DNA Instability in Bacterial Genomes: Causes and Consequences

13

Pedro H. Oliveira, Duarte M.F. Prazeres and Gabriel A. Monteiro

Abstract

DNA is a structurally dynamic molecule that is central to cellular processes such as replication, transcription and recombination. In order to maintain genomic integrity, bacteria have developed a finely tuned and interwoven network of mechanisms that operate at multiple levels, including damage recognition, signalling pathways, and DNA repair. On the other hand, without the capacity to accommodate genotypic variation up to a certain extent, bacteria would not be able to modify their fitness when faced with constantly changing environments. Herein we review our current knowledge on bacterial genome instability, with particular emphasis on findings gained from the often-studied Gram-negative model organism *Escherichia coli*. We will address topics such as spontaneous and stress-induced mutagenesis, major DNA repair pathways, and the design of more stable genomes. Major questions and future challenges will also be discussed.

Introduction

Bacterial DNA is recurrently exposed to insults of both endogenous and exogenous sources. Endogenous factors known to damage DNA in a spontaneous fashion, i.e. in actively dividing cells and in the absence of exogenous stressors, typically include DNA replication errors (Bzymek and Lovett, 2001; Hardin *et al.*, 2007), movement of mobile elements (Naas *et al.*, 1995), and lesions such as depurination, depyrimidation, deamination, alkylation and oxidation (Loeb and Preston, 1986; Cooke *et al.*, 2003; Drablos *et al.*, 2004; Gates, 2009). Many of these events are context specific, in the sense that they are highly dependent on the local DNA sequence and topology, as well as on the extent of replication, transcription and recombination activities. Typically, the overall rate of spontaneous mutations in wild-type *Escherichia coli* is maintained at low levels (around 10^{-3} per genome per generation) (Lee *et al.*, 2012), mostly due to the replication fidelity of the DNA polymerase III (Pol III) holoenzyme, as well as to the methyl-directed mismatch repair (MMR), base excision repair (BER) and nucleotide excision repair (NER) systems. Exogenous stresses, on the other hand, consist of alkylating agents (Drablos *et al.*, 2004), antimicrobials (Lopez and Blazquez, 2009), nutrient deprivation (Mandel and Silhavy, 2005), oxidative stress (Cooke *et al.*, 2003), ionizing (e.g. X-rays, gamma rays and atomic particles) (Tindall *et al.*, 1988; Bentchikou *et al.*, 2010) or ultraviolet (UV) radiation (Eichenbaum and Livneh, 1998), extreme temperature, pH and osmolarity shifts (Jolivet-Gougeon *et al.*, 2000; Ohtsubo *et al.*, 2005), among others. Under such stressful and adverse environments, the need to generate diversity, fuels maladapted cells to transiently up-regulate their mutation rates. This process of stress-induced mutagenesis varies according to the organism and persistence/intensity of the stress condition, and generally involves relaxing the consistency and fidelity of the DNA repair machinery, the use of error-prone polymerases, a shift to a transient hypermutable state, induced transposition of mobile elements and other rearrangements. For example, in the Gram-negative *E. coli* bacterium, a considerable number of mutations appearing in stressed cells results from error-prone repair of DNA double

strand breaks (DSBs), in a process controlled by the SOS DNA-damage response and the general stress response sigma factor RpoS (Galhardo *et al.*, 2009; Rosenberg *et al.*, 2012).

Making sense of the biological significance and utility of mutational events and associated mechanisms is therefore fundamental to answer long-standing questions in multiple fields such as those dealing with evolution and species diversity, cancer progression, synthetic biology, among many others. This book chapter is designed to present a brief overview of the causes and consequences of spontaneous and stress-induced genomic instability in bacteria. In addition, we will address topics that are of utmost relevance to those in the biopharmaceutical sector engaged in the development of streamlined genomes and DNA-based molecules for research and therapeutic applications.

Spontaneous DNA lesions, replication and transcription as sources of instability

Spontaneous DNA lesions

Several types of DNA damage can be generated endogenously as a result of intrinsic DNA instability or the presence of by-products of the cellular metabolism. Examples of this are depurination, depyrimidation, deamination, alkylation or oxidation. Depurination and depyrimidation consist in the hydrolysis of the *N*-glycosidic bond between the deoxyribose moiety and guanine or adenine (depurination) or cytosine and thymine (depyrimidation). This leads to the formation of the so-called apurinic or apyrimidinic (AP or abasic) sites, the latter being formed at roughly 5% the rate of the former (Lindahl, 1993). AP sites are highly mutagenic and constitute a roadblock to replication and transcription (Loeb and Preston, 1986). Additional base modifications may occur through the spontaneous hydrolytic deamination of adenine, guanine, cytosine, and the modified base 5-methylcytosine (5meC), respectively yielding hypoxanthine, xanthine, uracil and thymine (Coulondre *et al.*, 1978; Duncan and Miller, 1980) (Fig. 13.1). Deamination of cytosine in double-stranded DNA (dsDNA) proceeds at a rate of about 7×10^{-13}/s (Frederico *et al.*, 1990), while 5meC deaminates 2–3 times faster (Shen *et al.*, 1994). Both are capable of generating G:C>A:T transitions. On the other hand, deamination of adenine and guanine occur at much slower rates (approximately 2–3%) than those observed for cytosine (Lindahl, 1979). Adenine deamination may result in A:T>G:C transitions, whereas deamination of guanine to xanthine may lead to DNA replication arrest and G:C>A:T transitions due to its ability to mispair with thymine. Hypoxanthine may also pair with cytosine leading to A:T>G:C transitions. Interestingly, deamination occurs at faster rates in single-stranded DNA (ssDNA) than in dsDNA, meaning that cellular processes such as replication, transcription or recombination have the potential to exacerbate those events (Frederico *et al.*, 1990; Beletskii and Bhagwat, 1996).

Nucleophilic sites in DNA represent potential targets to both endogenous and exogenous electrophilic alkylating agents. Examples of endogenous alkylating agents include the ubiquitous methyl donor S-adenosylmethionine (SAM) (Rydberg and Lindahl, 1982; Taverna and Sedgwick, 1996) as well as nitrosated amines or amides (Taverna and Sedgwick, 1996), that are capable of generating lesions in dsDNA such as O6-methylguanine (O6meG), 3-methyladenine (3meA) and 7-methylguanine (7meG) (Fig. 13.1). O6meG is known to cause G:C>A:T transitions (Loechler *et al.*, 1984), whereas 3meA has the ability to block replication and generate AP sites (Boiteux *et al.*, 1984; Larson *et al.*, 1985). 7meG is *per se* regarded as non-cytotoxic and non-mutagenic, but is capable of destabilizing the *N*-glycosidic bond and generate AP sites (Shrivastav *et al.*, 2010). In *E. coli*, O6meG lesions can be repaired by two DNA methyltransferases, the inducible Ada protein and the constitutive Ogt protein, whereas 3meA and 7meG are repaired by the Tag and AlkA glycosylases of the BER pathway (see below).

Reactive oxygen species (ROS) such as hydroxyl radicals (•OH), singlet oxygen ($^1O_2$), superoxide ($O_2^{\bullet -}$), and hydrogen peroxide (H_2O_2) are by-products of the aerobic metabolism that are also capable of exerting a genotoxic effect on nucleic acids. The most intensively studied modified base is probably 8-oxo-7,8-dihydroguanine

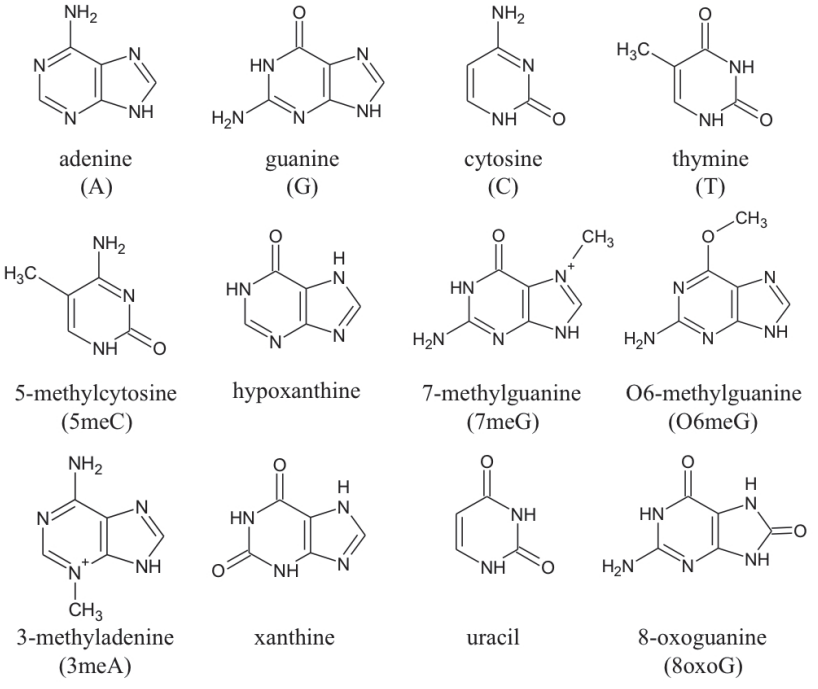

Figure 13.1 Structures of the DNA bases and of common endogenous base lesions.

(8-oxoguanine, 8oxoG), the product of guanine oxidation (Fig. 13.1). 8oxoG efficiently pairs with both cytosine and adenine mainly resulting in G:C>T:A transversions (Shibutani *et al.*, 1991). Many additional base lesions have been identified as the result of oxidative damage, and have been properly reviewed elsewhere (Cooke *et al.*, 2003).

The BER pathway is generally involved in the repair of structurally non-bulky damaged bases that do not cause major distortions to the double helix (e.g. after base excision, oxidation, alkylation) (reviewed by (Stivers and Jiang, 2003). In BER, mono- or bifunctional DNA glycosylases recognize and excise damaged bases by cleavage of the *N*-glycosidic bond, producing AP sites that are then processed by AP endonucleases, DNA polymerase and ligase to restore the original sequence (see Fig. 13.2 for further details). In order to mitigate the detrimental effect of 8oxoG, several bacteria make use of the oxidized guanine (GO) system, which typically comprises three proteins: MutT, MutM and MutY. MutT is a hydrolase that converts 8-oxoGTP to 8-oxoGMP and pyrophosphate in order to prevent the incorporation of the oxidized dGTP into the DNA (Maki and Sekiguchi, 1992). Conversely, MutM and MutY are glycosylases pertaining to the BER pathway. MutM excises 8oxoG when paired to cytosine, whereas MutY catalyses the excision of adenine when mismatched with 8oxoG (Michaels *et al.*, 1992; Tajiri *et al.*, 1995). The MMR system (see Fig. 13.3) corrects postreplicative errors but has also been shown to be involved in the direct excision of 8oxoG, mispaired adenines opposite 8oxoG, or both (Wyrzykowski and Volkert, 2003). Further details on these and other base lesions have been previously reviewed (Gates, 2009; Shrivastav *et al.*, 2010).

