Transcription increases multiple spontaneous point mutations in *Salmonella enterica*

Richard Ellis Hudson¹,², Ulfar Bergthorsson² and Howard Ochman¹,*

¹Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA and
²Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received April 2, 2003; Revised May 31, 2003; Accepted June 10, 2003

ABSTRACT

The spontaneous rate of G→C→A→T mutations and a hotspot T→A→G→C transversion are known to increase with the frequency of transcription—increases that have been ascribed primarily to processes that affect only these specific mutations. To investigate how transcription induces other spontaneous point mutations, we tested for its effects in repair-proficient *Salmonella enterica* using reversion assays of chromosomally inserted alleles. Our results indicate that transcription increases rates of all tested point mutations in the induced gene: induction significantly increased the individual rates of an A→T→T→A transversion, an A→T→G→C transition and the pooled rates of the three other point mutations assayed. Although the *S.enterica* genome is thought to have a mutational bias towards G→C base pairs, transitions creating A→T pairs were approximately 10 times more frequent than the reverse mutation, resulting in an overall mutation pressure to lower G+C contents. Transitions occurred at roughly twice the rate of transversions, similar to results from sequence comparisons; however, several individual transversions are more frequent than the least common transition.

INTRODUCTION

Spontaneous mutations, like those induced by mutagens, should become more frequent with increased transcription (1–3) because DNA becomes locally exposed during synthesis of RNA transcripts. In initial support of this supposition, reversions of an *Escherichia coli* trpA strain were found to be more common in a constitutive mutant (4). However, a similar transcription effect was not observed in a different strain, suggesting that the result was an artifact of the use of potentially misleading statistic—average mutants per culture—rather than more reliable alternatives. More robust studies have since shown that elevating transcription levels can influence varying types of mutations: for example, as transcription rates increase, frameshift reversions in yeast are more frequent (5,6), whereas chromosomal rearrangements and deletions in a human cell lines are less common (7).

Transcription is known to increase the spontaneous rates of certain point mutations in *E.coli*: C→T reversions in both plasmid-borne and chromosomal genes (8,9), and a T→A→G→C forward mutation at a mutation hotspot in an alternate plasmid-borne gene (10). In addition, transcription promotes the spontaneous formation of three forward mutations due to non-C→T point substitutions in *E.coli* cells situated in different sequence contexts (11). Although rates of point mutations were elevated by transcription in each of these cases, other mutations might be reduced or unaffected, because each substitution rate is set by mutagenic and antimutagenic processes that do not necessarily affect other substitutions. For example, the elevated C→T transition rates are attributable to the increased susceptibility of cytosines on single-stranded DNA to deamination (8,12), a process that does not significantly affect other mutations. If the rate of the predominant mutagenic processes for a particular mutation is not significantly altered by transcription, the antimutagenic effect of transcription-coupled repair (13,14) could overwhelm them and reduce mutations with increased expression (15), as appears to be the case in mammalian cells (7,16).

To investigate whether there are general effects of transcription on spontaneous base substitutions, we measured the rates of several point mutations under both induced and non-induced conditions. Our findings suggest that heightened transcription increases mutagenesis for all substitutions tested and that for this class of mutations, this effect always overrides that of transcription-coupled repair in bacteria.

Differences among point mutation rates can also affect the overall G+C content of a genome. For example, when C→T transitions occur at a higher rate than T→C transitions, A→T pairs will accumulate at neutral sites, such as non-coding DNA, and degenerate codon positions. Experimental data suggest a mutational bias towards A→T pairs in *E.coli* (17–21) making it difficult to explain the observed preponderence towards G→C base pairs at third codon positions (22). By measuring mutation rates for all possible point mutations, we found a highly significant bias towards the formation of AT pairs among transitions, but no significant bias among transversions, which is consistent with experimental rather than comparative sequence data.

*To whom correspondence should be addressed. Tel: +1 520 626 8355; Fax: +1 520 621 3709; Email: hochman@email.arizona.edu*
To test for differences in overall mutation rates among strains, we assayed for mutations resistant to 100 μg/ml rifampicin, using four batches per treatment of eighteen 0.3 ml cultures. Each culture was inoculated with ~10^6 cells and incubated for 6.5 h at 37°C with shaking (200 r.p.m.). Three cultures per batch were serially diluted and plated for total viable counts, and the rest were spread directly onto LB–rifampicin plates.

