

## MicroReview

# Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences

Susan T. Lovett

*Department of Biology and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.*

### Summary

**Mutations and rearrangements that occur by misalignment during DNA replication are frequent sources of genetic variation in bacteria. Dislocations between a replicating strand and its template at repetitive DNA sequences underlie the mechanism of these genetic events. Such misalignments can be transient or stable and can involve intramolecular or intermolecular DNA mispairing, even pairing across a replication fork. Paradoxically, these replication ‘slippage’ events both create and destroy repetitive sequences in bacterial genomes. This review catalogues several types of slippage errors, presents the cellular processes that act to limit them and discusses the consequences of this class of genetic events on the evolution of bacterial genomes and physiology.**

### Introduction

Mutations and genetic rearrangements are frequently associated with repetitive DNA sequences. These genetic errors share a common mechanism involving the mispairing of newly synthesized DNA strands with alternative complementary sequences in the local vicinity. Misalignment between repetitive sequences can occur between repeats that range from only a few nucleotides to hundreds of nucleotides in length. A characteristic feature of these misalignment errors is a requirement for close proximity of the repeats, most likely because they must reside together in a replication fork. Misalignments can occur in a number of configurations: between the nascent strand and a second site on its normal template, or to a second site on the nascent strand itself. Misalignments joining two

sister chromosomes across a replication fork have also been deduced to occur.

Both direct and inverted repeats act as sites for genetic rearrangements mediated by misalignment. Direct repeats can elicit rearrangements that cause deletions or duplications of genetic material, including the repeat and any intervening sequence. Slippage at inverted repeats can cause inversions between the repeats and any included sequence. One genetic hallmark of misalignment-mediated rearrangements is their independence of homologous recombination factors, including the RecA strand transfer protein of bacteria. RecA-independent slippage rearrangements constitute a subset of the events formerly described as ‘illegitimate’ recombination (Franklin, 1971). (Originally defined as chromosomal rearrangements that occur between regions sharing short or no homology, illegitimate recombination subsumes a number of distinct mechanisms, including slippage at short homologies as well as cut-and-join reactions at non-homologous sequences.)

Even transient misalignments that do not give rise to rearrangements can be sources of genetic mutation. If two repeats are not perfectly homologous, a genetic mutation can be templated by limited replication in the slipped and misaligned configuration, even if such misalignment is transient. Certain mutation hot-spots (discussed below) can be inferred to occur by this mechanism and can involve either direct or inverted repeat sequences.

Several different categories of mutations and genetic rearrangements, summarized below, can now be classified as ‘slippage’ events. Several of these have been studied systematically in some detail, and common mechanistic details are beginning to emerge. These misalignment mutational mechanisms have an important impact on the structure of the bacterial genome and, in some instances, may play an important role in the bacterial lifestyle.

### Frameshifts in nucleotide runs

George Streisinger was the first to note that frameshift mutations were not random but tended to occur in

stretches of repetitive DNA sequence (nucleotide 'runs'). To explain such events, he proposed the replication slippage model that stands as a paradigm for the larger family of mutational events discussed in this review. Streisinger's model proposed that frameshift mutations were produced by the misalignment of the nascent strand at the repetitive sequence on its template during DNA synthesis (Fig. 1). The misalignment can produce either addition or deletion of nucleotide bases. Slippage at simple sequence repeats can be observed readily *in vitro* with several different polymerases (Canceill *et al.*, 1999).

The result of frameshift strand slippage is a looped structure with unpaired bases on the template or on the nascent strand (Fig. 1). Such a structure, with up to three unpaired bases in the loop, is well recognized by the MutS protein of the post-replication mismatch repair pathway and targeted for excision (Parker and Marinus, 1992). The MutHLS mismatch repair system therefore acts as a major surveillance mechanism for frameshift mispairs, and frameshift mutations are greatly elevated when the pathway is inactivated (Schaaper and Dunn, 1987; Cupples *et al.*, 1990). However, frameshifts involving repeats of four or more nucleotides escape detection, presumably because these larger loops are poorly bound by MutS (Parker and Marinus, 1992).

Frameshift mutations in nucleotide runs are highly prevalent in mutational spectra and often constitute mutational hot-spots. For example, a mutational hot-spot in *lacI*, where frameshifts occur at a tetranucleotide sequence repeated three times, accounts for over 65% of mutations

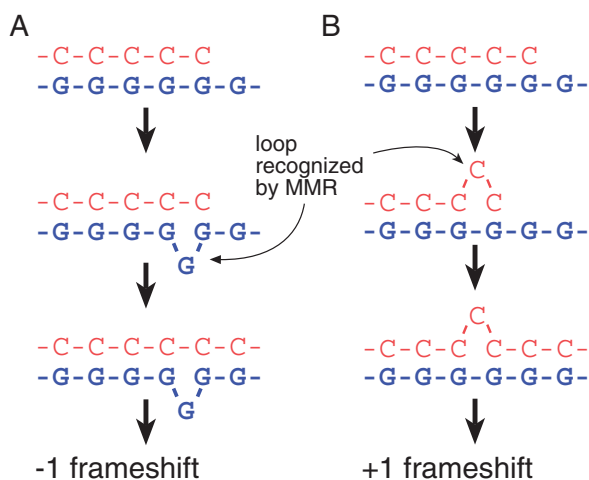
that inactivate the gene (Farabaugh *et al.*, 1978). Another frameshift hot-spot is the 8 nucleotide G run in the *xyiB* gene of *Escherichia coli*, at which frameshifts occur frequently in mismatch repair-deficient strains (Funchain *et al.*, 2000).

### Engineering frameshifts for profit – contingency loci

A number of pathogenic bacteria use frameshifting in nucleotide runs to engineer activation or inactivation of certain genes within the bacterial population (see reviews by van Belkum *et al.*, 1998; 1999). These 'contingency loci' (Bayliss *et al.*, 2001) often encode surface antigens that interact with the host immune response or target tissues. Inactivation of these loci in a subset of the population may be advantageous to thwart the immune system or to enable the bacterium to colonize new locations. Frameshifting in nucleotide runs is ideally suited for this role in that it is both frequent and reversible (for example, a  $-1$  inactivating frameshift can later be reverted by a subsequent  $+1$  frameshift). These properties ensure a dynamic population expressing variable phenotypes.

Two examples of well-studied contingency loci are found in *Neisseria gonorrhoeae*. A group of *opa* ('opacity') genes that control tissue attachment contains pentanucleotide repeats within their coding region. Frameshifts between these repetitive sequences cause variable expression of these loci (Makino *et al.*, 1991). Genetic variation of the *pilC* pilin gene, a target for the immune system, are likewise varied by frameshifts within a monotonic G run (Jonsson *et al.*, 1991). The occurrence of repetitive arrays within a gene, and hence probable frameshift regulation, can be predictive of a particular gene's role in pathogenesis. After the genomic sequence of *Haemophilus influenzae* was completed, nine loci with tetranucleotide repeats, repeated from six to 36 times, were identified that were likely to be regulated by frameshifting (Hood *et al.*, 1996). These loci showed similarity to known factors that play a role in pathogenesis in other organisms. Other pathogenic bacteria show a similar prevalence of simple sequence repeats (Hood *et al.*, 1996; Saunders *et al.*, 2000), and contingency loci regulated by frameshifting may be widespread among pathogenic bacteria.

