Secondary structures as predictors of mutation potential in the *lacZ* gene of *Escherichia coli*

Evan Burkala,¹ Jacqueline M. Reimers,² Karen H. Schmidt,¹ Nick Davis,¹ Ping Wei³ and Barbara E. Wright¹

Correspondence

Barbara E. Wright
barbara.wright@mso.umt.edu

¹Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

²EndoBiologics Inc., Missoula, MT 59808, USA

³University of Massachusetts, Worcester, MA 01605, USA

Four independent nonsense mutations were engineered into the *Escherichia coli* chromosomal *lacZ* gene, and reversion rates back to LacZ⁺ phenotypes were determined. The mutation potential of bases within putative DNA secondary structures formed during transcription was predicted by a sliding-window analysis that simulates successive folding of the ssDNA creating these structures. The relative base mutabilities predicted by the MFG computer program correlated with experimentally determined reversion rates in three of the four mutants analysed. The nucleotide changes in revertants at one nonsense codon site consisted of a triple mutation, presumed to occur by a templated repair mechanism. Additionally, the effect of supercoiling on mutation was investigated and, in general, reversion rates increased with higher levels of negative supercoiling. Evidence indicates that predicted secondary structures are in fact formed *in vivo* and that directed mutation in response to starvation stress is dependent upon the exposure of particular bases, the stability of the structures in which these bases are unpaired and the level of DNA supercoiling within the cell.

Received 22 December 2006

Revised 7 March 2007

Accepted 20 March 2007

INTRODUCTION

DNA mutations that occur in *Escherichia coli* under stress conditions (e.g. starvation) arise rapidly in derepressed genes, and most mutations occur during stationary phase when cell division and replication are halted (Cairns & Foster, 1991). The derepression and transcription of specific genes responsive to environmental stressors set the stage for mutagenic events to occur at sites that enable the organism to overcome the metabolic blockage and survive. Transcriptional activation as a mechanism for increasing mutation rates was first proposed in 1971 (Brock, 1971; Herman & Dworkin, 1971). Specifically induced transcription-enhanced mutations have since been described in *Saccharomyces cerevisiae* (Datta & Jinks-Robertson, 1995) and in starvation-induced derepressed genes of both *E. coli* (Wright, 1996; Wright *et al.*, 1999) and *Bacillus subtilis* (Rudner *et al.*, 1999). A direct dependence of mutation rate on the level of transcription has been demonstrated based upon mRNA accumulation and half-lives (Reimers *et al.*, 2004; Wright *et al.*, 1999). Such correlations of transcription and mutation have also been

documented for somatic hypermutation, which is triggered by antigen challenge during the immune response (Bachl *et al.*, 2001; Fukita *et al.*, 1998).

Transcription-directed mutagenesis occurs by the formation of ssDNA segments caused by transcription and supercoiling, the formation of stem-loop structures (SLSs) from these segments and the mutation of unpaired bases within these structures. The differential exposure of bases in secondary structures can explain why bases within the same gene have different mutation rates (Ripley & Glickman, 1983). Transcription drives localized supercoiling (Liu & Wang, 1987; Pruss & Drlica, 1986), and supercoiling causes both the formation and the stabilization of SLSs (Dayn *et al.*, 1992; Zheng *et al.*, 1991). SLSs are created in supercoiled DNA, in the wake of the transcription complex, as a result of the proximity in ssDNA of sequence segments that bond to form the stems of secondary structures. Higher levels of transcription initiated by starvation and gene derepression increase localized supercoiling and, consequently, increase the exposure of unpaired bases within the SLSs (Balke & Gralla, 1987; Krasilnikov *et al.*, 1999; Opel & Hatfield, 2001). Unpaired bases in these secondary structures are susceptible to mutation because of their intrinsic thermodynamic instability and their availability to nucleotide-altering enzymes and genotoxins (reviewed by Lindahl, 1993; Singer & Kusmierek, 1982; Wright *et al.*, 2002).

Abbreviations: acMI, average codon mutability index; MI, mutability index; SLS, stem-loop structure.

A figure showing a comparison of mutation rates for each supercoiling strain and mutant over time is available with the online version of this paper.

The mutability of a base during transcription is primarily determined by the extent to which it is unpaired. Therefore, the MFG computer program was designed to predict the intrinsic mutability of a base in its most stable SLSs in which it is unpaired (Wright *et al.*, 2003). MFG performs a sliding-window analysis of ssDNA using a chosen length of nucleotides that is folded successively by MFOLD (Markham & Zuker, 2005), which reports all SLSs that can form from any given sequence in the order of their stability. For each base in the sequence, MFG chooses the most stable SLS in which the base is unpaired and calculates the relative mutability index (MI) of the unpaired base. The MI is defined as the product of two key values: (a) the stability of the most stable SLS in which the base is unpaired and (b) the percentage of total folds in which it is unpaired during transcription.

