Slipped-Strand Mispairing Can Function as a Phase Variation Mechanism in *Escherichia coli*

Joshua Torres-Cruz and Marjan W. van der Woude*

Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received 6 June 2003/Accepted 10 September 2003

Slipped-strand mispairing (SSM) has not been identified as a mechanism of phase variation in *Escherichia coli*. Using a reporter gene, we show that sequences that cause phase variation by SSM in *Haemophilus influenzae* also lead to phase variation when introduced onto the chromosome of *E. coli*, and the frequencies of switching are in the biologically relevant range. Thus, the absence of SSM-mediated phase variation in *E. coli* does not appear to be due to a mechanistic constraint.

Phase variation is a heritable, but reversible, form of gene regulation in bacteria. One of the mechanisms that can result in phase variation is slipped-strand mispairing (SSM), which is a process that produces mispairing between the mother and daughter strand during DNA replication (11, 14). The genomic regions susceptible to SSM are those that contain a short, contiguous homogenous or heterogeneous repetitive DNA sequence of 6 bp or less, designated short sequence repeats (SSR), microsatellites, or “variable number of tandem repeats” (8, 27). SSM at these regions will result in a change in the number of unit repeats, which can lead to altered gene expression at either the transcriptional or translational level, depending on the position of the repeats relative to the promoter and coding sequence. Analyses of the *Haemophilus influenzae* strain Rd genome indicated that perhaps as many as 19 loci could be susceptible to SSM-mediated phase variation at tetranucleotide repeats (9, 12, 28). At least five of these are indeed subject to this regulation (6, 12, 30, 31). Similarly, in *Neisseria meningitidis* as many as 65 loci may be susceptible to SSM (22), and 27 loci in *Helicobacter pylori* may be susceptible (23, 26). In contrast, in the *Escherichia coli* genome, runs of significant length of repeat units of one to four nucleotides rarely occur. Even though slippage may occur at these rare sequences, they are not located in a promoter or coding sequence, suggesting this does not lead to phase variation (8, 13, 16). Slippage, specifically expansion, can occur at a four-unit repeat of a trinucleotide sequence in the *ahpC* gene at a frequency similar to that obtained in SSM-mediated phase variation. This one-unit expansion affects the specificity of the AhpC enzyme, changing it from a peroxiredoxin to a disulfide reductase (20). However, only site-specific recombination and DNA methylation have been identified as regulatory mechanisms for phase variation in *E. coli* (reviewed in references 4 and 11).

The apparent absence of SSM-mediated phase variation in *E. coli* raises the question whether phase variation as a result of SSM is possible in *E. coli* (6). A higher *mutSLH*-dependent repair fidelity or different *polI* characteristics in *E. coli*, compared to *H. influenzae*, could potentially result in a decrease of occurrence of SSM at mono- and dinucleotide tracts or at tetranucleotide repeat tracts, respectively (2, 19). Thus, the rate of SSM in *E. coli* may not be relevant for phase-variable genes. Studies on instability of repeat sequences in *E. coli* have mainly been as a model system for eukaryotic genomes (18, 21, 32). However, length of the repeat unit, sequence, and repeat numbers of naturally occurring SSR are different in bacterial genomes, and these characteristics affect stability (7, 8, 13). Therefore, we examined whether instability at di- and tetranucleotide repeat regions that are known to be relevant for phase variation in *H. influenzae* leads to phase variation with a biologically relevant switch frequency in *E. coli*.

Specifically, we examined phase variation in *E. coli* mediated by the SSR-containing regions of the *hif* and *mod* genes of *H. influenzae*. The expression of *H. influenzae* fimbiiae is subject to phase variation that is controlled at the level of transcription of *hifA* and *hifB*, encoding the major fimbrial subunit and the fimbrial chaperone, respectively. The *hifA* and *hifB* promoter regions overlap and contain repeats of the dinucleotide TA between the −35 and −10 sequences (Fig. 1A). Variation in the number of TA units changes the spacing between the −35 and −10 sequences and thus the promoter strength, which affects *hifA* and *hifB* transcription. Changes in the number of these dinucleotide repeats occurs on plasmids in *E. coli*, but the frequency of these changes was not determined (29). The *mod* gene of *H. influenzae* strain Rd, which has homology to DNA methyltransferases of type III restriction modification systems, has tetranucleotide repeats (5'-AGTC) in its coding region (Fig. 1B). Loss or gain of tetranucleotide repeats results in a frameshift, and thus Mod phase variation is regulated at the level of translation (6).

