Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome

Iwona J. Fijalkowska*, Piotr Jonczyk*, Magdalena Maliszewska Tkaczyk*, Malgorzata Bialoskorska*, and Roel M. Schaaper[†‡]

*Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02 106 Warsaw, Pawinskiego 5A, Poland; and †Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, PO Box 12233, Research Triangle Park, NC 27709

ABSTRACT We have investigated the question whether during chromosomal DNA replication in *Escherichia coli* the two DNA strands may be replicated with differential accuracy. This possibility of differential replication fidelity arises from the distinct modes of replication in the two strands, one strand (the leading strand) being synthesized continuously, the other (the lagging strand) discontinuously in the form of short Okazaki fragments. We have constructed a series of *lacZ* strains in which the *lac* operon is inserted into the bacterial chromosome in the two possible orientations with regard to the chromosomal replication origin *oriC*. Measurement of *lac* reversion frequencies for the two orientations, under conditions in which mutations reflect replication errors, revealed distinct differences in mutability between the two orientations. As gene inversion causes a switching of leading and lagging strands, these findings indicate that leading and lagging strand replication have differential fidelity. Analysis of the possible mispairs underlying each specific base pair substitution suggests that the lagging strand replication on the *E. coli* chromosome may be more accurate than leading strand replication.

The question as to how organisms duplicate their DNA with high accuracy is of fundamental interest. Previous studies have revealed the functioning of at least three separate steps, base selection, proofreading, and DNA mismatch repair, which, by their sequential action, are responsible for the low error rate of $\approx 10^{-10}$ per base replicated (1, 2). The most detailed information about this process is available for the bacterium *E. coli* based on both enzymological and genetical data. Replication of the *E. coli* chromosome is performed by DNA polymerase III holoenzyme, an asymmetric dimeric enzyme composed of 18 subunits (10 distinct) that simultaneously replicates the leading and lagging strand of the replication fork (for review, see ref. 3). It contains two polymerase core units, one for each strand, each consisting of three tightly associated subunits, α, ε, and θ. Of these, α is the polymerase (*dnaE* gene product), ε (*dnaQ* gene product) is a $3' \rightarrow 5'$ exonuclease that performs an editing function, and θ is a small subunit of unknown function. Additional components of the holoenzyme include the τ subunit (τ_2) that dimerizes the two cores, the β subunit (β_2) that encircles the DNA and tethers each DNA polymerase to the DNA to ensure high processivity, and the five-subunit γ complex (γ, δ, δ', χ, and ψ) that loads the β rings onto the DNA.

With regard to the fidelity of polymerase III holoenzyme, as studied both *in vivo* and *in vitro*, the main focus has been on the role of the α and ε subunits. The α (polymerase) subunit plays a critical role through the process of base selection, selecting with great preference correct nucleotides at the nucleotide insertion step. The ε subunit, in conjunction with the polymerase, is responsible for the subsequent proofreading step, in which by virtue of its $3'$ exonuclease activity incorrectly inserted nucleotides can be removed efficiently. These two steps together allow DNA synthesis to proceed at an average fidelity of 10^{-7}/bp replicated (1). The ultimate observed mutation rate of $\approx 10^{-10}$ (1, 2) is obtained by the subsequent action of postreplicative DNA mismatch repair, performed in *E. coli* by the products of the *mutH*, *mutL*, and *mutS* genes (4).

Although reasonable estimates have been made as to the overall fidelity of *in vivo* DNA replication, no insight exists into the important question of whether the two strands of DNA replication produce mutations at the same rate. This is an intriguing question because, due to their antiparallel nature, the two strands are replicated in different fashion (5). One strand (the leading strand) is synthesized continuously, whereas the complementary (lagging) strand is synthesized discontinuously in short Okazaki fragments, 1–2 kb in length. Lagging strand synthesis appears to be a more complicated process as it requires the cyclical repetition of several different reactions in a defined sequence, including priming of the Okazaki fragments, extension, and rapid recycling of the polymerase from a finished fragment to the next primer (3, 5). These differences suggest the possibility that the production of mutations may not be equal in the two strands. The question of differential replication fidelity is of relevance not only for a proper understanding of DNA replication *per se* but is also of importance to the area of molecular evolution in which strand-specific differences during evolution of DNA sequences have been observed (6). In fact, uncoupling of mutagenesis in leading and lagging strands has been proposed as one mechanism by which organisms might be able to evolve rapidly without suffering the deleterious consequences of high overall mutation rates (7).