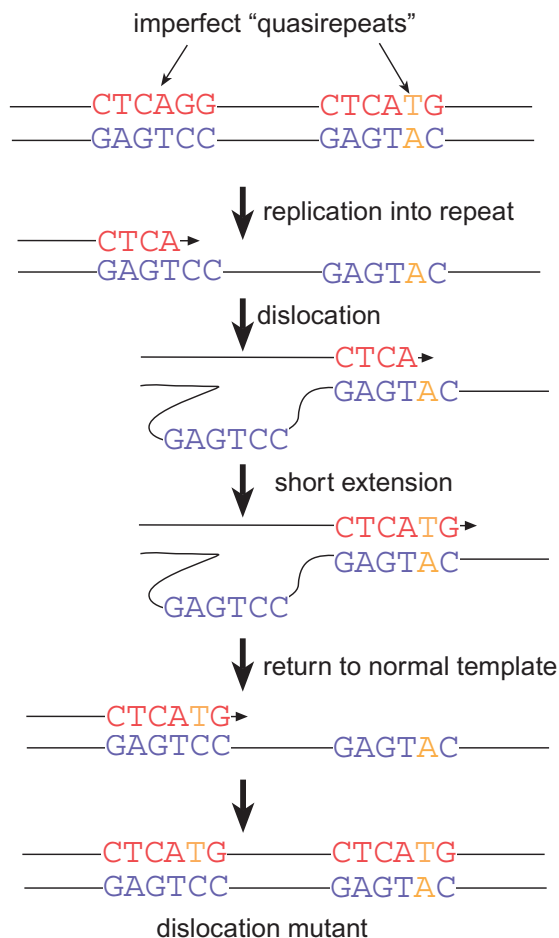
It is not known whether contingency loci have special features that promote efficient frameshifting. However, the frameshifts involved in creating variable expression at these loci tend to involve larger repeats, four nucleotides or more. The likely explanation for this size skew is that the larger frameshifts escape detection by the mismatch repair system (Parker and Marinus, 1992) and therefore occur at higher rates in mismatch repair-proficient bacteria.



**Fig. 1.** Replication slippage model for frameshift mutations. At a simple repetitive sequence, such as shown here, the nascent strand (in red) can misalign with its template (in blue), causing a loop to be formed, either on the template (A) or on the nascent strand (B). This loop can be recognized efficiently by the mismatch repair pathway and the nascent strand excised. If unrepaired, the slippaged misalignment will yield a  $-1$  frameshift (A) or a  $+1$  frameshift (B).

### Transient misalignment between short direct repeats: 'dislocation' mutagenesis

A number of genetic mutations can be deduced to occur by transient misalignment at slightly imperfect repeated sequences nearby. A limited amount of DNA synthesis in the misaligned configuration allows a mutational change to be transferred by subsequent realignment to the initial site (Fig. 2). This mechanism serves to extend and to perfect the homology between imperfect direct repeat sequences. This class of mutations, called 'dislocation' or 'quasi-repeat' mutations, were proposed to explain certain mutational hot-spots (Fowler *et al.*, 1974; Nelson *et al.*, 1981) and later observed in mutational spectra derived from *in vitro* polymerization by enzymes from eukaryotic sources (Kunkel and Soni, 1988; Kunkel, 1990). Dislocation mutations have been deduced from features of muta-



**Fig. 2.** Mechanism for dislocation mutagenesis. Mutations are generated in imperfect direct repeats or 'quasirepeats'. During the synthesis of one repeat, the nascent strand misaligns transiently at the second repeat. A short amount of DNA synthesis in the misaligned configuration causes a mutation to be templated by the second repeat. Realignment restores normal replication, but a mutation is left behind. Note that dislocation mutagenesis perfects nearby imperfect repeats.

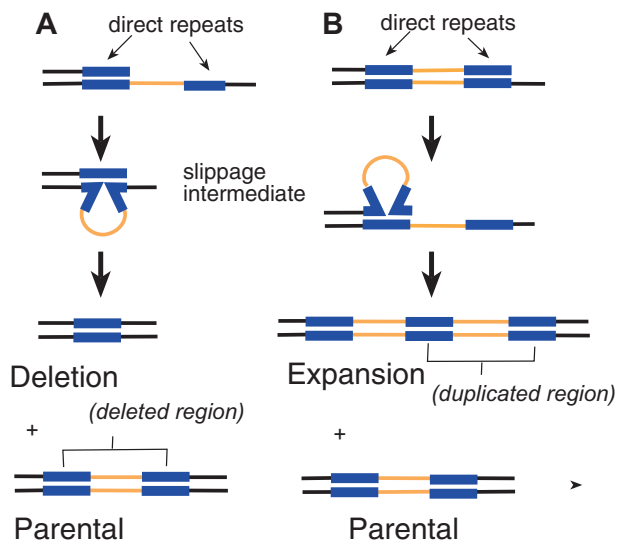
tional spectra from *E. coli in vivo* (Fowler *et al.*, 1986; Schaaper *et al.*, 1986; Schaaper, 1988).

Dislocation mutations can be difficult to recognize, as they often involve only a single nucleotide change and occur between very short and imperfect repeats. However, certain complex mutations involving two or more bases or frameshifts not within nucleotide runs (Ripley, 1990) have been inferred to occur by this mechanism. This class of mutation has not been studied systematically, and little is known about what cellular factors or sequence contexts promote or inhibit such events. Dislocation mutations seem to be more prominent in spectra derived from strains lacking polymerase-proofreading capacity (Fowler *et al.*, 1986; Schaaper, 1988). Presumably, proofreading deficiency helps to promote polymerase extension in the misaligned configuration. In one case, dislocation mutations were found to be adjacent to T runs (Kunkel and Soni, 1988), where relatively poorer basepairing of the nascent strand may aid either of the two template switch steps in the mechanism. In bacteriophage T4, dislocation mutagenesis is strongly enhanced by DNA polymerase mutants with impaired processivity (Ripley *et al.*, 1983).

### Long-range slippage: deletion and duplications

By a mechanism similar to frameshift mutagenesis but over longer distances, misalignment will produce rearrangements between dispersed directly repeated DNA sequences. These rearrangements can consist of either deletions or duplications of the repeat and any intervening DNA sequence (Fig. 3). Among the mutational hot-spots in *E. coli lacI* are deletions formed between dispersed repeats of eight or more nucleotides (Farabaugh *et al.*, 1978; Albertini *et al.*, 1982). Duplications or deletions between large repeats of several hundred bases occur independently of the RecA homologous recombination pathway at very high rates of  $10^{-5}$ – $10^{-4}$  per cell generation (Dianov *et al.*, 1991; Lovett *et al.*, 1993; 1994; Bi and Liu, 1994). In *E. coli*, there seems to be little or no bias to deletion- versus duplication-producing misalignments. In a system where identical repeats could be used to select either deletions or duplications, duplication and deletion occurred at roughly the same frequency and were stimulated by similar mutations affecting DNA polymerase and associated factors (Morag *et al.*, 1999).