In previous studies of pre-existing mutations, the MIs predicted by MFG have correlated well with experimentally determined mutation rates (Reimers *et al.*, 2004; Wright *et al.*, 2002, 2003, 2004). More recently, MFG predictions of MIs have also correlated with reversion rates for mutations engineered into specific contexts within the chloramphenicol-resistance gene using a plasmid system (Schmidt *et al.*, 2006). In the present study, reversion rates were determined in mutations placed in different mutational contexts within the chromosomal *lacZ* gene. Using MFG to predict base mutability, two mutations were placed at locations predicted to be highly unpaired within the same stable SLS. Two other mutations were placed in stems of SLSs in which these bases were predicted to be paired and not mutable. These mutations allowed investigations into the influence of different specific contexts on mutability, and also allowed comparisons of mutation rates determined experimentally and base mutability predicted by MFG.

The role of supercoiling was also examined in this study of four chromosomal mutations engineered into three *E. coli* strains with differing supercoiling levels. Reversion rates of mutations in these strains were determined in order to observe the effect of chromosomal supercoiling levels on reversion rates. This work confirms the participation of supercoiling in chromosomal DNA mutational events, since strains with increased negative supercoiling had higher mutation rates.

METHODS

Mutability analysis. The MFG program was created to predict the intrinsic mutability of unpaired bases in DNA secondary structures formed during transcription (Wright *et al.*, 2003). The relative MI of a base is calculated by multiplying the Gibbs free energy ($-\Delta G$) value of the most stable secondary structure in which the base is unpaired by the proportion (percentage) of total structures in which that base is unpaired during transcription. The MFG program is available at <http://biology.dbs.umt.edu/wright/upload/MFG.html>. MIs for particular bases within the *lacZ* gene (accession no. CG00850, NC_000913) were calculated using the sequence of the non-transcribed (sense) strand.

Mutagenesis. Mutations were initially introduced into the *lacZ* gene by 'megaprimer PCR' (Tyagi *et al.*, 2004), in which a small 400 bp PCR fragment was amplified with a primer containing the desired mutated sites, and the purified product was used as the primer for amplification of an approximately 1400 bp fragment incorporating the mutated site(s). Mutated *lacZ* PCR products were ligated into the pBluescript-SK(+) plasmid (Stratagene), which had been cut with appropriate restriction enzymes, and it was used for subsequent mutation of the bacterial genome *lacZ*. The subsequent pBluescript-*lacZ* mutant plasmid is referred to as the 'donor plasmid'. Gene replacement of chromosomal *lacZ* was performed using the two-plasmid 'gene gorging' procedure of Herring *et al.* (2003). The donor plasmid containing the engineered *lacZ* gene mutation was used to transform *E. coli* strains that had been previously transformed with the pACBSR-*SceI* mutagenesis plasmid, provided by Scarab Genomics. The pACBSR-*SceI* plasmid was originally isolated from pSCM525 and was a gift from Dr B. Dujon (Institut Pasteur, Paris, France); it expresses the I-*SceI* homing endonuclease and lambda Red genes necessary for gene replacement (Colleaux *et al.*, 1986). After induction of the gene replacement double recombination event between donor plasmid and the chromosomal *lacZ* gene, bacteria were screened for inability to use lactose as a carbon source (i.e. white on lactose MacConkey agar and no growth on minimal lactose medium) and the loss of plasmid-encoded antibiotic resistance, ensuring introduction of the mutation into chromosomal DNA. Candidate *LacZ*⁻ colonies were screened by PCR for the *lacZ* gene and sequenced to ensure correct mutation of the gene. All mutations were verified before gene gorging by sequencing donor plasmids, and after gene gorging by sequencing PCR products generated from genomic DNA using an ABI 3130xl Genetic Analyzer. To ensure observed mutations were not generated by PCR mutagenesis, reactions used high fidelity Pfu Turbo DNA polymerase (Stratagene).

Bacterial strains. Three strains of *E. coli* that have mutations of genes that affect supercoiling levels of DNA (JTT1, RS2 and SD7) were used to assess the effect of supercoiling on mutation rate. These strains have been described by Pruss *et al.* (1982), and were acquired from the *E. coli* Genetic Stock Center. The genotypes of the three strains used are: JTT1 *gal-25 λ^R pyrF287 fnr-1 rpsL195 iclR7 trpR72*; RS2 is isogenic to JTT1 but also has *topA10*; SD7 is isogenic to JTT1, but has both *topA10* and *gyrB226*.

Reversion rate analysis. Mutation rates were calculated using the classic method of Luria & Delbrück (1943). Briefly, bacteria were grown in individual cultures of complete liquid medium until glucose was depleted and there was no increase in OD₅₅₀ for 1 h. Bacteria were then plated onto solid minimal medium containing a new carbon source, lactose, to further activate the transcription of the *lac* gene and provide optimal conditions for transcription-mediated mutation (reversion). Only those cells that reverted to *LacZ*⁺ formed colonies on the lactose plates. The ratio of plates with revertants (i.e. those able to use lactose as a carbon source) to total plates was used to calculate reversion rate per generation (μ):

$$\mu = (-\ln[2]) / (\ln[P_0/N])$$

where N is the mean number of cells per culture and P_0 is the proportion of cultures without revertants. Reversion rate experiments were repeated four or more times. For all experiments, random representative colonies of revertants were sequenced to determine the nature of reversion.