Replication errors occurring during DNA biosynthesis

Some of the most common errors generated during DNA biosynthesis are point mutations, which can appear either in the form of base substitutions or insertions and deletions (indels). Among the latter, those that cause a change in the reading frame are termed frameshift mutations. Frameshift mutations can be found in

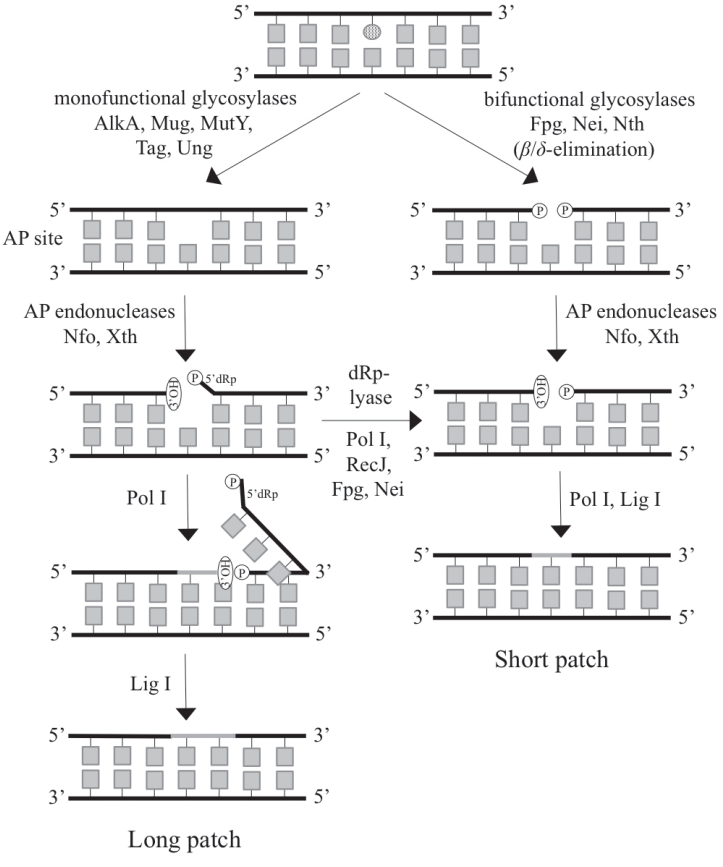

Figure 13.2 Scheme for the long-patch and small-patch BER pathways. The first step in BER involves the recognition and removal of base lesions (stippled circle) either by monofunctional glycosylases (for deaminated, alkylated or mismatched bases) or bifunctional glycosylases (for oxidized, or ring-saturated bases). Monofunctional glycosylases generate AP sites, which are subsequently cleaved at their 5′-side by an AP endonuclease leaving a 3′-hydroxyl (3′-OH) group and a 5′-deoxyribose phosphate (5′-dRp) terminus. Bifunctional glycosylases, on the other hand, possess AP lyase activity, with an associated β- or β/δ elimination activity (being the latter shown in the figure). In this situation, a single nucleotide gap is generated containing 5′- and 3′-phosphate (P) groups at the termini. Two distinct pathways are followed according to the type of glycosylases used in the process: if the repair is initiated by monofunctional glycosylases, it may follow both the long and short-patches, whereas repair initiated by bifunctional glycosylases may only follow the short-patch. In the long-patch pathway, the 5′-dRp and about 2–10 nucleotide patches are replaced during strand displacement by Pol I and subsequently ligated. On the other hand, in the short-patch pathway mediated by monofunctional glycosylases, the 5′-dRp terminus is removed by Fpg, Nei, RecJ or by the exonuclease activity of Pol I, generating a 5′-phosphate terminus. In the short-patch pathway mediated by bifunctional glycosylases, the 3′-phosphatase/phosphodiesterase activity of AP endonucleases will remove the corresponding blocking phosphate. Gap-filling and ligation steps then take place.

runs of repetitive DNA, and in this case, can be explained on the basis of misalignment between a nascent strand and its template, as proposed in the seminal work of Streisinger and colleagues (Streisinger *et al.*, 1966). Such misaligned structures allow the formation of loops which, according to their size, can be recognized and resolved by the MMR system (Lovett and Feschenko, 1996). In order to accommodate those situations in which frameshift mutations do not occur within homopolymeric DNA sequences, alternative models have been proposed. According to the so-called dislocation model, frameshift mutations could be initiated by direct misincorporation of a nucleotide (Kunkel and Soni, 1988) (Fig. 13.4a). It

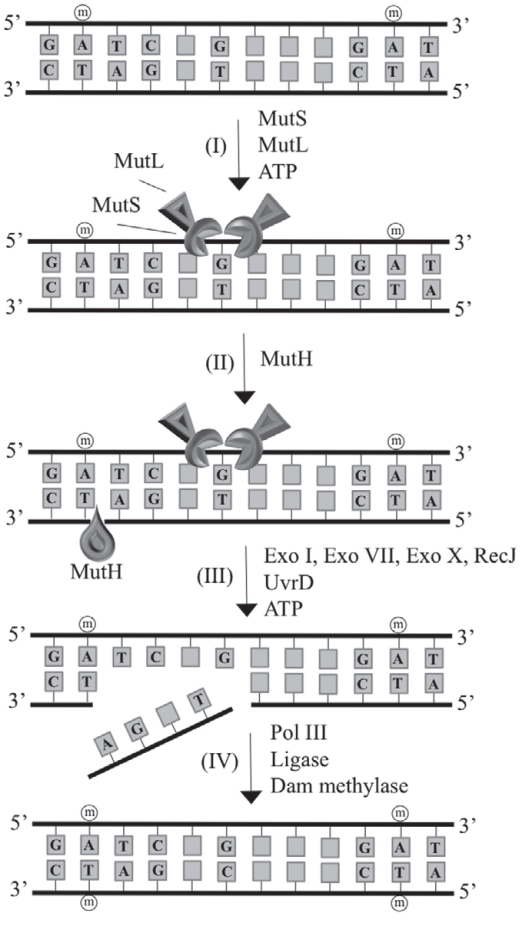

Figure 13.3 Scheme of the methyl-directed mismatch repair (MMR) pathway in *E. coli*. A base mismatch is recognized by the MutS homodimer, followed by recruiting of the MutL homodimer to the site of the lesion in a MutS- and ATP-dependent manner (I). Next, MutH is recruited to the nearest hemi-methylated (m) d(GATC) site and nicks the corresponding newly synthesized strand (non-methylated) (II). Because the nearest hemi-methylated d(GATC) site may be located in either sides of the lesion, the following step of strand degradation will involve UvrD (DNA helicase II) and different ssDNA exonucleases (III). Exo I, Exo VII and Exo X are involved in 3′→5′ degradation, whereas Exo VII and RecJ are involved in 5′→3′ degradation. DNA Pol III and ligase will fill the gap, and a Dam methylase will methylate the newly synthesized DNA strand (IV).

then follows that the misincorporated base may realign and pair with an adjacent complementary template, ultimately leading to an indel. Alternatively, base substitutions may originate if strand polymerization occurs in the absence of realignment (Fig. 13.4a). Many examples of spontaneous small-scale mutations have been reported to occur in a wide range of bacteria, such as in the *trpA* (Hardin *et al.*, 2007) and *lacI* (Schaaper and Dunn, 1991) genes of *E. coli*, in the *pyrE/pyrF loci* of *Thermus thermophilus* (Mackwan *et al.*, 2008), or in the *hisD3052* allele of *Salmonella typhimurium* (DeMarini *et al.*, 1998).

By a process resembling that of frameshift mutagenesis, slipped misalignment may take place over longer distances between interspersed DNA repeats. Such events can result in deletions or amplifications (deletion formation) of both repeat and spacer (intervening) sequence, in a process that is independent of the homologous recombination (HR) protein RecA, but highly dependent on repeat and spacer lengths (Dianov *et al.*, 1991; Mazin *et al.*, 1991; Bi and Liu, 1994; Chedin *et al.*, 1994) (Fig. 13.4b). Misalignment between direct repeats may also take place across a replication fork, in a process involving reciprocal sister-chromosome exchange (Lovett *et al.*, 1993; Bi and Liu, 1996; Feschenko and Lovett, 1998). Cross-fork slippage accompanied by sister-chromosome exchange involves an erroneous annealing between repeats from both nascent strands (see Fig. 13.4c for more details), and has been used to explain the formation of heterodimeric products in plasmids harbouring direct repeats (Morag *et al.*, 1999). In this regard, our group has previously reported the contamination of purified batches of a therapeutic non-viral plasmid DNA (pDNA) vector with deleted (monomeric, M) and head-to-tail heterodimeric (1+2, 1+3) forms (Ribeiro *et al.*, 2008; Oliveira *et al.*, 2010). The latter represent a safety concern, since there is the risk to administer small populations of unwanted plasmid molecules with unknown biological properties. These events may also lead to a loss of productivity during the expression of proteins of interest (Kawe *et al.*, 2009; Kim *et al.*, 2011). The frequency of such repeat-mediated recombination events can be estimated by using predictive mathematical tools (Oliveira *et al.*, 2008), and experimentally detected and/or quantified using techniques such as direct sequencing, PCR-based enrichment steps, temperature- or

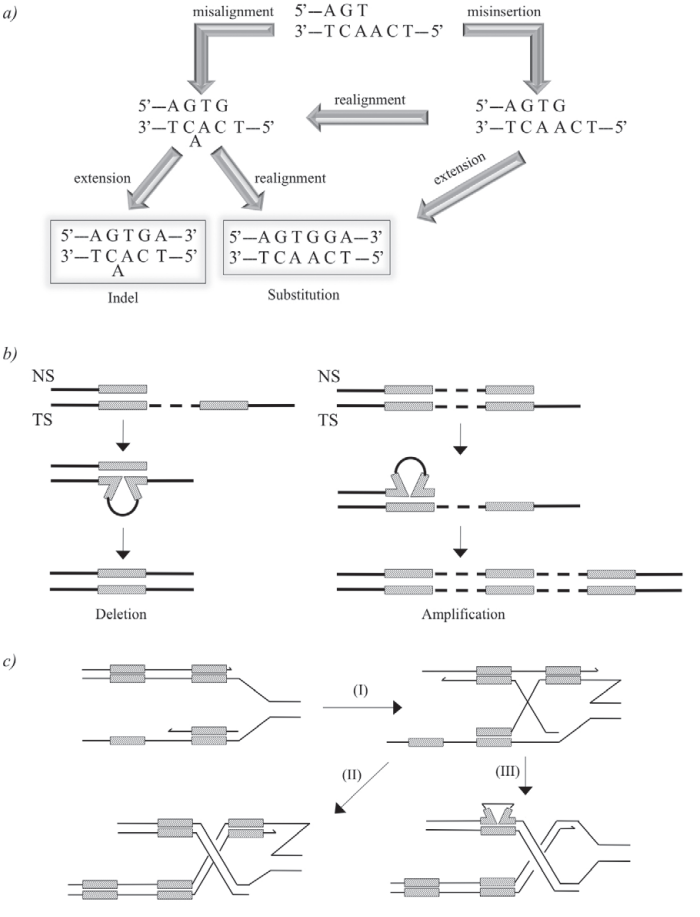

Figure 13.4 Slipped-misalignment events in DNA. (*a*) Indels and substitutions can be generated either by strand misalignment as initially proposed by Streisinger and co-workers (Streisinger *et al.*, 1966), or misinsertion as proposed in the dislocation model (Kunkel and Soni, 1988). (*b*) Slipped misalignment between interspersed direct repeats (rectangles) may take place when the nascent strand (NS) disengages from the template strand (TS) and re-anneals at a new downstream or upstream position, respectively generating deletion or amplification of genetic material. The spacer (or intervening) sequence between repeats is shown as an intermittent line. (*c*) Model of slipped misalignment accompanied by sister chromosome exchange. In a stalled replication fork, unequal pairing between nascent 3′-ends may take place involving repeated sequences (I). The intermediate form yielding a Holliday junction can be further processed in the absence of realignment (II) or with realignment of parental strands (III). This respectively leads to the formation of one deletion/triplication product or a deletion/duplication product. In circular molecules, the latter are termed 1+3 and 1+2 respectively. (b) and (c) were adapted from Bzymek and Lovett (2001).

denaturing-gradient gel electrophoresis, and real-time PCR (reviewed in (Oliveira *et al.*, 2011). In a very recent study using 454 pyrosequencing technology, large-scale spontaneous inversions, deletions and duplications were found at high frequency (20–40%) in growing cultures of *S. enterica* (Sun *et al.*, 2012), which highlights the importance of performing these types of analysis on a more frequent basis.

Instability mediated by non-B structures

Depending on its sequence and surrounding conditions, a DNA molecule has the ability to occasionally adopt several conformations beside the canonical right-handed double stranded B-form. These commonly include bent-inducing sequences, slipped structures, cruciforms, G-quadruplexes, triplexes and Z-DNA (Fig. 13.5).