In this study, we have developed a system to measure this potential difference in replication fidelity between leading and lagging strand replication on the *E. coli* chromosome. Previous *E. coli* studies using plasmid-contained target genes have indeed suggested differences between leading and lagging strands (8–13). However, the *ColE1* plasmids used in these studies replicate in a manner quite distinct from the *E. coli* chromosome (14), requiring for example extensive synthesis by DNA polymerase I, which likely complicates the comparisons. Furthermore, most of these studies have not directly addressed the intrinsic fidelity of replication because they focused on the probability of DNA lesion bypass (8, 9) or the production of deletion and duplication mutations requiring specific misaligned intermediates that can be formed in one strand but not in the other (11, 12).

We have constructed strains containing a series of *lacZ* mutations on the chromosome in the two orientations with respect to the origin of replication (*oriC*). Gene inversion is equivalent to a switching of the direction of replication fork movement through the gene, thus causing nucleotides in the gene that were replicated previously as leading strand to be replicated as lagging strand and vice versa. If the leading and lagging strand modes of replication do have different fidelity, this may be observable as a difference in mutation frequencies between the two orientations. We report here that, for all four *lacZ* alleles tested, gene inversion changes the observed mutation frequencies by severalfold, strongly suggesting that on the *E. coli* chromosome the fidelities of leading and lagging strand replication are intrinsically different.

MATERIALS AND METHODS

Strains and Media. The *E. coli* strains and plasmids used are listed in Table 1. Solid and liquid media [both Luria–Bertani (LB)] and minimal medium) were as described (17). Minimal plates were supplemented with 0.4% glucose or 0.4% lactose as a carbon source, 5 μg/ml thiamine, and 50 μg of amino acids per ml, as required. Antibiotics were added as follows: tetracycline, 12.5 μg/ml; chloramphenicol, 10 μg/ml; kanamycin, 25 μg/ml; ampicillin, 25 μg/ml, and rifampicin, 100 μg/ml.

Strain Constructions. Strains containing the *lac* operon inserted in two orientations in the phage λ *attB* attachment site were created by using the method of Diederich *et al.* (18) in strains that had the operon deleted from its normal location near 8 min on the *E. coli* map (19). A 12-kb *Pst*I fragment containing the entire *lacIZYA* operon was obtained from F'*prolac* isolated from strains CC102, CC104, CC105, and CC106, each carrying a different *lacZ* mutation (ref. 15 and Table 1). The *Pst*I fragments were inserted into the *Pst*I site of plasmid pLDR10 (18) in the two possible orientations. The resulting plasmids were used, in conjunction with helper plasmid pLDR8, to integrate the *lac* operon into *attL* (phage

λ attachment site) of the recipient strains MC4100 and CD4 (Table 1) both of which are Δ*lac*. The orientation of *lac* in plasmid pLDR10 determines the ultimate orientation in the chromosome. In this manner, four pairs of strains were obtained, each pair containing a particular *lacZ* allele in the two orientations [left (L) and right (R), see Fig. 1] with regard to *oriC*. The expected chromosomal orientation was confirmed by PCR, DNA sequencing, and Southern hybridization for three independent integrants for each plasmid orientation (data not shown). Additional markers were introduced into these strains by P1 transduction by using P1*virA*. The *mutL*::Tn5 allele was introduced based on its kanamycin resistance; the *mutD5* allele was introduced by cotransduction with *zaf-13*::Tn*10*; *dnaQ49* was introduced by cotransduction with *zae-502*::Tn*10*; and *ung-152*::Tn*10* was introduced based on its tetracycline resistance. The *ung*::Tn*10* allele was kindly provided by D. Fix, Southern Illinois University, Carbondale, IL.

Mutant Frequency Measurements. Mutant frequencies were determined by toothpicking a total of 30 single colonies for each strain into 1 ml of LB or minimal medium and growing them to saturation at 37°C. The colonies were from three independent integrants for each orientation and usually several independent transductants (*mutL*, *dnaQ49*, etc.) derived from each integrant. Appropriate dilutions of the saturated cultures were plated on minimal Lac plates and LB Rif plates to determine the number of *lac*$^+$ and Rif^r mutants, respectively, and on LB and minimal medium plates to determine the total cell count. To calculate mutant frequencies, the median number of mutants per plate was determined and divided by the average number of total cells. *P* values for statistical difference in mutant frequencies between L and R orientations were determined by using the nonparametric Mann–Whitney criterion (20) applied to the mutant yield distributions of the 30 independent cultures for the two compared orientations by using STATMOST Ver. 2.50 (DataMost, Salt Lake City, UT).

