Chapter 4

Testing for R-loop occurrence in rho and nusG mutants by bisulphite mutagenesis in the absence of denaturation
INTRODUCTION

The investigations described in this chapter were performed to standardize the assay of in vitro bisulphite mutagenesis and to employ this approach in testing for the occurrence of R-loops on the chromosome of the *rho* and *nusG* mutants of *E. coli*. The studies described in Chapters 3-6 were performed in an endeavour to understand the molecular consequences of the inability to terminate synthesis of translation-uncoupled transcripts in the *rho* and *nusG* mutants of *E. coli*.

As per the hypothesis proposed in the laboratory, based on previous genetic investigations (Harinarayanan and Gowishankar, 2003), R-loops occur at increased frequency in the *rho* and *nusG* mutants of *E. coli* as a consequence of defective Rho-dependent transcription termination. One instance of R-loop occurrence in these mutants is the protein coding loci when translation uncoupled transcripts fail to be terminated efficiently (Gowrishankar and Harinarayanan, 2004). Following this model, a set of isogenic wild-type, *rho* and *nusG* mutant strains were constructed, as described in Chapter 3, by a series of P1-mediated transductions from the common laboratory strain, MG1655. These strains had *lacZ*<sub>U118</sub> and *trpE*<sub>9777</sub> mutant alleles that bear early nonsense mutations, leading to premature termination of translation at these loci.

In the wild-type strain, transcription was expected to terminate immediately downstream of the premature translation stop codon while longer transcripts that extend well beyond the nonsense mutations were expected to be produced in the *rho* and *nusG* mutants. These untranslated stretches of mRNA would render these loci susceptible to R-loop formation in the *rho* and *nusG* mutant strains. Thus in the current study, the hypothesis of R-loop occurrence on the chromosome of *rho* and *nusG* mutants of *E. coli* was tested downstream of the premature termination codons at mutant *lacZ* and *trpE* genes.

Additionally, mutations in the *lacI* and *trpR* regulatory genes that render the expression of the respective structural genes constitutive were also introduced into these strains to ensure continuous transcription, thus increasing the possibility of detection of R-loops at these loci by molecular approaches (construction of strains discussed in detail in Chapter 3, Part A).

Part B of chapter 3 elaborated various biochemical approaches attempted to test for the presence of single-strandedness on the DNA of the test strains in the absence of denaturation. Presence of a signal corresponding to single-strandedness on the non-template strand downstream of the premature termination codons but not on the template strand, at the *lacZ*<sub>U118</sub> and *trpE*<sub>9777</sub> loci that would disappear upon RNase H treatment was proposed to be taken as evidence of R-loop occurrence at these loci.

Among the various assays tested, it was observed that bisulphite mutagenesis in the absence of denaturation was extremely sensitive in detecting single-strandedness in the chromosomal preparations. Thus this approach was sought to be standardized further to test for R-loop occurrence in the wild-type, *rho* and *nusG* strains of *E. coli*. A detailed description of the systematic approach employed to adapt the bisulphite assay to the current requirements is outlined in this chapter.

Sodium bisulphite treatment causes cytosine to uracil changes on ssDNA. Shortle and Nathans (1978) found that local bisulphite mutagenesis of SV40 DNA at a single-strand gap resulted in C-to-U conversions specifically in the gap. This process was further described later to induce C-to-T conversions on ssDNA by
bisulphite (Shortle and Botstein, 1983). This property of the compound makes bisulphite mutagenesis the classical approach in the field of epigenetics to detect methylation patterns in eukaryotes, since methylated cytosine in ssDNA is resistant to attack by sodium bisulphite. The standard protocol for bisulphite conversion in epigenetic assays for detection of methylation involves denaturation of DNA prior to bisulphite treatment so as to ensure maximum conversion of unmethylated cytosines to uracils (Hajkova et al., 2002). Thus bisulphite mutagenesis after denaturation of DNA followed by PCR and sequencing reveals methylation patterns of the gene/genome being tested. Presence of T residues in place of C residues in the final sequence indicates absence of methylation and hence bisulphite susceptibility while retention of C residues in the PCR product obtained from bisulphite treated DNA is considered as evidence of methylation.

However, lately bisulphite mutagenesis has been increasingly used to demonstrate the occurrence of R-loops. This approach has been successfully used to demonstrate the occurrence of R-loops on the chromosome at murine immunoglobulin class switch regions (Yu et al., 2003; Huang et al., 2006). Furthermore, R-loops were also demonstrated to occur in alternate splicing factor/splicing factor 2 (ASF/SF2) depleted Chicken DT-40 cell line (Li and Manley, 2005).