Such non-canonical or non-B structures arise from separation of duplex DNA and negative supercoiling induced by several biological processes such as replication, transcription or protein binding. In addition, they have also been considered as 'fragile regions' prone to rearrangements.

Slipped-strand structures originate whenever a displaced nascent strand misaligns and pairs with an alternative complementary region in the template strand (previously shown in Fig. 13.4b). Cruciforms are typically generated by two main mechanisms (Lilley, 1985). The most common one, which is observed under physiological ionic conditions (S-type), involves a pre-nucleation unwinding step of 8 to 10 central base pairs, intrastrand hydrogen bonding between cognate bases flanking the symmetry centre (nucleation), and a final extrusion and complete intrastrand annealing between inverted repeats (Fig. 13.5a). Alternatively, the C-type mechanism has been observed under low-salt concentrations, and requires a long AT-rich DNA region flanking the inverted repeats, that allows the formation of a large bubble where nucleation and extrusion may take place. If ssDNA becomes available (e.g. during replication, DNA repair, or during uptake of exogenous DNA), single hairpins may be formed instead (Fig. 13.5b), with a stability that is proportional to its length and sequence context (Gacy *et al.*, 1995). Long inverted repeats have been shown to be replication stall sites both in prokaryotes and eukaryotes (Voineagu *et al.*, 2008). The authors demonstrate that hairpins, rather than cruciforms are preferentially formed, and thus should be regarded as the main agents of instability. Notwithstanding, hairpin DNA structures can be recognized and cleaved by the SbcCD nuclease (Connelly *et al.*, 1998), leaving a DSB for subsequent repair by the RecBCD pathway of HR (Eykelenboom *et al.*, 2008).

Concerning triplex structures, they can form at homopurine–homopyrimidine stretches, where a single strand portion can bind to the purine-rich strand of the duplex via non-canonical hydrogen bonding (Fig. 13.5c). Different triplexes can thus be formed according to the source of the triplex-forming strand: intermolecular if it originates from a second DNA molecule (e.g.

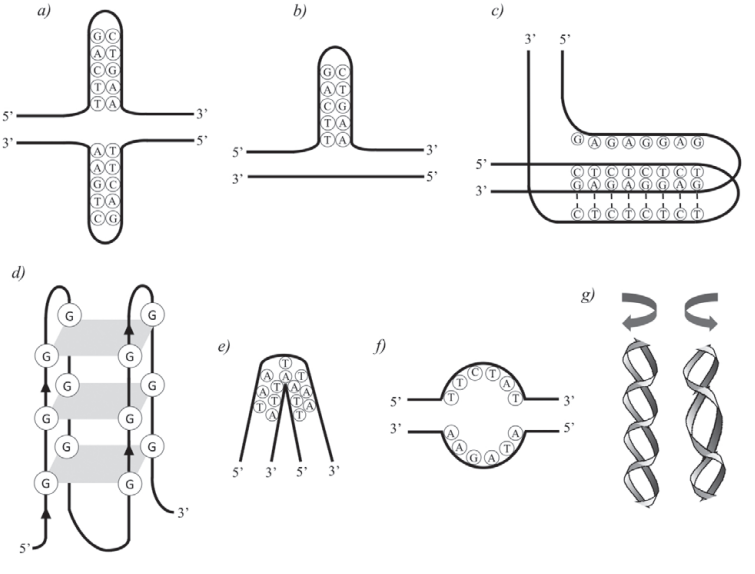

Figure 13.5 Examples of non-B (non-canonical) DNA structures. (a) Cruciform; (b) hairpin; (c) triplex or H-DNA (shown is the Y*R:Y structure formed by Hoogsteen hydrogen bonding); (d) intramolecular tetraplex, G4-DNA or G-quadruplex with diagonal loop; (e) bent DNA; (*f*) unwound DNA; (g) comparison between B-DNA (left, clockwise) and Z- or left-handed DNA (right, anticlockwise). Figure was partially adapted from (Oliveira *et al.*, 2012).

triplex-forming oligonucleotides (TFOs)), or intramolecular if it involves regions with mirror symmetry (H-DNA) that originate from the same duplex strand. Irrespectively of the scenario, the third strand can be either pyrimidine-rich and bind with a parallel orientation to the purine strand in the duplex (Y*R:Y) via Hoogsteen base pairing, or purine-rich and bind with an anti-parallel orientation to the purine strand in the duplex (R*R:Y) via reverse Hoogsteen base pairs. It has been documented that H-DNA induces a yet unclear cascade of events encompassing the formation of DSBs, processing via non-homologous end-joining (NHEJ), and eventual formation of large-scale deletions or translocations (Jain *et al.*, 2008). Triple-helical structures have also been recognized as hot spots for the integration of mobile elements such as Tn7 (Rao *et al.*, 2000).

Tetraplex, G4-DNA or G-quadruplex structures (Fig. 13.5d) are stable stacks of coplanar G-quartets stabilized by monovalent cations, originating from transient single-stranded DNA bearing a signature motif of the type $G \geq_3 N_{1-7} G \geq_3 N_{1-7} G \geq_3 N_{1-7} G \geq_3$ or $C \geq_3 N_{1-7} C \geq_3 N_{1-7} C \geq_3 N_{1-7} C \geq_3$. Based on the number of DNA strands involved in quadruplex formation, these structures may be classified as intermolecular (two or four strands) or intramolecular (one strand), whereas loops may be classified as edgewise, diagonal or double-chain reversal. G-quadruplexes have been shown to interfere with important biological processes like replication, transcription, recombination and telomere protection, thus being regarded as potential agents of genetic instability (Han and Hurley, 2000; Belotserkovskii *et al.*, 2010; Boan and Gomez-Marquez, 2010). In addition, several genes involved in secondary metabolite biosynthesis and signal transduction, seem to be under the control of G4 motifs in *E. coli* (Rawal *et al.*, 2006).

The presence of properly phased $A_n T_m$ ($n + m \geq 4$) tracts often leads to intrinsically curved regions (Fig. 13.5e), which play an important role in modulating replication and transcription initiation (Linial and Shlomai, 1988; Kahn and Crothers, 1993), and serve as hot targets for the insertion of mobile elements (Kobori *et al.*, 2009; Lewis *et al.*, 2012). Conversely, extremely AT-rich sequences are able to promote unwinding of DNA with increasing negative superhelical stress (Fig. 13.5f), and are highly hypersensitive to attack by nucleases (Kowalski *et al.*, 1988).

Alternating purine-pyrimidine repeat tracts such as $(CA:TG)_n$ and $(CG:CG)_n$ may fold into a left-handed structure with about 12 bp per turn called Z-DNA (Fig. 13.5g). Such structures have been shown to block transcription (Ditlevson *et al.*, 2008), and induce the formation of DSBs, resulting in gross deletions and complex rearrangements in mammalian cells, but only small deletions or expansions within the repetitive sequences in *E. coli*. Further processing either by NHEJ or HR is thought to be behind the different outcomes (Kha *et al.*, 2010).

Triplexes, G-quadruplexes and three- or four-way junctions have been described to be resolved by RecQ helicases, a ubiquitous subgroup of helicases highly conserved from prokaryotes to higher eukaryotes that ensures the faithful progression of the replication fork (Sharma *et al.*, 2006; Sharma, 2011).

A number of *ab initio* and consensus-based bioinformatics resources have been developed aiming at providing information about the presence of non-B elements. Several *k*-mer and seed extension approaches have been developed to detect repeated regions. Some examples include Reputer (Kurtz and Schleiermacher, 1999), Repeatoire (Treangen *et al.*, 2009) and Repseek (Achaz *et al.*, 2007) for the detection of exact or degenerate repeats, or Tandem Repeats Finder (Benson, 1999) for tandem repeats. Others focus on identifying elements such as triplex DNA, Z-DNA and quadruplexes (e.g. TTS (Jenjaroenpun and Kuznetsov, 2009), QuadParser (Huppert and Balasubramanian, 2005)). More recently, a database named Non-B DB has been developed, which takes into consideration all classes of non-B DNA-forming motifs, and provides the user with a search option for any input sequence as well as access to data in its genomic context (Cer *et al.*, 2011). Also, several experimental techniques such as two-dimensional gel electrophoresis, circular dichroism, nuclease cleavage, and others, have been used to detect non-B structures (Wang *et al.*, 2009).

Replication–transcription conflicts

In bacteria, the rate of replication is about 12–30 times higher than that of transcription (respectively 600 nucleotides/s versus 20–50 nucleotides/s) (Mirkin and Mirkin, 2007; Rudolph *et al.*, 2007), which leads to frequent physical conflicts between the replication and transcription machineries as well as occasional genomic rearrangements. While replisome and RNA polymerase (RNAP) collisions may occur either in a head-on or co-directional fashion, a greater impairment and stronger pausing of the replication fork is often seen in head-on collisions rather than in co-directional conflicts (Mirkin and Mirkin, 2005). To this end, bacterial genomes have evolved towards an over-representation of long, essential, and highly transcribed genes in the leading strand, presumably in an effort to limit head-on collisions (Brewer, 1988; Rocha and Danchin, 2003; Omont and Kepes, 2004; Price *et al.*, 2005). Studies performed *in vitro* show that co-directional collisions are often resolved by rapidly dislodging the RNAP from the template and resuming elongation using mRNA as a primer (Pomerantz and O'Donnell, 2008) (Fig. 13.6). However, *in vivo* studies performed in *E. coli*, revealed a less benign face of co-directional encounters, by showing that such thermodynamically stable DNA:RNA hybrids (R-loops) may cause replication fork stalling (Gan *et al.*, 2011) and favour the accumulation of DSBs mutations, particularly as a result of collisions with backtracked elongation complexes (Dutta *et al.*, 2011). RNAP backtracking allows the formation of longer and more stable R-loops, which provide accessible 3′-OH termini that can be used as primers for DNA synthesis (Dutta *et al.*, 2011). This leads to the formation of a single-strand break

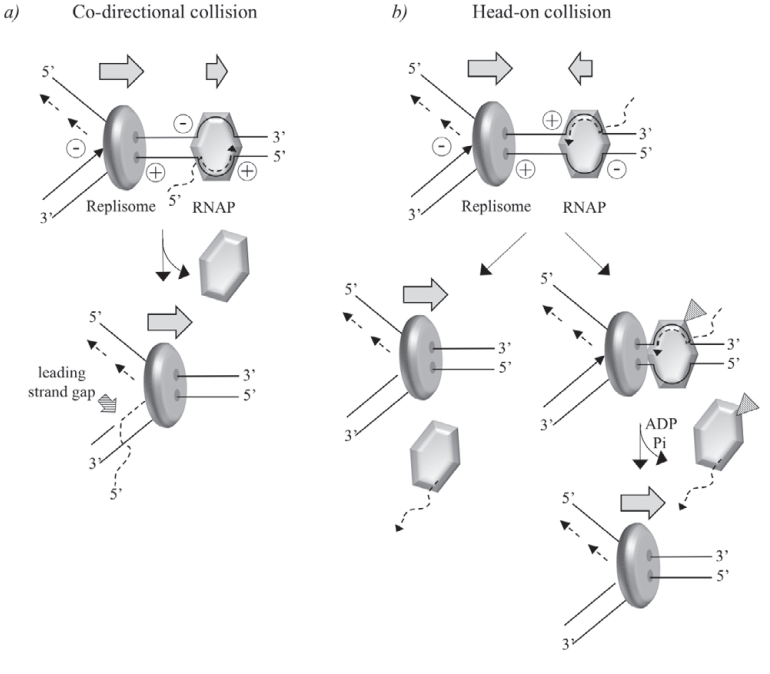

Figure 13.6 Replication fork arrest as the result of co-directional and head-on collisions between the replisome and RNAP. (*a*) Co-directional collisions occur in the leading strand and typically have a less detrimental effect on fork progression, since RNAP is rapidly dislodged and the newly synthesized displaced RNA is used as primer to resume elongation leaving behind a nick or gap. (*b*) Upon a head-on (frontal) collision in the lagging strand, the replisome stalls for longer periods of time because of the physical interaction with RNAP and the excess of positive supercoiling between both complexes (shown as plus signals). The replisome may then resume elongation after slowly displacing RNAP (left), or alternatively, the transcription-repair coupling factor Mfd (shown as a triangle) may bind to the halted transcription complex and force its dissociation in a process catalysed by ATP hydrolysis.

