

Mutations Induced by Bacteriophage T7 RNA Polymerase and their Effects on the Composition of the T7 Genome

Anton Beletskii¹, Andrei Grigoriev², Susan Joyce³
and Ashok S. Bhagwat^{1*}

¹Department of Chemistry
Wayne State University
Detroit, MI 48202, USA

²Genome Pharmaceuticals
Corp., Fraunhoferstr., 20
Martinsried, 82152, Germany

³Laboratoire de Génétique
Moléculaire, Ecole Normal
Supérieure, 46 rue D'Ulm
75005 Paris, France

We show here that transcription by the bacteriophage T7 RNA polymerase increases the deamination of cytosine bases in the non-transcribed strand to uracil, causing C to T mutations in that strand. Under optimal conditions, the mutation frequency increases about fivefold over background, and is similar to that seen with the *Escherichia coli* RNA polymerase. Further, we found that a mutant T7 RNA polymerase with a slower rate of elongation caused more cytosine deaminations than its wild-type parent. These results suggest that promoting cytosine deamination in the non-transcribed strand is a general property of transcription in *E. coli* and is dependent on the length of time the transcription bubble stays open during elongation. To see if transcription-induced mutations have influenced the evolution of bacteriophage T7, we analyzed its genome for a bias in base composition. Our analysis showed a significant excess of thymine over cytosine bases in the highly transcribed regions of the genome. Moreover, the average value of this bias correlated well with the levels of transcription of different genomic regions. Our results indicate that transcription-induced mutations have altered the composition of bacteriophage T7 genome and suggest that this may be a significant force in genome evolution.

© 2000 Academic Press

Keywords: cytosine deamination; uracil; transcription; mutagenesis; spontaneous mutations

*Corresponding author

Introduction

The spontaneous mutational spectra of many genes in organisms as diverse as *Escherichia coli*, yeast *Saccharomyces cerevisiae* and humans are dominated by C to T mutations (Coulondre *et al.*, 1978; Greenblatt *et al.*, 1994; Hutchinson & Donnellan, 1998; Kunz *et al.*, 1998). The likely reason for this phenomenon is that cytosine bases hydrolytically deaminate to uracil bases at significant rates (Lindahl, 1993). Unless removed by ura-

cil-DNA glycosylase (UDG), uracils base-pair with adenine during DNA replication, causing C to T mutations. Cytosine molecules in single-stranded DNA deaminate at a rate 140-fold faster than in double-stranded DNA (Frederico *et al.*, 1990). For this reason, cellular processes such as transcription and replication, which transiently separate DNA strands, have the potential to promote C to T mutations.

We have showed that transcription by *E. coli* RNA polymerase (RNAP) promotes C to T mutations in the non-transcribed strand of DNA, but not in the transcribed strand (Beletskii & Bhagwat, 1996). We refer to this phenomenon as transcription-induced mutations (TIM). The mutations resulted from C to U deamination and were largely, but not completely, eliminated by the presence of UDG in cells. Further, the frequency of C to T mutations increased in rough proportion to the level of transcription (Beletskii & Bhagwat, 1998). Under optimal conditions, the level of C to

Present addresses: A. Beletskii, Harvard Institute of Human Genetics, 4 Blackfan Circle, Room 441, Boston, MA 02143, USA; S. Joyce, Department of Microbiology, Moyne Institute of Preventive Medicine, Trinity College Dublin, Dublin 2, Republic of Ireland.

Abbreviations used: UDG, uracil-DNA glycosylase; TIM, transcription-induced mutations; RNAP, RNA polymerase.

E-mail address of the corresponding author:
axb@chem.wayne.edu

T mutations in the non-transcribed strand increased up to tenfold compared to the transcriptionally uninduced state (Beletskii & Bhagwat, 1998). These results are explained by a model in which cytosine bases in the non-template DNA strand within the transcription bubble are exposed to the aqueous environment, promoting their deamination. In contrast, the bases in the transcribed DNA strand are better protected by the RNAP or by the mRNA (Beletskii & Bhagwat, 1996).

It should be noted that Jinks-Robertson and colleagues have shown that frame-shift mutations in the *LYS2* locus increase when the gene is transcribed, and they call these transcription-associated mutations (Datta & Jinks-Robertson, 1995; Morey *et al.*, 2000). The underlying mechanism for this effect is likely to be different from that of TIM, because the increase in the level of yeast mutations is insensitive to the presence of UDG in cells (Morey *et al.*, 2000). Further, the *LYS2* mutations increase in cells deficient in nucleotide-excision repair (Morey *et al.*, 2000), while TIM is not affected by this DNA repair pathway (A. Johnson & A.S.B., unpublished results).

We wished to determine whether other RNA polymerases were capable of causing TIM, and chose bacteriophage T7 RNAP as a test enzyme. This RNAP is encoded by a single gene and has very strong specificity for its own promoters (Studier & Moffatt, 1986). Further, some of the T7 promoters are known to be very strong and this property is advantageous for promoting mutations. Finally, the T7 RNAP is non-essential for cellular growth and hence it is possible to work with mutant RNAPs without affecting cellular growth. Our results suggest that T7 RNAP does promote mutations and that this has affected the evolution of the T7 genome.