The protocol employed for bisulphite treatment for R-loop detection differs from that employed for methylation analysis in that the DNA analyzed for R-loop detection is exposed to sodium bisulphite in the absence of denaturation. Loci involved in R-loop formation will have a three stranded structure in which one strand of DNA (the template strand of transcription) is heteroduplexed with the complementary RNA molecule while the other DNA strand (the non-template strand of transcription) is left unpaired. The principle of R-loop detection using bisulphite mutagenesis involves the detection of C-to-T conversions on the unpaired non-template DNA strand after bisulphite treatment in the absence of denaturation, followed by PCR and sequencing, while similar conversions are not expected to be observed on the template DNA strand heteroduplexed with the complementary RNA.

Biochemical testing of the hypothesis of R-loop occurrence at the lacZU118 and the trpE9777 loci downstream of the translation termination mutations in rho and nusG backgrounds was performed and evidence for increased occurrence of R-loops in these mutants was obtained (current chapter and Chapter 5). Global analysis of the propensity of R-loop occurrence in the nusG mutant which is compromised for Rho-dependent transcription termination led to the revelation that R-loops occur at increased frequency throughout the genome in this mutant (Discussed in elaborate depth in Chapter 6). Remarkably, in this analysis R-loops were also revealed to occur at low frequencies even in the wild-type strain that is proficient for Rho-dependent transcription termination.

**PRINCIPLE**

**4.1 Part A: Chemistry of DNA modification by sodium bisulphite**

The reactivity of cytosines with sodium bisulphite causing conversion to uracil residues was first reported nearly 40 years ago (Hayatsu, et al., 1970; Shapiro, et al., 1973). Treatment of DNA by sodium bisulphite leads to the conversion of cytosine residues to uracils that are revealed as C-to-T conversions upon
subsequent PCR amplification and sequencing. Owing to its high sensitivity and the option of addressing the non-symmetrical modification pattern on the two strands of DNA, bisulphite mutagenesis is rendered the credible method for various applications.

The scheme of the bisulphite reaction can be depicted as follows:

The reaction is comprised of three principal steps (Figure 4.1):

i. *Sulphonation of cytosine to cytosine-6-sulphonate:* This reversible reaction is favoured at low temperature and pH, with the equilibrium state being achieved in 20 minutes at 0°C.

ii. *Hydrolytic deamination of cytosine-6-sulphonate to uracil-6-sulphonate:* At an optimum pH of 5-6 and higher temperatures, high concentrations of sodium bisulphite favour irreversible conversion of cytosine-6-sulphonate to uracil-6-sulphonate. Presence of hydroquinone in the reaction mix serves to inhibit the unwanted bisulphite-oxidation reaction and also prevents DNA strand breakage that can occur through depurination.

iii. *Desulphonation of uracil-6-sulphonate to uracil:* This is an elimination reaction that is reversible and favoured at high pH. Hence an alkali such as NaOH is included at this step.

*Figure 4.1:* Diagrammatic representation of the scheme of bisulphite reactivity of unpaired cytosine residues. Reversible sulphonation of cytosine at C-6 to give rise to cytosine-6-sulphate occurs at step I, followed by irreversible deamination of cytosine-6-sulphate at C-4 to yield uracil-6-sulphate in step II. Subsequent reversible desulphonation of uracil-6-sulphate to uracil occurs in step III in the presence of alkali.

Thus sodium bisulphite causes C-to-U conversions on a DNA single strand being probed. Such C-to-U conversions can be detected by PCR using specially designed converted primers, either as a converted pair or in combination with native primers, followed by cloning and sequencing the individual clones. The converted primers are derived by replacing the C residues with T residues or G residues with A residues depending on the strand being probed. During PCR, the polymerase will incorporate an A residue opposite the U residue. In the next round of amplification a T residue is incorporated opposite the A residue, resulting in a C-to-T transition in the final pool of PCR products arising from such a template. For a duplex that has a region
of single-strandedness, the C-to-T conversions are expected to be limited to that region.

Under non-denaturing conditions the cytosine residues in double-stranded DNA (i.e., cytosines that are base paired with guanine residues) are resistant to conversion by sodium bisulphite. In a cytosine residue that is base paired with guanine on the complementary strand, the amino group on the C-4 carbon of cytosine has one of its hydrogen bonds in the O…..H bond formation with the O-6 of guanine. Thus it is unavailable for deamination and subsequent uracil formation.

