes	N	-/+	V	Q/-	G/R/S/+	C/-	T/A	S/H/+	I/F	A/+	G	I

Position	213	214	215	216	217	218	219	220	221	222	223
Wild-type	G	I	S	A	P	D	Q	V	K	A	A
<i>trpA540</i>	G	I	S	A	R	I	R	STOP			
Revertant Changes	W	Y/N/-	F/-	R/C/S/+ /++/-	P/L/A/-	D/Y/N/H/R/-	Q/++/+/-	V/Y/++/-	K/-	A/F/-	A/-

Fig. 4.

The composite amino acid replacements for *trpA9813* and *trpA540* revertants are shown below. All the amino acids found at each position that differ from the mutant peptide sequence are included as well as deletions and insertions. A “+” indicates the insertion of a single amino acid at a position, “++” the insertion of two amino acids, and “-” a deletion. The 175, 177, 179, 211, and 213 sites were shown to be critical for enzymatic activity in previous studies (see text). Revertants of *trpA9813* had wild-type sequences before position 172 and after 179; *trpA540* revertants had wild-type sequences before position 201 and after 224.

Table 1*E. coli* strains used in this study

Strain	Relevant genotype	Frameshift in <i>trpA</i> gene	Source or reference
SR1282	<i>trpA21</i> [A] ^a	⁶⁴⁷ CCCC → ⁶⁴⁷ CCC	[23]
SR1283	<i>trpA540</i> [A]	⁶⁴⁷ CCCC → ⁶⁴⁷ CCC	[23]
SR1284	<i>trpA9813</i> [A]	⁵²³ TAT → ⁵²³ TT	[23]
KD1093	<i>mutL218::Tn10</i>		[24]
KD1601	<i>trpA540, mutL218::Tn10</i>		SR1283 x P1 <i>virA</i> KD1093
KD1602	<i>trpA9813, mutL218::Tn10</i>		SR1284 x P1 <i>virA</i> KD1093

^a[A] = *argE, hisG4, leuB6, thr-1, thi-1, ara-14, galK2, lacY1, mtl-1, xyl-5, tsx-33, rfbD1, mgl-51, rpsL31, Δ(gpt-proA), supE44*.

Table 2DNA sequence changes associated with reversion of the *trpA9813* allele in wild-type and *mutL218* strains.

Revertant	Sequence change	Mutational Change	Number of Occurrences	
			wild type	<i>mutL218</i>
R1	⁵¹⁸ ACA → ⁵¹⁸ ACCA	+C	31	1
R2	⁵²² CTTTG → ⁵²² CTTTTG	+T	33	89
R3	⁵¹⁷ TACACC → ⁵¹⁷ TACC	-AC	20	5
R4	⁵¹⁸ ACA → ⁵¹⁸ ACAA	+A	26	8
R5	⁵²⁴ TTT → ⁵²⁴ TTAT	+A	6	0
R6	⁵³³ ACGA → ⁵³³ AA	-CG	4	0
R7	⁵³² CACG → ⁵³² CG	-AC	3	0
R8	⁵²⁰ ACCTTTTGCT → ⁵²⁰ AT	-CCTTTTGC	1	0
R9	⁵²⁷ GCT → ⁵²⁷ GCCT	+C	1	0
R10	⁵¹⁵ GTTA → ⁵¹⁵ GTTTA	+T	1	0
R11	⁵²³ TTTT → ⁵²³ TTCTT	+C	1	0
R12	⁵²⁸ CTG → ⁵²⁸ CTTG	+T	1	0
R13	⁵¹⁸ ACACCTTTTGC → ⁵¹⁸ ACC	-ACCTTTTG	1	1
R14	⁵²⁰ ACCT → ⁵²⁰ AT	-CC	1	0
R15	⁵¹⁹ CACCTTT → ⁵¹⁹ CACCTACCTTT	+ACCT	1	0
R16	⁵²⁶ TGCT → ⁵²⁶ TT	-GC	2	0
R17	⁵¹³ TGGT → ⁵¹³ TT	-GG	0	9
R18	⁵¹⁸ ACA → ⁵¹⁸ ACTA	+T	0	1
R19	⁵²⁰ ACCTTTTG → ⁵²⁰ ATG	-CCTTT	0	1
			133	115

Table 3DNA sequence changes associated with reversion of the *trpA540* allele in wild-type and *mutL218* strains.