Results

Genetic system for studying transcription-induced C to T mutations

We have previously described alleles of the *kan* gene from the transposon Tn5 that revert to *kan*⁺ as

a result of a specific C to T change (Bandaru *et al.*, 1995; Wyszynski *et al.*, 1994). We cloned one of these alleles, *kanS-D94*, under the control of the T7 gene 10 promoter. This arrangement places the mutable C in the non-template strand with respect to the promoter. The transcription unit containing *kanS-D94* allele was then placed on a mini-F plasmid and this episome (pAB6) was maintained in the cells by selecting for the ampicillin-resistance gene within it. Maintaining the phage promoter at (or near) single copy levels was expected to prevent the inhibition of cell growth frequently seen with transcription from strong T7 promoters. Kanamycin-resistant (Kan^R) revertants were scored on plates with kanamycin and the ratio, the number of Kan^R revertants:total number of cells, was determined under different experimental conditions.

An *E. coli* strain used in these experiments, BH161, harbors a single copy of the T7 RNAP gene under control of the *lacUV5* promoter. It also contains the *lacI* gene, which controls expression from the *lacUV5* promoter. Consequently, addition of IPTG to the growth medium induces expression of the T7 RNAP gene and initiates transcription of the *kanS-D94* allele. We used only 100 μM IPTG for the induction of *kanS-D94* transcription because use of IPTG concentrations above 100 μM significantly decreased the growth rate of cells containing pAB6 (not shown).

Transcription by T7 RNA polymerase increases C to T mutations

The addition of IPTG to the growth medium had the expected effect on transcription of pAB6. The amount of *kanS-D94*-specific RNA in BH161 cells increased approximately sevenfold after the addition of IPTG, while IPTG had little effect on RNA levels in cells lacking T7 RNAP (Figure 1).

In BH161, addition of IPTG to the growth medium caused a fivefold increase in the frequency of revertants (Table 1). This increase was reproducible and was similar to the fourfold increase previously seen with the *E. coli tac* promoter (Beletskii & Bhagwat, 1996). As expected, in the host lacking the T7 RNAP gene (strain BH156), IPTG had little

Table 1. Mutations caused by T7 RNA polymerase

T7 RNA polymerase	Culture	Kan ^R revertant frequency		Ratio ^a
		Without IPTG	With IPTG	
No	1	2.0 × 10 ⁻⁸	1.9 × 10 ⁻⁸	1.0
	2	7.9 × 10 ⁻⁹	1.4 × 10 ⁻⁸	1.8
	3	6.4 × 10 ⁻⁸	2.4 × 10 ⁻⁸	0.4
Mean ^b		5.4(±3.0) × 10 ⁻⁸	1.9(±0.5) × 10 ⁻⁸	1.1 ± 0.7
Yes	1	3.5 × 10 ⁻⁸	1.3 × 10 ⁻⁷	3.6
	2	1.8 × 10 ⁻⁸	1.3 × 10 ⁻⁷	6.8
	3	2.6 × 10 ⁻⁸	1.5 × 10 ⁻⁷	5.7
Mean ^b		2.6(±0.9) × 10 ⁻⁸	1.4(±0.1) × 10 ⁻⁷	5.4 ± 1.6

^a Ratio of revertant frequency in the presence of IPTG to revertant frequency in the absence of IPTG for cultures grown from the same colony.

^b Mean (± SD) for three independent cultures is reported.

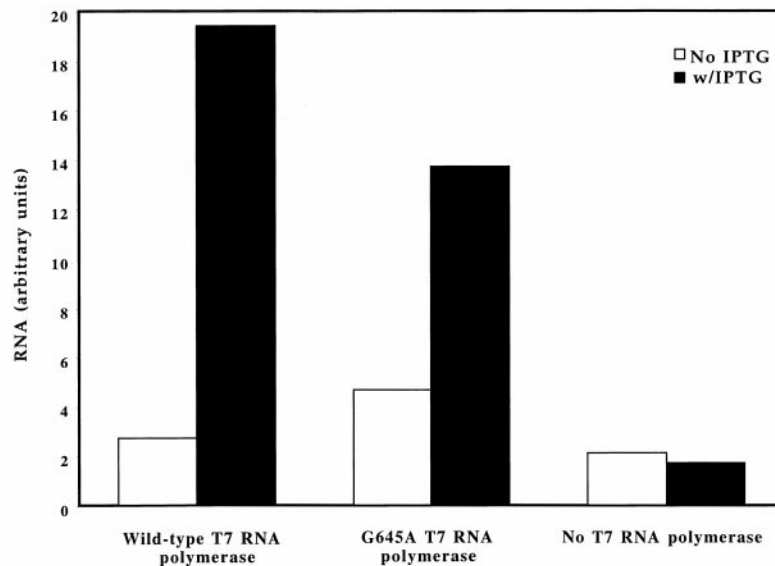


Figure 1. Transcripts produced from the non-coding strand of the *kanS-D94* gene. The levels of RNA in the absence of T7 RNAP, are shown for wild-type RNAP and for the G645A mutant. The RNA amounts are expressed in arbitrary units because they reflect phosphorescence intensities measured by the phosphorimager. The data represent the mean of two independent experiments.

effect on revertant frequency (Table 1). This demonstrates that, similar to the *E. coli* RNAP, transcription by the T7 RNAP also promotes C to T mutations in the non-transcribed strand.