### Calculation of mutation rates

Mutants arising from exponential-phase mutations are expected to follow a Luria–Delbrück (LD) distribution (26), whereas those arising in stationary phase and after plating are expected to be Poisson distributed. Because we anticipate both exponential and post-exponential-phase mutations, we investigated the fit of combined Poisson–LD distributions to the data of each batch. Such distributions are characterized by two parameters: m, the overall number of mutations per culture, and θ, the fraction of mutations following the Poisson rather than the LD distribution. When compared to pure LD distributions, combined distributions can fit the data better even if no post-exponential-phase mutations actually occur because they employ an additional parameter. Therefore, our default assumption was that there were no post-exponential-phase mutations (θ = 0). However, if the best-fitting combined distribution had a likelihood 10-fold greater than that of the pure LD distribution, we applied its value of θ. To determine θ for each treatment—a given strain grown under a particular IPTG concentration—we averaged estimates derived from batches containing >36 cultures. (Estimates from smaller batches were ignored because they yielded much less accurate estimates.) The values of θ thus determined were then used in generating maximum likelihood estimates of overall (m) and exponential-phase (m_1 = 1 − θ) × m) mutations per culture for each batch. Overall and exponential-phase mutation rates for each treatment were calculated using the following formula:

$$\mu_{\text{treatment}} = \exp \left[ \frac{1}{N} \sum_{i=1}^{N} \ln \left( m_i / n_i \right) \right]$$
Calculating T’A posteriori comparisons of means, the previously observed increases in spontaneous reversion frequency, was only measured under non-induced conditions. The C’G the rates of these three mutations were indistinguishable statistically. The G’C transition increased 2.0 rather than 4.9 times. However, both of the results remained statistically significant whether or not we included post-exponential-phase mutations. The three remaining mutations—the G’C→A’T, C’G→A’T and T’A→G’C reversions—were unaffected by the requirement that \( \theta = 0 \). For these mutations, exponential-phase and overall rates were higher in the presence of IPTG, although individually none reached statistical significance. However, there is a statistically significant overall effect of induction in an analysis restricted to these mutations (\( P < 0.01 \), two-factor ANOVA). To test whether IPTG, in addition to inducing lac expression, might cause a genome-wide increase in mutation rates, we also assayed mutations to rifampicin resistance in the presence and absence of IPTG (Table 2). None of the strains tested has a significantly different mutation rate to rifampicin resistance in the presence of IPTG nor was there a significant effect when the data from different strains were pooled, or subjected to an analysis of variance (\( P = 0.52 \)) (Table 2).

In the presence and absence of inducer, the relative frequencies of mutations (except T’A→A’T) were the same for both the exponential-phase and overall mutation rates: \( \mu_{G’C\rightarrow A’T} > \mu_{C’G\rightarrow A’T} = \mu_{T’A\rightarrow G’C} > \mu_{A’T\rightarrow G’C} \), where all differences were significant (Tukey–Kramer test for a posteriori comparisons of means, \( P < 0.05 \)). The rate of the T’A→A’T exponential-phase mutation in the absence of inducer was significantly lower than that of the C’G→A’T and T’A→G’C mutations; however, in the presence of inducer, the rates of these three mutations were indistinguishable statistically. The C’G→G’C mutation rate was virtually undetectable in our system and, due to its extremely low frequency, was only measured under non-induced conditions.