(SSB) that can be converted into a DSB after a second round of replication. To prevent excessive instability, the cells make use of several anti-backtracking mechanisms, namely termination factors and anti-termination complexes, among others (Dutta *et al.*, 2011). *In vivo* studies performed with *B. subtilis* also describe replication disruption and restart upon co-directional collisions at highly expressed ribosomal genes (Merrikh *et al.*, 2011). *In vivo* head-on collisions, on the other hand, induce a profound fork arrest eventually followed by fork collapse and recombination (Vilette *et al.*, 1996; Prado and Aguilera, 2005). This fact results from the strong physical contact and the positively supercoiled constraining domain generated between both machineries, being the former more detrimental than the latter. Further evidence on the fate of both replisome and RNAP after head-on collisions came from a study performed *in vitro* in *E. coli* (Pomerantz and O'Donnell, 2010). In this situation the replisome pauses but slowly resumes elongation after displacing RNAP from DNA (Fig. 13.6). This process of RNAP displacement and fork progression may be aided by the transcription-repair coupling factor Mfd in an ATP-driven reaction (Fig. 13.6). Recently it was shown that blocked forks arising from head-on collisions, may be reversed and further processed by HR or DNA degradation, allowing the reloading of a new replisome further away from the RNAP (De Septenville *et al.*, 2012).

Stress-induced instability

The concept of adaptive or stationary-phase mutation

Cells trying to cope with a changing environment find themselves in the need to effectively fine-tune the balance between genomic integrity and the chance of acquiring beneficial changes. According to the more classical view of DNA mutation, newly arising mutations occur spontaneously in growing cells, in a constant and stochastic manner, and in the absence of an environment in which such mutations might prove advantageous. This view has been challenged by the finding that mutations may also appear in non-growing or slowly growing cells under conditions that favour such mutant subpopulation (Shapiro, 1984; Cairns *et al.*, 1988). Nevertheless, these so-called adaptive or stationary-phase mutations accumulate at the same time that unselected mutations, and therefore, should not be regarded as directed to a particular advantageous gene (Foster, 1997; Torkelson *et al.*, 1997; Rosenberg, 2001). While several mechanisms of adaptive genetic changes have been described in bacteria, those occurring at the lactose catabolism (*lac*) operon in starving *E. coli* (adaptive point mutation and gene amplification) are among the best studied (Foster and Trimarchi, 1994; Hastings *et al.*, 2000; Powell and Wartell, 2001; Slack *et al.*, 2006; Gonzalez *et al.*, 2008). In this assay, cells carrying deleted chromosomal *lac* genes and an episomal *lac* +1 frameshift allele, respond to starvation on a lactose medium either by generating a compensatory frameshift (point mutation) Lac^+ allele (Foster and Trimarchi, 1994), or by allowing the amplification of the leaky *lac* allele to 20–100 copies (Hastings *et al.*, 2000). The mechanism underlying the appearance of adaptive point mutations presumes the differentiation towards a hypermutable subpopulation of cells (i.e. mutagenesis becomes up-regulated), during the repair of DNA lesions such as DSBs or double stranded ends (DSEs) via HR (Ponder *et al.*, 2005; Gonzalez *et al.*, 2008). The mutations are permanent and not directed preferentially to *lac* genes, as many unselected mutations also appear (Foster, 1997; Torkelson *et al.*, 1997). In this scenario there is an activation of the SOS and RpoS (σ^S)-mediated general stress responses, which leads the cell to shift to a mutagenic mode that makes use of the error-prone DNA Pols II (*polB*), IV (*dinB*) and V (*umuDC*) instead of the high-fidelity DNA Pol III (Lombardo *et al.*, 2004; Ponder *et al.*, 2005) (Fig. 13.7a). In this regard, it was recently shown that DSBs concentrate stress-inducible mutations within vicinal hotspots: the so called 'strong ones' are located at approximately 60 kb and arise via RecD-dependent resection of DSBs followed by gap-filling synthesis, and the 'weak ones' are located at roughly 1 Mb distance from a DSB and are formed via break-induced replication (BIR) (Shee *et al.*, 2012). To this matter, the authors also added that single-base differences observed preferentially at poorly expressed genes of the *E. coli* genome may result from the fact that

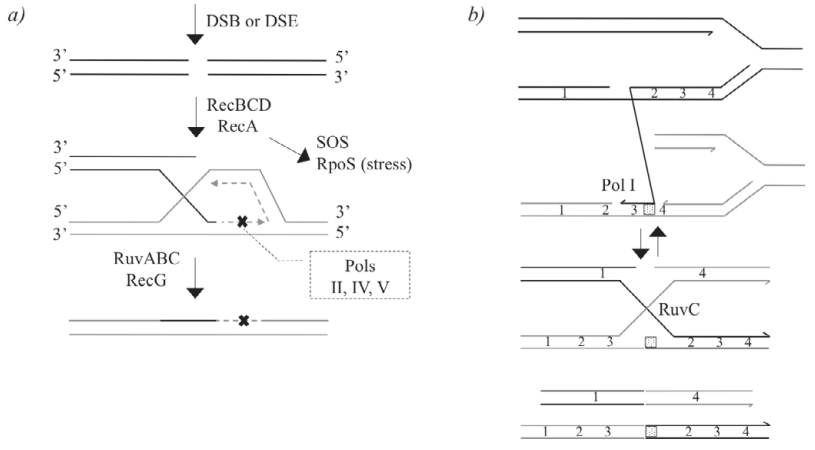

Figure 13.7 Models for DSB-dependent stress-induced point mutations (*a*) and gene amplification (*b*). In the initial stages involved in the repair of a DSB or DSE by homologous recombination (HR), RecBCD produces ssDNA and the RecA protein is loaded. The presence of DSB or DSE induces the SOS response, and the presence of a second stress (e.g. starvation) induces the general stress response mediated by the RpoS activator. This concatenation of events leads to a switch to a mutagenic mode that makes use of the error-prone DNA Pols II, IV and IV, instead of the high-fidelity DNA Pol III. Strand invasion then takes place with a sister molecule or duplicated region, and replication resumes, allowing for the accumulation of mutations (shown as X). On the other hand, the long-distance template-switching model has been proposed as a possible mechanism for adaptive gene amplification. According to this model, stalled replication forks may originate under starvation and in the presence of DSBs. In this situation, a 3′-end from a lagging strand might invade a template strand from a nearby fork, in a process aided by the flap-endonuclease function of Pol I and by a region of microhomology (stippled square). Such invasion is further stabilized by ligation to the 5′-extremity of the Okazaki fragment located ahead. The switch-strand intermediate then forms a Holliday junction that can be resolved by RuvC, yielding a duplicated and a deleted molecule. Schemes (a) and (b) were respectively adapted from Rosenberg *et al.* (2012) and Slack *et al.* (2006).

the latter are more DSB-prone, and thus more mutagenic (Shee *et al.*, 2012). In the mechanism of DSB-dependent adaptive gene amplification, the cells acquire an iterated duplication of the *lac* region. Unlike the *lac* point mutation mechanism, the cell does not enter a hypermutable state, does not require the activation of the SOS response nor Pol IV, and does not make use of HR to generate tandem duplications (Hastings *et al.*, 2000; Lombardo *et al.*, 2004; Slack *et al.*, 2006) (Fig. 13.7b). Amplifications appear to be the result of non-HR and are temporary, in the sense that they can revert to a single copy status by recombination.

Alternative mutagenic pathways have been proposed to take place in both growing and stationary-phase cells. In resting organisms in structured environments (ROSE), *E. coli* cells growing in colonies were found to acquire a rifampicin-resistant phenotype, in a process dependent of cAMP and the SOS regulon (Taddei *et al.*, 1997). In the growth advantage in stationary-phase (GASP) phenotype, a minor population of *E. coli* cells is able to outcompete their counterparts, presumably through an increased ability to catabolize some amino acids. Mutations in the GASP phenotype were found to require Pols II, IV and V, although not requiring the induction of the SOS system and DNA damage (Yeiser *et al.*, 2002).

Environmentally stressed bacteria have also evolved systems allowing for a local increase of mutagenesis at specific sites of the genome. In the next sections we will briefly comment on these 'contingency *loci*', and enumerate several environmental stressors, which have been described to induce mutagenesis.

Contingency loci

Contingency *loci* refer to specific hypermutable sites in a genome that are implicated in stochastic

and reversible on–off phase variation events. Simple sequence repeats (SSRs) are the bacterial equivalents to eukaryotic microsatellites, being composed of tandem repeats of motifs sizing 1–6 bp. SSR-mediated indels arise by slipped-strand mispairing during replication (see Fig. 13.4a and b), and have the potential to, under environmental stress, favourably modulate gene expression by altering reading frames or promoter strength (Li *et al.*, 2004; Martin *et al.*, 2005; Vinces *et al.*, 2009). In the recent decades, whole-genome sequencing of several prokaryotes highlighted curious asymmetries concerning the length, abundance and distribution of SSRs. Particularly, it was observed that (i) longer SSRs are much less frequent in prokaryotes than eukaryotes and tend to be excluded from coding regions; (ii) smaller and AT-rich genomes tend to have an overrepresentation of SSRs when compared to larger and more GC-rich genomes; (iii) tetranucleotide SSRs are typically overrepresented in pathogenic bacteria (Coenye and Vandamme, 2005; Mrazek *et al.*, 2007). As an example, the small 1.8-Mb genome of the commensal bacterium *Haemophilus influenzae* has revealed a surprisingly high density of tetranucleotide SSRs, particularly associated with genes coding for virulence-like factors (Fleischmann *et al.*, 1995; Hood *et al.*, 1996).

Certain SSRs described as variable number of tandem repeats (VNTR) have been used as genetic markers owing to their inter-individual polymorphism. In particular, multilocus VNTR analysis (MLVA) typing assays are particularly useful in epidemiological and evolutionary studies, and essentially consist of a PCR amplification of a predefined set of VNTRs, for which the length and copy number can be estimated. To aid in this process, several bioinformatic resources are currently available that focus not only on the search for repetitive DNA (e.g. Tandem Repeats Finder; Benson, 1999), but also on assisting with the comparison between strains (Denoeud and Vergnaud, 2004; Chang *et al.*, 2007).

Use of antibiotics

Antibiotics are commonly used selectors of resistant individuals. Yet, compelling evidence has built up showing that they are also agents of bacterial evolvability. Antibiotic-induced mutagenesis seems to proceed through the induction of the SOS response, increased transposition of mobile elements, recombination and lateral gene transfer. As an example, studies involving exposure of *E. coli* cells to the broad-spectrum fluoroquinolone antibiotic ciprofloxacin have described the onset of adaptive mutations (Riesenfeld *et al.*, 1997) and provided insight into the mechanisms governing the emergence of resistance-conferring mutations. This response is mediated by cleavage of the LexA repressor, followed by transcription of genes encoding the SOS-regulated error-prone Pols II, IV and V (Cirz *et al.*, 2005). It has also been shown that ciprofloxacin is able to stimulate intrachromosomal recombination in a SOS-independent manner via either RecBCD or RecFOR (Lopez *et al.*, 2007). In another study that involved incubation of *Pseudomonas aeruginosa* for four days in the presence of the bacteriostatic antibiotic tetracycline, the authors found that the mutation rate per cell per day had increased by 10^5 times (Alonso *et al.*, 1999).