Table 1. Bacterial strains and plasmids

Name	Relevant genotype, description	Reference
Strains		
CC102	F'*lacIZ*(G·C → A·T)*	15
CC104	F'*lacIZ*(G·C → T·A)*	15
CC105	F'*lacIZ*(A·T → T·A)*	15
CC106	F'*lacIZ*(A·T → G·C)*	15
MC4100	Δ(*argF-lac*)U169	*E. coli* Genetic Stock Center
CD4	Δ(*lacI-Y*)6	16
NR9458	*mutD5 zaf-13*::Tn*10*	16
NR9559	*mutL*::Tn5	17
NR9695	*dnaQ49 zae-502*::Tn*10*	16
NR11515	*ung-152*::Tn*10*	This work
EC3114	*attB*::*lacIZYA*(CC102)Right	This work
EC3120	*attB*::*lacIZYA*(CC102)Left	This work
EC3126	*attB*::*lacIZYA*(CC104)Left	This work
EC3132	*attB*::*lacIZYA*(CC104)Right	This work
EC3138	*attB*::*lacIZYA*(CC105)Left	This work
EC3144	*attB*::*lacIZYA*(CC105)Right	This work
EC3150	*attB*::*lacIZYA*(CC106)Left	This work
EC3156	*attB*::*lacIZYA*(CC106)Right	This work
Plasmids		
pLDR8	*int* gene expression vector, *neo*	18
pLDR10	integration vector, *cat*	18

EC3114 through EC3156 are derivatives of MC4100. The designations CC102, CC104, CC105, and CC106 in their genotype indicates the *lacZ* allele from the corresponding CC strain. The designation 'right' or 'left' refers to the orientation of *lac* relative to *oriC*, as shown in Fig. 1.
*The base pair substitution in parentheses indicates the specific substitution by which the *lacZ* allele reverts to wild type.
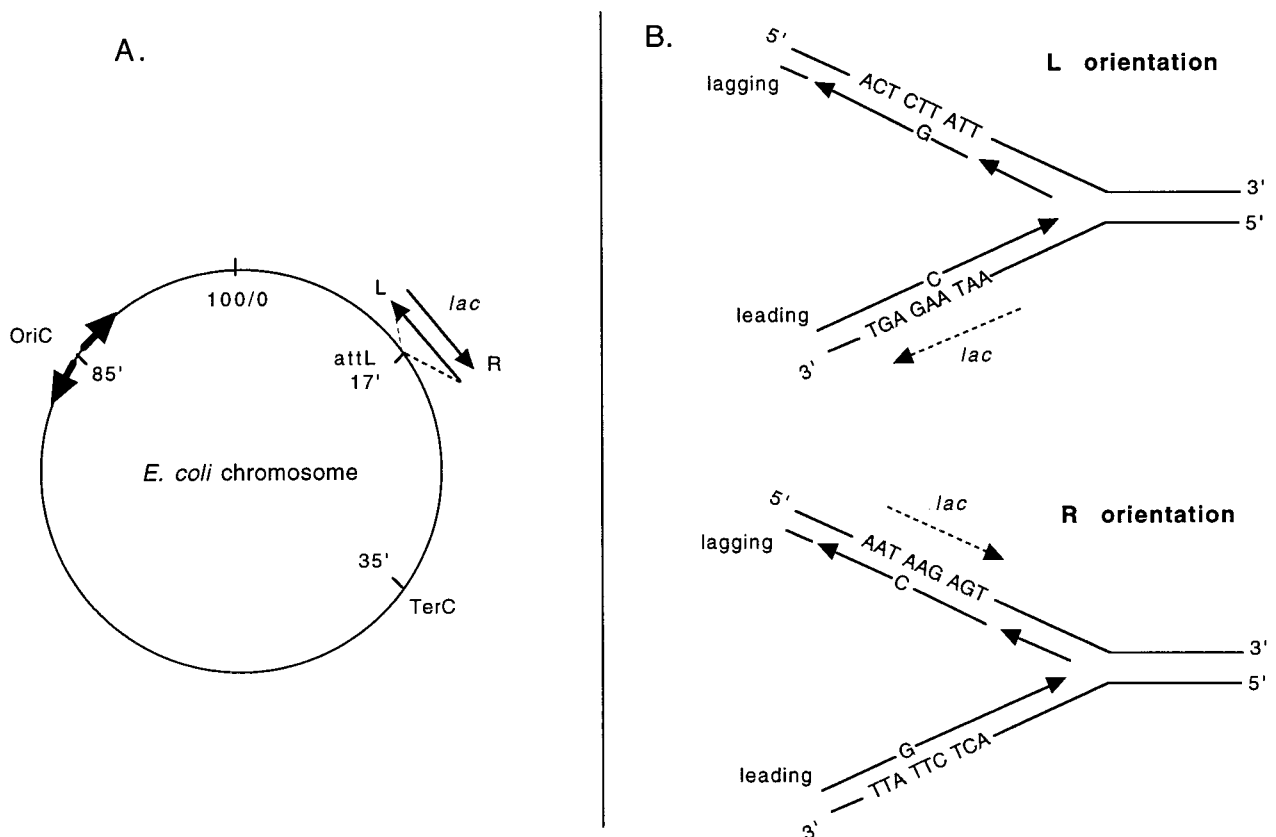
Fig. 1.   (*A*) Insertion of the *lac* operon into the *attL* site of the *E. coli* chromosome in two orientations with regard to the chromosomal replication origin *oriC*. The orientation in which the *lac* operon is transcribed in the same direction as the movement of the replication fork through the target is designated the right (R) orientation, whereas the left (L) orientation indicates *lac* transcription in a direction opposite to the movement of the replication fork. The thick arrows at OriC represent the two forks initiated at this site. (*B*) A more detailed drawing of the R and L orientations for the case of the *lacZ* CC106 allele that reverts by A·T → G·C transition (AAG → GAG codon change), along with the potential T·G and A·C mispairs that cause this transition in either orientation. The dashed arrow indicates the direction of *lac* transcription.