The parameters that govern the efficiency of tandem repeat deletion have been studied systematically in *E. coli* (reviewed by Bzymek and Lovett, 2001a). The length, perfection and proximity of the repeats are important determinants of their propensity to rearrange. Although rearrangements can be detected involving only very short homologies (Albertini *et al.*, 1982; Mazin *et al.*, 1991; Bi and Liu, 1994), misalignment is nonetheless homology dependent (Dianov *et al.*, 1991; Bi and Liu, 1994). This is



**Fig. 3.** Misalignment deletion or duplication.

A. Deletion slippage intermediates are formed by the forward misalignment of the nascent strand at the direct repeats (shown in blue). After replication of this structure, a deletion of one repeat and the intervening DNA (in orange) results.

B. Backward slippage of the nascent strand at the direct repeats (shown in blue) will produce an expansion of the number of repeats, including a duplication of the intervening region (shown in orange).

because larger homologies provide slip-paired intermediates with more extensive heteroduplex and, hence, greater stability. Tandem repeats of over a hundred nucleotides in length delete at very high rates, more reminiscent of recombination ( $\approx 10^{-4}$ ) than of mutational ( $\approx 10^{-8}$ ) rates. In addition to the length, the perfection of the homology between repeats governs deletion between repeated sequences. Small sequence differences between repeats reduce deletion efficiency by several orders of magnitude (Lovett and Feschenko, 1996). This reduction is caused by at least two factors: (i) mismatch repair recognition and destruction of the heteroduplex slippage intermediate; and (ii) exonuclease sensitivity of a more unstable slippage intermediate (Feschenko *et al.*, 2003). Unlike the case of frameshift mutations, the mismatch repair system does not recognize the loop of the long-range slippage intermediate (which is too large) but, rather, mismatches within the heteroduplex formed by the misalignment. Finally, misalignment between repeats is highly proximity dependent. There is an exponential dependence of deletion rate on proximity of the repeats (Bi and Liu, 1994; Chedin *et al.*, 1994; Lovett *et al.*, 1994), presumably because the two repeats must interact within a single replication fork. Slippage is therefore limited to repeats that are less than 10 or so kilobases apart.

Although several other mechanisms can contribute, in theory, to tandem repeat deletion, there is good evidence that most repeat deletion in *E. coli* occurs by misalignment during replication (see arguments in Lovett and Fes-

chenko, 1996; Bzymek and Lovett, 2001a). Homologous recombination can be ruled out because repeat rearrangements occur efficiently in the absence of RecA and other homologous recombination factors in *E. coli*. Rearrangements between large repeats do occur, in part, by RecA-dependent recombination; however, recombination does not contribute to rearrangements between repeats shorter than 200 bp in length (Bi and Liu, 1994) and therefore does not mediate rearrangements between short, spurious repeats in bacterial genes. The impact of mismatch repair on deletion between engineered imperfect direct repeats argues that the majority of deletion events occur in hemimethylated, that is newly replicated, DNA (Lovett and Feschenko, 1996). A possible alternative mechanism for RecA-independent deletion, involving breakage and annealing at short repeats, is not consistent with these mismatch repair effects (see Bzymek and Lovett, 2001a). Furthermore, defects in many components of the replication complex in both *E. coli* and *Bacillus subtilis* strongly elevate the rates of tandem repeat rearrangements (Saveson and Lovett, 1997; Bruand *et al.*, 2001), suggesting that they are the outcomes of aberrant or stalled replication.

### Longer range slippage and bacterial genomes

Long-range slippage has an important impact on the bacterial genome. By mediating deletion at spurious direct repeats, slippages trim and streamline the bacterial genome by the removal of DNA that is unnecessary for survival. Slippage can also create genetic duplications that can be amplified further to larger copy number under selection for elevated gene expression (see below.)

Misalignment-mediated rearrangements between large or dispersed repeats also change the structure or the expression of certain proteins in pathogenic bacteria, thereby creating genetic diversity (van Belkum *et al.*, 1998; 1999; Puopolo *et al.*, 2001), in a manner similar to 'contingency' loci discussed earlier, where short frameshifts are involved (Bayliss *et al.*, 2001). Small bacterial genomes (<2 Mb) are more likely to possess direct repeats in close proximity (Aras *et al.*, 2003a), possibly because proximity-dependent slippage rearrangements at these repeats provide a measure of genetic plasticity that cannot otherwise be achieved with a limited genome size. *Helicobacter pylori* exhibits a high degree of host-specific genetic variation and possesses extensive non-randomly distributed repetitive DNA in its genome (Saunders *et al.*, 1998; 2000; Aras *et al.*, 2003a). Most of *H. pylori*'s repeats greater than 24 bp are found at a distance less than 5 kb from each other, in the range feasible for replication slippage (Aras *et al.*, 2003a). A number of direct repeat clusters occur within open reading frames such as *amiA*, *cagY* and *cagA*, where in frame deletions or duplications alter

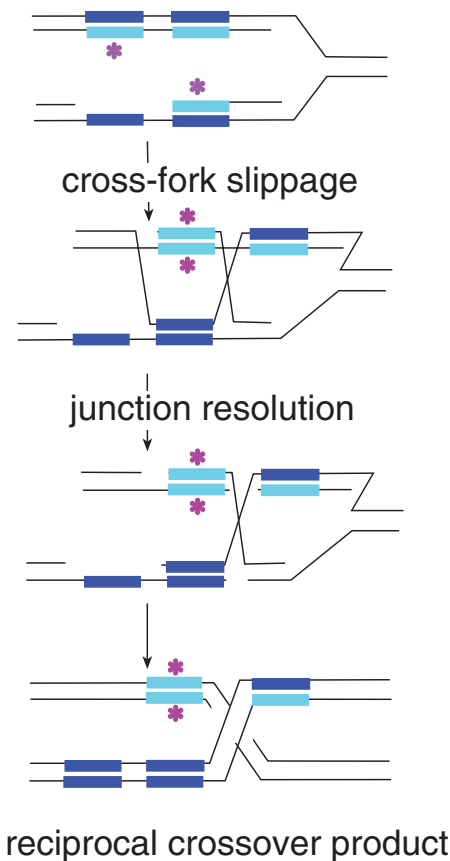
protein properties such as antigenicity or phosphorylation status (Aras *et al.*, 2003a,b,c). These rearrangements may therefore contribute to the persistence of the pathogen by antigenic evasion and adaptation to individual hosts.