Different concentrations of selector also influence mutation rate and the spectra of mutations formed. We have previously shown that increasing concentrations of the aminoglycoside antibiotic kanamycin are able to shape the relative frequencies of deletions versus amplifications of genetic material in a bacterial plasmid model (Oliveira *et al.*, 2009). A shift towards the amplification of genetic material (heterodimeric forms) preferentially occurred at higher selector concentrations, in what constitutes an evidence of the advantage of gene amplification under stressful environments (Oliveira *et al.*, 2009). Increased transposition frequency of the mobile element IS*2* was additionally observed to take place in a bacterial plasmid in the presence of kanamycin. The latter event was found to be region specific and capable of generating a novel hybrid promoter upstream of the kan^R gene (Oliveira *et al.*, 2009). Subinhibitory concentrations of antibiotics can also be the cause of increased genetic variability. A recent example describes the deletion of a haemolysin operon flanked by IS*91* sequences from an *E. coli* plasmid (Pedro *et al.*, 2011). Also, exposure to sublethal levels of a representative group of bactericidal antibiotics was found to increase the minimum inhibitory

concentrations and to induce mutagenesis and resistance to antibiotics other than those used to treat cells, via formation of ROS species (Kohanski *et al.*, 2010).

From the standpoint of preventing the emergence of multiresistant bacteria in clinical settings, it becomes necessary to consider the antibiotic-induced increase in horizontal gene transfer and mutation rates, prior to devising the best strategy for treatment. It was recently shown that the combination of antibiotics is ideal to fight an infection, but detrimental to avoid a scenario of double resistance, in which case the use of cycling and mixing strategies proved to be more effective (Obolski and Hadany, 2012).

Starvation-induced mutations

Upon experiencing periods of nutrient limitation (e.g. when entering stationary phase), bacteria undergo transcriptional (e.g. RpoS-dependent general stress response), metabolic (e.g. stringent response), morphological and translational changes (reviewed in Navarro Llorens *et al.*, 2010). Mutation rate is also increased, in a SOS and RpoS-dependent manner. In this regard, we have already mentioned earlier in this chapter the example of *lac* point mutations and amplification events that take place in situations of cell starvation. Yet, other not so thoroughly studied examples merit our attention. For example, *E. coli* cells grown under carbon starvation on a lactose minimal medium were shown to acquire β-lactam resistance through the accumulation of *ampD* base substitutions and frameshift mutations. Similarly to what is known from ciprofloxacin-induced mutagenesis, loss of function of *ampD* also occurs during DSB repair and requires Pols IV and V (Petrosino *et al.*, 2009). An increasing body of evidence suggests that transposition of mobile elements is favoured at later growth stages under nutrient deprivation. Transposition of Tn*552* and IS*903* was found to occur at later stages of colony growth and to require the presence of GTP, despite the underlying mechanism remaining unclear (Coros *et al.*, 2005). Twiss and co-workers (Twiss *et al.*, 2005) generated random Tn5 insertion mutant libraries in *E. coli*, and selected several mutations capable of affecting transposition of other mobile elements. Those mutations disrupting the *aspA* gene of the TCA cycle were found to be of particular relevance, as they were able to promote transposition of IS*903* at earlier stages of growth. In another study in *P. putida*, the transposition of the toluene degradation transposon Tn*4652* also increased under starvation (Ilves *et al.*, 2001).

Temperature-induced mutations

For many microbes, changes in temperature can profoundly shape their growth, development, virulence profiles, and alter genomic integrity. A genome-wide study performed in several *E. coli* cell lines adapted for 2,000 generations to a temperature of 41.5°C, describes the occurrence of complex rearrangements involving deletions, duplications and insertion of mobile elements (Riehle *et al.*, 2001). Among the regions affected, the authors highlight duplications affecting *rpoS* as well as other genes involved in stress survival and adaptation to high temperature (Riehle *et al.*, 2001). According to a recent study performed in *E. coli* O157:H7 cell lines, the number of mutations involving tandem repeats more than triplicated when temperature increased from 10°C to 43°C (Cooley *et al.*, 2010). Only single repeat changes confined to a single locus were observed at lower temperatures, contrasting to double- and triple-step mutations in multiple *loci* at higher temperatures. In another study, growing *B. multivorans* at a high temperature (42°C) instead of a normal temperature (30°C), led to a sevenfold increase in the transposition activity of seven IS elements belonging to four different families (e.g. IS*401*, IS*402*, IS*406*, among others) (Ohtsubo *et al.*, 2005). Since transposase activity depends on temperature, the authors suggest that changes in the conformation of the donor or target DNA may take place at higher temperatures, favouring a more effective transposition. Also, consecutive submission of several *E. coli* cell lines to freeze–thaw–growth cycles for 1000 generations, led to the accumulation of small deletions and IS*150* insertions in the *cls* and *uspA-uspB* genes, which were found to increase membrane fluidity and enhance survival and recovery (Sleight *et al.*, 2008).

Oxidative stress

We have mentioned before the role of ROS from endogenous origin as a catalyst of spontaneous DNA lesions, in particular involving the modified base 8-oxoG. However, several studies have focused on the adaptation of bacteria to oxidative stress resulting from exogenous sources such as ionizing and UV radiation, or chemical compounds. For example, exposure of *Helicobacter pylori* to ROS and reactive nitrogen species, which are usually found in its natural environment, induced a hypermutable state characterized by the onset of point mutations, intergenomic recombination, and direct repeat-mediated rearrangements (Kang *et al.*, 2006). Exposure of *Burkholderia cenocepacia* to H_2O_2 also stimulated IS-mediated genomic rearrangements, which may consist in an adaptation mechanism to the ROS-rich environment typically found in the lungs of a cystic fibrosis patient (Drevinek *et al.*, 2010). Another interesting example involves the resistance of some enteric bacteria such as *E. coli* or *S. enterica* to the DNA damaging salts contained in bile, an antibacterial secretion produced by the liver and stored in the gall bladder. It has been shown that such salts cause oxidative damage (G:C>A:T transitions), activating a cascade of repair functions such as BER, the SOS response and DinB-mediated translesion synthesis (Prieto *et al.*, 2004, 2006).

UV radiation-induced mutations

UV light is typically classified according to the wavelength of its components: UVA (320–400 nm), UVB (290–320 nm) and UVC (<290 nm). Cell exposure to UV light (particularly to more energetic wavelengths) frequently results in the formation of dimers between adjacent pyrimidine bases, such as *cis-syn* cyclobutadipyrimidines (CPDs), pyrimidine 6–4 pyrimidone photoproducts (6–4PPs), and corresponding Dewar isomers (Fig. 13.8a) (Tyrrell, 1973). In this regard it has been found that T-T and T-C sequences are typically more photoreactive than C-C and C-T (Douki and Cadet, 2001). UV radiation is also known to cause base substitutions, frameshifts and deletions in bacterial DNA (Schaaper *et al.*, 1987). It has also been reported the selection of advantageous induced mutations in genes related to repair and replication of DNA upon submitting *E. coli* to cyclic UV light conditions (Alcantara-Diaz *et al.*, 2004). Such lesions may lead to replication fork arrest if allowed to persist in the genome, and for this reason bacteria make use of several mechanisms to correct them. Several bacteria, archaea, virus, fungi and others, have the ability to correct UV-induced pyrimidine dimers by means of a process named photoreactivation, which uses the light-dependent (300–500 nm) enzyme photolyase encoded by the *phr* gene. Alternatively, pyrimidine dimers as well as other bulky lesions can be corrected in a 'dark mode' by means of excision repair pathways such as BER (described before in this chapter) and NER. NER was first identified in *E. coli* (Pettijohn and Hanawalt, 1964), but it is now known to be highly ubiquitous among the entire tree of life. Recognition and repair of a lesion typically involve the proteins UvrA, UvrB, UvrC and UvrD as follows: first the ternary complex $UvrA_2UvrB$ recognizes and binds the DNA lesion; UvrA proteins dissociate from lesion site and UvrB recruits UvrC that cleaves the DNA at the 4th–5th nucleotide upstream the lesion and at the 7th–8th nucleotide downstream the lesion; finally UvrD (helicase II) removes the single stranded fragment carrying the lesion, and Pol I and DNA ligase complete the repair (Fig. 13.8b).

Exposure to UV light also seems to induce the transposition of mobile elements. In *E. coli* cells, a 28-fold increase in transposition of IS*10* was observed after irradiation with UV light, in a process dependent of the SOS response (Eichenbaum and Livneh, 1998). The authors of this study suggest that the cut-and-paste mechanism of IS*10* transposition may be favoured by the presence of ssDNA gaps resulting from stalled replication forks or DNA repair enzymes.

Cell storage and shipment

Studies on the influence of cell storage and shipment as potential sources of stress are scarce. Among the few studies available, it was found that long starvation periods in resting *E. coli* cells stored in agar stabs promoted increased genetic variation by transposition of mobile elements, namely IS*2*, IS*3*, IS*5* and IS*30* (Naas *et al.*, 1994, 1995). In a more recent work, it was found that

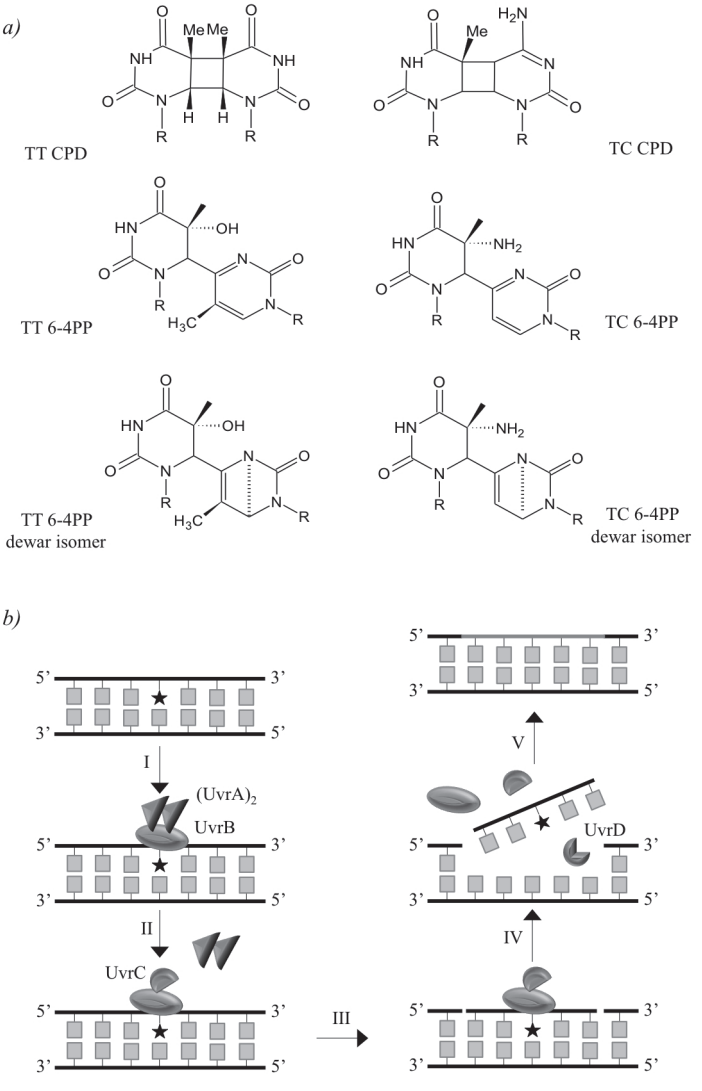

Figure 13.8 Main UV-induced pyrimidine photoproducts and nucleotide excision repair (NER) in eubacteria. (*a*) Three major classes of DNA lesions are formed upon exposure to UV radiation: cyclobutane pyrimidine dimers (CPDs), pyrimidine 6–4 pyrimidone photoproducts (6–4PPs), and corresponding Dewar isomers. (*b*) In NER, a complex formed by two UvrA molecules and one UvrB molecule is involved in the recognition of bulky lesions in the DNA such as pyrimidine dimers (black star) (I). After recognition, both UvrA molecules dissociate, and UvrC binds UvrB (II). In the next step, UvrC cuts 5′ and 3′ to the lesion site (III), and the helicase action of UvrD allows removal of the fragment and bound UvrB and UvrC (IV). Extension by DNA polymerase I and ligation restore the integrity of the fragment (V).

maintaining *E. coli* K12 cells for periods of time as short as a few days in LB-stabs is enough to confer mutations in the *rpoS* locus similar to the ones found in the GASP phenotype (Spira *et al.,* 2011). The authors of this study suggest resuspending bacteria in 15% glycerol (v/v) and applying it to a filter disk, as this strategy appears to maintain a high RpoS phenotype in all cells for a three-week period.