Our approach to investigate phase variation mediated by SSM in *E. coli* was to determine the effect of the *H. influenzae hif* and *mod* SSR sequences in *E. coli* using single-copy LacZ reporter fusions. Standard genetic manipulations and techniques were used (1, 15, 17). Sequencing of PCR products and plasmids was performed by the Genetics Core Facility at the University of Pennsylvania. The SSR-containing regions were amplified from *H. influenzae* genomic DNA and cloned into either pRSS550 or pRSS552 and recombined with ARS45. Single-copy lysogens were isolated which contain the relevant *lacZ*...
transcriptional or translational fusion as a recombinant phage integrated at the att site on the chromosome (25). In E. coli strain MV792, a 1,071-bp fragment (nucleotide [nt] 1218 to nt 2277 of GenBank accession number Z33502) with the TA-repeat-containing mod region of H. influenzae AM30, drives expression of a single-copy transcriptional lacZ fusion (Fig. 1A). In strain MV759, a 1,030-bp fragment containing the mod regulatory region of strain Rd and part of the coding sequence containing the AGTC repeats in the coding sequence (Fig. 1A). In strain MV759, a 1,030-bp fragment containing the mod regulatory region of strain Rd and part of the coding sequence containing the AGTC repeats in the coding sequence (Fig. 1A). In strain MV759, a 1,030-bp fragment containing the mod regulatory region of strain Rd and part of the coding sequence containing the AGTC repeats in the coding sequence (Fig. 1A).

To examine Lac phenotypes, strains were grown on M9 medium with glycerol as a carbon source and with 5-bromo-4-chloro-3-indolyl-β-D-galactoside and antibiotics as necessary (10, 15, 24). Colonies of E. coli strain MV792 carrying the hifA-lacZ transcriptional fusion showed phase-variable LacZ expression. The frequency of the switch of Lac+ (ON) to Lac- (OFF) was determined as previously described (5), and results are shown in Table 1. In E. coli, the switch frequencies for this sequence containing a dinucleotide repeat were two- to threefold lower than in H. influenzae (Table 1) (29). The H. influenzae switch frequencies were calculated based on fimbrial production, and those frequencies reflect not only transcriptional regulation but also regulation of translation and fimbrial assembly and, thus, a higher number of cells may be regarded as OFF. Therefore, the two approaches may contribute to the observed minor differences in the switch frequencies between the two species. Alternatively, the differences can be attributed to the different genetic backgrounds.

The number of TA repeats in the hif promoter region of ON

and OFF phase colonies was determined by sequencing the repeat region from several Lac+ (ON) and Lac- (OFF) colonies. Changes in the number of repetitive TA base pairs spanning the promoter region were correlated with the expression phase. Specifically, four ON colonies each contained 10 repeats of the TA dinucleotide sequence, which results in an optimal spacing of 16 bp between the −10 and −35 sequence of the hifA promoter. In contrast, four sequenced OFF colonies contained nine TA repeats, which results in a suboptimal spacing of 14 bp. The results were not dependent on the Lac status of the parent colony. The same correlation was previously found in H. influenzae (29). In H. influenzae, 11 unit repeats were also found, which led to an intermediate level of fimbriae expression (29). At a frequency that was lower than the ON-to-OFF switch, colonies of strain MV792 were identi-
fied with intermediate levels of LacZ expression. Cultures inoculated with these colonies had 853 ± 34 Miller units (MU) of LacZ activity, in between that of a predominantly OFF culture (21 ± 4 MU) and ON culture (1,719 ± 194 MU). However, sequence analysis showed 10×(TA) repeats, the same as found in the ON colonies. Either the 11×(TA) repeat is very unstable, or this colony phenotype was a result of a different regulatory event. The data nevertheless showed that changes in the number of unit repeats in a dinucleotide (TA) tract can cause phase variation in E. coli.