Supercoiling. To confirm that each strain used in mutation assays had the predicted level of DNA supercoiling, we performed supercoiling assays as described by Mojica *et al.* (1994). Supercoiling determinations were made on cell lines representing the highest and lowest reversion rates among the engineered mutants. Mutated strains were transformed with a standard plasmid (pBR322), and

late-exponential-phase cultures (250 ml) were centrifuged and transferred to 100 ml starvation medium formulated to mimic the plating conditions for reversion assays. Plasmid DNA was extracted at 2 and 10 min intervals post-starvation from 50 ml culture using a Wizard miniprep kit (Promega). Topoisomers (1 µg total) were run on a 0.7% (w/v) agarose gel with 1 × TAE running buffer, containing 5 µg ml⁻¹ chloroquine, in a circulating electrophoresis apparatus. Gels were washed four times with water for 1 h, post-stained with ethidium bromide, photographed and scanned using a Bio-Rad GelDoc 1000 densitometer to determine peak band intensity.

RESULTS

Analysis of mutation potential by MFG involves examination of secondary structures

To study some of the intrinsic factors that can increase the reversion rate of *E. coli* mutants, nonsense mutations were introduced at selected sites within the *lacZ* gene of chromosomal DNA. Following MFG analyses using several window sizes to form *lacZ* secondary structures, a window size of 40 nt from the non-transcribed strand sequence was chosen. Appropriate sites in SLSs were identified where it was possible to create nonsense codons by making a single nucleotide change. Four sites were chosen, representing four different mutability contexts within predicted secondary structures. Two mutant bases are highly unpaired in two different loops of the same very stable SLS (Fig. 1a, c), and one mutant base was placed in the stem of the same SLS (Fig. 1e). The last mutation was placed 1500 bp downstream, in the stem of a structure predicted to have lower stability (Fig. 1i). The LacZ1 mutation was located within an internal loop formed by a mismatch with opposing bases (Fig. 1a). The LacZ2 mutation (Fig. 1c) was located in the apical loop of this structure and the LacZ3 mutant codon (Fig. 1e) was located within a stem next to the LacZ1 internal loop. Since the predicted structure modified to create LacZ1–3 mutants had an atypically high $-\Delta G$ value, it was useful for comparative purposes to engineer another mutation (LacZ4) in a SLS with lower stability, and hence a lower predicted base mutability (Fig. 1i). In the LacZ3 and LacZ4 engineered mutations, it was also necessary to mutate the pairing partner base in order to maintain the paired stem and thus the stability of the secondary structure in which these mutations were located.

Analysis by MFG gave the subset of structures predicted to be the most stable SLSs in which each mutant base is unpaired (Fig. 1, right). In the case of LacZ1 and LacZ2, these codons were expected to be the most unpaired and mutable in the same stable structures (Fig. 1b, d, respectively). The mutated base T of LacZ3, as well as the other two bases of this codon, are paired (Fig. 1e). The most stable SLS in which the T is unpaired is shown in Fig. 1(f); this SLS has a shorter stem than LacZ3 and also a lower stability (-8.1 kcal mol⁻¹). The middle position (A) of the stop codon in LacZ3 is shown in the most stable structure in which it is unpaired (-12.7 kcal mol⁻¹) in

Fig. 1(g). The final A in this codon has a different most stable structure in which it is unpaired (-10.4 kcal mol⁻¹; Fig. 1h). The MFOLD program predicted that the most stable structure for the LacZ4 engineered mutant was one in which all three bases of the codon are paired in the same structure (Fig. 1i); the most stable structure in which each base is unpaired (Fig. 1j) was slightly less stable (-4.1 and -3.5 kcal mol⁻¹, respectively). The LacZ1 mutant was predicted to be most unpaired, followed closely by LacZ2; LacZ3 was intermediate; and LacZ4 was predicted to be the least unpaired (Table 1).

MFG uses percentage unpaired and stability values to predict relative base MIs. However, as MFG calculates one MI for each base (at a given window size), and since each base within the nonsense codon can mutate to create a functional codon, the three MIs of individual bases within a codon are averaged to produce the 'average codon mutability index' (acMI) (Schmidt *et al.*, 2006). As summarized in Table 1, the engineered LacZ1 nonsense codon was predicted to have the highest intrinsic mutation potential (acMI=13.9); LacZ2 was also predicted to be high (acMI=12.8). The LacZ3 codon had an intermediate acMI of 6.2 and LacZ4 had a very low acMI of 0.8.

Reversion rates of *lacZ* mutants

Reversion rates were determined by plating a known number of glucose-starved LacZ⁻ cells onto multiple plates containing lactose as the only carbon source. Growth of LacZ⁺ revertant colonies was observed over time, at 48, 72 and 96 h post-plating. Cells able to utilize lactose as a carbon source were defined as phenotypic revertants (LacZ⁺), regardless of the genetic nature of the reversion. LacZ1 showed the highest reversion rate of all four mutants. LacZ3 had a reversion rate approximately four-fold lower than LacZ1, and LacZ4 reversion rates were approximately 10-fold lower than those of LacZ3 (Table 1, Fig. 2). Thus, a qualitative correlation was seen between observed mutation rates and predicted acMIs for the engineered nonsense codons LacZ1, 3 and 4. However, for all strains, reversion rates observed for LacZ2 were lower than that predicted by MFG.