The problem of acquired bacterial resistance upon exposure to relatively mild stressors or certain growth-permissive environments gains particular relevance in the food industry. In an effort to adopt a more cost-effective and

environmental-friendly approach, as well as to retain the sensory characteristics of certain food products, many companies are employing mild preservation approaches in a synergistic way, instead of using harsh preservation techniques such as heat sterilization. The former include washing steps, irradiation with UV light, controlled atmosphere packaging, among others. It turns out that exposure to these stresses at a non-detrimental level, confers the cells what is known as 'cross protection', i.e. an enhanced resistance against other stresses. Concomitantly, this has obvious implications in the assessment of food safety. Readers are referred to Capozzi *et al.* (2009) for a recent review on this topic.

More stable bacterial genomes and DNA-based therapeutic molecules

Severely reduced genomes can be found in nature (e.g. the human parasite *Mycoplasma genitalium* with a genome size of 0.58 Mb), being usually associated with unique and highly specific and stable ecological niches. Such minimal genomes are typically compositionally biased, highly compact, with few intergenic and non-functional regions, and preserve genes involved in core processes of replication, transcription and translation (reviewed in McCutcheon and Moran, 2012)). These naturally occurring examples of genome reduction and the notion that certain genome subsets may be dispensable under a particular environment immediately raised the question on what would be the minimal genomic content necessary to support life. Moreover, the wealth of sequencing data currently available has fuelled the development of projects dealing with minimal or streamlined chassis, particularly from industrial and academic parties. The practical advantages seem obvious: increased genomic stability, construction of microbial power-horses with streamlined pathways for production of high-value bioproducts, reduced production of contaminants by knocking out competing pathways, among many others.

Genome deconstruction by top-down approaches has been successfully achieved through the combinatorial removal of nonessential elements, namely mobile elements or transporter and cryptic genes (Posfai *et al.*, 2006; Ara *et al.*, 2007; Kato and Hashimoto, 2007; Lee *et al.*, 2009). In order to overcome IS-mediated structural instability, IS-free cellular chassis have been devised (Umenhoffer *et al.*, 2010). Such *E. coli* MDS42 cell lines allow the propagation of otherwise unstable clones, show reduced evolutionary potential, and represent a safer option to be used for amplification of DNA-based therapeutics. More recently, it was reported an *E. coli* MDS42-derived and stabilized strain lacking mobile elements and harbouring a triple deletion in the three error prone DNA Pols II, IV and V (Csorgo *et al.*, 2012).

Alternatively, bottom-up engineering attempts to specifically incorporate modular parts to form larger and more complex systems. In this regard, promising tailor-made DNA assembly strategies have been unveiled. For example, a successive hybridization assembly (SHA) relies on a simple *in vitro* hybridization of partially overlapping DNA fragments without the need of restriction enzymes, DNA ligases or recombinases (Jiang *et al.*, 2012). However, the prospect of constructing novel reduced genomes raises important issues that are worth some reflection. Namely, how will the current legislation on intellectual property cope with each novel assembled genome? How do these findings affect our view of life and the way we interact with it?

Within the biopharmaceutical industry, a similar tendency for the minimization of DNA-based therapeutics has been observed in the last couple of years. These strategies try to tackle or at least reduce structural instability and immunogenicity to a minimum, by simply removing those portions that are not needed for the therapeutic application. One possibility relies on the elimination of antibiotic resistance markers, since they pose a metabolic burden to the cells and may constitute a threat if disseminated (Cranenburgh *et al.*, 2001; Pfaffenzeller *et al.*, 2006; Goh and Good, 2008; Marie *et al.*, 2010; Peubez *et al.*, 2010; Vandermeulen *et al.*, 2011). Other strategies based on further plasmid minimization are also gaining wide acceptance mainly due to their small size, low content of unmethylated CpGs, increased transfection efficiency and higher levels

of transgene expression in several tissues. This minimal vector design concept includes both the Minimalistic Immunogenically Defined Gene Expression (MIDGE) vectors and minicircles. MIDGE vectors are linear covalently closed dumbbell-shaped DNA fragments comprising the eukaryotic expression unit and terminal loop structures that protect against exonuclease degradation. In a similar way, minicircles are also devoid of bacterial elements, but are generated by site-specific recombination in *E. coli*, in a process that separates them from miniplasmids (which contain all bacterial elements). These constructs have demonstrated a great clinical significance when used as vectors for gene therapy (Chang *et al.*, 2008) and DNA vaccination (Lopez-Fuertes *et al.*, 2002; Moreno *et al.*, 2004), when used to control cancer growth (Schakowski *et al.*, 2001; Xu *et al.*, 2011), and as a non-viral carriers for reprogramming of stem cells (Jia *et al.*, 2010).

We can reasonably anticipate that in the next decades, the rational design of improved DNA-based biopharmaceuticals, whole synthetic pathways and even genomes will largely profit from the use of standardized biological parts (Lee *et al.*, 2011), as well as computational tools that must take into account important structural aspects that go beyond the simple functionality and assembly.

Concluding remarks

The purpose of this chapter was to provide the reader with a brief view on aspects dealing with genome instability in bacteria, with a focus on spontaneous and stress-induced mutations. Genomes have evolved mechanisms that limit the accumulation of potentially detrimental lethal lesions, while still harnessing beneficial genetic changes to their own profit. Genomes are now known to be highly responsive to the surrounding environment, and capable of using a *potpourri* of strategies to cope with adversity. These include the use of error-prone polymerases, activation of stress-response pathways, downregulation of DNA repair systems, increasing the rate of random mutagenesis, and decreasing the barriers of inter-species gene transfer. In the years to come, a deeper understanding of such mechanisms and the use of improved methods for analysing genetic variation, will pave the way to a more unified view on evolution, as well as to the development of streamlined genomes.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) (project PTDC/EBB-EBI/113650/2009 and Ph.D. grant BD/22320/2005 to P.H.O.) and by the MIT-Portugal Program.

References

Achaz, G., Boyer, F., Rocha, E.P., Viari, A., and Coissac, E. (2007). Repseek, a tool to retrieve approximate repeats from large DNA sequences. Bioinformatics *23*, 119–121.

Alcantara-Diaz, D., Brena-Valle, M., and Serment-Guerrero, J. (2004). Divergent adaptation of *Escherichia coli* to cyclic ultraviolet light exposures. Mutagenesis *19*, 349–354.

Alonso, A., Campanario, E., and Martinez, J.L. (1999). Emergence of multidrug-resistant mutants is increased under antibiotic selective pressure in Pseudomonas aeruginosa. Microbiology *145*, 2857–2862.

Ara, K., Ozaki, K., Nakamura, K., Yamane, K., Sekiguchi, J., and Ogasawara, N. (2007). Bacillus minimum genome factory: effective utilization of microbial genome information. Biotechnol. Appl. Biochem. *46*, 169–178.

Beletskii, A., and Bhagwat, A.S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. *93*, 13919–13924.

Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M., and Hanawalt, P.C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. Proc. Natl. Acad. Sci. U.S.A. *107*, 12816–12821.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Bentchikou, E., Servant, P., Coste, G., and Sommer, S. (2010). A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in Deinococcus radiodurans. PLoS Genet. *6*, e1000774.

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., and Liu, L.F. (1996). A replicational model for DNA recombination between direct repeats. J. Mol. Biol. *256*, 849–858.

Boan, F., and Gomez-Marquez, J. (2010). In vitro recombination mediated by G-quadruplexes. Chembiochem *11*, 331–334.

Boiteux, S., Huisman, O., and Laval, J. (1984). 3-Methyladenine residues in DNA induce the SOS

function sfiA in *Escherichia coli*. EMBO J. *3*, 2569–2573.

Brewer, B.J. (1988). When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell *53*, 679–686.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U.S.A. *98*, 8319–8325.

Cairns, J., Overbaugh, J., and Miller, S. (1988). The origin of mutants. Nature *335*, 142–145.

Capozzi, V., Fiocco, D., Amodio, M.L., Gallone, A., and Spano, G. (2009). Bacterial stressors in minimally processed food. Int. J. Mol. Sci. *10*, 3076–3105.

Cer, R.Z., Bruce, K.H., Mudunuri, U.S., Yi, M., Volfovsky, N., Luke, B.T., Bacolla, A., Collins, J.R., and Stephens, R.M. (2011). Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. Nucleic Acids Res. *39*, D383–391.

Chang, C.H., Chang, Y.C., Underwood, A., Chiou, C.S., and Kao, C.Y. (2007). VNTRDB: a bacterial variable number tandem repeat locus database. Nucleic Acids Res. *35*, D416–421.

Chang, C.W., Christensen, L.V., Lee, M., and Kim, S.W. (2008). Efficient expression of vascular endothelial growth factor using minicircle DNA for angiogenic gene therapy. J. Control Release *125*, 155–163.

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.

Cirz, R.T., Chin, J.K., Andes, D.R., de Crecy-Lagard, V., Craig, W.A., and Romesberg, F.E. (2005). Inhibition of mutation and combating the evolution of antibiotic resistance. PLoS Biol. *3*, e176.

Coenye, T., and Vandamme, P. (2005). Characterization of mononucleotide repeats in sequenced prokaryotic genomes. DNA Res. *12*, 221–233.

Connelly, J.C., Kirkham, L.A., and Leach, D.R. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. Proc. Natl. Acad. Sci. U.S.A. *95*, 7969–7974.

Cooke, M.S., Evans, M.D., Dizdaroglu, M., and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. *17*, 1195–1214.

Cooley, M.B., Carychao, D., Nguyen, K., Whitehand, L., and Mandrell, R. (2010). Effects of environmental stress on stability of tandem repeats in *Escherichia coli* O157:H7. Appl. Environ. Microbiol. *76*, 3398–3400.

Coros, A.M., Twiss, E., Tavakoli, N.P., and Derbyshire, K.M. (2005). Genetic evidence that GTP is required for transposition of IS903 and Tn552 in *Escherichia coli*. J. Bacteriol. *187*, 4598–4606.

Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. Nature *274*, 775–780.

Cranenburgh, R.M., Hanak, J.A., Williams, S.G., and Sherratt, D.J. (2001). *Escherichia coli* strains that allow antibiotic-free plasmid selection and maintenance by repressor titration. Nucleic Acids Res. *29*, E26.

Csorgo, B., Feher, T., Timar, E., Blattner, F.R., and Posfai, G. (2012). Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. Microb. Cell Fact. *11*, 11.

De Septenville, A.L., Duigou, S., Boubakri, H., and Michel, B. (2012). Replication fork reversal after replication-transcription collision. PLoS Genet. *8*, e1002622.

DeMarini, D.M., Shelton, M.L., Abu-Shakra, A., Szakmary, A., and Levine, J.G. (1998). Spectra of spontaneous frameshift mutations at the hisD3052 allele of Salmonella typhimurium in four DNA repair backgrounds. Genetics *149*, 17–36.

Denoeud, F., and Vergnaud, G. (2004). Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. BMC Bioinformatics *5*, 4.

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 mediated by long direct repeats. Mol. Gen. Genet. *228*, 153–159.

Ditlevson, J.V., Tornaletti, S., Belotserkovskii, B.P., Teijeiro, V., Wang, G., Vasquez, K.M., and Hanawalt, P.C. (2008). Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. Nucleic Acids Res. *36*, 3163–3170.