![R-loop detection by Bisulphite Modification Assay](image)

**Figure 4.2:** R-loop detection by bisulphite mutagenesis. Treatment of R-looped DNA by sodium bisulphite in the absence of denaturation causes conversion of C residues on the looped out unpaired non-template strand to U’s. Subsequent PCR amplification followed by sequencing will reveal C-to-T changes.

The converted DNA is then amplified with modified or converted primers in which thymine residues have been incorporated in place of cytosine residues (where uracil corresponds to thymine in its base pairing behaviour) in the original parent DNA sequence. The amplified product is then cloned into a TA cloning vector and individual clones corresponding to single molecules in the original amplified pool are isolated. Sequencing of these clones would be expected to reveal C-to-T transitions specific to the original parent DNA, thereby providing an indirect estimate of the stretch of DNA that was unpaired during bisulphite treatment.

**RESULTS**

**4.2 Part B: Optimization of protocol for bisulphite treatment of total nucleic acids**

The initial set of experiments was done after embedding the total nucleic acids in beads of low-melting-point (LMP) agarose. The beads were exposed to sodium bisulphite-hydroquinone mixture as
described by Hajkova et al. (2002), with the difference that the denaturation of DNA was avoided so that only R-looped regions would exhibit bisulphite reactivity. Briefly, nucleic acids were mixed with 1% molten LMP agarose at 45°C to a final DNA concentration of 2 ng/µl. 50 µl aliquots of this mixture were added to ice cold mineral oil using a prewarmed pipette tip to form beads (100 ng DNA per 50 µl bead). 8-10 beads were then exposed to 2.5 M sodium bisulphite (pH=5.0) containing 1 M hydroquinone. The tubes were wrapped in aluminium foil and incubated at 37°C for 5 hours. The bisulphite-hydroquinone mixture was then removed with a pipette and the beads were washed once with sterile water for 10 minutes. 500 µl of NaOH was added per tube and left at room temperature for 15 minutes for alkaline treatment. The beads were then washed thrice with sterile water (10 minutes each wash) and were stored at 4°C in TE. The beads were stable up to two weeks.

One bead was used per PCR reaction to score for the presence of amplification. However the biggest drawback of this approach was the frequent disintegration of beads and the irreproducibility of results in subsequent PCR amplifications with converted primers that led to inconsistency in quantification of PCR products. This inconsistency was probably due to unequal distribution of DNA among the beads leading to considerable variations among replicate experiments.

Subsequently, the agarose embedding step was eliminated and the DNA was directly exposed to sodium bisulphite-HQ solution and incubated in the dark at 37°C for 5 hours. The DNA was extracted from the bisulphite solution using QiaEX II Reaction Cleanup kit. It was then treated with NaOH for 15 minutes, precipitated with ethanol and resuspended in sterile water. The solution was stable up to two months at −20°C.

Later reactions were done using the Epitect Bisulphite Kit (Qiagen) as per the manufacturer’s instructions, with the modification that the heat-denaturation step was substituted by incubation at a constant temperature of 37°C to prevent both unwinding of DNA as well as disruption of the RNA-DNA heteroduplexes. 20 ng of treated DNA was used as a template for subsequent PCR reactions.

**4.2.1 Optimization of PCR amplification protocol for bisulphite modified DNA**

Sodium bisulphite is known to cause large scale shearing of DNA due to depurination (Raizis et al. 1995, Grigg, 1996). As a consequence, amplification of larger fragments becomes progressively difficult. So an optimum amplicon size of 250-300 bp of DNA was chosen to try and ensure reproducible results.

**4.2.1.1 Primer design**

One specific advantage of treatment of DNA with sodium bisulphite is that the unpaired cytosines are converted to uracils while the complementary guanines remain unchanged. This renders the sequences of the Watson and Crick strands non-complementary after exposure to bisulphite mutagenesis. This allows information on strand specificity of bisulphite conversions to be preserved, so that strand specific primers can be designed to amplify the strand of interest. The amplicon band intensity can be considered as a direct estimate of the proportion of the parent molecules modified by bisulphite.
For the lac operon, four distinct 250-300 bp stretches of DNA downstream of the chain terminating mutation in lacZ were demarcated to be tested for signal corresponding to R-loops after bisulphite treatment. Three of these stretches were in the lacZ gene, while one was in the lacY gene. These were designated as lacZ1, lacZ2, lacZ3 and lacY1 (and will be referred to as such in the text), lacZ1 being closest to the amber mutation and lacZ3 being farthest in the lacZ gene. lacY1 was situated in the lacY gene. The regions so selected were expected to give information of the extent of R-loop stretch in the strains tested, lacZ1-3 will reveal whether R-loops occur immediately downstream of the chain terminating mutation or stretch up to the 3’ end of the gene. Probing the lacY1 region will give information whether R-loops extend into the downstream lacY gene or terminate at the natural stop codon of lacZ (since lacY translation is expected to remain coupled to transcription even in the lacZ mutants). In a manner similar to the lac operon, trpE1, trpE2 and trpD1 regions were selected in the trp operon.