Reversion rates and supercoiling

The effect of chromosomal supercoiling on reversion rate was examined using three strains of *E. coli* with differing levels of supercoiling. Of the strains used in this study, one (RS2) has a reported deficiency in the topoisomerase I gene (*topA*), resulting in a higher level of negative supercoiling, and another strain (SD7) has both *topA* and gyrase gene (*gyrB*) mutations, resulting in less negative supercoiling than the corresponding wild-type *E. coli* JTT1 (Pruss *et al.*, 1982). Each *lacZ* mutation was introduced into the chromosomal DNA of each of the three parental supercoiling strains. Supercoiling phenotypes were later confirmed in these strains by introducing plasmid DNA and

Fig. 1. SLs of *lacZ* predicted by MFOLD and MFG. (a), (c), (e) and (i) The most energetically stable structures predicted by MFOLD for each 40 nt sequence containing an engineered mutation: (a) LacZ1, (c) LacZ2, (e) LacZ3 and (i) LacZ4. (b), (d), (f), (g), (h) and (j) The structures predicted by MFG to be the most stable structures in which each base within each engineered codon is unpaired and, therefore, the most mutable. Bold nucleotides represent all engineered positions; bracketed bases indicate the nonsense codon for each mutant. Arrows indicate the location of each unpaired mutable base in the structure in which it is the most stable. ΔG values are given within each structure, and the numbers exterior to each structure indicate the nucleotide positions within the *lacZ* gene.

Table 1. Engineered mutations in the chromosomal *lacZ* gene and reversion rates compared with base mutabilities predicted by MFG

Mutant (nt position)*	Codon mutation†	Nucleotide†	Fig. 1 SLS	Unpaired (%)	−ΔG (kcal mol ^{−1})	Base MI	acMI‡	10 ^{−9} ×Reversion rate§
LacZ1 (188)	TGG→TAG	T	b	100	13.9	13.9	13.9	3.91 ± 1.66
		A	b	100	13.9	13.9		
		G	b	100	13.9	13.9		
LacZ2 (202)	GAA→TAA	T	d	100	13.7	13.7	12.8	0.83 ± 0.27
		A	d	90	13.7	12.3		
		A	d	90	13.7	12.3		
LacZ3 (214)	GAA→TAA	T	f	52	8.1	4.3	6.2	0.88 ± 0.34
		A	g	62	12.7	7.9		
		A	h	62	10.4	6.5		
LacZ4 (1720)	CAG→TAG	T	j	40	3.5	1.4	0.8	0.26 ± 0.12
		A	j	15	3.5	0.5		
		G	j	15	3.5	0.5		

*Additional engineered silent mutations include: LacZ2 at position 201 (A→T), LacZ3 at position 193 (C→A) and LacZ4 at position 1728 (G→A).

†Mutated bases for the creation of the nonsense codons are indicated in bold.

‡acMI for a window size of 40 nt.

§Reversion rate of wild-type JTT1 strain at 72 h.

examining relative levels of supercoiling of these plasmids (data not shown).

Reversion rates were compared for each mutant in each supercoiling phenotype and showed that the number of revertant colonies increased constantly over time (Fig. 2). However, the frequency and the extent to which revertants accumulated varied according to the supercoiling *E. coli* host strain. The mutation rates of JTT1- and RS2-derived strains were not statistically different for most mutants (Table 1, supplementary Fig. S1, available with the online version of this paper). However, RS2 strains had the highest mutation rates for three of four mutants (LacZ1–3; Fig. 2). Mutation rates for SD7-derived strains were always significantly lower than those seen for the other two strains. Thus, statistically significant higher reversion rates were

consistently associated with increased or wild-type levels of negative supercoiling compared with the SD7 lower negative supercoiling strains. Reversion rates for all three strains with the LacZ4 mutation did not appear to be influenced by supercoiling, but it is likely that the reversion potential of this mutant is too low to be affected by changes in supercoiling (Fig. S1, available with the online version of this paper).

The acMI of a codon is correlated with reversion rate

A linear relationship has been demonstrated previously between the acMI and mutation rates in an episomal *E. coli* system (Schmidt *et al.*, 2006). In the present study, a

Fig. 2. Mutation rates of *lacZ* gene mutants. Data are separated by engineered mutation and *E. coli* host strain with differing levels of supercoiling. Mutation rates for three time points are shown for all mutants and strains. Supercoiling phenotypes: JTT1, wild-type (hatched bars); RS2, increased negative supercoiling (grey bars); SD7, decreased negative supercoiling (white bars). All data are mean values and error bars represent SD.

correlation was also seen between acMIs predicted by MFG and experimentally determined mutation rates (Fig. 3). These correlations were only possible by the exclusion of the LacZ2 data. The experimental reversion rate of this mutant was much lower than expected and presumed to be influenced by factors other than those taken into account by MFG. Nevertheless, comparison of reversion rates and acMIs for LacZ1, 3 and 4 at 72 h post-plating showed good correlations for each *E. coli* supercoiling strain (Fig. 3; JTT1, $r^2=0.942$; RS2, $r^2=0.933$; and SD7, $r^2=0.831$). Similar correlations were also observed at 48 and 96 h for all strains (data not shown).