**RESULTS**

**Effects of transcription on mutation rates**

To determine the extent to which transcription affects particular point mutations, reversion rates were calculated for mutant lacZ alleles in the presence and absence of inducer (Fig. 1, Table 1). Exponential-phase and overall mutation rates were significantly higher in the presence of IPTG for the T’A→A’T transition and the A’T→G’C transition. When the assumption that some mutations occurred after exponential-phase was relaxed, that is when \( \theta \) was required to equal zero, some increases with induction were less pronounced: the rate of the T’A→A’T transition rose 1.9- rather than 2.6-fold, and that of the A’T→G’C transition increased 2.0 rather than 4.9 times. However, both of the results remained statistically significant whether or not we included post-exponential-phase mutations. The three remaining mutations—the G’C→A’T, C’G→A’T and T’A→G’C reversions—were unaffected by the requirement that \( \theta = 0 \). For these mutations, exponential-phase and overall rates were higher in the presence of IPTG, although individually none reached statistical significance. However, there is a statistically significant overall effect of induction in an analysis restricted to these mutations (\( P < 0.01 \), two-factor ANOVA).

To test whether IPTG, in addition to inducing lac expression, might cause a genome-wide increase in mutation rates, we also assayed mutations to rifampicin resistance in the presence and absence of IPTG (Table 2). None of the strains tested has a significantly different mutation rate to rifampicin resistance in the presence of IPTG nor was there a significant effect when the data from different strains were pooled, or subjected to an analysis of variance (\( P = 0.52 \)) (Table 2).

In the presence and absence of inducer, the relative frequencies of mutations (except T’A→A’T) were the same for both the exponential-phase and overall mutation rates: \( \mu_{G’C\rightarrow A’T} > \mu_{C’G\rightarrow A’T} = \mu_{T’A\rightarrow G’C} > \mu_{A’T\rightarrow G’C} \), where all differences were significant (Tukey–Kramer test for a posteriori comparisons of means, \( P < 0.05 \)). The rate of the T’A→A’T exponential-phase mutation in the absence of inducer was significantly lower than that of the C’G→A’T and T’A→G’C mutations; however, in the presence of inducer, the rates of these three mutations were indistinguishable statistically. The C’G→G’C mutation rate was virtually undetectable in our system and, due to its extremely low frequency, was only measured under non-induced conditions.

**DISCUSSION**

The previously observed increases in spontaneous reversion rates of C→T transitions and a particular T’A→G’C transversion with transcription have both been attributed to specific mutagenic processes: cytosine deamination in the first case (8,12) and a mechanism involving local six-base pseudo-inverted repeats in the second (10). We have found that transcription also causes a highly significant increase in the rate of the T’A→A’T transversion and the A’T→G’C transition, mutations that cannot be attributed to deamination or to the presence of local sequence features. Our findings complement and extend previous work showing that spontaneous mutation rates of ble and kan gene constructs that mutate by the G’C→T’A transversion (and other non-C→T base substitutions) are elevated by transcription (11). Different spontaneous mutations were primarily affected in our system, and we were able to statistically test the effects of transcription on individual mutations. Moreover, by assaying chromosomal mutations in repair-proficient strains (as opposed to plasmid-borne mutations in uracil-DNA-glycosylase-deficient strain), we were able to investigate mutation rates that were lower, and more representative, of natural spontaneous mutations.

To test for a general effect of transcription on the remaining substitutions, we combined the reversion rates of the G’C→A’T, T’A→G’C and T’A→G’C mutations into a single ANOVA. Whereas the increases in the rates of these mutations did not reach statistical significance when analyzed individually, there is a statistically significant effect when these data are pooled. Cumulatively, this outcome, along with the results for other mutations in this and other studies, suggests that there are general mechanisms by which transcription augments rates of spontaneous point mutations.

The increase in mutation rates with gene expression is observed whether or not post-plating mutations were estimated to have occurred. In all but two of the cases, the frequency of post-plating mutations (\( \theta \)) was calculated to be zero, arguing against the possibility that substantial numbers of mutants arose from plated cells sustained by nutrients derived from dying bacteria. However, the frequency of non-induced T’A→A’T transversions had a non-zero value of \( \theta \): the likelihood of the distribution assuming \( \theta = 0.77 \) was over 700 times greater than that of the distribution where \( \theta = 0 \), thus indicating that post-exponential-phase mutations were detectable. In this case, the presence of Poisson-distributed mutations likely reflects stationary-phase rather than post-plating mutations because the growth rate (under non-induced conditions) of the strain harboring this allele was among the highest of all strains tested, and cultures had entered stationary phase prior to plating.