Effect of RNA polymerase elongation rate on TIM

The current model for TIM predicts that the length of time the mutable cytosine base spends in an unpaired state determines the probability of its mutation to thymine. This length of time can be increased in two ways: by increasing the number of times the transcription bubble goes past the cytosine, and by increasing the length of time the base remains unpaired during each passage of the bubble. We previously showed (Beletskii & Bhagwat, 1998) that increasing the frequency of

transcription initiation leads to an increase in C to T mutations. To demonstrate that TIM also depends on the length of time it takes for the transcription bubble to go past a base, we used a mutant T7 RNAP with a reduced rate of elongation.

T7 RNA polymerase mutant G645A has a two- to threefold slower rate of elongation than the wild-type enzyme (Bonner *et al.*, 1994, 1992), but its processivity is close to 80% of wild-type (Makarova *et al.*, 1995). The gene for the G645A mutant was placed under the control of the *lacUV5* promoter by cloning it into the bacteriophage λ cloning vector D69 (Studier & Moffatt, 1986) and then inserting the recombinant bacteriophage into the *E. coli* chromosome as a lysogen (strain BH200). Upon induction of the G645A RNAP, the amount of *kanS-D94*-specific mRNA increased to

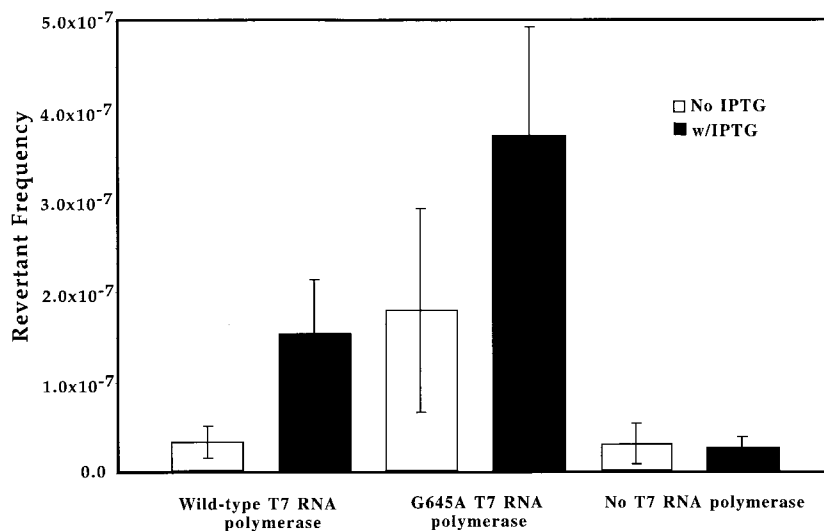


Figure 2. Frequency of Kan^R revertants in the presence and absence of T7 RNAP transcription. The Kan^R revertant frequencies in the absence of T7 RNAP, are shown for wild-type RNAP and for the G645A mutant. The data are means of at least six or more independent colonies from two or more experiments.