### Cross-fork (SCE-associated) slippage

In addition to slipped alignment events between a nascent strand and its template, there is evidence that misalignment between direct repeats can also occur across a replication fork. Such events have been invoked to explain RecA-independent sister chromosome exchange (SCE), which is often associated with repeat deletion or duplication (Lovett *et al.*, 1993; Morag *et al.*, 1999). In plasmid replicons, cross-fork slippage can be detected by the formation of dimeric plasmids coincident with selected tandem repeat deletion or duplication (Yi *et al.*, 1988; Dianov *et al.*, 1991; Mazin *et al.*, 1991; Lovett *et al.*, 1993; Bi *et al.*, 1995; Morag *et al.*, 1999). We estimate that such cross-fork events account for  $\approx 10$ –40% of all deletions or duplications. Models for cross-fork slippage (Lovett *et al.*, 1993; Feschenko and Lovett, 1998; Morag *et al.*, 1999) propose mispairing across the replication fork to produce a Holliday junction intermediate that is resolved to dimeric final products with one of the monomers bearing a deletion or duplication between the direct repeats (Fig. 4). Mutations in DNA exonuclease I (*sbcB*) and DNA polymerase III (*dnaE*) specifically elevate the frequency of cross-fork slippage deletion events (Saveson and Lovett, 1997; Bzymek *et al.*, 1999). The biological consequence of this cross-fork template switching is still unclear, but we have proposed that such events may function normally to repair blocked replication forks (Lovett *et al.*, 1993).

### Replication slippage and inverted repeats

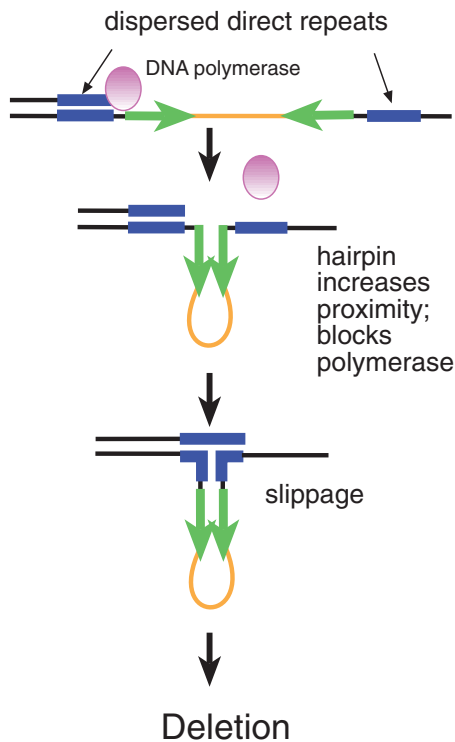
The presence of inverted repeats that form secondary structure strongly enhances misalignment between flanking direct repeats (Albertini *et al.*, 1983; Weston-Hafer and Berg, 1989; Pierce *et al.*, 1991; Sinden *et al.*, 1991; Bzymek and Lovett, 2001a,b). Spontaneous deletion hot-spots between short direct sequence repeats often possess such associated inverted repeats. For example, a deletion hot-spot in *E. coli lacI* is associated with an 18 bp palindrome (Albertini *et al.*, 1982). Where it has been studied, inverted repeats affect deletion primarily on the lagging strand (Trinh and Sinden, 1991; Rosche *et al.*, 1995; Pinder *et al.*, 1998; Bzymek and Lovett, 2001b). Presumably, a hairpin formed by the inverted repeats is favoured on the lagging strand template with greater single-stranded character. Secondary structure formation can stimulate deletion by several means (Fig. 5). One is to increase the local proximity of otherwise dispersed



**Fig. 4.** Model for cross-fork or SCE-associated slippage. After blocked DNA replication, mispairing occurs across the replication fork at direct repeats on both nascent DNA strands (light blue). Additional pairing of the repeats on the template strands (dark blue) produces a Holliday junction structure across the replication fork. After resolution of the crossed strands, repair DNA synthesis and ligation, a cross-over can be produced. On plasmids, this cross-over produces a plasmid dimer with one deleted repeat allele and one triplicated repeat allele.

direct repeats. However, even perfectly juxtaposed direct repeats can be stimulated by the inclusion of inverted repeats. In this latter case, the secondary structure might act as a block to DNA replication (LaDuca *et al.*, 1983), allowing the nascent strand to realign. *In vivo*, deletion end-points abut the inverted repeat, consistent with a block to replication and a slippage event on the lagging strand (Bzymek and Lovett, 2001b). *In vitro*, misalignment is favoured by a lack of processivity and dissociation of the polymerase at the hairpin (Canceill *et al.*, 1999; Viguera *et al.*, 2001). In agreement with this, mutations in the polymerase subunit of *E. coli* DNA polymerase III strongly enhance inverted repeat-stimulated slippage *in vivo* (Bzymek and Lovett, 2001a,b).

It should be noted that inverted repeats stimulate deletion not only by favouring replication slippage, but also by a mechanism distinct from those featured in this review, involving cleavage of the secondary structure by the



**Fig. 5.** Stimulation of slippage by inverted repeats. Dispersed direct repeat sequences (shown in blue) can be juxtaposed by the formation of a secondary structure by the inverted repeats (green arrows). Furthermore, this structure might act as a block to DNA replication, freeing the 3' nascent strand so that it may engage in misalignment at the direct repeats.

SbcCD nuclease, followed by annealing at the repeats (Bzymek and Lovett, 2001a,b).

### Template switching between inverted repeats

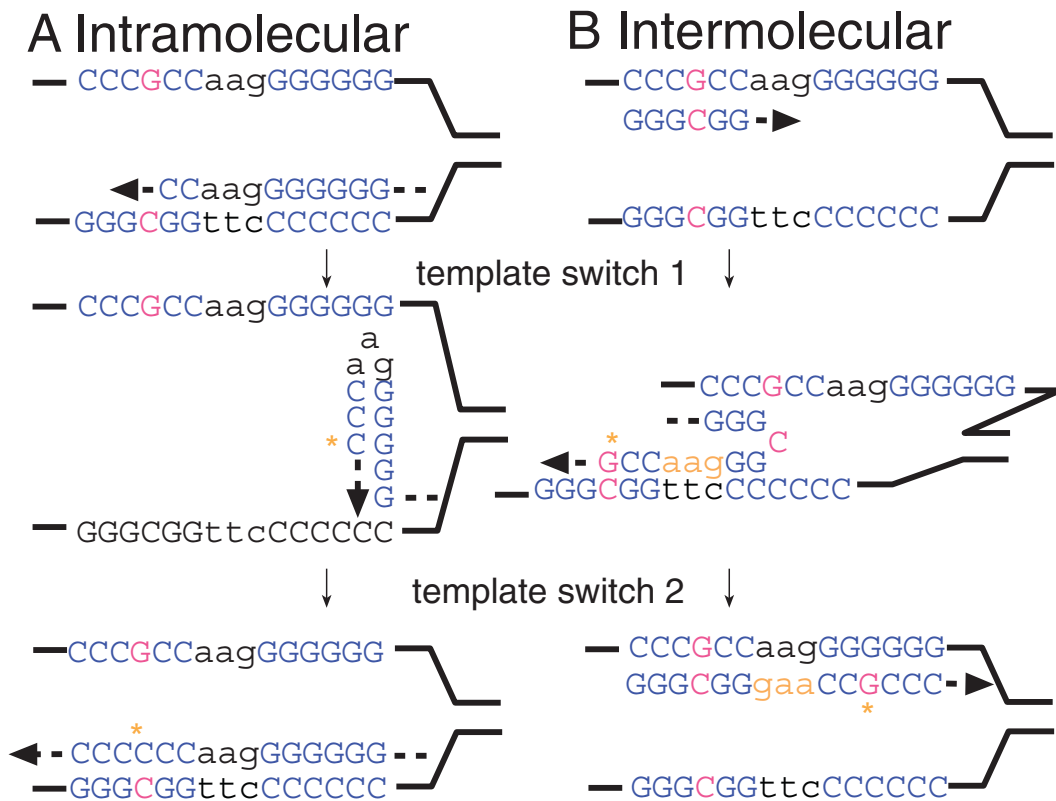
Like 'dislocation' mutagenesis between imperfect direct repeats, transient misalignment between inverted repeats with slight sequence differences ('quasipalindromes') can be a frequent source of mutation. This misalignment allows one repeat to template mutations into the other. This mechanism converts imperfect inverted repeats into more perfect ones and can lengthen repeats. Quasipalindrome-associated mutations were first noted by Lynn Ripley in bacteriophage T4 (Ripley, 1982). They are often recognized as complex mutations with multiple sequence changes or as frequent frameshift mutations that are not associated with nucleotide runs. Quasipalindrome-associated mutations have been noted in many organisms, from bacteria to humans, and can be potent mutation hot-spots (reviewed by Ripley, 1990; Bissler, 1998). For example, a natural quasipalindrome in the *thyA* gene of *E. coli* accounts for 60% of the mutations that inactivate the gene (Viswanathan *et al.*, 2000).