For any given locus a pair of “native” primers (designated as N–N primer pair) was designed, the sequence of which was identical to the non-bisulphite treated DNA in that region. This N–N pair would amplify both strands of DNA which were not modified by bisulphite. Hence N–N amplicons would serve as controls for input unmodified DNA.

After bisulphite treatment, the template and non-template strands of DNA at the test loci were expected to become non-complementary. To selectively amplify such modified strands, unique “converted” primers were designed specific to each strand (see table 4.1a and table 4.1b).

**Figure 4.3:** Schematic representation of the 250 bp regions, which were intended to be amplified by various native and converted primer pairs after bisulphite treatment to test for signal corresponding to R-loops (Gene length annotations not to scale). All regions have been demarcated downstream of the premature translation termination codon. The amplicons in the lac operon have been designated as lacZ1, lacZ2, and lacZ3 on the basis of their location on the lacZ gene, lacZ1 being nearest to the termination mutation and lacZ3 being farthest. Another region in the lacY gene, designated as lacY1 was intended to be tested for R-loop signal downstream of the lacZ gene. Similar to the scheme of the lac operon, trpE1, trpE2 and trpD1 were designed for the trp operon.

To amplify the converted non-template strand (nts), nts-specific “converted” primers were designed. Assuming complete conversion by bisulphite, all C’s on the non-template strand would be converted to U’s. The forward converted primer sequence derived from this template will thus have C-to-T base changes. Thus
the converted forward primer \((C_{nts})\) sequence was obtained by replacing the C’s of the native forward primer with T’s. The reverse “converted” primer \((-C_{nts})\) would be converse to this strand and hence have G-to-A base changes. Thus this primer sequence was derived by replacing G’s of the native reverse primer with A’s. Hence the \(C_{nts}-C_{nts}\) primer pair was generated. Likewise for the template strand \((ts)\), the \(C_{nts}-C_{nts}\) primer pair sequences were generated by replacing G’s with A’s in the native forward primer and C’s with T’s in the native reverse primer.

The length of the primers was ensured to be between 20 and 25 nucleotides. Extensive T and A stretches were avoided in both primers wherever feasible so as to minimize the formation of primer dimers.

The nomenclature of various primer combinations designed at each locus is listed in tables 4.1a and 4.1b below:

### \(lac\) operon

<table>
<thead>
<tr>
<th>Primer Description (nomenclature in text)</th>
<th>(lacZ1)</th>
<th>(lacZ2)</th>
<th>(lacZ3)</th>
<th>(lac)Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native forward ((N))</td>
<td>JGBPMA(lacZF)1</td>
<td>JGBPMA(lacZF)2</td>
<td>JGBPMA(lacZF)3’</td>
<td>JGBPMA(lac)Y1</td>
</tr>
<tr>
<td>Native reverse ((-N))</td>
<td>JGBPMA(lacZR)1</td>
<td>JGBPMA(lacZR)2</td>
<td>JGBPMA(lacZR)3’</td>
<td>JGBPMA(lac)Y1</td>
</tr>
<tr>
<td>Converted nts-specific forward ((C_{nts}))</td>
<td>JGBPBM(lacZF)1</td>
<td>JGBPBM(lacZF)2</td>
<td>JGBPBM(lacZF)3’</td>
<td>JGBPBM(lac)Y1</td>
</tr>
<tr>
<td>Converted nts-specific reverse ((-C_{ints}))</td>
<td>JGBPBM(lacZR)1</td>
<td>JGBPBM(lacZR)2</td>
<td>JGBPBM(lacZR)3’</td>
<td>JGBPBM(lac)Y1</td>
</tr>
<tr>
<td>Converted ts-specific forward ((C_{ts}))</td>
<td>JGBPBM(lacZCF)1</td>
<td>JGBPBM(lacZCF)2</td>
<td>JGBPBM(lacZCF)3’</td>
<td>JGBPBM(lac)YCF1</td>
</tr>
<tr>
<td>Converted ts-specific reverse ((-C_{tts}))</td>
<td>JGBPBM(lacZCR)1</td>
<td>JGBPBM(lacZCR)2</td>
<td>JGBPBM(lacZCR)3’</td>
<td>JGBPBM(lac)YCR1</td>
</tr>
</tbody>
</table>

**Table 4.1a:** Scheme of nomenclature of various primers designed to be employed in amplification of bisulphite modified DNA at the \(lac\) locus. The sequences of the primers are given in section 2.1.1.3 of Chapter 2.