Codon preference and diversity of reversion fate

Table 2 shows the results of sequencing random LacZ⁺ revertants that arose at all time points. Many different codons can substitute for the wild-type codon to produce functional β -galactosidase necessary for lactose utilization. Reversion of the nonsense mutations often occurred at positions within the codon other than the site of the engineered mutation; in most cases, there was no preference for reversion at the exact point mutation site.

A low frequency of reversions other than single point mutations was also seen. Several deletions ranging from 3–12 nt were observed. These resulted in the removal of the nonsense mutation plus adjacent nucleotides. In LacZ1, mutation of the nonsense codon often resulted in three simultaneous mutations that were complementary to, and presumably templated by, the bases in the opposing loop (i.e. TAG mutated to AGC to complement the opposing GCT). This triple mutant was found to be the predominant mutation for both RS2 and SD7 strains of LacZ1 (63 and 48 % of mutations, respectively). For the wild-type JTT1 strain, however, the TGG (Trp) codon and the triple mutation were the most prevalent mutations (32 and 28 %, respectively; Table 2). As the number of mutations arising

after 72 h was greater than anticipated, a small number of the revertants at 96 h were sequenced to determine the nature of these reversions. The majority of these were triple revertants, but, due to the low number analysed, this does not preclude the emergence of single nucleotide revertants after 72 h. These analyses revealed that templated triple mutations occurred at a relatively low frequency at early time points, but were the predominant revertant type after 72 h.

Mutations acquired during LacZ4 reversion favoured transversions of the TAG nonsense codon to either GAG (Glu) or TCG (Ser) (50 and 30 %, respectively). Among the revertants sequenced, there was no clear evidence of a preferred nucleotide at the mutation site, nor was there a codon preference, regardless of the supercoiling host strain (Table 2). Growth rates of representative revertants for all mutation contexts showed that none had a significant growth advantage (data not shown).

DISCUSSION

This study describes the reversion rates of several chromosomal LacZ[−] mutants with different predicted mutation potentials in three strains of *E. coli* with differing levels of supercoiling. The reversion rates of these different mutations clearly highlight the importance of the mutation site context with respect to the ability of a nonsense codon to revert to a functional state. All four of the engineered nonsense mutations resulted in the loss of β -galactosidase activity. As the *lacZ* gene becomes derepressed in response to starvation for a carbon source, the increased rate of transcription will increase the probability of mutations that restore enzyme activity (Reimers *et al.*, 2004). Thus, gene derepression and associated increases in mutation rate act as an evolutionary salvage pathway critical for overcoming environmental stress (Wright, 2000, 2004). Starvation for a carbon source leads to the derepression and transcriptional

Fig. 3. Association of predicted acMI for a 40 nt window and experimental mutation rates for LacZ⁺ revertants. Standard error bars are given for each data point. Relative supercoiling phenotypes: JTT1, wild-type (diamonds); RS2, increased negative supercoiling (squares); SD7, decreased negative supercoiling (triangles). All data are mean values and error bars represent SD.

Table 2. Nucleotide changes of nonsense codons for representative LacZ⁺ revertants

LacZ1	From TAG to:	JTT1	RS2	SD7
1st position	AAG (Lys)	5	1	1
	CAG (Gln)	2	1	0
	GAG (Glu)	0	1	1
2nd position	TCG (Ser)	2	0	0
	TGG (Trp)*	15	4	5
	TTG (Leu)	3	0	1
3rd position	TAA (Stop)	NA	NA	NA
	TAC (Tyr)	3	0	1
	TAT (Tyr)	3	2	4
Others	Suppressor (TAG)†	0	0	0
	Triple AGC (Ser)	13	15	12
	GTT deletion	1	0	0
LacZ2	From TAA to:	JTT1	RS2	SD7
1st position	AAA (Lys)	3	4	2
	CAA (Gln)	3	2	4
	GAA (Glu)*	4	1	5
2nd position	TCA (Ser)	3	5	3
	TGA (Stop)	NA	NA	NA
	TTA (Leu)	4	0	3
3rd position	TAC (Tyr)	2	3	2
	TAG (Stop)	NA	NA	NA
	TAT (Tyr)	5	2	1
Others	Suppressor (TAA)†	0	4	1
	ATA (Ile)	0	0	1
	GCTTAA (insertion/deletion)	0	1	3
	TTA deletion	0	0	1
LacZ3	From TAA to:	JTT1	RS2	SD7
1st position	AAA (Lys)	2	2	2
	CAA (Gln)	4	0	4
	GAA (Glu)*	1	0	3
2nd position	TCA (Ser)	2	5	7
	TGA (Stop)	NA	NA	NA
	TTA (Leu)	1	3	3
3rd position	TAC (Tyr)	0	3	1
	TAG (Stop)	NA	NA	NA
	TAT (Tyr)	7	3	2
Others	Suppressor (TAA)†	1	0	1
	G/TA deletion (part of 2 codons)	0	1	0
	CCG/TAA/AGC/TGG deletion	0	1	0
LacZ4	From TAG to:	JTT1	RS2	SD7
1st position	AAG (Lys)	0	0	0
	CAG (Gln)*	4	3	1
	GAG (Glu)	7	5	10
2nd position	TCG (Ser)	6	7	5
	TGG (Trp)	0	1	2
	TTG (Leu)	1	3	1

Table 2. cont.