The simplest mode of quasipalindrome mutagenesis

invokes the formation of a DNA hairpin from the imperfect repeat in which one strand of the hairpin acts as a template for synthesis of the other (Fig. 6A). The *thyA* hot-spot mutation, an AT to TA transversion within a 17 bp stem-loop structure, can be accounted for by this type of simple intramolecular template switch at a natural quasipalindrome (Viswanathan *et al.*, 2000). In theory, the DNA synthesis required for this mechanism can be initiated by localized DNA repair within the hairpin (perhaps elicited by mismatch repair of the unpaired bases in the hairpin; Ripley, 1982) or, alternatively, by a template switch reaction during normal chromosomal duplication. Our study of the effect of mismatch repair on the *thyA* quasipalindrome mutational hot-spot in *E. coli* (Viswanathan *et al.*, 2000) allowed us to conclude that most mutations at this site occur during normal replication, rather than localized DNA repair synthesis. The mutation-prone repeat of the inverted pair in the *thyA* hot-spot is 3' on the lagging strand, which would be consistent with the idea that hairpin formation by the inverted repeats is more favourable on the lagging strand.

Another type of quasipalindrome mutagenesis can occur by slipped mispairing between the inverted repeats across the replication fork (sometimes termed 'intermolecular' misalignment). In this model (Ripley, 1982; 1990), mispairing across the replication fork templates limited DNA synthesis at the repeat before a second-strand switch restores normal replication (Fig. 6B). If the switch occurs during replication of the first repeat of an inverted pair, it can cause an inversion of any intervening sequence between the two inverted repeats (as in Fig. 6B). Using this feature, Rosche *et al.* (1995; 1997; 1998) devised a clever assay for cross-fork mispairing and noted that it occurred more frequently on the nascent leading strand than on the nascent lagging strand. The explanation for this bias is that the leading strand's target for mispairing is the lagging strand template, more likely to be single stranded and therefore available for pairing.

Quasipalindrome-associated mutations can be seen in mutational spectra in *E. coli*. Quasipalindrome-associated complex mutations in a plasmid-borne *rpsL* gene were elevated more than 200 000-fold by a mutator allele of DNA polymerase III, *dnaE173*, and comprise 8% of all detected mutations in this strain (Mo *et al.*, 1991). The effect of *dnaE173* may be to increase the frequency of template switch events or to favour the likelihood that DNA synthesis occurs in the slipped configuration. However, other quasipalindrome-associated mutations are not affected by *dnaE173* (Yoshiyama *et al.*, 2001), suggesting that different factors may initiate the events that culminate in this class of mutations. During *in vitro* DNA polymerization, quasipalindrome-associated mutations were highly correlated with replication pause sites (Papanicolaou and Ripley, 1991). In *rpsL*, most quasipalindrome-associated



**Fig. 6.** Quasipalindrome-associated mutations. Schematics of intramolecular and intermolecular template switches that give rise to mutations at sites of imperfect inverted repeats are shown.

A. Formation of a hairpin at the inverted sequence (shown in blue) allows the nascent strand to copy one strand of the hairpin, causing a mutational change (yellow asterisk).

B. Cross-form mispairing at the inverted repeat allows the templating of a mutation (yellow asterisk). If the template switch occurs during synthesis of the first repeat, an inversion of the intervening DNA is produced after realignment (sequence in yellow). Two template switch reactions are required in both mechanisms, and the mutations restore perfection to the inverted repeat sequence.

mutations occur more frequently in one orientation relative to the replication fork (Yoshiyama *et al.*, 2001). Unlike that of the *thyA* hot-spot, most *rpsL* quasipalindromes carry the mutable repeat 3' on the replicating leading strand, which may implicate a cross-fork mechanism.

Surveillance by single-strand DNA exonucleases with activity on 3' ends may abort many misalignment events *in vivo*, including quasipalindrome-associated mutations. In the absence of the major 3' single-strand DNA exonucleases, exonuclease I and exonuclease VII, the frequency of the *thyA* quasipalindrome-associated mutation rises 40-fold (Viswanathan *et al.*, 2000). These exonucleases may deter such mutations by degrading 3' ended single strands that are displaced during the two template switch reactions, in either intramolecular or intermolecular misalignment mechanisms.

### Replication slippage and genome dynamics

The processes of misalignment discussed above will both create and destroy repetitive sequences in the genome.

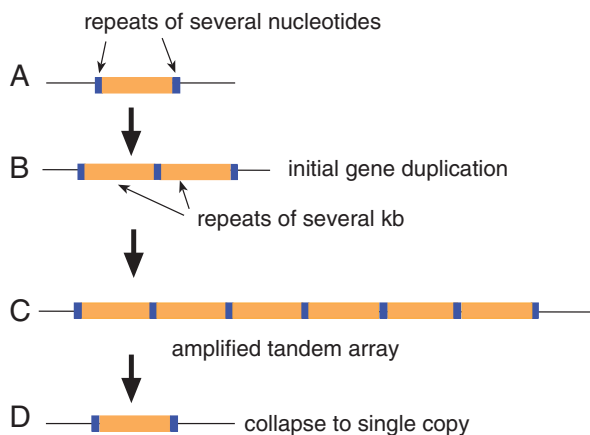
Transient slipped templating of imperfect direct or inverted repeats will tend to extend and to perfect nearby repeat sequences. Data for the natural *thyA* imperfect palindrome suggest that mutations that enhance the palindrome are at least 100 times more frequent than those that destabilize it (Viswanathan *et al.*, 2000), an indication of the force towards acquisition of repetitive DNA. Transient misalignment mutagenesis ('dislocation' or quasipalindrome associated) will therefore tend to drive the formation of tandem repetitive sequences, especially in intergenic regions where they pose less selective detriment.