RESULTS

To investigate the possibility of differential replication fidelity during leading and lagging strand DNA synthesis, we developed a system to measure mutagenesis in *lacZ* target sequences. We constructed pairs of strains containing a *lacZ* sequence of interest in the two possible orientations with respect to the origin of DNA replication. In such a system, a given nucleotide sequence within the gene is replicated as a leading strand in one strain but as a lagging strand in the other. Under conditions in which the observed mutation frequency represents the frequency of DNA replication errors, a difference in the mutation frequency between the two orientations, would indicate differential replication fidelity. For example, when measuring *lac* reversion proceeding via a defined A·T → G·C base pair substitution, the observed mutation frequency is the sum of T·G mispairs in one strand and A·C mispairs in the other (we follow the convention of stating the template base first). In most cases, these two complementary mispairs will be made at different frequencies, and the mutant frequency reflects the most frequent one. Switching this more frequent mispair (likely T·G, see *Discussion*) from the more accurate to the less accurate strand will cause an increase in the observed error rate and vice versa. This method can be used for any case in which the two complementary mismatches occur at different frequencies and, in fact, does not require knowledge *per se* about which of the two complementary mispairs is the most frequent.

We used four *lacZ* alleles that are part of a set of six that have been widely used for studies of mutational specificity (15). We used the two transition alleles (derived from strains CC102 and

CC106, which revert specifically by G·C → A·T and A·T → G·C transitions, respectively; see Table 1) and two transversion alleles (derived from strains CC104 and CC105, which revert specifically by G·C → T·A and A·T → T·A transversions, respectively). The entire *lacIZYA* operons containing each of these four *lacZ* mutations were removed from the F'*prolac* on which they normally reside and inserted into the *attL* phage λ attachment site near 17 min of the *E. coli* chromosome in the two possible orientations, using the method of Diederich *et al.* (see *Materials and Methods* and ref. 18 for details). Two recipient strains were used (MC4100 and CD4, see Table 1), both of which have the *lac* operon deleted from its normal position. Fig. 1*A* illustrates relevant aspects. *E. coli* replication starts at a unique site termed *oriC* located near 85 min of the *E. coli* map (19). From this origin, replication proceeds bidirectionally, the two forks meeting at the termination site *terC* located near 35 min. We have designated the two orientations of the *lac* operon with respect to the replication fork moving through the target R and L. The R orientation reflects the *lac* operon being transcribed in the same direction as movement of the replication fork, whereas the L orientation indicates *lac* transcription in a direction opposite to the replication fork movement.

To specifically measure the *in vivo* replication fidelity, we introduced the *mutL* mismatch-repair deficiency. Because of the absence of mismatch repair, mutations in these strains directly reflect replication errors (1, 21). Mutations in these strains are mostly transitions because these are the predominant type of replication errors made by pol III HE *in vivo* (1, 21). Table 2 shows that significantly different mutant frequencies were observed when testing the effect of orientation on the

Table 2. Mutant frequencies (per 10^6 cells) in *mutL* or *ung* strains containing the *lac* operon in opposite (L and R) orientations on the *E. coli* chromosome

		mutL		*ung*
lac allele (mutation measured)	*lac* orientation	$lac^- \rightarrow lac^+$	$rif^s \rightarrow rif^r$	$lac^- \rightarrow lac^+$
CC102 (G·C → A·T)	L	2.1	7.0	0.35
	R	0.50*(4.2)	6.4	0.34
CC106 (A·T → G·C)	L	0.27	nd	nd
	R	0.51*(0.53)	nd	nd

Mutant frequencies were determined as described in *Materials and Methods*. Each entry is based on the median value of 30 independent cultures, comprising three independent *lacZ* integrants for each orientation. Statistically significant differences between the two orientations are indicated by an asterisk and by a calculated L/R ratio in parentheses ($P = 5 \times 10^{-6}$ and 2×10^{-3} for the CC102 and CC106 allele, respectively). The difference for the rif^r frequency of the CC102 strains is not statistically significant ($P = 0.50$). The presented results are for the MC4100 background; the *mutL* experiment also was performed in the CD4 background yielding identical results (data not shown). For comparison, the wild-type levels for the reversion of the CC102 and 106 alleles are $\sim 2 \times 10^{-8}$ and 0.1×10^{-8}, respectively. nd, not determined.

two *lacZ* transition alleles. For example, the number of G·C → A·T transitions was 4-fold higher for the L-oriented *lac* operon than for the R-oriented *lac* operon. For the A·T → G·C transition, the R-oriented *lac* operon showed a 2-fold higher mutant frequency. These differences are statistically significant (see Table 2 legend) and were observed in several repeated experiments. These data strongly suggest that leading and lagging strand replication on the *E. coli* chromosome are not equal.