LacZ4	From TAG to:	JTT1	RS2	SD7
3rd position	TAA (Stop)	NA	NA	NA
	TAC (Tyr)	0	1	0
	TAT (Tyr)	2	0	1
Others	Suppressor (TAG)†	0	0	0

NA, Not applicable

*Wild-type codon sequences are in bold.

†Suppressor mutations contained the nonsense codon despite being LacZ⁺.

activation of genes specifically required for the metabolism of alternative energy sources, such as lactose (Balke & Gralla, 1987). Increased transcriptional activity and the participation of cellular topoisomerases result in supercoiling, the formation of secondary structures in ssDNA and the consequent exposure of unpaired bases vulnerable to mutagenesis (Balke & Gralla, 1987; Wright *et al.*, 1999).

The reversion rate of the LacZ1 mutant was four- to fivefold higher than that in LacZ3 and at least 10-fold higher than that in LacZ4. With the exception of LacZ2, these experimental reversion rates were qualitatively consistent with the relative acMIs of these mutants predicted by MFG (Table 1), but analyses of more *lacZ* mutants in similar and different contexts may have supported these trends more conclusively. The reversion rate of LacZ2 was clearly much lower than that predicted by its acMI. This nonsense codon is located in a terminal loop formed by 6 nt at the end of an 11 bp stem (Fig. 1c). Nucleotide residues within such a loop may be protected from modification by either nucleotide pairing or protein interactions that render the bases less accessible for mutation. Since these types of structures are typical in gene promoter regions that bind proteins involved in the regulation of transcription and replication, such protein interactions within this structure may protect the apical bases from exposure and mutagenesis (reviewed by Pearson *et al.*, 1996). Additionally, tertiary structure (such as loop-loop and pseudoknot interactions) may be involved in pairing these bases with unpaired bases in other secondary structures, thereby decreasing their intrinsic mutation potential *in vivo*. Moreover, bases in an apical loop do not typically have opposing bases for enzymes to use as templates for repair. Further experiments are necessary to determine how other molecular interactions might limit the ability of MFG to predict mutability of bases in apical loops.

Supercoiling, especially negative supercoiling, can have a number of effects: for example, on transcriptional pausing (Theissen *et al.*, 1990), the creation and stabilization of secondary structures (Dayn *et al.*, 1992) and, now, on the mutation potential of a gene. Supercoiling of the *E. coli* chromosome is important for the regulation of gene

expression in response to environmental changes, such as increased osmolarity (Higgins *et al.*, 1988), nutrient limitation (Balke & Gralla, 1987), pH changes (Karem & Foster, 1993), anaerobiosis (Dorman *et al.*, 1988) and temperature variations (Goldstein & Drlica, 1984). Increased levels of DNA supercoiling driven by transcription both create and stabilize SLSs, thereby significantly increasing the relative susceptibility of unpaired bases to mutate within that structure. It is likely that under starvation conditions, both increased transcription and, consequently, supercoiling cause an increase in the mutation potential of genes specifically involved in overcoming the metabolic blockage.

Previous studies have shown that mutability is influenced by the level of supercoiling in plasmids (Schmidt *et al.*, 2006). Evidence that the *E. coli* genome contains domains with similar degrees of negative supercoiling to plasmid DNA (Miller & Simons, 1993; Postow *et al.*, 2004; Sinden & Pettijohn, 1981) suggests that the mutability of chromosomal DNA could indeed be similar to that of episomal DNA. The present study found that the level of DNA supercoiling had a significant effect on mutation rate for *lacZ* mutations within the *E. coli* chromosome. Generally, increased levels of negative supercoiling coincided with increased reversion rate of mutants. These data suggest that supercoiling of derepressed genes is a major contributor to the mutation potential of nucleotides within secondary structures.

Mutations can occur by either enzymic or non-enzymic reactions. Unpaired bases in ssDNA are intrinsically mutable and non-enzymic mutations (e.g. hydrolytic deamination, oxidation, depurination and depyrimidation) occur at significant levels under physiological conditions (Lindahl, 1993). In the present study, predictions of base MI by MFG were determined by the gene-encoded location of a base in secondary structures and by the extent to which the base was unpaired during transcription. Since enzyme activity (e.g. mismatch repair proteins and deaminases) also depends upon the availability of ssDNA, their effect on mutation rates would be superimposed upon, and difficult to distinguish from, those caused only by base exposure.