However, when repeats get large enough, they become less stable and are deleted, in part as a result of replication slippage. Direct repeats are deleted by misalignment and, if they reach the size of several hundred bases, can be excised by homologous recombination. Direct repeats can be stabilized against deletion by one of several mechanisms. Introduction of mutations in one of the repeats will inhibit both slippage and homologous recombination. In addition, separation of the repeats by intervening DNA can move them out of the range of efficient replication

slippage; inclusion of essential genes between the repeats can offer a counterselection to their deletion. For repeats that offer selective advantage (contingency or hypervariable loci, for example), repeats that have collapsed to single copy can be restored by homologous recombination with DNA derived from other individuals retaining the repeats, for instance by transduction or transformation.

Inverted repeats also suffer loss from the genome. Large inverted repeats (*c.* hundreds of bases) form secondary structures that interfere with DNA replication and, for this reason, are particularly deleterious to the cell. The SbcCD nuclease of bacteria acts to cleave and to promote their removal by break-mediated deletion processes (Leach, 1994). In addition, the ability of even modest-sized inverted repeats to stimulate slippage-mediated deletion at short, spurious direct repeats that flank the element (Weston-Hafer and Berg, 1989) will tend to encourage the removal of inverted repeats.

An example of the dynamic nature of repetitive sequences is seen in the process of gene amplification in bacteria (Anderson and Roth, 1978; Edlund and Normark, 1981; Mekalanos, 1983; Tlsty *et al.*, 1984). When amplified expression of particular genes is advantageous, misalignment can be the initiating and rate-limiting process in the generation of large arrays of direct tandem repeats (Fig. 7). Such tandem array amplification underlies at least part or all of the mutational phenomenon known as 'adaptive mutation' in bacteria, in which mutations arise in non-dividing cells under selection (Hastings *et al.*, 2000; Hendrickson *et al.*, 2002). The first step of amplification is



**Fig. 7.** Bacterial gene amplification by formation of tandem repeat arrays.  
 A. Short spurious repeats, shown in blue, may flank a gene segment.  
 B. Replication slippage at the short direct repeats can create a much larger duplication (orange plus blue segments).  
 C. Homologous recombination can act on the larger homology to create tandem arrays by unequal crossing over.  
 D. In the absence of selection, the array collapses by recombination to a single copy.

often a slippage between short spurious repeats to generate a larger genomic duplication of the short repeat and any intervening DNA sequence (Edlund and Normark, 1981). These larger repeats can then easily be expanded by unequal RecA-dependent homologous recombination (Anderson and Roth, 1978; Mattes *et al.*, 1979; Chandler and Galas, 1983; Huffman and Rownd, 1984; Goldberg and Mekalanos, 1986; Romero *et al.*, 1995; Volf and Altenbuchner, 1998; Barten and Meyer, 2001) or, in rare cases, by RecA-independent processes (Mattes *et al.*, 1979; Barten and Meyer, 2001), presumably by additional slipped duplications. These large tandem arrays can be maintained under selection but, when selection is removed, the array will tend to collapse to single copy by homologous recombination.

The properties of slippage-promoted rearrangements and mutations have important implications for bacterial evolution. The high frequency of such events allows genetic adaptation of small founding bacterial populations, as well as the creation of genetic diversity within larger populations. This may be especially important for bacteria with small genomes, where genetic plasticity can be created with little excess sequence (Aras *et al.*, 2003a). In addition, slippage-promoted mutations (dislocation and quasipalindrome-associated mutations) can produce multiple sequence changes in a single event, allowing greater coding variation than can be obtained by simple point mutation. Quasipalindrome template switch mutagenesis also creates inverted repeats that can be recruited as sites for regulatory protein binding, allowing variation of gene regulation.

### Questions for the future

Despite the prevalence and importance of misalignment errors in genomic variation, certain mechanistic questions remain. Recent studies suggest that replication forks stall frequently and are repaired and restarted (Cox *et al.*, 2000). The connection between replication fork failure, fork repair and misalignment events in bacteria remains to be investigated. Certainly, many mutations that interfere with the processivity and co-ordination of DNA replication also stimulate slippage-mediated deletions and duplications (Saveson and Lovett, 1997; Bruand *et al.*, 2001). Are certain slippage events actively promoted by attempted replication fork repair? Long-range slippage events involve the interaction of repeated sequences hundreds or thousands of bases apart – active unwinding of the nascent strand as part of replication fork repair might accidentally lead to misalignment events if repeats are present in the stalled fork (as we have proposed; Lovett *et al.*, 1993). If many slippage events occur as a mistake in replication fork repair, what is the genetic basis of such repair? Does recruitment of alternative DNA repair poly-

merases (*E. coli* DNA polymerases II, IV and V) that might be less accurate or processive play a role in promoting misalignment-mediated rearrangements and mutations? Are there additional cellular factors that act to limit misalignment errors? Are there factors to promote slippage under conditions where it may be advantageous, as for contingency loci?

### Unifying principles

During DNA replication, nascent strands appear to wander and mispair with nearby repetitive sequences, leading to a host of mutations and rearrangements. Such events can be quite frequent and often appear to be stimulated by failure of the replication apparatus. 3' ssDNA-specific exonucleases play an important role in removing these wandering strands as, in the absence of the major 3' exonucleases in *E. coli*, a number of slippage-mediated mutations and rearrangements are substantially elevated. Mismatch repair also aborts short frameshifts and misalignment between imperfectly homologous sequences. Slippage events are not always disadvantageous but can act to promote genetic diversity within bacterial populations, streamline the genome and provide raw material for gene amplification.

### Acknowledgements

I thank all the members of my laboratory who have contributed throughout the last 10 years to the study of slipped alignments, and the General Medical Institute of the NIH for funding (GM43889 and GM51759).