### \(trp\) operon

<table>
<thead>
<tr>
<th>Primer Description (nomenclature in text)</th>
<th>(trpE1)</th>
<th>(trpE2)</th>
<th>(trpD1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native forward ((N))</td>
<td>JGBPMA(trpEF)1</td>
<td>JGBPMA(trpEF)2</td>
<td>JGBPMA(trpDF)1</td>
</tr>
<tr>
<td>Native reverse ((-N))</td>
<td>JGBPMA(trpERP)1</td>
<td>JGBPMA(trpERP)2</td>
<td>JGBPMA(trpDRP)1</td>
</tr>
<tr>
<td>Converted nts-specific forward ((C_{nts}))</td>
<td>JGBPBM(trpEF)1</td>
<td>JGBPBM(trpEF)2</td>
<td>JGBPBM(trpDFP)1</td>
</tr>
<tr>
<td>Converted nts-specific reverse ((-C_{ints}))</td>
<td>JGBPBM(trpERP)1</td>
<td>JGBPBM(trpERP)2</td>
<td>JGBPBM(trpDRP)1</td>
</tr>
<tr>
<td>Converted ts-specific forward ((C_{ts}))</td>
<td>JGBPBM(trpECF)1</td>
<td>JGBPBM(trpECF)2</td>
<td>JGBPBM(trpDCF)1</td>
</tr>
<tr>
<td>Converted ts-specific reverse ((-C_{tts}))</td>
<td>JGBPBM(trpECR)1</td>
<td>JGBPBM(trpECR)2</td>
<td>JGBPBM(trpDCR)1</td>
</tr>
</tbody>
</table>

**Table 4.1b:** Scheme of nomenclature of various primers designed to be employed in amplification of bisulphite modified DNA at the \(trp\) locus. The annotations in parenthesis are the nomenclatures used in the main text for the indicated primers. The sequences of the primers are given in section 2.1.1.3 of Chapter 2.

### 4.2.1.2 Determination of the annealing and extension temperatures

The optimum annealing temperature for all the primer pairs was determined to be 50ºC using
Chapter 4

Eppendorf Mastercycler Gradient PCR machine. Extension time was kept at 1 minute and the number of cycles was set at 30 for all subsequent reactions.

Since the DNA is expected to become 25% more AT rich after bisulphite treatment (assuming equal ATGC contents of the parent DNA and also assuming complete conversion of cytosines to uracils), it was apprehended that the AT rich template might get denatured at the regular extension temperature ($T_{\text{ext}}$) of 72°C leading to dissociation of primers from the template. A reason to presume this occurrence was a report describing the use of a $T_{\text{ext}}$ of 60°C for PCR amplification from the *Plasmodium falciparum* genome since it is highly AT rich (Su *et al.*, 1996), thus suggesting that AT rich sequences may require lower $T_{\text{ext}}$ to be amplified with greater efficiency.

To test this possibility, PCRs were set up with denatured bisulphite treated DNA as template at $T_{\text{ext}} = 60^\circ \text{C}$ and $T_{\text{ext}} = 72^\circ \text{C}$ with trpE2 N–N, N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs as a preliminary test. There was no difference in the intensity of the PCR product obtained between the two $T_{\text{ext}}$ alternatives tested. Hence the default $T_{\text{ext}}$ of 72°C was chosen to perform subsequent experiments.