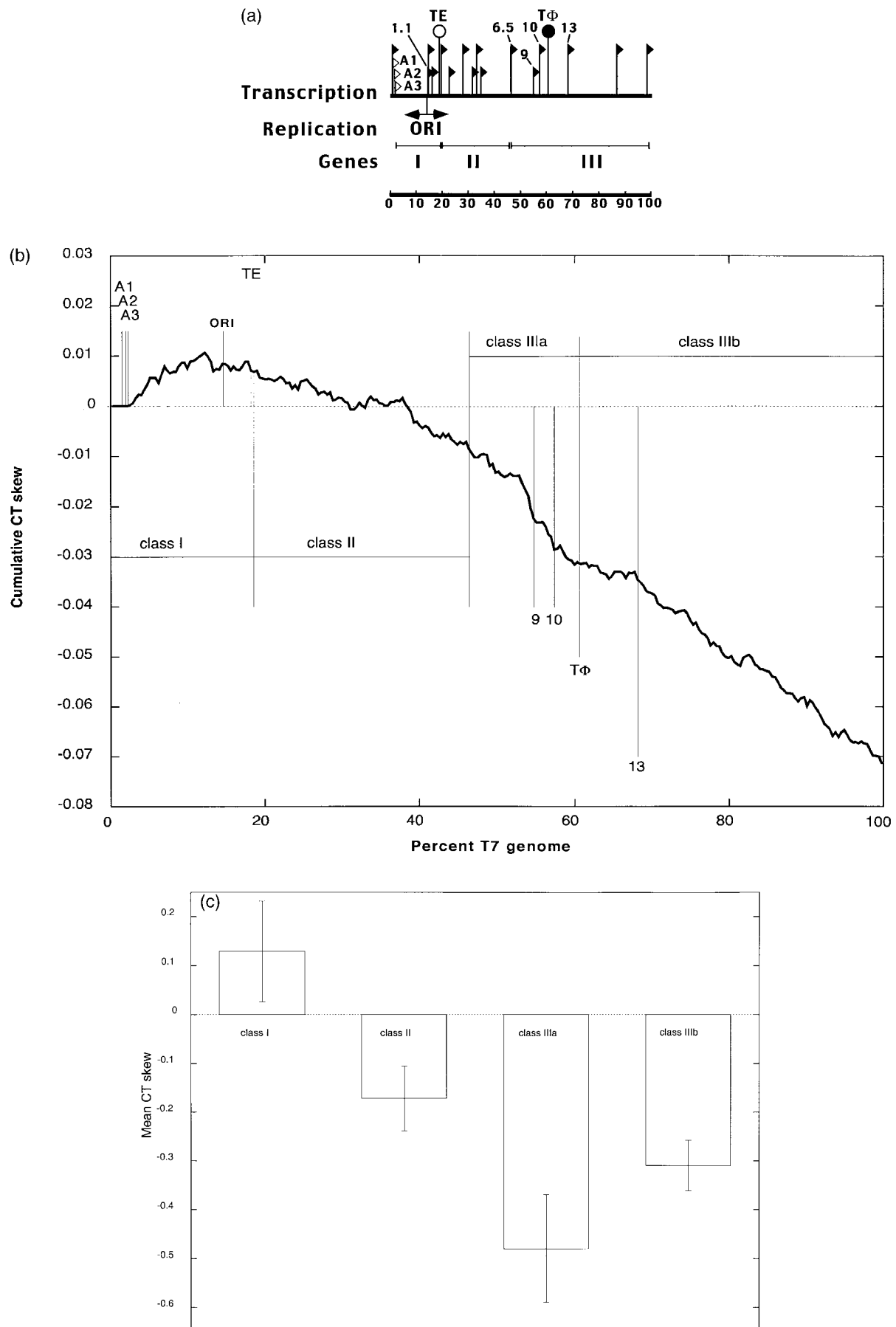


Figure 3 (legend opposite)

~70% of the level found in cells containing the wild-type RNAP (Figure 1). However, in the absence of IPTG, the amount of *kanS-D94*-specific mRNA found in cells with G645A mutant was 1.7-fold higher than that produced by the wild-type polymerase (Figure 1). The high background level of transcription in the mutant RNAP-containing cells can be explained by the fact that while the wild-type RNAP gene is in the DE3 orientation in the λ D69 vector, the G645A mutant gene is in the DE4 orientation (not shown). As a result, the p_1 and p_L promoters can direct the expression of the G645 RNAP gene, but not of the wild-type gene (Studier & Moffatt, 1986).

Regardless, the G645A mutant caused more C to T mutations than the wild-type enzyme and this was independent of the state of induction of the *lacUV5* promoter (Figure 2). It is particularly striking that in the induced state the mutant RNAP produced only 70% *kanS-D94*-specific RNA compared to the wild-type polymerase, but it promoted 2.4 times as many mutations. When normalized with respect to the amount of transcript produced, the G645A mutant promotes ~3.5 times as many mutations as the wild-type enzyme. This is roughly consistent with the two- to threefold lower rate of elongation of the G645A RNAP enzyme (Bonner *et al.*, 1992, 1994).

The effect of TIM on the T7 genome composition

The nucleotide sequence of the T7 genome is known and its genetic organization and overall pattern of gene expression are understood at the nucleotide level (Dunn & Studier, 1983; Studier & Dunn, 1983). The phage DNA contains a number of strong promoters, and the genome is transcribed from left to right on its genetic map (Figure 3(a)). We wondered whether the ability of the T7 RNAP to promote C to T mutations had affected the base composition of the T7 genome.

The total nucleotide count at the third position of fourfold degenerate codons in the T7 genome is 1137 A, 2405 T, 994 G and 1194 C, suggesting that there may be a bias in the non-template strand in favor of thymine bases. To evaluate this further, we defined a novel function, cumulative CT skew, which calculates the excess of cytosine over thymine bases (see Materials and Methods). The pyri-

midines at the third positions of codons in the coding strand were used in this analysis because they are not strongly affected by selective pressures. The CT skew was calculated in non-overlapping windows along the T7 sequence, moving from the left end of the genetic map to the right along the non-transcribed strand (see Materials and Methods). Further, the CT skew values in successive windows were added together, and this cumulative skew was plotted against the physical map of the genome (Figure 3(b)).

The T7 genome can be divided into three classes according to the level of transcription (Dunn & Studier, 1983; Studier & Dunn, 1983). The class I genes are transcribed the least and lie at the left end of the genome (Figure 3(a)). The class II genes are transcribed moderately, while the class III genes, mostly coding for virion proteins and virion assembly proteins, are transcribed at extremely high levels (Studier & Dunn, 1983). We further divided class III genes into class IIIa and class IIIb; the former genes are upstream of the transcriptional terminator T ϕ , while the latter are downstream from it.

The cumulative CT skew diagram shows that the excess of T over C is non-random, and is most pronounced in the highly transcribed regions of the T7 genome. One of the striking features of the diagram is the sharp drop in the CT skew in class IIIa genes (Figure 3(b)). Just as important, the CT skew levels off immediately downstream of the T ϕ transcription terminator. This may be explained by the fact that genes 11 and 12 do not have their own promoters and are transcribed entirely by read-through transcription past T ϕ (Studier & Dunn, 1983). On the other hand, gene 13 and other genes downstream of T ϕ have strong class III promoters, and show a significant decrease in CT skew (Figure 3(b)). The cumulative plot diagram for all three codon positions in T7 genome was qualitatively similar to fourfold degenerate site diagram, but the magnitude of the skew was smaller (not shown).