In the same experiments, we also determined the frequency of rifampicin-resistant mutations but observed no difference between L and R strains (Table 2). This result is as expected because the target gene for rifampicin resistance, *rpoB*, is not subject to inversion. In an additional control experiment, no effect of inversion was observed for the *lac* G·C → A·T transitions in an *ung* strain, defective in uracil-*N*-glycosylase (22) (Table 2). Because these strains are mismatch-repair proficient, they produce few replication errors but, instead, they produce high levels of G·C → A·T transitions resulting from deamination of cytosine to uracil. Because uracil (which remains unrepaired in *ung* strains) will always pair correctly with A regardless of its presence in leading or leading strand, no effect is expected of gene inversion. These results show that the effect of gene inversion is specific for replication errors.

We also tested the effect of gene orientation in proofreading-deficient strains. These strains are strong mutators in which DNA replication errors are specifically enhanced (23, 24). The proofreading deficiency enhances both transition and transversion errors and the strains are therefore useful for investigating the effects of strand biases on both kinds of mutations. We tested the effect of gene orientation in strains

carrying the *mutD5* (25) or the *dnaQ49* allele (26), both carrying a known defect in the ε proofreading subunit of pol III HE (27–29). For the A·T → T·A transversion, the *dnaQ49* strain with the L-oriented *lac* operon was 2.6-fold more mutable than the same strain with the R-oriented *lac* operon (Table 3). Similarly, the *mutD5* strain displayed a 3.8-fold higher level of these transversions for the L-oriented *lac* operon. For the *mutD5* strain, we also tested the effect on the G·C → T·A transversion: the L orientation was 3-fold more mutable than the R orientation. The A·T → G·C transition also was tested for both mutators. An ≈2-fold difference in favor of the R-oriented target was observed in both strains, very similar to the results obtained with the *mutL* strain. Again, no significant differences were observed when the strains with the two *lac* orientations were compared for rif^r mutations (Table 3).

Based on the above data, we conclude that both transition and transversion errors on the *E. coli* chromosome are subject to strand-specific differences in replication fidelity. The observed effects are in the range of 2- to 5-fold based on multiple experiments for each of the strains. These numbers may be an underestimate if both complementary mispairs contribute to the observed mutation rate. In that case, any decrease in mutant frequency caused by the more frequent mispair moving to the more accurate strand would be compensated for in part by the parallel move of the less frequent mispair to the more error-prone strand. Sample calculations show, e.g., that a 10-fold difference in the replication accuracy between the two strands would be observed as only a 3.4-fold change in mutant frequency if the two mispairs were to contribute intrinsically (i.e., when compared in the same strand) in a 5:1 ratio

Table 3. Mutant frequencies (per 10^6 cells) in proofreading-deficient *dnaQ49* and *mutD5* strains containing the *lac* operon in L and R orientations on the chromosome

		dnaQ49		*mutD5*	
lac allele (mutation measured)	*lac* orientation	$lac^- \rightarrow lac^+$	$rif^s \rightarrow rif^r$	$lac^- \rightarrow lac^+$	$rif^s \rightarrow rif^r$
CC104 (G·C → T·A)	L	nd	nd	1.9	7.3
	R	nd	nd	0.67* (2.8)	6.6
CC105 (A·T → T·A)	L	1.3	82	3.6	80
	R	0.50* (2.6)	80	0.94* (3.8)	89
CC106 (A·T → G·C)	L	2.7	nd	6.6	nd
	R	6.7* (0.40)	nd	10* (0.66)	nd

Mutant frequencies were determined as described in *Materials and Methods*. Each entry is based on the median value of 30 independent cultures, comprising 3 independent *lacZ* integrants for each orientation. Statistically significant differences are indicated by an asterisk and a calculated L/R ratio in parentheses. *P* values were ≤5×10^{-6} for the two *dnaQ49* experiments, and 5×10^{-6}, 0.03, and 0.002 for the three *mutD5* experiments (CC104, CC105, and CC106, respectively). None of the rif^r differences were statistically significant ($P > 0.05$). The *dnaQ49* experiments were performed at 37° in LB; the *mutD5* experiments at 37° in LB (CC105) or minimal medium (CC104, CC106). The choice of growth medium strongly affects the overall mutability of *mutD5* strains (23, 25), as can be seen here for the differential rif^r frequency in the two media (CC104 vs. CC105). However, the observed strand bias for the *lac* alleles is not affected by this medium effect (this Table and data not shown). The wild-type levels for the CC104, CC105, and CC106 alleles are ~ 2, 2, and 0.1×10^{-8}, respectively. nd, not determined.