Strain MV757, containing the mod-lacZ fusion but not the AGTC repeat region, gave rise to homogenous, Lac+ colonies (Fig. 1B). The level of LacZ expression from the promoter was 44 ± 3 MU, showing the H. influenzae mod promoter is functional in E. coli. Strain MV759 contains the mod-lacZ fusion with the repeat region. If there are 32 repeats of the AGTC region, an in-frame translation product with LacZ is formed, leading to a Lac+ colony phenotype. Any change in unit repeat numbers in multimers of three will also be in frame, whereas all other changes lead to an out-of-frame fusion and a Lac− phenotype. Strain MV759 showed phase variation of LacZ expression. A culture with 93% ON cells had an average of 5.7 ± 0.7 MU of β-galactosidase activity, whereas a culture with 90% OFF cells had an approximately fivefold decrease in activity to 1.3 ± 0.4 MU. As a comparison, a lysogen obtained with the pRS552 vector with no insert, and thus no promoter or ribosome binding site preceding lacZ, gives rise to less than 0.3 MU of LacZ activity. The level of activity in ON cells of strain MV759 is lower than the level of expression from the short fusion in strain MV757. This suggests that the repeat region may also affect LacZ expression of an in-frame mod-lacZ fusion by different means, for example, by mRNA stability or translational efficiency as a result of the specific codon usage in the repeat region.

The switch frequencies for strain MV759 were determined and are shown in Table 1. Shown also are the rates that were previously reported for phase variation of the mod region with 37 and 38 unit repeats in H. influenzae (6). The rates of phase variation as a result of SSM at the tetranucleotide repeat region in mod in the two species are different, for example, the ON-to-OFF rate in E. coli is 50-fold higher (Table 1). The rates in H. influenzae were derived using different calculation methods than the one used here, including the Drake method (6). However, frequencies calculated with the Drake method using our data yield only a twofold higher rate than with the method used here. The calculation method, therefore, cannot account for the observed difference in switching frequencies. The molecular basis for the observed difference is not known, but promoter activity, differences in the DNA polymerase, or different involvement of recombination or DNA repair may be involved. Regardless, the results show that the frequency of phase variation at a tetranucleotide repeat in the E. coli background is in the order of magnitude of the naturally occurring switch frequency in E. coli of, for example, the phase-varying gene agr43 (10). Thus, SSM at this tetranucleotide sequence can lead to phase variation in E. coli with switch frequencies that are biologically relevant.

The variable Lac phenotype of strain MV759 indicated that a change in the number of repeats had occurred, presumably as a result of SSM. The repeat tract was amplified with a total of 102 bp flanking DNA, so that an in-frame 32×(AGTC)-repeat-containing region yields a 230-bp fragment. Analysis of the size of these PCR-derived products on an 8% polyacrylamide gel of the mod-lacZ region of multiple individual colonies confirmed that the number of AGTC repeat units had changed (Fig. 2). The precise number of repeat units was determined for several samples by sequence analysis of the PCR products, and in each case the results showed that the number of unit repeats led to an in- or out-of-frame protein fusion that correlated to the Lac colony phenotype. For example, a randomly chosen Lac+ colony had 38 unit repeats. This is higher than the range of 2 to