This correlation can also be seen by comparing the average CT skews of the different classes of genes (Figure 3(c)). As expected, class IIIa genes exhibited the lowest CT skew among all T7 genes. They were followed by the class IIIb, class II, and class I genes (Figure 3(c)). The average CT skew for the class IIIb genes is less negative compared to

Figure 3. Organization of the T7 genome and its sequence bias. (a) Genetic organization of the T7 genome. The important genetic elements in the T7 genome are schematically shown. Top line, the promoters and transcription terminators. The *E. coli* RNAP promoters and terminator are represented by open symbols, while phage T7 RNAP promoters and terminator are shown by filled symbols. Promoters are depicted as triangles, while terminators are represented by circles. Second line, the position of the bidirectional origin of replication. Third line, the locations of the three classes of T7 genes are shown. Bottom line, percentage scale for the genome. The complete genome is ~40 kbp. (b) Cumulative CT skew for the T7 genome. Positions of the *E. coli* promoters A1, A2 and A3, the principal origin of replication (ORI), transcriptional terminators TE and T ϕ , and T7 promoters 9, 10 and 13 are shown as vertical lines. The segments of the T7 genome corresponding to the three classes of genes are also indicated. The class III genes have been divided into IIIa and IIIb subclasses as indicated in the text. (c) The CT skew was calculated in non-overlapping 50 bp windows in each promoter class and the mean skew for the different classes is shown.

that of class IIIa genes, because the former group includes genes 11 and 12, which are not strongly transcribed.

Discussion

We have shown here that the T7 RNAP can cause C to T mutations in the non-transcribed strand of DNA and this finding extends our previous results with the *E. coli* RNAP (Beletskii & Bhagwat, 1996). In particular, it shows that the structure of the RNAP may not be important for this phenomenon. The two RNAPs do not share significant sequence homology, and while the *E. coli* RNAP is a large multi-subunit complex, T7 RNAP is a single polypeptide.

We have suggested that TIM is an unavoidable consequence of transcription, and the feature of the transcription bubble that causes cytosine deaminations is the exposure of the non-template strand to the aqueous environment (Beletskii & Bhagwat, 1996). The data presented here are consistent with this idea. Further, the structure of the transcription bubble may be similar in all organisms (Mooney & Landick, 1999), and hence TIM is likely to be a general property of transcription elongation by all RNA polymerases.

Our results also confirm a prediction based on this model for TIM. Because cytosine deamination occurs in the strand displaced by the RNAP, a slower RNAP should cause more C to T mutations than a faster one (Beletskii & Bhagwat, 1996). Consistent with this idea, the G645A mutant of T7 RNAP transcribes at a slower rate and yields a higher level of mutations than the wild-type enzyme.

The demonstration of TIM for T7 RNAP suggested to us that heavily transcribed genes in the T7 genome may have accumulated C to T mutations in the non-transcribed strand over the evolutionary time-scale. To examine this possibility, we introduced a new integral skew function, the cumulative CT skew. The cumulative skew diagram is easier to interpret than plots of the skew value itself, since fluctuations in the small sequence windows used to calculate the skew in the latter case often hide the general plot trends. We observed that changes in CT skew in T7 genome regions are roughly proportional to the level of transcription of these regions.

It should be noted that the cumulative CT skew may have wide applicability in the study of genome evolution. To test its usefulness, we used it to analyze the H-strand of the human mitochondrial genome, which has a pronounced excess of thymine over cytosine bases, due to a strand-specific delay in DNA replication. The shape of the cumulative CT diagram in mitochondria (not shown) indicates that the relative excess of T over C decreases linearly from oriH to oriL, in full agreement with the results of a previously published

study (Reyes *et al.*, 1998) which used a less direct method of analysis to arrive at this conclusion.

A notable feature of the cumulative CT skew is that it peaks just before the bidirectional origin of replication in the T7 genome and decreases in either direction from it (Figure 3(b)). This pattern is similar to the behavior of the GC skew in different organisms where the global minimum and maximum of cumulative GC skew coincide with replication origin and terminus, respectively (Grigoriev, 1998, 1999). We believe that the shape of the T7 CT skew diagram is consistent with both transcription and replication, contributing to strand bias observed in the T7 genome.

It is generally believed that during DNA replication, the coordinated synthesis of the leading and lagging strands creates single-stranded regions in the lagging strands that are bound by the single-strand DNA-binding protein (Kornberg & Baker, 1992). This has been specifically shown to be true for DNA synthesis by T7 DNA polymerase (Lee *et al.*, 1998; Park *et al.*, 1998). Assuming that the T7-coded single-stranded DNA-binding protein, gp2.5, does not protect the exposed cytosine bases in the lagging strand templates, they should deaminate at a higher rate. As a result, when replication and transcription have the same directionality, as is true for 85% of the T7 genome that lies to the right of the origin of replication, both replication and transcription should create identical bias in pyrimidine composition. This was proposed in the context of the T4 bacteriophage genome evolution (Kano-Sueoka *et al.*, 1999) and is schematically shown in Figure 4.

Consequently, the observed negative CT skew in class II and class III genes of T7 is likely to be an additive effect of replication and transcription. However, there is good correlation between the level of transcription in different subgroups of genes within these classes and their average CT skews (Figure 3(c)). This suggests that TIM makes a substantial contribution to the CT skew of class II and class III genes.