In addition to inducing transcription of the lac operon, the addition of IPTG reduced the total number of cells by 23%; but it is not likely that this decrease contributes to the increased lacZ reversion rates for three reasons. (i) Control mutation rates tested at a non-induced locus did not increase (Table 2). (ii) Spontaneous reversion rates of lacZ alleles are not affected when S.enterica cells are stressed or starved (27). (iii) There is no significant negative correlation between total cell counts and mutation rates among batches of the same strain (\( r = -0.06, N = 44, P = 0.71 \)).

The rates of both transitions investigated, the A’T→G’C and G’C→A’T mutations, were elevated by transcription, and in both cases, the pyrimidine, the base where the mutation was likely initiated, was situated on the transcribed strand. Although the transcribed strand spends less time than its counterpart in the vulnerable single-stranded state [and is therefore considered to be less prone to damage (8)], these
results suggest that it also exhibits higher rates of damage during transcription. The most plausible explanation for the observed increase of multiple types of spontaneous substitutions is an elevation in rates of a variety of damages with transcription on both DNA strands, and not just those caused by deamination or errors generated by local repeats. For example, the increase in the forward mutation rates of the \textit{ble} and \textit{kan} gene constructs (11) are mainly attributable to more frequent G$\rightarrow$T transversions on the non-transcribed strand caused by either oxidative damage or the hydrolysis of the glycosidic linkage of guanosine residues. Because neither of these processes, nor those already mentioned, can account for the significant increase in the rate of the T$\rightarrow$A$\rightarrow$T transversion detected in our system, this mutation must be due to yet another transcription-sensitive damage process. These results therefore imply that the increased susceptibility to damage due to the disassociation of the DNA strands normally overrules any reduction attributable to transcription-coupled repair.

Both chromosomal location and immediate sequence context are also known to affect rates of point mutations (23,28). In particular, substitutions at specific hot spots occur at significantly higher rates than those at other positions and can dominate mutational spectra (28). However, most mutable sites are not hotspots and have spontaneous mutation rates of similar magnitude, as observed for the 91 of 93 detectable sites in the \textit{rpsL} gene of \textit{E.coli} (28). This phenomenon, coupled with the observation that the reversion rates that we estimated are much lower than those typical of a mutational hotspot, suggests that the single-base substitutions we assayed are characteristic of other mutations in this class.

In our system, there was a significant cumulative bias towards the formation of A$\rightarrow$T pairs, which would elevate the A+T content of the genome. This is due to the G$\rightarrow$C$\rightarrow$A$\rightarrow$T transition, which occurs at 10-fold higher than the reverse mutation. This trend among transitions towards A$\rightarrow$T base pairs has previously been observed in reversion-assay studies of the same alleles when borne on episomes (18,25,29) and on the \textit{E.coli} chromosome (30). In contrast, there is no compositional bias among transversions because the rate of C$\rightarrow$G$\rightarrow$A$\rightarrow$T is approximately equal to that of the reverse mutation, and the other transversions do not alter base composition.

It is also possible to deduce mutational biases from the spectrum of point mutations that inactivate a particular gene. Such studies provide complementary information to reversion assays because the same point mutation can be observed in numerous sequence positions and contexts. Although it is difficult to statistically compare mutation rates derived from such studies (because computing mutation rates would require the identification of all potential target sites), the trends detected among spectra of spontaneous mutations match those detected by reversion assays. For example, in studies employing the \textit{lacI}, \textit{supF}, \textit{rpsL} and \textit{tonB} genes of \textit{E.coli}, C$\rightarrow$T transitions occur at higher per-target frequencies and are
Table 3. Mutational spectra of enterobacterial genes