Revertant	Sequence change	Mutational Change	Number of Occurrences	
			wild type	<i>mutL218</i>
R1	⁶⁴³ TCCG → ⁶⁴³ TCCCG	+C	76	67
R2	⁶²¹ TCCA → ⁶²¹ TCCCA	+C	11	2
R3	⁶¹⁹ CCTCCA → ⁶¹⁹ CCTCCCTCCA	+CTCC	11	0
R4	⁶¹⁷ CACCTC → ⁶¹⁷ CACCTACCTC	+ACCT	7	0
R5	⁶²⁰ CTC → ⁶²⁰ CTTC	+T	18	9
R6	⁶⁴⁵ CGCC → ⁶⁴⁵ CC	-GC	8	0
R7	⁶⁵¹ GATCAGGTA	+ATCAGGT	4	0
	→ ⁶⁵¹ GATCAGGTATCAGGTA			
R8	⁶⁴⁰ ATTTC → ⁶⁴⁰ ATTTTC	+T	9	6
R9	⁶⁵⁸ TAAAA → ⁶⁵⁸ TATAAA	+T	5	3
R10	⁶⁴⁶ GCCCGGA → ⁶⁴⁶ GA	-CCCGG	10	2
R11	⁶⁴³ TCCGCC → ⁶⁴³ TCCGCCCGCCC	+CGCC	4	0
R12	⁶⁵² ATC → ⁶⁵² ATTC	+T	1	0
R13	⁶⁴⁹ CG → ⁶⁴⁹ CGGATCCCG	+GGATCCC	2	0
R14	⁶³⁹ TATT → ⁶³⁹ TT	-AT	1	0
R15	⁶⁵⁷ GTAA → ⁶⁵⁷ GA	-TA	1	0
R16	⁶¹⁸ ACCT → ⁶¹⁸ ACCCT	+C	1	1
R17	⁶²² CCA → ⁶²² CACA	+A	1	0
R18	⁶⁵¹ GAT → ⁶⁵¹ GTAT	+T	1	0
R19	⁶²⁴ ATTG → ⁶²⁴ ATTTG	+T	1	2
R20	⁶⁵⁵ AGGT → ⁶⁵⁵ AGGGT	+G	5	1
R21	⁶²⁷ GCAGGG → ⁶²⁷ GCAGGCAGGG	+CAGG	1	0
R22	⁶⁵¹ GAT → ⁶⁵¹ GAAT	+A	2	0
R23	⁶⁴⁹ CGGA → ⁶⁴⁹ CA	-GG	1	1
R24	⁶⁴⁶ GCCCC → ⁶⁴⁶ GCG	-CC	1	0
R25	⁶⁵⁷ GTAAAAGCAGCGA → ⁶⁵⁷ GA	-TAAAAGCAGCG	1	0
R26	⁶⁵⁵ AGGTAAAAGC → ⁶⁵⁵ AC	-GGTAAAAG	1	0
R27	⁶⁴³ TCCG → ⁶⁴³ TG	-CC	1	4
R28	⁶¹² CGC → ⁶¹² CGGC	+G	1	0
R29	⁶³⁶ TGGT → ⁶³⁶ TGGGT	+G	1	0
R30	⁶³⁹ TAT → ⁶³⁹ TAAT	+A	1	0
R31	⁶⁵⁹ AAA → ⁶⁵⁹ ACAA	+C	1	0
R32	⁶⁴⁶ GCCCG → ⁶⁴⁶ GCCCTG	+T	1	0
R33	⁶⁴⁰ ATTTCG → ⁶⁴⁰ AG	-TTTCC	1	0
R34	⁶⁰⁴ GAGT → ⁶⁰⁴ GT	-AG	1	0
R35	⁶⁴⁰ ATTTCG → ⁶⁴⁰ ATTTCCTTCG	+TTCC	1	0
R36	⁶⁵³ TCAGGT → ⁶⁵³ TCAGGCAGGT	+CAGG	1	0
R37	⁶²⁶ TGCAGG → ⁶²⁶ TGCAGGCAGG	+GCAG	1	0
R38	⁶⁵¹ GATCAG → ⁶⁵¹ GATCAATCAG	+ATCA	2	0
R39	⁶⁴⁶ GCCCGGATCAGGT → ⁶⁴⁶ GT	-CCCGGATCAGG	1	0
R40	⁶⁴³ TCCG → ⁶⁴³ TG	-CC	1	0
R41	⁶¹⁷ CACC → ⁶¹⁷ CC	-AC	1	0
R42	⁶⁰⁹ CAACGC → ⁶⁰⁹ CAACGAACGC	+AACG	1	0
R43	⁶⁵⁰ GGATCAGGT	+GATCAGG	1	0
	→ ⁶⁵⁰ GGATCAGGGATCAGGT			
R44	⁶⁴⁶ GCCCG → ⁶⁴⁶ GCCCCG	+C	0	17
R45	⁶⁵⁰ GGAT → ⁶⁵⁰ GT	-GA	0	2
R46	⁶¹⁸ ACCTCC → ⁶¹⁸ ACCTCCCTCC	+CCTC	0	2
R47	⁶⁵¹ GATCAGGTA	+ATCAGGT	0	3
	→ ⁶⁵¹ GATCAGGTATCAGGTA			
R48	⁵⁹⁸ CTGAAA → ⁵⁹⁸ CTGAATGAAA	+TGAA	0	1
R49	⁶⁵⁸ TAAAAG → ⁶⁵⁸ TAGAAAAG	+G	0	1
R50	⁶³⁶ TGGT → ⁶³⁶ TGGGT	+G	0	2
R51	⁶⁵¹ GATC → ⁶⁵¹ GC	-AT	0	1
R52	⁶³³ ATTTG → ⁶³³ ATTTTG	+T	0	1
R53	⁶⁴⁶ GCCCGGA → ⁶⁴⁶ GCCCGGCCGGA	+GGCC	0	1
R54	⁶¹⁴ CTG → ⁶¹⁴ CTTG	+T	0	1
R55	⁶¹⁰ AACG → ⁶¹⁰ AG	-AC	0	1
R56	⁶⁴⁰ ATTTCGCCCCG → ⁶⁴⁰ ATTG	-TCCGCCCCG	0	1
R57	⁶²² CCATTG → ⁶²² CCATTTCATTG	+CATT	0	1
			202	133