The situation is more complex to the left of the T7 replication origin. Here the lagging strand template and the non-transcribed strand do not coincide, and C to U conversions in the lagging strand template should not affect the CT skew for the non-template strand of T7 genome (Figure 4). Further, TIM should have little influence on the cumulative CT skew in this region because this region is transcribed at a low level for a brief period by the *E. coli* RNAP (McAllister *et al.*, 1981). Despite this, the cumulative skew increases in this region (Figure 3(b)). There is no good explanation for the increase.

Finally, the observed strand bias in C to T mutations in the evolution of enteric bacterial genes has been attributed to transcription-coupled repair (Francino *et al.*, 1996; Francino & Ochman, 1997). There is no evidence that T7 RNAP is able to support transcription-coupled DNA repair machinery. Further, the proposed mechanism of

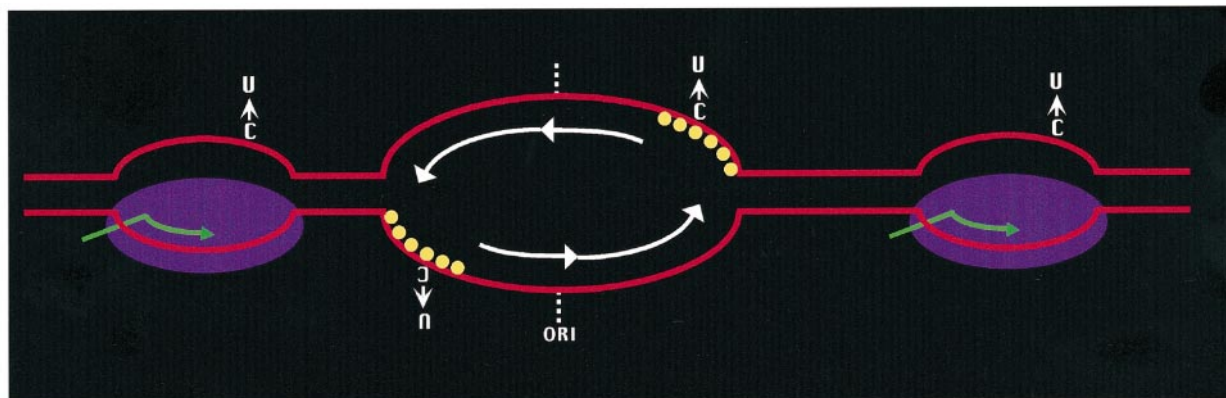


Figure 4. Cytosine deaminations induced by DNA replication and transcription of phage T7. Bidirectional replication of the T7 phage and its transcription are schematically shown. Replication is shown as white arrows, while transcription is shown as green arrows. RNA polymerase is shown as a purple ellipse and gp2.5 is shown as yellow circles. ORI, principal origin of replication of T7.

transcription-repair coupling (Selby & Sancar, 1993) and the lack of sequence similarity between the two polymerases makes this unlikely. Consequently, it is unlikely that observed sequence bias in the T7 genome is related to the action of transcription-coupled repair.

There is considerable interest in the existence of strand asymmetry of DNA sequences and its causes (Francino & Ochman, 1997; Frank & Lobry, 1999). Specifically, GC and AT skew diagrams were introduced by Lobry (1996b) for the analysis of short regions of bacterial chromosomes and were later used for the analysis of complete genomes of bacteria (Blattner *et al.*, 1997; Grigoriev, 1998; Lobry, 1996a) and viruses (Grigoriev, 1999). We have used a CT skew to analyze the T7 genome, and our results suggest that transcription should contribute to the evolution of genomes in two ways. First, the non-template strand of genes should suffer more C to T mutations than the template strand. Second, highly transcribed genes should experience more C to T mutations than the less transcribed genes. Both these predictions can be tested by sequencing genes of natural T7-like phages and identifying the sequence changes that are responsible for their evolution. The well-characterized biology of the T7 phage makes this an attractive model system for the study of molecular evolution.

Materials and Methods

Construction of the genetic reversion system

The strains, GM31 and BH156 (=GM31 *ung*) have been described (Lutsenko & Bhagwat, 1999). A cassette containing the mutant allele *kanS-D94* was excised from pKanS-D94 (Wyszynski *et al.*, 1994) with *Bam*HI and cloned into pET3 (Rosenberg *et al.*, 1987) to generate pAB1. Plasmid pAB1 was partially digested with *Bgl*III to release the *Bgl*III fragment containing *kanS-D94* flanked by the T7 gene 10 promoter and terminator. This fragment was ligated into the pUC18 *Bam*HI site and the

orientation in which the *kanS-D94* was co-linear with *lacZ'* gene was called pAB2. A *Kpn*I-Asp718 fragment of pAB2 encompassing the *kanS-D94* gene and the T7 gene 10 promoter and terminator was treated with Klenow fragment to fill-in protruding 5' termini and ligated into RSF1050H3 (Lutsenko & Bhagwat, 1999) plasmid which was cut with *Hind*III and treated with Klenow fragment. The resulting plasmid pAB4 contains an active transposon Tn3 with the T7 promoter-driven *kanS-D94* transcription unit co-linear with the β -lactamase gene. Plasmid pAB4 was transformed into the RR1023 strain (Patterson *et al.*, 1986) containing mini-F' episome pOX38 (Guyer *et al.*, 1981). The cells were grown at 30°C and mated with GM31 to separate the cointegrates between pAB4 and mini-F plasmid generated by Tn3 transposition (Lee *et al.*, 1983). Streptomycin was used to select against the donor. Following spontaneous resolution of the cointegrate in the recipient, the mini-F::Tn3-*kanS-D94*, called pAB6, was isolated using a plasmid isolation kit (Qiagen) and used for the reversion assays.