Evidence in Table 2 shows that several engineered mutations had a codon reversion preference to which the nonsense codon reverted. LacZ1 mutated predominantly to either the wild-type Trp codon (via A→G mutation at the second position, resulting in TGG) or the triple ACG mutant. While growth rates of LacZ1 revertants showed no significant differences (data not shown), a marked increase occurred in the proportion of triple revertants compared with all revertants arising between 72 and 96 h (approaching 100 % of revertants; data not shown). Additionally, an increase in the rate of reversion (versus a steady-state increase) was observed for this mutant between 72 and 96 h (Fig. S1, available with the online version of this paper). The increase in reversion rate for LacZ1 after 72 h

was likely to be due to two concurrent, but independent, mechanisms: point mutations that are generated early upon starvation and templated repair mutations that have a delayed appearance. However, it should be noted that regardless of the occurrence of triple mutations, the single nucleotide reversion rate for LacZ1 was still relatively higher in these strains than any other mutant at 72 h.

The LacZ4 mutation showed a preference for reverting to either Glu or Ser codons (GAG and TCG, respectively). The LacZ4 codon is in close proximity to the active site pocket; this is likely to restrict the choice of amino acid and, therefore, the codon to which the site can mutate in order to regain fully functional enzyme activity. This restriction could in part account for the low reversion rate observed for this mutant. However, all possible reversions were observed at a low frequency, with the exception of the first position T-to-A (Lys). These data highlight the significance of constraints on reversion at the protein level and the importance of their effects during mutation rate experiments.

The LacZ1 mutant showed a remarkably high occurrence of the ACG triple mutation. These mutations cannot arise by chance, as the single site reversion rate is approximately 10^{-9} , and three such independent mutations would have the highly unlikely frequency of 10^{-27} . This suggests that triple mutations arise as the result of a single event, especially since no double mutants were observed. The repair of mismatched bases within SLSs evidently results from using the opposing mismatched bases in the internal loop of the opposing strand as a template for repair (Ripley, 1982; Ripley & Glickman, 1983; Viswanathan *et al.*, 2000; Weiss & Wilson, 1987; Wright *et al.*, 2003). The presence of triple mutations is compelling evidence for the existence of such large SLSs *in vivo*. Presumably in these large SLSs, repair mechanisms mistake stable SLSs with long paired stems for chromosomal dsDNA that appears to be mismatched. Excision repair mechanisms then remove the unpaired mismatched bases and replace them with bases complementary to the opposing sequence, which is used as a template.

Taken together with our other studies (Reimers *et al.*, 2004; Schmidt *et al.*, 2006), the data presented here suggest that directed mutational events that overcome metabolic stress are affected by both transcription and supercoiling. Thus, the transcriptional activity of a gene influences the creation and stability of SLSs, and the context of unpaired bases within these structures determines their potential to mutate. We propose that base mutability within genes is primarily determined by the intrinsic mutability of unpaired bases and the extent to which they are exposed in transcriptionally driven secondary structures. It is likely that this inherent mutagenic mechanism is linked with basic cellular processes allowing organisms to recover rapidly from potentially catastrophic errors. As most mutations that are critical to survival occur in chromosomal

DNA, it was valuable to confirm these findings in a native gene system. Thus, the examination of chromosomal mutation has more relevance to higher organisms than previously reported for episomal DNA. This study demonstrates the utility of MFG for analysing a prokaryotic chromosomal system and illustrates the important contribution of transcription-induced mutagenesis in response to selective stress.

ACKNOWLEDGEMENTS

We thank the Murdock DNA Sequencing Core Facility for expert service, and Scott Samuels for equipment and expertise related to supercoiling analysis. This research was funded by NIH grant R01CA099242 and the Stella Duncan Memorial Research Institute.