Douki, T., and Cadet, J. (2001). Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. Biochemistry *40*, 2495–2501.

Drablos, F., Feyzi, E., Aas, P.A., Vaagbo, C.B., Kavli, B., Bratlie, M.S., Pena-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H.E. (2004). Alkylation damage in DNA and RNA – repair mechanisms and medical significance. DNA Repair (Amst) *3*, 1389–1407.

Drevinek, P., Baldwin, A., Lindenburg, L., Joshi, L.T., Marchbank, A., Vosahlikova, S., Dowson, C.G., and Mahenthiralingam, E. (2010). Oxidative stress of Burkholderia cenocepacia induces insertion sequence-mediated genomic rearrangements that interfere with macrorestriction-based genotyping. J. Clin. Microbiol. *48*, 34–40.

Duncan, B.K., and Miller, J.H. (1980). Mutagenic deamination of cytosine residues in DNA. Nature *287*, 560–561.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. Cell *146*, 533–543.

Eichenbaum, Z., and Livneh, Z. (1998). UV light induces IS10 transposition in *Escherichia coli*. Genetics *149*, 1173–1181.

Eykelenboom, J.K., Blackwood, J.K., Okely, E., and Leach, D.R. (2008). SbcCD causes a double-strand break at a DNA palindrome in the *Escherichia coli* chromosome. Mol. Cell. *29*, 644–651.

Feschenko, V.V., and Lovett, S.T. (1998). Slipped misalignment mechanisms of deletion formation: analysis of deletion endpoints. J. Mol. Biol. *276*, 559–569.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., *et al.* (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science *269*, 496–512.

Foster, P.L. (1997). Nonadaptive mutations occur on the F' episome during adaptive mutation conditions in *Escherichia coli*. J. Bacteriol. *179*, 1550–1554.

Foster, P.L., and Trimarchi, J.M. (1994). Adaptive reversion of a frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. Science *265*, 407–409.

Frederico, L.A., Kunkel, T.A., and Shaw, B.R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry *29*, 2532–2537.

Gacy, A.M., Goellner, G., Juranic, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. Cell *81*, 533–540.

Galhardo, R.S., Do, R., Yamada, M., Friedberg, E.C., Hastings, P.J., Nohmi, T., and Rosenberg, S.M. (2009). DinB upregulation is the sole role of the SOS response in stress-induced mutagenesis in *Escherichia coli*. Genetics *182*, 55–68.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev. *25*, 2041–2056.

Gates, K.S. (2009). An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. Chem. Res. Toxicol. *22*, 1747–1760.

Goh, S., and Good, L. (2008). Plasmid selection in *Escherichia coli* using an endogenous essential gene marker. BMC Biotechnol. *8*, 61.

Gonzalez, C., Hadany, L., Ponder, R.G., Price, M., Hastings, P.J., and Rosenberg, S.M. (2008). Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in *Escherichia coli*. PLoS Genet. *4*, e1000208.

Han, H., and Hurley, L.H. (2000). G-quadruplex DNA: a potential target for anti-cancer drug design. Trends Pharmacol. Sci. *21*, 136–142.

Hardin, A., Villalta, C.F., Doan, M., Jabri, M., Chockalingham, V., White, S.J., and Fowler, R.G. (2007). A molecular characterization of spontaneous frameshift mutagenesis within the trpA gene of *Escherichia coli*. DNA Repair (Amst) *6*, 177–189.

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.

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. U.S.A. *93*, 11121–11125.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

Ilves, H., Horak, R., and Kivisaar, M. (2001). Involvement of sigma(S) in starvation-induced transposition of Pseudomonas putida transposon Tn4652. J. Bacteriol. *183*, 5445–5448.

Jain, A., Wang, G., and Vasquez, K.M. (2008). DNA triple helices: biological consequences and therapeutic potential. Biochimie *90*, 1117–1130.

Jenjaroenpun, P., and Kuznetsov, V.A. (2009). TTS mapping: integrative WEB tool for analysis of triplex formation target DNA sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome. BMC Genomics *10* (Suppl 3), S9.

Jia, F., Wilson, K.D., Sun, N., Gupta, D.M., Huang, M., Li, Z., Panetta, N.J., Chen, Z.Y., Robbins, R.C., Kay, M.A., *et al.* (2010). A nonviral minicircle vector for deriving human iPS cells. Nat. Methods *7*, 197–199.

Jiang, X., Yang, J., Zhang, H., Zou, H., Wang, C., and Xian, M. (2012). In vitro assembly of multiple DNA fragments using successive hybridization. PLoS One *7*, e30267.

Jolivet-Gougeon, A., David-Jobert, S., Tamanai-Shacoori, Z., Menard, C., and Cormier, M. (2000). Osmotic stress-induced genetic rearrangements in *Escherichia coli* H10407 detected by randomly amplified polymorphic DNA analysis. Appl. Environ. Microbiol. *66*, 5484–5487.

Kahn, J.D., and Crothers, D.M. (1993). DNA bending in transcription initiation. Cold Spring Harb. Symp. Quant. Biol. *58*, 115–122.

Kang, J.M., Iovine, N.M., and Blaser, M.J. (2006). A paradigm for direct stress-induced mutation in prokaryotes. FASEB J. *20*, 2476–2485.

Kato, J., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. Mol. Syst. Biol. *3*, 132.

Kawe, M., Horn, U., and Pluckthun, A. (2009). Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors. Microb. Cell Fact. *8*, 8.

Kha, D.T., Wang, G., Natrajan, N., Harrison, L., and Vasquez, K.M. (2010). Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. J. Mol. Biol. *398*, 471–480.

Kim, M., O'Callaghan, P.M., Droms, K.A., and James, D.C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol. Bioeng. [Epub ahead of print]

Kobori, S., Ko, Y., and Kato, M. (2009). A target site for spontaneous insertion of IS10 element in pUC19 DNA located within intrinsically bent DNA. Open Microbiol. J. *3*, 146–150.

Kohanski, M.A., DePristo, M.A., and Collins, J.J. (2010). Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. Mol. Cell. *37*, 311–320.

Kowalski, D., Natale, D.A., and Eddy, M.J. (1988). Stable DNA unwinding, not 'breathing,' accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. Proc. Natl. Acad. Sci. U.S.A. *85*, 9464–9468.

Kunkel, T.A., and Soni, A. (1988). Mutagenesis by transient misalignment. J. Biol. Chem. *263*, 14784–14789.

Kurtz, S., and Schleiermacher, C. (1999). REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics *15*, 426–427.

Larson, K., Sahm, J., Shenkar, R., and Strauss, B. (1985). Methylation-induced blocks to *in vitro* DNA replication. Mutat. Res. *150*, 77–84.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc. Natl. Acad. Sci. U.S.A. *109*, E2774–2783.

Lee, J.H., Sung, B.H., Kim, M.S., Blattner, F.R., Yoon, B.H., Kim, J.H., and Kim, S.C. (2009). Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb. Cell Fact. *8*, 2.

Lee, T.S., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., and Keasling, J.D. (2011). BglBrick vectors and datasheets: A synthetic biology platform for gene expression. J. Biol. Eng. *5*, 12.

Lewis, L.A., Astatke, M., Umekubo, P.T., Alvi, S., Saby, R., Afrose, J., Oliveira, P.H., Monteiro, G.A., and Prazeres, D.M. (2012). Protein–DNA interactions define the mechanistic aspects of circle formation and insertion reactions in IS2 transposition. Mob. DNA *3*, 1.

Li, Y.C., Korol, A.B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. *21*, 991–1007.

Lilley, D.M. (1985). The kinetic properties of cruciform extrusion are determined by DNA base-sequence. Nucleic Acids Res. *13*, 1443–1465.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. Prog. Nucleic Acid Res. Mol. Biol. *22*, 135–192.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Linial, M., and Shlomai, J. (1988). Bent DNA structures associated with several origins of replication are recognized by a unique enzyme from trypanosomatids. Nucleic Acids Res. *16*, 6477–6492.

Loeb, L.A., and Preston, B.D. (1986). Mutagenesis by apurinic/apyrimidinic sites. Annu. Rev. Genet. *20*, 201–230.

Loechler, E.L., Green, C.L., and Essigmann, J.M. (1984). In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. Proc. Natl. Acad. Sci. U.S.A. *81*, 6271–6275.

Lombardo, M.J., Aponyi, I., and Rosenberg, S.M. (2004). General stress response regulator RpoS in adaptive mutation and amplification in *Escherichia coli*. Genetics *166*, 669–680.

Lopez, E., and Blazquez, J. (2009). Effect of subinhibitory concentrations of antibiotics on intrachromosomal homologous recombination in *Escherichia coli*. Antimicrob. Agents Chemother. *53*, 3411–3415.

Lopez, E., Elez, M., Matic, I., and Blazquez, J. (2007). Antibiotic-mediated recombination: ciprofloxacin stimulates SOS-independent recombination of divergent sequences in *Escherichia coli*. Mol. Microbiol. *64*, 83–93.

Lopez-Fuertes, L., Perez-Jimenez, E., Vila-Coro, A.J., Sack, F., Moreno, S., Konig, S.A., Junghans, C., Wittig, B., Timon, M., and Esteban, M. (2002). DNA vaccination with linear minimalistic (MIDGE) vectors confers protection against Leishmania major infection in mice. Vaccine *21*, 247–257.

Lovett, S.T., Drapkin, P.T., Sutera, 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., 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. U.S.A. *93*, 7120–7124.

Mackwan, R.R., Carver, G.T., Kissling, G.E., Drake, J.W., and Grogan, D.W. (2008). The rate and character of spontaneous mutation in Thermus thermophilus. Genetics *180*, 17–25.

Maki, H., and Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. Nature *355*, 273–275.

Mandel, M.J., and Silhavy, T.J. (2005). Starvation for different nutrients in *Escherichia coli* results in differential modulation of RpoS levels and stability. J. Bacteriol. *187*, 434–442.

Marie, C., Vandermeulen, G., Quiviger, M., Richard, M., Preat, V., and Scherman, D. (2010). pFARs, plasmids free of antibiotic resistance markers, display high-level transgene expression in muscle, skin and tumour cells. J. Gene Med. *12*, 323–332.

Martin, P., Makepeace, K., Hill, S.A., Hood, D.W., and Moxon, E.R. (2005). Microsatellite instability regulates transcription factor binding and gene expression. Proc. Natl. Acad. Sci. U.S.A. *102*, 3800–3804.

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.

McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26.

Merrikh, H., Machon, C., Grainger, W.H., Grossman, A.D., and Soultanas, P. (2011). Co-directional replication-transcription conflicts lead to replication restart. Nature *470*, 554–557.

Michaels, M.L., Cruz, C., Grollman, A.P., and Miller, J.H. (1992). Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc. Natl. Acad. Sci. U.S.A. *89*, 7022–7025.

Mirkin, E.V., and Mirkin, S.M. (2005). Mechanisms of transcription-replication collisions in bacteria. Mol. Cell. Biol *25*, 888–895.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication fork stalling at natural impediments. Microbiol. Mol. Biol. Rev. *71*, 13–35.

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.

Moreno, S., Lopez-Fuertes, L., Vila-Coro, A.J., Sack, F., Smith, C.A., Konig, S.A., Wittig, B., Schroff, M., Juhls, C., Junghans, C., *et al.* (2004). DNA immunization with minimalistic expression constructs. Vaccine *22*, 1709–1716.

Mrazek, J., Guo, X., and Shah, A. (2007). Simple sequence repeats in prokaryotic genomes. Proc. Natl. Acad. Sci. U.S.A. *104*, 8472–8477.

Naas, T., Blot, M., Fitch, W.M., and Arber, W. (1994). Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics *136*, 721–730.

Naas, T., Blot, M., Fitch, W.M., and Arber, W. (1995). Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. Mol. Biol. Evol. *12*, 198–207.