Construction of *E. coli* strains with the T7 RNA polymerase

The T7 RNAP gene was introduced into BH156 by lysogenizing the strain with λ DE3 phage (Studier & Moffatt, 1986) using published protocols (Silhavy *et al.*, 1984). The resulting strain BH161 was transformed with plasmid pAB6 for use in reversion assays.

To construct an *E. coli* strain with the T7 RNAP mutant G645A, the 4.3 kb *Bam*HI fragment of pAR1219-like plasmid (Davanloo *et al.*, 1984) containing the G645A mutant of T7 RNA (Bonner *et al.*, 1992; Makarova *et al.*, 1995) along with the *lacI* gene was cloned into the unique *Bam*HI site of the λ D69 vector. The final DNA ligation mix was packaged with Gigapack III Gold Packaging Extract (Stratagene) according to the manufacturer's instructions and plaques screened for β -galactosidase activity directed from the T7 promoter and restriction analysis of the phage DNA. We selected for further work the λ G645A/DE4 construct in which the T7 RNA polymerase gene is colinear with the λ *int* gene. Strain BH156 was lysogenized with this phage in the same way as λ DE3. PCR was used to confirm the presence and the orientation of *kanS-D94* in the resulting

lysogens. The resulting strain, BH200, was used in reversion assays with pAB6 plasmid as described below.

Total RNA isolation and quantification

Total RNA was isolated from different cells carrying pAB6 plasmid as described (Beletskii & Bhagwat, 1996) and transferred to nylon membranes for quantification. The membranes were successively probed with *kan*-specific and *rpoC*-specific probes. The latter is a constitutively expressed gene and was used to correct any pipetting errors. The bound radioactivity from the *kan*-specific probe was quantified using a phosphorimager (Molecular Dynamics) and normalized with respect to the amount of signal obtained using the *rpoC*-specific probe.

Reversion assay for C to T mutations

For each reversion assay, a midlog (absorbance at 550 nm = 0.3) culture of cells with pAB6 plasmid was diluted 100-fold into two tubes containing LB medium; one containing 100 μ M IPTG and the other without IPTG. The resulting cultures were again grown to an absorbance at 550 nm of 0.3 and plated on kanamycin plates to determine the level of revertants. The total cell count was determined by plating dilutions of the cultures on LB plates.

Cumulative CT skew analysis of the T7 genome

The complete T7 genome sequence was obtained from GenBank (accession numbers V01146, J02518, X00411). Because some of the T7 genes overlap (some in the same frame or not), an analysis of the third codon position (abbreviation 3cp) in all the coding sequences of T7 would account for some bases multiple times. It would also include 3cp bases in one gene which may be in 1cp or 2cp in an overlapping gene. To analyze only the 3cp bases and to prevent repeated counting of them, the overlaps were treated as follows. Genes 6A (part of gene 6B), 4B (part of 4A), 5.5-5.7 (fusion of 5.5 and 5.7), 10A (part of 10B) as well as short genes generating out-of-frame overlaps (4.1, 19.2 and 19.3) were excluded from the analysis. Other out-of-frame overlaps (between genes 2.8 and 3, 3.8 and 4A, 5.9 and 6, 6 and 6.3, 18.5 and 18.7) and all start and stop codons and intergenic regions were replaced by the equivalent number of Ns in the sequence.

Base composition analysis was performed using the method of cumulative skew diagrams (Grigoriev, 1998). Consider a genome of length G , divided into k non-overlapping windows $W_1 \dots W_k$, each w base-pairs long. For a window W_i , containing [C] cytosine and [T] thymine bases, CT skew is calculated as $s_i = ([C] - [T]) / ([C] + [T])$. We chose a window size of 50 bp, hence $k = 799$ for the T7 genome. The w/G ratio is small, so a cumulative CT skew, or a sum $S(m) = \sum_{i=1}^m s_i w / G$ in adjacent windows $W_1 \dots W_m$ ($m \leq k$), represents a numerical integration of s_i over the contiguous sequence segment covered by the windows $W_1 \dots W_m$. In addition, mean CT skew values were calculated for each promoter class region (class I to class IIIb). In addition, mean (non-cumulative) CT skew values were calculated by averaging across the genomic regions corresponding to promoter classes I to IIIb (the same 50 bp windows were used). Thus, for a region covering R such windows, the mean skew was $(R) = (1/R) \sum_{i=1}^R s_i$.

Acknowledgements

The authors thank Dr R. Weisberg (NCI, Frederick, MD) for gift of *E. coli* strains for the red plaque test; Dr M. Gottesman (Columbia University, New York, NY) for phage lambda D69 and Dr M. Dreyfus (Ecole Normal Supérieure, Paris, France) for advice on the use of T7 RNA polymerase mutants. This work was supported by NIH grant GM57200 to A.S.B.