![Figure 4.4: Determination of optimum $T_{\text{ext}}$ for amplification of bisulphite modified DNA with trpE2 N–N, N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs at 60°C and 72°C. It was observed that there was no difference in the efficiency of amplification at the two $T_{\text{ext}}$ alternatives tested. Hence the default $T_{\text{ext}}$ of 72°C was used to perform subsequent PCR reactions.](image)

### 4.2.1.3 Testing of various primer pairs for amplification of bisulphite modified DNA

From the three pairs of primers designed (native: N–N, converted nts-specific: C<sub>nts</sub>–C<sub>nts</sub> and converted ts-specific: C<sub>ts</sub>–C<sub>ts</sub>), theoretically seven primer combinations are possible (the first letters below before the hyphen represent the forward primer and the second letter after the hyphen represent the reverse primer): N–N, N–C<sub>nts</sub>, C<sub>nts</sub>–N, C<sub>nts</sub>–C<sub>nts</sub>, N–C<sub>ts</sub>, C<sub>ts</sub>–N and C<sub>ts</sub>–C<sub>ts</sub>.

Each of these primer pairs were tested to determine the optimum primer combinations for further work with for each locus. While C<sub>nts</sub>–C<sub>nts</sub> and C<sub>ts</sub>–C<sub>ts</sub> primer pairs are theoretically ideal to test for C-to-T conversion on either strand, one drawback of working with these primer pairs is the extremely faint and nearly undetectable amplification of DNA with these primers. In their studies Li and Manley (2005) employed N–C and C–N primer pairs to increase the efficiency of PCR (Personal communication). Incorporation of one native primer appeared to increase the annealing efficiency in the pioneer round of PCR, thereby generating more template molecules per cycle.

Thus in the present study when various nts primer pairs were analyzed for amplification of bisulphite treated DNA, it was observed that the N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs gave better amplification than the C<sub>nts</sub>–C<sub>nts</sub>.
C<sub>nts</sub> primer pair. It was frequently observed that amplification with the N–C<sub>nts</sub> primer pair was more consistent than the C<sub>nts</sub>–N primer pair. One possible explanation for this observation is that in the N–C<sub>nts</sub> primer pair, in the pioneer round of PCR, the converted reverse primer C<sub>nts</sub> selectively amplifies the converted non-template strand of transcription in the DNA. The native forward primer N then amplifies both modified and unmodified template strands of DNA with varying efficiencies in subsequent rounds. While probing the non-template strand of transcription, which if single-stranded is modified by bisulphite, inclusion of a converted reverse primer would be better suited to increase amplification efficiency. In the first round of PCR, the converted reverse primer favours the selective amplification of the bisulphite-modified strand of DNA (which is the transcriptional non-template strand), whose efficiency is rate limiting. In subsequent rounds the product so generated allows perfect annealing of the converted primer favouring accumulation of C-to-T converted PCR products in the final amplified pool. On the other hand, in the C<sub>nts</sub>–N primer pair, the native reverse primer anneals to the bisulphite converted non-template strand with reduced efficiency. This may lead to inefficient PCR amplification. Hence the N–C<sub>nts</sub> primer pair was chosen to amplify bisulphite modified DNA on the non-template strand for later experiments and the cognate N–C<sub>nts</sub> primer pair was chosen to amplify the template strand.

It was reasoned that including one native primer may have both a desired and undesired effect compared to a converted primer pair. The former would be more efficient amplification of bisulphite-modified template but the latter would be misprimed amplification of the unmodified DNA. To test whether including one native primer will cause amplification of unmodified DNA, amplification of MG1655 genomic DNA not treated with sodium bisulphite was performed at lacZ<sub>3</sub> and trpE<sub>2</sub> regions with the respective N–N, N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs. It was observed that while the N–N primer pairs gave good amplification at the regions tested, the amplification with the N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs could not be detected on gel.

Figure 4.5: Amplification of DNA untreated with sodium bisulphite with various primer pairs at lacZ<sub>3</sub> and trpE<sub>2</sub> regions. As per the expectation, N–N primer pairs amplified the DNA efficiently at both loci while amplification with N–C<sub>nts</sub> and C<sub>nts</sub>–N primer pairs was undetectable on agarose gel.

4.3 Part C: Cloning of amplified products into TA vectors

The PCR products were checked for the presence of amplification on agarose gel. Amplification with normal Taq DNA polymerase after bisulphite treatment was generally observed to be very faint in initial experiments. This may have been due to extensive nonspecific priming during PCR reactions owing to reduced sequence complexity of the bisulphite treated genome. To overcome this problem, PCR reactions were performed thereafter using Illustra Hotstart<sup>TM</sup> PCR Master Mix from GE Healthcare Ltd. This mastermix
contains a recombinant hotstart activator protein in a proprietary reaction buffer, which sequesters primers in the reaction mix thereby preventing the formation of primer dimers and nonspecific priming in PCR. Robust PCR products were observed upon amplification with this kit.