Navarro Llorens, J.M., Tormo, A., and Martinez-Garcia, E. (2010). Stationary phase in Gram-negative bacteria. FEMS Microbiol. Rev. *34*, 476–495.

Obolski, U., and Hadany, L. (2012). Implications of stress-induced genetic variation for minimizing multidrug resistance in bacteria. BMC Med. *10*, 89.

Ohtsubo, Y., Genka, H., Komatsu, H., Nagata, Y., and Tsuda, M. (2005). High-temperature-induced transposition of insertion elements in burkholderia multivorans ATCC 17616. Appl. Environ. Microbiol. *71*, 1822–1828.

Oliveira, P.H., Lemos, F., Monteiro, G.A., and Prazeres, D.M.F. (2008). Recombination frequency in plasmid DNA containing direct repeats – predictive correlation with repeat and intervening sequence length. Plasmid *60*, 159–165.

Oliveira, P.H., Prather, K.J., Prazeres, D.M.F., and Monteiro, G.A. (2010). Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. Appl. Microbiol. Biotechnol. *87*, 2157–2167.

Oliveira, P.H., Prather, K.L., Prazeres, D.M.F., and Monteiro, G.A. (2011). Mutation detection in plasmid-based biopharmaceuticals. Biotechnol. J. *6*, 378–391.

Oliveira, P.H., Prazeres, D.M.F., and Monteiro, G.A. (2009). Deletion formation mutations in plasmid expression vectors are unfavored by runaway amplification conditions and differentially selected under kanamycin stress. J. Biotechnol. *143*, 231–238.

Oliveira, P.H., Prazeres, D.M.F., and Monteiro, G.A. (2012). Structural and segregational instability in plasmid biology. In Plasmids: Genetics, Applications and Health, Gonzalez, F.E.R., ed. (New York, Nova Science Publishers), pp. 79–99.

Omont, N., and Kepes, F. (2004). Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. Bioinformatics *20*, 2719–2725.

Pedro, L., Banos, R.C., Aznar, S., Madrid, C., Balsalobre, C., and Juarez, A. (2011). Antibiotics shaping bacterial genome: deletion of an IS91 flanked virulence determinant upon exposure to subinhibitory antibiotic concentrations. PLoS One *6*, e27606.

Petrosino, J.F., Galhardo, R.S., Morales, L.D., and Rosenberg, S.M. (2009). Stress-induced beta-lactam antibiotic resistance mutation and sequences of stationary-phase mutations in the *Escherichia coli* chromosome. J. Bacteriol. *191*, 5881–5889.

Pettijohn, D., and Hanawalt, P. (1964). Evidence for repair-replication of ultraviolet damaged DNA in bacteria. J. Mol. Biol. *9*, 395–410.

Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., De Luca, K., Speck, D., and Sodoyer, R. (2010). Antibiotic-free selection in E. coli: new considerations for optimal design and improved production. Microb. Cell Fact. *9*, 65.

Pfaffenzeller, I., Mairhofer, J., Striedner, G., Bayer, K., and Grabherr, R. (2006). Using ColE1-derived RNA I for suppression of a bacterially encoded gene: implication for a novel plasmid addiction system. Biotechnol. J. *1*, 675–681.

Pomerantz, R.T., and O'Donnell, M. (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. Nature *456*, 762–766.

Pomerantz, R.T., and O'Donnell, M. (2010). Direct restart of a replication fork stalled by a head-on RNA polymerase. Science *327*, 590–592.

Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. Mol. Cell. *19*, 791–804.

Posfai, G., Plunkett, G. 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., *et al.* (2006). Emergent properties of reduced-genome *Escherichia coli*. Science *312*, 1044–1046.

Powell, S.C., and Wartell, R.M. (2001). Different characteristics distinguish early versus late arising adaptive mutations in *Escherichia coli* FC40. Mutat. Res. *473*, 219–228.

Prado, F., and Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J. *24*, 1267–1276.

Price, M.N., Alm, E.J., and Arkin, A.P. (2005). Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. Nucleic Acids Res. *33*, 3224–3234.

Prieto, A.I., Ramos-Morales, F., and Casadesus, J. (2004). Bile-induced DNA damage in Salmonella enterica. Genetics *168*, 1787–1794.

Prieto, A.I., Ramos-Morales, F., and Casadesus, J. (2006). Repair of DNA damage induced by bile salts in Salmonella enterica. Genetics *174*, 575–584.

Rao, J.E., Miller, P.S., and Craig, N.L. (2000). Recognition of triple-helical DNA structures by transposon Tn7. Proc. Natl. Acad. Sci. U.S.A. *97*, 3936–3941.

Rawal, P., Kummarasetti, V.B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K., and Chowdhury, S. (2006). Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. Genome Res. *16*, 644–655.

Ribeiro, S.C., Oliveira, P.H., Prazeres, D.M., and Monteiro, G.A. (2008). High frequency plasmid recombination mediated by 28 bp direct repeats. Mol. Biotechnol. *40*, 252–260.

Riehle, M.M., Bennett, A.F., and Long, A.D. (2001). Genetic architecture of thermal adaptation in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. *98*, 525–530.

Riesenfeld, C., Everett, M., Piddock, L.J., and Hall, B.G. (1997). Adaptive mutations produce resistance to ciprofloxacin. Antimicrob. Agents Chemother. *41*, 2059–2060.

Rocha, E.P., and Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet. *34*, 377–378.

Rosenberg, S.M. (2001). Evolving responsively: adaptive mutation. Nat. Rev. Genet. *2*, 504–515.

Rosenberg, S.M., Shee, C., Frisch, R.L., and Hastings, P.J. (2012). Stress-induced mutation via DNA breaks in *Escherichia coli*: A molecular mechanism with implications for evolution and medicine. Bioessays *34*, 885–892.

Rudolph, C.J., Dhillon, P., Moore, T., and Lloyd, R.G. (2007). Avoiding and resolving conflicts between DNA replication and transcription. DNA Repair (Amst) *6*, 981–993.

Rydberg, B., and Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J. *1*, 211–216.

Schaaper, R.M., and Dunn, R.L. (1991). Spontaneous mutation in the *Escherichia coli* lacI gene. Genetics *129*, 317–326.

Schaaper, R.M., Dunn, R.L., and Glickman, B.W. (1987). Mechanisms of ultraviolet-induced mutation. Mutational spectra in the *Escherichia coli* lacI gene for a wild-type and an excision-repair-deficient strain. J. Mol. Biol. *198*, 187–202.

Schakowski, F., Gorschluter, M., Junghans, C., Schroff, M., Buttgereit, P., Ziske, C., Schottker, B., Konig-Merediz, S.A., Sauerbruch, T., Wittig, B., *et al.* (2001). A novel minimal-size vector (MIDGE) improves transgene expression in colon carcinoma cells and avoids transfection of undesired DNA. Mol. Ther. *3*, 793–800.

Shapiro, J.A. (1984). Observations on the formation of clones containing araB-lacZ cistron fusions. Mol. Gen. Genet. *194*, 79–90.

Sharma, S. (2011). Non-B DNA Secondary Structures and Their Resolution by RecQ Helicases. J. Nucleic Acids *2011*, 724215.

Sharma, S., Doherty, K.M., and Brosh, R.M., Jr. (2006). Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. Biochem. J. *398*, 319–337.

Shee, C., Gibson, J.L., and Rosenberg, S.M. (2012). Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. Cell Rep. *2*, 714–721.

Shen, J.C., Rideout, W.M. 3rd, and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. *22*, 972–976.

Shibutani, S., Takeshita, M., and Grollman, A.P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature *349*, 431–434.

Shrivastav, N., Li, D., and Essigmann, J.M. (2010). Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. Carcinogenesis *31*, 59–70.

Slack, A., Thornton, P.C., Magner, D.B., Rosenberg, S.M., and Hastings, P.J. (2006). On the mechanism of gene amplification induced under stress in *Escherichia coli*. PLoS Genet. *2*, e48.

Sleight, S.C., Orlic, C., Schneider, D., and Lenski, R.E. (2008). Genetic basis of evolutionary adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth. Genetics *180*, 431–443.

Spira, B., de Almeida Toledo, R., Maharjan, R.P., and Ferenci, T. (2011). The uncertain consequences of transferring bacterial strains between laboratories – rpoS instability as an example. BMC Microbiol. *11*, 248.

Stivers, J.T., and Jiang, Y.L. (2003). A mechanistic perspective on the chemistry of DNA repair glycosylases. Chem. Rev. *103*, 2729–2759.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb. Symp. Quant. Biol. *31*, 77–84.

Sun, S., Ke, R., Hughes, D., Nilsson, M., and Andersson, D.I. (2012). Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. PLoS One *7*, e42639.

Taddei, F., Halliday, J.A., Matic, I., and Radman, M. (1997). Genetic analysis of mutagenesis in aging *Escherichia coli* colonies. Mol. Gen. Genet. *256*, 277–281.

Tajiri, T., Maki, H., and Sekiguchi, M. (1995). Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*. Mutat. Res. *336*, 257–267.

Taverna, P., and Sedgwick, B. (1996). Generation of an endogenous DNA-methylating agent by nitrosation in *Escherichia coli*. J. Bacteriol. *178*, 5105–5111.

Tindall, K.R., Stein, J., and Hutchinson, F. (1988). Changes in DNA base sequence induced by gamma-ray mutagenesis of lambda phage and prophage. Genetics *118*, 551–560.

Torkelson, J., Harris, R.S., Lombardo, M.J., Nagendran, J., Thulin, C., and Rosenberg, S.M. (1997). Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation. EMBO J. *16*, 3303–3311.

Treangen, T.J., Darling, A.E., Achaz, G., Ragan, M.A., Messeguer, X., and Rocha, E.P. (2009). A novel heuristic for local multiple alignment of interspersed DNA repeats. IEEE/ACM Trans. Comput. Biol. Bioinform. *6*, 180–189.

Twiss, E., Coros, A.M., Tavakoli, N.P., and Derbyshire, K.M. (2005). Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress. Mol. Microbiol. *57*, 1593–1607.

Tyrrell, R.M. (1973). Induction of pyrimidine dimers in bacterial DNA by 365 nm radiation. Photochem. Photobiol. *17*, 69–73.

Umenhoffer, K., Feher, T., Baliko, G., Ayaydin, F., Posfai, J., Blattner, F.R., and Posfai, G. (2010). Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. Microb. Cell Fact. *9*, 38.

Vandermeulen, G., Marie, C., Scherman, D., and Preat, V. (2011). New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials. Mol. Ther. *19*, 1942–1949.

Vilette, D., Ehrlich, S.D., and Michel, B. (1996). Transcription-induced deletions in plasmid vectors: M13 DNA replication as a source of instability. Mol. Gen. Genet. *252*, 398–403.

Vinces, M.D., Legendre, M., Caldara, M., Hagihara, M., and Verstrepen, K.J. (2009). Unstable tandem repeats in promoters confer transcriptional evolvability. Science *324*, 1213–1216.

Voineagu, I., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2008). Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. Proc. Natl. Acad. Sci. U.S.A. *105*, 9936–9941.

Wang, G., Zhao, J., and Vasquez, K.M. (2009). Methods to determine DNA structural alterations and genetic instability. Methods *48*, 54–62.

Wyrzykowski, J., and Volkert, M.R. (2003). The *Escherichia coli* methyl-directed mismatch repair system repairs base pairs containing oxidative lesions. J. Bacteriol. *185*, 1701–1704.

Xu, B.L., Yuan, L., Wu, J.X., Xu, N., Fang, W.J., Zhao, P., and Huang, W.L. (2011). Mc-hES, a novel plasmid carrying human endostatin gene, inhibits nasopharyngeal carcinoma growth. Cancer Gene Ther. *19*, 110–117.

Yeiser, B., Pepper, E.D., Goodman, M.F., and Finkel, S.E. (2002). SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. Proc. Natl. Acad. Sci. U.S.A. *99*, 8737–8741.