References

- Bandaru, B., Wyszynski, M. & Bhagwat, A. S. (1995). *HpaII* methyltransferase is mutagenic in *Escherichia coli*. *J. Bacteriol.* **177**, 2950-2952.
- Beletskii, A. & Bhagwat, A. S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **93**, 13919-13924.
- Beletskii, A. & Bhagwat, A. S. (1998). Correlation between transcription and C to T mutations in the non-transcribed DNA strand. *Biol. Chem.* **379**, 549-551.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453-1474.
- Bonner, G., Patra, D., Lafer, E. M. & Sousa, R. (1992). Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J*, **11**, 3767-3775.
- Bonner, G., Lafer, E. M. & Sousa, R. (1994). Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269**, 25120-25128.
- Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. (1978). Molecular basis of substitution hotspots in *Escherichia coli*. *Nature*, **274**, 775-780.
- Datta, A. & Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science*, **268**, 1616-1619.
- Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **81**, 2035-2039.
- Dunn, J. J. & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477-535.
- Francino, M. P. & Ochman, H. (1997). Strand asymmetries in DNA evolution. *Trends Genet.* **13**, 240-245.
- Francino, M. P., Chao, L., Riley, M. A. & Ochman, H. (1996). Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science*, **272**, 107-109.
- Frank, A. C. & Lobry, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*, **238**, 65-77.
- Frederico, L. A., Kunkel, T. A. & Shaw, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, **29**, 2532-2537.

- Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54**, 4855-4878.
- Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucl. Acids Res.* **26**, 2286-2290.
- Grigoriev, A. (1999). Strand-specific compositional asymmetries in double-stranded DNA viruses. *Virus Res.* **60**, 1-19.
- Guyer, M. S., Reed, R. R., Steitz, J. A. & Low, K. B. (1981). Identification of a sex-factor-affinity site in *E. coli* as gamma delta. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 135-140.
- Hutchinson, F. & Donnellan, J. E., Jr. (1998). A mutation spectra database for bacterial and mammalian genes: 1998. *Nucl. Acids Res.* **26**, 290-291.
- Kano-Sueoka, T., Lobry, J. R. & Sueoka, N. (1999). Intra-strand biases in bacteriophage T4 genome. *Gene*, **238**, 59-64.
- Kornberg, A. & Baker, T. A. (1992). *DNA Replication*, 2nd edit., W.H. Freeman, New York.
- Kunz, B. A., Ramachandran, K. & Vonarx, E. J. (1998). DNA sequence analysis of spontaneous mutagenesis in *Saccharomyces cerevisiae*. *Genetics*, **148**, 1491-505.
- Lee, C. H., Bhagwat, A. & Heffron, F. (1983). Identification of a transposon Tn3 sequence required for transposition immunity. *Proc. Natl Acad. Sci. USA*, **80**, 6765-6769.
- Lee, J., Chastain, P. D., II, Kusakabe, T., Griffith, J. D. & Richardson, C. C. (1998). Coordinated leading and lagging strand DNA synthesis on a minicircular template. *Mol. Cell*, **1**, 1001-1010.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.
- Lobry, J. R. (1996a). Origin of replication of *Mycoplasma genitalium*. *Science*, **272**, 745-746.
- Lobry, J. R. (1996b). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660-665.
- Lutsenko, E. & Bhagwat, A. S. (1999). Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications. *Mutat. Res.* **437**, 11-20.
- Makarova, O. V., Makarova, E. M., Sousa, R. & Dreyfus, M. (1995). Transcribing of *Escherichia coli* genes with mutant T7 RNA polymerases stability of *lacZ* mRNA inversely correlates with polymerase speed. *Proc. Natl Acad. Sci. USA*, **92**, 12250-12254.
- McAllister, W. T., Morris, C., Rosenberg, A. H. & Studier, F. W. (1981). Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.* **153**, 527-544.
- Mooney, R. A. & Landick, R. (1999). RNA polymerase unveiled. *Cell*, **98**, 687-690.
- Morey, N. J., Greene, C. N. & Jinks-Robertson, S. (2000). Genetic analysis of transcription-associated mutation in *Saccharomyces cerevisiae*. *Genetics*, **154**, 109-120.
- Park, K., Debyser, Z., Tabor, S., Richardson, C. C. & Griffith, J. D. (1998). Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins. *J. Biol. Chem.* **273**, 5260-5270.
- Patterson, T. A., Court, D. L., Dubuc, G., Michniewicz, J. J., Goodchild, J., Bukhari, A. I. & Narang, S. A. (1986). Transposition studies of mini-Mu plasmids constructed from the chemically synthesized ends of bacteriophage Mu. *Gene*, **50**, 101-109.
- Reyes, A., Gissi, C., Pesole, G. & Saccone, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* **15**, 957-966.
- Rosenberg, A. H., Lade, B. N., Chui, D. S., Lin, S. W., Dunn, J. J. & Studier, F. W. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, **56**, 125-135.
- Selby, C. P. & Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science*, **260**, 53-58.
- Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984). *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Studier, F. W. & Dunn, J. J. (1983). Organization and expression of bacteriophage T7 DNA. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 999-1007.
- Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.
- Wyszynski, M., Gabbara, S. & Bhagwat, A. S. (1994). Cytosine deaminations catalyzed by DNA cytosine methyltransferases are unlikely to be the major cause of mutational hot spots at sites of cytosine methylation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 1574-1578.

Edited by M. Gottesman

(Received 3 April 2000; received in revised form 24 May 2000; accepted 6 June 2000)