Table 4

Molecular models to account for the occurrence during DNA replication of the frameshift events shown in Tables 2 and 3. These models are shown in detail in Figs. 2 and 3.

Frameshift mechanism	Observed Outcome	Specific Revertants	
		<i>trpA9813</i>	<i>trpA540</i>
a. Simple slippage of nascent strand in mononucleotide runs.	+1 insertions	R2, R10 [*]	R1, R2, R8, R16, R19, R20, R29, R44, R50, R52
b. Misincorporation of primer nucleotide followed by misalignment of nascent strand.	+1, +4, +7 insertions	R1, R4, R9, R12, R15	R4 [*] , R5, R7, R12, R22, R28, R30, R35, R36, R38, R42, R48, R53, R54, R57
c. Misincorporation of primer nucleotide followed by misalignment of template strand.	-2, -5, -8 deletions	R6 [*] , R7, R14, R16, R17, R19	R15, R23, R26, R27, R33, R40, R45, R51, R55, R56
d. Misalignment of nascent strand with no misincorporation.	+4, +7 insertions	None	R3 [*] , R11, R13, R21, R37, R43, R46, R47
e. Misalignment of template strand with no misincorporation.	-2, -5, -11 deletions	R3 [*]	R6, R10, R14, R24, R25, R34, R39, R41
f. Misincorporation of primer nucleotide followed by correct incorporation of the next nucleotide followed by misalignment of nascent strand.	+1 insertions	R5, R11	R9 [*] , R18, R31, R32, R49
Complex—see text.	+1 insertions -8 deletions	R8, R13, R18	R17

^{*} This particular revertant is used as an example of the model in Fig. 2.