Amplified products whenever obtained were gel eluted using Qiagen gel extraction kit as per the manufacturer’s instructions. The eluate was concentrated to 10 µl. 6 µl of this was mixed with pCR™II-TOPO TA cloning vector from Invitrogen. The ligation reaction was set up as per the manufacturer’s instructions. The ligation mixture was incubated at room temperature for up to two hours.

4.3.1 Transformation and screening of clones

5 µl of the ligation mix was transformed into ultra-competent E. coli DH5α cells and the transformants were selected on LB Kan X-gal at 37ºC. The plasmid vector carried kanamycin and ampicillin resistance markers. Of these two antibiotics kanamycin was used as a selectable marker to select for transformants to avoid the problem of satellite colony growth that is associated with ampicillin selection. X-gal was also added to the medium to allow blue-white screening of the transformants. The strain DH5α was used as it allowed convenient blue-white screening owing to its lacZ– phenotype. This strain encodes only the ω fragment of beta galactosidase and allows α complementation with a plasmid that supplies the alpha fragment of the enzyme in trans. Also this strain is recA+, and thus prevents the integration of the plasmid into the host chromosome. The transformants generally took 14-16 hours to grow into visible colonies.

It was observed that three types of colonies were recovered on the transformation plate based on their X-gal phenotypes:

a. White,

b. Light blue and
c. Dark blue.

The colonies were tested for the presence of insert by colony PCR using M13Forward and M13rev18 primers. The empty vectors were expected to give a PCR product of 150 bp length while the vector with insert was expected to give a product that was the sum of the 150 bp backbone flanked by the primers and the expected length of the insert, i.e., approximately 400-450 bp.

All the white and light blue colonies appeared to carry the insert, even when obtained from the same amplicon as indicated by a PCR product size larger than 150 bp. It was presumed that inframe insertion of the α fragment of β-galactosidase in the vector owing to the small size of the insert gave rise to light blue coloured colonies. It was noted that all the dark blue colonies gave an amplification product corresponding to 150 bp and hence did not carry the insert.

The colonies with the plasmid clones of interest were grown overnight in 3 ml LB Kan broth at 37ºC and the plasmids were extracted the next morning using the Qiagen plasmid extraction kit as per the manufacturer’s instructions.
4.3.2 Sequencing of the positive clones and sequence analysis

The positive clones were sequenced using M13Forward and M13rev18 primers. The sequences of the inserts obtained were aligned against the original sequence from the published *E. coli* K-12 genome using ClustalW multiple sequence alignment software (URL: http://align.genome.jp).

C-to-T changes were graphically depicted by modifying the sequences as per the protocol described in Appendix B. Briefly, the A’s, T’s and G’s in the sequences were replaced with black dots. The C’s that were unmodified were replaced with red circles while the C’s that were converted to T’s were replaced with blue circles. This enabled quick visual perception of the C-to-T changes as blue circles when compared with the positions containing the red circles in the original sequence depiction. The graphical representation of the various clones obtained is depicted in the respective sections.

4.4 Part D: Testing for R-loop occurrence in the rho and nusG mutants using bisulphite mutagenesis

According to the hypothesis under investigation, the *rho* and *nusG* mutants suffer increased R-loops due to accumulation of untranslated transcripts. To test this hypothesis, a set of three isogenic strains was constructed to test for R-loop occurrence which is listed as here under (construction and phenotypic verification described in detail in Chapter 3):

GJ6504 = MG1655 lacI lacZU118trpR trpE9777
GJ6509 = GJ6504 rho-4
GJ6511 = GJ6504 nusG-G146D

Thus according to the hypothesis proposed, there is increased R-loop formation at the lacZU118 and trpE9777 loci downstream of the premature termination codons, specifically in strains in which *rho* or *nusG* is mutated compared to that in the wild-type strain.

Total nucleic acid preparations from GJ6504 (wild-type), GJ6509 (*rho*) and GJ6511 (*nusG*) were prepared as per the protocol given in the materials and methods section 2.2.2.3. The preparations were treated with sodium bisulphite and subjected to PCR with various native and converted primer pairs at the lacZ3 and trpE2 regions.

It was expected that in the *rho* and *nusG* strains, the template strand of the lacZ and trpE regions downstream of the premature termination codon is heteroduplexed with the untranslated mRNA leading to formation of R-loops. As a consequence the non-template strand is displaced as a single strand. This unpaired non-template DNA strand is susceptible to mutagenesis by bisulphite whereas the heteroduplexed template strand (as also DNA in regions that are not R-looped) is resistant to bisulphite treatment. The non-template strand of the wild-type strain was expected to be resistant to bisulphite treatment since R-loops were not expected to occur in this strain.