<table>
<thead>
<tr>
<th>Mutation</th>
<th>lacI</th>
<th>rpsL</th>
<th>supF</th>
<th>tonB</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-G→T-A</td>
<td>182</td>
<td>78</td>
<td>3.3</td>
<td>6</td>
</tr>
<tr>
<td>T-A→C-G</td>
<td>43</td>
<td>29</td>
<td>1.5</td>
<td>2.9</td>
</tr>
<tr>
<td>G-C→A-T</td>
<td>36</td>
<td>74</td>
<td>0.5</td>
<td>4.0</td>
</tr>
<tr>
<td>A-T→C-G</td>
<td>55</td>
<td>43</td>
<td>1.3</td>
<td>0.9</td>
</tr>
<tr>
<td>C-G→T-C</td>
<td>10</td>
<td>50</td>
<td>0.2</td>
<td>3.5</td>
</tr>
<tr>
<td>Transitions</td>
<td>225</td>
<td>107</td>
<td>2.1</td>
<td>172</td>
</tr>
<tr>
<td>Transversions</td>
<td>148</td>
<td>220</td>
<td>0.7</td>
<td>149</td>
</tr>
<tr>
<td>Total</td>
<td>373</td>
<td>321</td>
<td>45</td>
<td>17</td>
</tr>
</tbody>
</table>

*Columns present numbers of mutants (M) isolated for the specified number, numbers of target sites (T) for each mutation and the fraction of mutants per target sites (M/T). Mutants were determined using wild type strains and target sites were identified using both mutant and wild type strains (although some targets may still escape detection), un, values undefined. Data compiled from: Schaaper and Dunn (19), Halliday and Glickman (20 and references therein) (for lacI), Mo et al. (38), Timms et al. (39), Yoshiyama et al. (40), Yoshiyama and Maki (10) (for rpsL), Akasaka et al. (21), Obata et al. (22) (for supF), Yamamura et al. (41) (for tonB).*

more common than T→C transitions (Table 3). A similar A-T bias has also been observed in studies of human and yeast cells, and in a coliphage (31-34).

We detected no significant bias for (or against) A-T pairs among transversions. Previous reversion assay studies yield similar results (18,25,29) (Table 3); however, the evidence from mutational spectra is contradictory; in the supF and rpsL genes, there appears to be a bias for A-T pairs, but in the lacI gene the opposite bias is observed (Table 3).

The spontaneous transitions assayed in this study occur at nearly twice the rate of transversions (Table 1). As observed previously for these lacZ alleles (18,25,29), ranges of the transition and transversion rates overlap. The C-G→G-C transition occurs at a significantly lower rate than all other reversions examined and is also the least common point mutation detected in two of four mutational spectra (Table 3).

We have documented a general mutagenic effect of transcription on spontaneous point mutations in the induced gene and a mutational bias towards A-T pairs among spontaneous transitions. Though such results reflect and extend many of the trends observed in other experimental studies, both results conflict with previously held views about the mutational process in enteric bacteria based on the analysis of sequence data in two ways:

First, comparative analyses of *E.coli* and *S.enterica* homologs suggest that transcription reduces spontaneous point mutation rates (35,36) because synonymous substitution rates, after accounting for the effects of adaptive codon bias, decrease as indicators of transcription frequency increase. In contrast, our experimental data indicate that the rates of potentially all point mutations increase with transcription. Our results suggest that theoretical studies have overestimated the degree of selection acting on synonymous sites in highly expressed genes, thereby underestimating their mutation rates.

Second, the sequence compositions of *E.coli* and *S.enterica* genes suggest an intrinsic mutational bias against A-T pairs (23,37) because third codon positions, where most sites are under no selective constraints and should reflect the underlying mutational process, are G+C rich. It is possible that the A-T mutational bias observed in rapidly growing, aerobic, experimental populations is not representative of the mutational process in nature. Alternatively, these enteric species could be undergoing a shift in base composition that is not apparent from the cumulative measures of G+C contents at codon positions. Changes in the patterns of mutations within a species can be exposed by reconstructing the substitutions in a phylogenetic context, and such studies, as well as the analysis of mutational patterns under more natural growth conditions, will help reconcile the differences between the experimental and genomic measures of mutational bias.

**ACKNOWLEDGEMENTS**

The authors thank John Roth for support, advice and bacterial strains, and Frank Stewart and Phil Gerrish for providing computer programs. This research was supported by NIH grants GM55535 and GM56120 (to H.O.), and U.B. was supported by NIH grant GM27068 (to J.R.Roth).

**REFERENCES**