REFERENCES

- Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M. & Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol* **166**, 5051–5057.
- Balke, V. L. & Gralla, J. D. (1987). Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J Bacteriol* **169**, 4499–4506.
- Brock, R. D. (1971). Differential mutation of the β -galactosidase gene of *Escherichia coli*. *Mutat Res* **11**, 181–186.
- Cairns, J. & Foster, P. L. (1991). Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* **128**, 695–701.
- Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* **44**, 521–533.
- Datta, A. & Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616–1619.
- Dayn, A., Malkhosyan, S. & Mirkin, S. M. (1992). Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res* **20**, 5991–5997.
- Dorman, C. J., Barr, G. C., Bhriani, N. N. & Higgins, C. F. (1988). DNA supercoiling and the anaerobic and growth phase regulation of *tonB* gene expression. *J Bacteriol* **170**, 2816–2826.
- Fukita, Y., Jacobs, H. & Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* **9**, 105–114.
- Goldstein, E. & Drlica, K. (1984). Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc Natl Acad Sci U S A* **81**, 4046–4050.
- Herman, R. K. & Dworkin, N. B. (1971). Effect of gene induction on the rate of mutagenesis by ICR-191 in *Escherichia coli*. *J Bacteriol* **106**, 543–550.
- Herring, C. D., Glasner, J. D. & Blattner, F. R. (2003). Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311**, 153–163.
- Higgins, C. F., Dorman, C. J., Stirling, D. A., Waddell, L., Booth, I. R., May, G. & Bremer, E. (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* **52**, 569–584.
- Karem, K. & Foster, J. W. (1993). The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. *Mol Microbiol* **10**, 75–86.
- Krasilnikov, A. S., Podtelezhnikov, A., Vologodskii, A. & Mirkin, S. M. (1999). Large-scale effects of transcriptional DNA supercoiling *in vivo*. *J Mol Biol* **292**, 1149–1160.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.
- Liu, L. F. & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* **84**, 7024–7027.
- Luria, S. E. & Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511.
- Markham, N. R. & Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33**, W577–W581.
- Miller, W. G. & Simons, R. W. (1993). Chromosomal supercoiling in *Escherichia coli*. *Mol Microbiol* **10**, 675–684.
- Mojica, F. J., Charbonnier, F., Juez, G., Rodriguez-Valera, F. & Forterre, P. (1994). Effects of salt and temperature on plasmid topology in the halophilic archaeon *Haloferax volcanii*. *J Bacteriol* **176**, 4966–4973.
- Opel, M. L. & Hatfield, G. W. (2001). DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the *ilvYC* operon of *Escherichia coli* is proportional to promoter strengths and transcript lengths. *Mol Microbiol* **39**, 191–198.
- Pearson, C. E., Zorbas, H., Price, G. B. & Zannis-Hadjopoulos, M. (1996). Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J Cell Biochem* **63**, 1–22.
- Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. (2004). Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* **18**, 1766–1779.
- Pruss, G. J. & Drlica, K. (1986). Topoisomerase I mutants: the gene on pBR322 that encodes resistance to tetracycline affects plasmid DNA supercoiling. *Proc Natl Acad Sci U S A* **83**, 8952–8956.
- Pruss, G. J., Manes, S. H. & Drlica, K. (1982). *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell* **31**, 35–42.
- Reimers, J. M., Schmidt, K. H., Longacre, A., Reschke, D. K. & Wright, B. E. (2004). Increased transcription rates correlate with increased reversion rates in *leuB* and *argH* *Escherichia coli* auxotrophs. *Microbiology* **150**, 1457–1466.
- Ripley, L. S. (1982). Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proc Natl Acad Sci U S A* **79**, 4128–4132.
- Ripley, L. S. & Glickman, B. W. (1983). Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation. *Cold Spring Harb Symp Quant Biol* **47**, 851–861.
- Rudner, R., Murray, A. & Huda, N. (1999). Is there a link between mutation rates and the stringent response in *Bacillus subtilis*? *Ann N Y Acad Sci* **870**, 418–422.
- Schmidt, K. H., Reimers, J. M. & Wright, B. E. (2006). The effect of promoter strength, supercoiling and secondary structure on mutation rates in *Escherichia coli*. *Mol Microbiol* **60**, 1251–1261.
- Sinden, R. R. & Pettijohn, D. E. (1981). Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc Natl Acad Sci U S A* **78**, 224–228.
- Singer, B. & Kusmierek, J. T. (1982). Chemical mutagenesis. *Annu Rev Biochem* **51**, 655–693.
- Theissen, G., Pardon, B. & Wagner, R. (1990). A quantitative assessment for transcriptional pausing of DNA-dependent RNA polymerases *in vitro*. *Anal Biochem* **189**, 254–261.
- Tyagi, R., Lai, R. & Duggleby, R. G. (2004). A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol* **4**, 2.

- Viswanathan, M., Lacirignola, J. J., Hurley, R. L. & Lovett, S. T. (2000).** A novel mutational hotspot in a natural quasipalindrome in *Escherichia coli*. *J Mol Biol* **302**, 553–564.
- Weiss, U. & Wilson, J. H. (1987).** Repair of single-stranded loops in heteroduplex DNA transfected into mammalian cells. *Proc Natl Acad Sci U S A* **84**, 1619–1623.
- Wright, B. E. (1996).** The effect of the stringent response on mutation rates in *Escherichia coli* K-12. *Mol Microbiol* **19**, 213–219.
- Wright, B. E. (2000).** A biochemical mechanism for nonrandom mutations and evolution. *J Bacteriol* **182**, 2993–3001.
- Wright, B. E. (2004).** Stress-directed adaptive mutations and evolution. *Mol Microbiol* **52**, 643–650.
- Wright, B. E., Longacre, A. & Reimers, J. M. (1999).** Hypermutation in derepressed operons of *Escherichia coli* K-12. *Proc Natl Acad Sci U S A* **96**, 5089–5094.
- Wright, B. E., Reimers, J. M., Schmidt, K. H. & Reschke, D. K. (2002).** Hypermutable bases in the *p53* cancer gene are at vulnerable positions in DNA secondary structures. *Cancer Res* **62**, 5641–5644.
- Wright, B. E., Reschke, D. K., Schmidt, K. H., Reimers, J. M. & Knight, W. (2003).** Predicting mutation frequencies in stem-loop structures of derepressed genes: implications for evolution. *Mol Microbiol* **48**, 429–441.
- Wright, B. E., Schmidt, K. H. & Minnick, M. F. (2004).** Mechanisms by which transcription can regulate somatic hypermutation. *Genes Immun* **5**, 176–182.
- Zheng, G. X., Kochel, T., Hoepfner, R. W., Timmons, S. E. & Sinden, R. R. (1991).** Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells. *J Mol Biol* **221**, 107–122.

Edited by: S. J. W. Busby