### References

- Albertini, A.M., Hofer, M., Calos, M.P., and Miller, J.H. (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* **29**: 319–328.
- Albertini, A.M., Hofer, M., Calos, M.P., Tlsty, T.D., and Miller, J.H. (1983) Analysis of spontaneous deletions and gene amplification in the *lac* region of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* **47**: 841–850.
- Anderson, R.P., and Roth, J.R. (1978) Tandem genetic duplications in *Salmonella typhimurium*: amplification of the histidine operon. *J Mol Biol* **126**: 53–71.
- Aras, R.A., Kang, J., Tschumi, A.I., Harasaki, Y., and Blaser, M.J. (2003a) Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proc Natl Acad Sci USA* **100**: 13579–13584.
- Aras, R.A., Fischer, W., Perez-Perez, G.I., Crosatti, M., Ando, T., Haas, R., and Blaser, M.J. (2003b) Plasticity of repetitive DNA sequences within a bacterial (type IV) secretion system component. *J Exp Med* **198**: 1349–1360.
- Aras, R.A., Lee, Y., Kim, S.K., Israel, D., Peek, R.M., Jr and Blaser, M.J. (2003c) Natural variation in populations of persistently colonizing bacteria affect human host cell phenotype. *J Infect Dis* **188**: 486–496.
- Barten, R., and Meyer, T.F. (2001) DNA circle formation in *Neisseria gonorrhoeae*: a possible intermediate in diverse genomic recombination processes. *Mol Gen Genet* **264**: 691–701.
- Bayliss, C.D., Field, D., and Moxon, E.R. (2001) The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J Clin Invest* **107**: 657–662.
- van Belkum, A., Scherer, S., van Alphen, L., and Verbrugh, H. (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**: 275–293.
- van Belkum, A., van Leeuwen, W., Scherer, S., and Verbrugh, H. (1999) Occurrence and structure-function relationship of pentameric short sequence repeats in microbial genomes. *Res Microbiol* **150**: 617–626.
- Bi, X., and Liu, L.F. (1994) *recA*-independent and *recA*-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. *J Mol Biol* **235**: 414–423.
- Bi, X., Lyu, Y.L., and Liu, L.F. (1995) Specific stimulation of *recA*-independent plasmid recombination by a DNA sequence at a distance. *J Mol Biol* **247**: 890–902.
- Bissler, J.J. (1998) DNA inverted repeats and human disease. *Front Biosci* **3**: d408–418.
- Bruand, C., Bidnenko, V., and Ehrlich, S.D. (2001) Replication mutations differentially enhance RecA-dependent and RecA-independent recombination between tandem repeats in *Bacillus subtilis*. *Mol Microbiol* **39**: 1248–1258.
- Bzymek, M., and Lovett, S.T. (2001a) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci USA* **98**: 8319–8325.
- Bzymek, M., and Lovett, S.T. (2001b) Evidence for two mechanisms of palindrome-stimulated deletion in *Escherichia coli*: Single-strand annealing and replication slipped mispairing. *Genetics* **158**: 527–540.
- Bzymek, M., Saveson, C.J., Feschenko, V.V., and Lovett, S.T. (1999) Slipped misalignment mechanisms of deletion formation: *in vivo* susceptibility to nucleases. *J Bacteriol* **181**: 477–482.
- Cancelli, D., Viguera, E., and Ehrlich, S.D. (1999) Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. *J Biol Chem* **274**: 27481–27490.
- Chandler, M., and Galas, D.J. (1983) IS1-mediated tandem duplication of plasmid pBR322. Dependence on *recA* and on DNA polymerase I. *J Mol Biol* **165**: 183–190.
- Chedin, F., Dervyn, E., Dervyn, R., Ehrlich, S.D., and Noirot, P. (1994) Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol Microbiol* **12**: 561–569.
- Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000) The importance of repairing stalled replication forks. *Nature* **404**: 37–41.
- Cupples, C.G., Cabrera, M., Cruz, C., and Miller, J.H. (1990) A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of specific frameshift mutations. *Genetics* **125**: 275–280.
- Dianov, G.L., Kuzminov, A.V., Mazin, A.V., and Salganik, R.I. (1991) Molecular mechanisms of deletion formation in *Escherichia coli* plasmids. I. Deletion formation medi-

- ated by long direct repeats. *Mol Gen Genet* **228**: 153–159.
- Edlund, T., and Normark, S. (1981) Recombination between short DNA homologies causes tandem duplication. *Nature* **292**: 269–271.
- Farabaugh, P.J., Schmeissner, U., Hofer, M., and Miller, J.H. (1978) Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J Mol Biol* **126**: 847–857.
- Feschenko, V.V., and Lovett, S.T. (1998) Slipped misalignment mechanisms of deletion formation: analysis of deletion endpoints. *J Mol Biol* **276**: 559–569.
- Feschenko, V.V., Rajman, L.A., and Lovett, S.T. (2003) Stabilization of perfect and imperfect tandem repeats by single-strand DNA exonucleases. *Proc Natl Acad Sci USA* **100**: 1134–1139.
- Fowler, R.G., Degnen, G.E., and Cox, E.C. (1974) Mutational specificity of a conditional *Escherichia coli* mutator *mutD5*. *Mol Gen Genet* **133**: 179–191.
- Fowler, R.G., Schaaper, R.M., and Glickman, B.W. (1986) Characterization of mutational specificity within the *lacI* gene for a *mutD5* mutator strain of *Escherichia coli* defective in 3′–5′ exonuclease (proofreading) activity. *J Bacteriol* **167**: 130–137.
- Franklin, N. (1971) Illegitimate recombination. In *The Bacteriophage Lambda*. Hershey, A.D. (ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 175–194.
- Funchain, P., Yeung, A., Stewart, J.L., Lin, R., Slupska, M.M., and Miller, J.H. (2000) The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154**: 959–970.
- Goldberg, I., and Mekalanos, J.J. (1986) Effect of a *recA* mutation on cholera toxin gene amplification and deletion events. *J Bacteriol* **165**: 723–731.
- Hastings, P.J., Bull, H.J., Klump, J.R., and Rosenberg, S.M. (2000) Adaptive amplification: an inducible chromosomal instability mechanism. *Cell* **103**: 723–731.
- Hendrickson, H., Slechta, E.S., Bergthorsson, U., Anderson, D.I., and Roth, J.R. (2002) Amplification-mutagenesis: evidence that ‘directed’ adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc Natl Acad Sci USA* **99**: 2164–2169.
- Hood, D.W., Deadman, M.E., Jennings, M.P., Bisercic, M., Fleischmann, R.D., Venter, J.C., and Moxon, E.R. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc Natl Acad Sci USA* **93**: 11121–11125.
- Huffman, G.A., and Rownd, R.H. (1984) Transition of deletion mutants of the composite resistance plasmid NR1 in *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* **159**: 488–498.
- Jonsson, A.B., Nyberg, G., and Normark, S. (1991) Phase variation of gonococcal pili by frameshift mutation in *pilC*, a novel gene for pilus assembly. *EMBO J* **10**: 477–488.
- Kunkel, T.A. (1990) Misalignment-mediated DNA synthesis errors. *Biochemistry* **29**: 8003–8011.
- Kunkel, T.A., and Soni, A. (1988) Mutagenesis by transient misalignment. *J Biol Chem* **263**: 14784–14789.
- LaDuca, R.J., Fay, P.J., Chuang, C., McHenry, C.S., and Bambara, R.A. (1983) Site-specific pausing of deoxyribonucleic acid synthesis catalyzed by four forms of *Escherichia coli* DNA polymerase III. *Biochemistry* **22**: 5177–5188.
- Leach, D.R. (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioassays* **16**: 893–900.
- Lovett, S.T., and Feschenko, V.V. (1996) Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. *Proc Natl Acad Sci USA* **93**: 7120–7124.
- Lovett, S.T., Drapkin, P.T., Sutura, V.A., Jr, and Gluckman-Peskind, T.J. (1993) A sister-strand exchange mechanism for *recA*-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics* **135**: 631–642.
- Lovett, S.T., Gluckman, T.J., Simon, P.J., Sutura, V.A., Jr, and Drapkin, P.T. (1994) Recombination between repeats in *Escherichia coli* by a *recA*-independent, proximity-sensitive mechanism. *Mol Gen Genet* **245**: 294–300.
- Makino, S., van Putten, J.P., and Meyer, T.F. (1991) Phase variation of the opacity outer membrane protein controls invasion by *Neisseria gonorrhoeae* into human epithelial cells. *EMBO J* **10**: 1307–1315.
- Mattes, R., Burkardt, H.J., and Schmitt, R. (1979) Repetition of tetracycline resistance determinant genes on R plasmid pRS1 in *Escherichia coli*. *Mol Gen Genet* **168**: 173–184.
- Mazin, A.V., Kuzminov, A.V., Dianov, G.L., and Salganik, R.I. (1991) Mechanisms of deletion formation in *Escherichia coli* plasmids. II. Deletions mediated by short direct repeats. *Mol Gen Genet* **228**: 209–214.
- Mekalanos, J.J. (1983) Duplication and amplification of toxin genes in *Vibrio cholerae*. *Cell* **35**: 253–263.
- Mo, J.Y., Maki, H., and Sekiguchi, M. (1991) Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. *J Mol Biol* **222**: 925–936.
- Morag, A.S., Saveson, C.J., and Lovett, S.T. (1999) Expansion of DNA repeats in *Escherichia coli*: effects of recombination and replication functions. *J Mol Biol* **289**: 21–27.
- Nelson, M.A., Singer, B.S., Gold, L., and Pribnow, D. (1981) Mutations that detoxify an aberrant T4 membrane protein. *J Mol Biol* **149**: 377–403.
- Papanicolaou, C., and Ripley, L.S. (1991) An *in vitro* approach to identifying specificity determinants of mutagenesis mediated by DNA misalignments. *J Mol Biol* **221**: 805–821.
- Parker, B.O., and Marinus, M.G. (1992) Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc Natl Acad Sci USA* **89**: 1730–1734.
- Pierce, J.C., Kong, D., and Masker, W. (1991) The effect of the length of direct repeats and the presence of palindromes on deletion between directly repeated DNA sequences in bacteriophage T7. *Nucleic Acids Res* **19**: 3901–3905.
- Pinder, D.J., Blake, C.E., Lindsey, J.C., and Leach, D.R. (1998) Replication strand preference for deletions associated with DNA palindromes. *Mol Microbiol* **28**: 719–727.
- Puopolo, K.M., Hollingshead, S.K., Carey, V.J., and Madoff, L.C. (2001) Tandem repeat deletion in the alpha C protein