$[(50+1)/(5+10) = 3.4]$. Thus, all observed effects are to be considered minimum values.

DISCUSSION

The data obtained in this study demonstrate that inversion of gene orientation on the *E. coli* chromosome leads to a significant change in the observed mutation frequency in each of the four cases tested. We interpret these results to mean that on the chromosome replication fidelity must be different in leading and lagging strands. This conclusion is rooted in the realization that gene inversion preserves all aspects DNA metabolism, with the direction of replication fork movement relative to the gene as the only exception. In addition, because these experiments were performed in strains in which mutations directly reflect replication errors, the interpretation in terms of replication errors is facilitated.

We have considered the possibility that the differential mutability of the two orientations might be a consequence of transcription affecting the fidelity of DNA replication: in the R orientation, transcription opposes the movement of the replication fork, whereas in the L orientation transcription and fork movement are in the same direction. However, we note that of the four *lacZ* alleles tested (Tables 2 and 3), three are more mutable in the R orientation whereas one is more mutable in the L orientation, not consistent with such model. In addition, transcription activity of the *lac* operon is relatively low, and encounters between the replication fork and the transcriptional apparatus must be considered rare. Nevertheless, this possibility cannot be entirely excluded. One other formal possibility is that the different modes of replication of leading and lagging strands do not result in an intrinsic fidelity difference, but rather in a differential susceptibility of replication fidelity to DNA sequence context.

Assuming that the two strands are replicated with differential fidelity, which of the two is replicated more accurately? Although our experiments do not measure this directly, an indirect method may be used to make an inference, as exemplified by the case of the A·T → G·C transitions. As illustrated in Fig. 1*B*, the L-oriented construct measures the sum of $(T·G)_{lagging} + (A·C)_{leading}$, whereas the R-oriented construct measures $(T·G)_{leading} + (A·C)_{lagging}$. Given that T·G mispairs are generally much more frequent than A·C mispairs (30–37), the L orientation measures $(T·G)_{lagging}$, and the R orientation measures $(T·G)_{leading}$. Our observation that the R orientation is more mutable than the L orientation (Tables 2 and 3) thus indicates that the leading strand produces more T·G errors and therefore that lagging strand replication is the more accurate one.

In Table 4, we have analyzed each of the four base pair substitutions in terms of the underlying mispairs, the strand in which each mispair occurs in the two orientations, and the orientation observed to be the most mutable. In deciding which mispair may be predominantly responsible for each base

pair substitution, we surveyed the literature for direct measurements of misinsertion frequencies as well as the relative efficiencies by which the various mispairs can be extended by DNA polymerases (30–37). The latter aspect is important because reduced extension efficiency correlates with increased removal by exonucleolytic proofreading (e.g., see ref. 38). Only limited data is available for DNA polymerase III, but sufficient consensus has been demonstrated among a wide variety of polymerases in a number of different DNA sequence contexts to validate extrapolation to this enzyme (e.g., see ref. 36). The combined data are most straightforward for the transition mismatches: T·G and G·T are more frequent at the misinsertion step than the complementary A·C and C·A, and their extension is more efficient as well. Thus, in Table 4, we assume that the two transitions are mediated primarily by G·T and T·G mismatches. The data for transversion mismatches are more heterogeneous, particularly at the misinsertion stage. As a general rule, pyrimidine·pyrimidine (Py·Py) oppositions are similar to purine·purine (Pu·Pu) oppositions at the misinsertion stage. However, at the extension step, Pu·Pu pairs are much more difficult to extend than the Py·Py pairs and hence will suffer significantly more from proofreading. Based on these data, we suggest that most transversions, at least in case of proofreading enzymes, result largely from Py·Py mismatches, as we have also indicated in Table 4. Remarkably, when using this information, it follows that, in each of the four cases considered, the lagging strand replication is the more accurate.