![FIG. 2. Analysis of the size of the mod-lacZ region in strain MV759 shows variation in the number of tetranucleotide repeats. (A) Size analysis of PCR products of the mod regulatory region of strain MV759 of a Lac+ parent and eight Lac+ and two Lac− daughter colonies. (B) Diagram showing lineage of PCR products of the mod region shown in panel C. Lac+ colonies are depicted as solid circles, and Lac− colonies are depicted as open circles. Single colonies were serially passaged nine times based on Lac phenotype, according to the four strategies shown. The numbers within circles correspond to the lane numbers in panel C, in which the mod PCR product of a single colony with the given heritage is shown. (C) Size analysis of PCR products from colonies of strain MV759 as specified in panel B. The phenotype of the colony and its lineage are also given. Based on sequence analysis, a 230-bp PCR product containing a 32×(AGTC) repeat tract is shown in lane 3, a 286-bp fragment [46×(AGTC)] is in lane 10, a 258-bp fragment [39×(AGTC)] is in lane 14, and a 254-bp fragment [38×(AGTC)] is in lane 15 (also see text). Sizes of the fragments in the standard (lane 9) are denoted to the left.
23 unit repeats that was found in various *H. influenzae* strains, but when 38 unit repeats were introduced onto the *H. influenzae* genome this resulted in phase variation as well (6). Analysis of eight Lac- daughter colonies of this MV759 Lac+ parent colony showed that, in three daughters, a one-unit expansion had occurred, based on sequence analysis of the lane 2 product (Fig. 2A). In the other five daughter colonies, a contraction had occurred (Fig. 2A). Based on sequence analysis of the sample shown in lane 5, this was a two-unit contraction. The fluctuation in number of repeat units was also followed over nine serial passages, with an average of 24 generations per passage. Different strategies for passage were used, and the correspondence between the lineage and length of PCR product of randomly chosen, Lac+ or Lac- colonies at the second and ninth passage is shown in Fig. 2B and C. Based on sequence analysis, the products from the ON colonies shown in Fig. 2B and C. Based on sequence analysis, the products from the ON colonies shown in lanes 6 and 15 had 32× and 38×(AGTC) repeat tracts, respectively, which resulted in an in-frame translational product. Conversely, the products from the OFF colony shown in lanes 10 and 13 contained 46×(AGTC) repeat tracts and the product shown in lane 14 contained 39×(AGTC) repeat tracts, both resulting in an out-of-frame fusion to LacZ. The larger PCR products did not run exactly true to size, perhaps due to DNA bending induced by the repeat sequence. The data in Fig. 2C suggested that over time a strong bias existed toward a net increase in the length of the AGTC tract. However, of 10 Lac- daughter colonies of a Lac+ parent in the 13th passage, 4 showed a decrease in the number of repeat units (data not shown), indicating that this is not the case. Taken together, the data show that both expansion and contraction of tetranucleotide repeats occur in an *E. coli* background.

To summarize, our results show that variation in length of short di- and tetranucleotide repeat sequences can occur in *E. coli* and can lead to phase variation. Our data are consistent with the models for transcriptional and translational regulation of phase variation at *hif* and *mod* in *H. influenzae* (6, 29). The phenotypic switch frequencies at the dinucleotide repeat sequence in *E. coli* are similar to those determined for this sequence in the native host, *H. influenzae*. In contrast, the instability of the tetranucleotide repeat region was higher in *E. coli* than in *H. influenzae*. Importantly, the switch frequencies at both repeat sequences in *E. coli* were in the range of the naturally occurring phase-varying systems in *E. coli* (3). Thus, SSM in *E. coli* occurs at a rate that can lead to biologically relevant switch frequencies. Furthermore, our data show that the changes in the number of repeat units are not limited to either expansion or contraction. Taken together, the results suggest that the apparent absence of naturally occurring SSM-mediated phase variation in *E. coli* is not a result of mechanistic limitations.

A small genome, like that of *H. influenzae* and *H. pylori*, may place a limitation on the acquisition of regulatory mechanisms that require complex regulatory networks, like the DNA methylation and site-specific recombination-dependent systems that control *E. coli* pap and fim fimbral phase variation, respectively (4). SSM could, therefore, be an advantage to organisms with small genomes. It has also been suggested that bacteria, like these two species that live in a restricted but unpredictable environment, may benefit from random, unregulated phenotypic diversity obtained by SSM (19). Thus, it may also be that the absence or paucity of SSM-mediated phase variation in *E. coli* is driven by a necessity or desirability to superimpose environmental regulation upon the random events of phase variation. This would allow a change in switch frequency in different environments, which could be facilitated by the larger genome. Therefore, the mechanism of choice may be related to the diversity of environments in which the bacterium can survive and grow.

This work was supported in part by grant MCB-0077501 and a Research Experience Undergraduates supplement to J.T.-C. from the National Science Foundation.

We thank Sol Goodgal, Emmanuelle Binet, Anu Wallecha, and Mark Goulian for critical reading of the manuscript and technical guidance.

REFERENCES