- of group B streptococcus is *recA* independent. *Infect Immun* **69**: 5037–5045.
- Ripley, L.S. (1982) Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proc Natl Acad Sci USA* **79**: 4128–4132.
- Ripley, L.S. (1990) Frameshift mutation: determinants of specificity. *Annu Rev Genet* **24**: 189–213.
- Ripley, L.S., Glickman, B.W., and Shoemaker, N.B. (1983) Mutator versus antimutator activity of a T4 DNA polymerase mutant distinguishes two different frameshifting mechanisms. *Mol Gen Genet* **189**: 113–117.
- Romero, D., Martinez-Salazar, J., Girard, L., Brom, S., Davilla, G., Palacios, R., *et al.* (1995) Discrete amplifiable regions (amplicons) in the symbiotic plasmid of *Rhizobium etli* CFN42. *J Bacteriol* **177**: 973–980.
- Rosche, W.A., Trinh, T.Q., and Sinden, R.R. (1995) Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J Bacteriol* **177**: 4385–4391.
- Rosche, W.A., Trinh, T.Q., and Sinden, R.R. (1997) Leading strand specific spontaneous mutation corrects a quasipalindrome by an intermolecular strand switch mechanism. *J Mol Biol* **269**: 176–187.
- Rosche, W.A., Ripley, L.S., and Sinden, R.R. (1998) Primer-template misalignments during leading strand DNA synthesis account for the most frequent spontaneous mutations in a quasipalindromic region in *Escherichia coli*. *J Mol Biol* **284**: 633–646.
- Saunders, N.J., Peden, J.F., Hood, D.W., and Moxon, E.R. (1998) Simple sequence repeats in the *Helicobacter pylori* genome. *Mol Microbiol* **27**: 1091–1098.
- Saunders, N.J., Jeffries, A.C., Peden, J.F., Hood, D.W., Tettelin, H., Rappuoli, R., and Moxon, E.R. (2000) Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol Microbiol* **37**: 207–215.
- Saveson, C.J., and Lovett, S.T. (1997) Enhanced deletion formation by aberrant DNA replication in *Escherichia coli*. *Genetics* **146**: 457–470.
- Schaaper, R.M. (1988) Mechanisms of mutagenesis in the *Escherichia coli* mutator mutD5: role of DNA mismatch repair. *Proc Natl Acad Sci USA* **85**: 8126–8130.
- Schaaper, R.M., and Dunn, R.L. (1987) 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* **84**: 6220–6224.
- Schaaper, R.M., Danforth, B.N., and Glickman, B.W. (1986) Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli* *lacI* gene. *J Mol Biol* **189**: 273–284.
- Sinden, R.R., Zheng, G.X., Brankamp, R.G., and Allen, K.N. (1991) On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation *in vivo*. *Genetics* **129**: 991–1005.
- Tlsty, T.D., Albertini, A.M., and Miller, J.H. (1984) Gene amplification in the *lac* region of *E. coli*. *Cell* **37**: 217–224.
- Trinh, T.Q., and Sinden, R.R. (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**: 544–547.
- Viguera, E., Canceill, D., and Ehrlich, S.D. (2001) Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* **20**: 2587–2595.
- Viswanathan, M., Lacirignola, J.J., Hurley, R.L., and Lovett, S.T. (2000) A novel mutational hotspot in a natural quasipalindrome in *Escherichia coli*. *J Mol Biol* **302**: 553–564.
- Volff, J.N., and Altenbuchner, J. (1998) Genetic instability of the *Streptomyces* chromosome. *Mol Microbiol* **27**: 239–246.
- Weston-Hafer, K., and Berg, D.E. (1989) Palindromy and the location of deletion endpoints in *Escherichia coli*. *Genetics* **121**: 651–658.
- Yi, T.M., Stearns, D., and Demple, B. (1988) Illegitimate recombination in an *Escherichia coli* plasmid: modulation by DNA damage and a new bacterial gene. *J Bacteriol* **170**: 2898–2903.
- Yoshiyama, K., Higuchi, K., Matsumura, H., and Maki, H. (2001) Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in *Escherichia coli*. *J Mol Biol* **307**: 1195–1206.