What could be the mechanism by which the lagging strand replication is more accurate than leading strand replication? Because the mutational effect of gene inversion is observed in both *dnaQ(mutD)* and *mutL* strains, the mechanism cannot involve differential proofreading or differential mismatch repair. The effect may be mediated by differences in base selection in the two strands. However, this seems unlikely because the same DNA polymerase (*dnaE* gene product) is responsible for base selection in the two cases. Instead, we suggest that the effect reflects differential processing in the two strands of terminal mismatches that arise as result of misinsertion errors. We have previously suggested based on studies of *dnaE* antimutator alleles (17, 28) that, in addition to exonucleolytic proofreading, dissociation of the DNA polymerase from the terminal mismatch is an alternative mode of error removal. The abandoned mismatch may then become substrate for the exonuclease activity of pol I, which cleans up the ends of Okazaki fragments (5), or of other polymerase-associated or free exonucleases. This mechanism can in principle operate in both strands. However, because the lagging strand polymerase must dissociate each time when reaching the end of an Okazaki fragment, this polymerase may have a greater tendency to dissociate from terminal mismatches than the highly processive leading strand polymerase. At least two studies (39, 40) have indicated that higher processivity may be associated with higher mutability.

Table 4. The base·base mispairs in leading and lagging strands for each of the four *lacZ* base pair substitutions and inference of the more mutable strand based on observed mutant frequencies in the two orientation with respect to *oriC*

lacZ allele	Mutation	Mispairs and strand		Most frequent mispair	Observed error-prone orientation	Error-prone strand
		Orientation L	Orientation R			
CC102	G·C → A·T	$(G·T)_{leading} + (C·A)_{lagging}$	$(G·T)_{lagging} + (C·A)_{leading}$	G·T	L	Leading
CC104	G·C → T·A	$(G·A)_{lagging} + (C·T)_{leading}$	$(G·A)_{leading} + (C·T)_{lagging}$	C·T	L	Leading
CC105	A·T → T·A	$(A·A)_{lagging} + (T·T)_{leading}$	$(A·A)_{leading} + (T·T)_{lagging}$	T·T	L	Leading
CC106	A·T → G·C	$(A·C)_{leading} + (T·G)_{lagging}$	$(A·C)_{lagging} + (T·G)_{leading}$	T·G	R	Leading

Which mispair occurs in each strand is deduced from the *lacZ* missense codon (GGG, GCG, GTG, and AAG, for CC102, 104, 105, and 106, respectively) that must revert back to the wild-type GAG glutamic acid codon (15) and the orientation (L or R) of the *lacZ* gene as described in Fig. 1. Which of the two possible mispairs likely predominates in each case was deduced from a literature analysis of measured misinsertion and mispair extension efficiencies, as described in the Text. The CC designation of the *lacZ* alleles indicates the CC strain (Table 1) from which each allele was derived.

Our data are a demonstration of a difference in intrinsic replication accuracy between leading and lagging strands during chromosomal DNA replication. The *E. coli* system that we used is particularly suitable to address this question because the same polymerase is responsible for replication of either strand. This is in contrast to eukaryotic systems in which certain reported strand biases are difficult to interpret because of the likely operation of more than one DNA polymerase at the replication fork (41, 42). Our results further contrast with results obtained previously with plasmid-based systems in *E. coli*, which have suggested that the lagging strand may be more error-prone (8, 10). However, some of these effects appeared dependent on the proximity of the target to the plasmid origin (9) and may reflect the involvement of DNA polymerase I—which synthesizes several hundreds of nucleotides of leading strand after initiation at the plasmid origin (14)—rather than the differential mutability of leading and lagging strand replication by DNA polymerase III. The plasmid based system also has revealed differential mutability for deletion/duplication mutagenesis based on differential availability of the premutagenic intermediates in the two strands (11, 12), but these studies have not addressed the intrinsic accuracy within the two strands.

Experiments with the *lac* alleles also have been performed in wild-type (i.e., mismatch-repair-proficient) background (data not shown). These experiments also suggested differences between the two orientations, although the comparisons are complicated by significantly lower mutant yields. Nevertheless, statistically significant differences have been found for the CC105 (A·T → T·A) allele, in which the same strand bias (L > R) is observed as in the *mutL* background (a 3.2-fold effect: 1.2×10^{-8} vs. 0.41×10^{-8} for L and R orientations, respectively; data not shown). In a mismatch-repair-proficient background, mutations accumulate not only from uncorrected replication errors but also from a variety of other sources (43), and they may thus be of particular relevance for studies of molecular evolution. Our data therefore suggest that a mutational strand bias also could be observable on an evolutionary time scale, as suggested previously (6, 7).

1. Schaaper, R. M. (1993) *J. Biol. Chem.* **268**, 23762–23765.
2. Drake, J. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7160–7164.
3. Kelman, Z. & O'Donnell, M. (1995) *Annu. Rev. Biochem.* **64**, 171–109.
4. Modrich, P. & Lahue, R. (1996) *Annu. Rev. Biochem.* **65**, 101–133.
5. Kornberg, A. & Baker, T. A. (1992) *DNA Replication* (Freeman, New York).
6. Wu, C.-I. & Maeda, N. (1987) *Nature (London)* **327**, 169–170.
7. Furusawa, M. & Doi, H. (1992) *J. Theor. Biol.* **157**, 127–133.
8. Veaute, X. & Fuchs, R. P. P. (1993) *Science* **261**, 598–600.
9. Wagner, J., Kamiya, H. & Fuchs, R. P. P. (1997) *J. Mol. Biol.* **265**, 302–309.
10. Iwaki, T., Kawamura, A., Ishino, Y., Kohno, K., Kano, Y., Goshima, N., Yara, M., Furusawa, M., Doi, H. & Imamoto, F. (1996) *Mol. Gen. Genet.* **251**, 657–664.
11. Trinh, T. Q. & Sinden, R. R. (1991) *Nature (London)* **352**, 544–547.
12. Rosche, W. A., Trinh, T. W. & Sinden, R. R. (1995) *J. Bacteriol.* **177**, 4385–4391.
13. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. (1995) *Nat. Genet.* **10**, 213–218.
14. Marians, K. J. (1992) *Annu. Rev. Biochem.* **61**, 673–719.
15. Cupples, C. G. & Miller, J. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5345–5349.
16. Schaaper, R. M. & Cornacchio, R. (1992) *J. Bacteriol.* **174**, 1974–1982.
17. Fijalkowska, I. J. & Schaaper, R. M. (1995) *J. Bacteriol.* **177**, 5979–5986.
18. Diederich, L., Rasmussen, L. J. & Messer, W. (1992) *Plasmid* **28**, 14–25.
19. Berlyn, M. K. B., Low, K. B. & Rudd, K. E. (1996) *Escherichia coli and Salmonella*, ed. Neidhardt, F. C. (Am. Soc. Microbiol, Washington, DC), 2nd Ed., pp. 1715–1902.
20. Sokal, R. R. & Rohlf, F. J. (1981) *Biometry* (Freeman, San Francisco), 2nd Ed., pp. 429–445.
21. Schaaper, R. M. & Dunn, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **262**, 6220–6224.
22. Duncan, B. K. & Weiss, B. (1982) *J. Bacteriol.* **151**, 750–755.
23. Schaaper, R. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8126–8130.
24. Echols, H. & Goodman, M. F. (1991) *Annu. Rev. Biochem.* **60**, 477–511.
25. Degnen, G. E. & Cox, E. C. (1974) *J. Bacteriol.* **117**, 477–487.
26. Horiuchi, T., Maki, H. & Sekiguchi, M. (1978) *Mol. Gen. Genet.* **163**, 277–283.
27. Scheuermann, R., Tam, S., Burgers, P. M. J., Lu, C. & Echols, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7085–7089.
28. Fijalkowska, I. J. & Schaaper, R. M. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 2856–2861.
29. Maruyama, M., Horiuchi, T., Maki, H. & Sekiguchi, M. (1983) *J. Mol. Biol.* **167**, 757–771.
30. Kunkel, T. A., Hamatake, R. K., Motto-Fox, J., Fitzgerald, M. P. & Sugino, A. (1989) *Mol. Cell. Biol.* **9**, 4447–4458.
31. Mendelman, L. V., Boosalis, M. S., Petruska, J. & Goodman, M. F. (1989) *J. Biol. Chem.* **264**, 14415–14423.
32. Perrino, F. W. & Loeb, L. A. (1989) *J. Biol. Chem.* **264**, 2898–2905.
33. Bebenek, K., Joyce, C. M., Fitzgerald, M. P. & Kunkel, T. A. (1990) *J. Biol. Chem.* **265**, 13878–13887.
34. Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265**, 2338–2346.
35. Creighton, S., Huang, M.-M., Cai, H., Arnheim, N. & Goodman, M. F. (1992) *J. Biol. Chem.* **267**, 2633–2639.
36. Joyce, C. M., Sun, X. C. & Grindley, N. D. F. (1992) *J. Biol. Chem.* **267**, 24485–24500.
37. Poltev, V. I., Shulyupina, N. V. & Bruskov, V. I. (1995) *Mol. Biol. (New York)* **29**, 591–599.
38. Johnson, K. A. (1993) *Annu. Rev. Biochem.* **62**, 685–713.
39. Kunkel, T. A., Patel, S. S. & Johnson, K. A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6830–6834.
40. Mozzherin, D. J., McConnell, M., Jasko, M. V., Krayevsky, A. A., Tan, C.-K., Downey, K. M. & Fisher, P. A. (1996) *J. Biol. Chem.* **271**, 31711–31717.
41. Shcherbakova, P. V. & Pavlov, Y. I. (1996) *Genetics* **142**, 717–726.
42. Izuta, S., Roberts, J. D. & Kunkel, T. A. (1995) *J. Biol. Chem.* **270**, 2595–2600.
43. Schaaper, R. M. & Dunn, R. L. (1991) *Genetics* **129**, 317–326.