The PCR products were cloned and the insert-carrying clones were identified, and sequenced. The DNA sequence so obtained was compared against the original DNA sequence. The C residues were
subsequently depicted graphically as described in Appendix B.

4.4.1 Demonstration of single-strandedness of non-template strand at the \( \text{lacZ}_{\text{U118}} \) and \( \text{trpE}_{9777} \) in \( \text{rho} \) and \( \text{nusG} \) mutants

Total nucleic acids that were isolated from GJ6504 (wild-type) GJ6509 (\( \text{rho} \)) and GJ6511 (\( \text{nusG} \)) strains were subjected to bisulphite treatment and the 3’ end of the \( \text{lacZ} \) gene was amplified using the \( \text{lacZ3} \) N–N primer pair and the \( \text{lacZ3} \) N–C\text{nts} primer pair. Similarly the 3’ end of the \( \text{trpE} \) gene was also amplified with the \( \text{trpE2} \) N–N primer pair and the \( \text{trpE2} \) N–C\text{nts} primer pair.

As expected it was observed that C-to-T conversions occur at reasonable frequencies in the clones derived from the non-template strand of the \( \text{lacZ}_{\text{U118}} \) and \( \text{trpE}_{9777} \) loci beyond the premature transcription termination codon. The following figures illustrate the observations made from the sequence analysis of various clones obtained from the \( \text{rho} \) and \( \text{nusG} \) mutants after bisulphite treatment at the 3’ end of the \( \text{lacZ} \) and \( \text{trpE} \) mutant alleles. Thus C-to-T changes were observed in the non-template strand at the \( \text{lacZ} \) and \( \text{trpE} \) loci tested.

**Figure 4.6:** Graphical depiction of C-to-T changes on the non-template strand after bisulphite treatment in the absence of denaturation and PCR in fifteen different \( \text{lacZ} \) clones derived from the \( \text{rho} \) mutant strain, GJ6509. The top stretch represents the parental unmodified sequence followed by the sequences derived from the clones. The large red circles at the ends denote native primers where C’s and G’s have not been modified. The large blue circles at the ends denote primers where C’s or G’s have been modified, as described in the text. The small red circles denote the C’s that were not modified by bisulphite while the small blue circles denote the C’s that were modified to T’s in the sequence analyzed.

**Figure 4.7:** Graphical depiction of C-to-T changes on the non-template strand after bisulphite treatment in the absence of denaturation and PCR in ten different \( \text{lacZ} \) clones derived from the \( \text{nusG} \) mutant strain, GJ6511. Annotations in the figure are explained in legend to figure 4.6.
Figure 4.8: Graphical depiction of C-to-T changes on the non-template strand after bisulphite treatment in the absence of denaturation and PCR in twenty three different trpE clones derived from the rho mutant strain, GJ6509. Annotations in the figure are explained in legend to figure 4.6.

Figure 4.9: Graphical depiction of C-to-T changes on the non-template strand after bisulphite treatment in the absence of denaturation and PCR in eight different trpE clones derived from the nusG mutant strain, GJ6511. Annotations in the figure are explained in legend to figure 4.6.

4.4.2 Detection of single-strandedness even of the transcriptional template strand in bisulphite treated samples

The expectation from the bisulphite mutagenesis experiments was that there would be differential amplification when at least one nts-specific converted primer was employed for PCR such that the quantity of amplified product from rho and nusG would be greater than that from the wild-type strain. Also, no amplification in any sample was expected to occur when amplified with a primer pair having at least one ts-specific converted primer in any of the strains tested.

However, it was occasionally observed that the N–C_{ts} and the C_{ts}–N primer pairs at both loci gave amplification comparable to the mutant strains even in the wild-type samples, as also with the N–C_{ns} and C_{ns}–N primer pairs, which are expected to amplify C-to-T changes on the template strand. Two possibilities were considered for this observation:
• Mispriming in round 1, which would not give further C-to-T changes elsewhere in the amplicon
• Actual occurrence of C-to-T changes on the template DNA strand.

In one such instance a prominent band was observed when the wild-type sample was amplified with trpE2 N–C<sub>n</sub> primer pair. To test the nature of the amplified product, the band was gel eluted, cloned into TA vector and the clones obtained were isolated and sequenced.

**Figure 4.10:** Graphical depiction of C-to-T changes on the template strand after bisulphite treatment in the absence of denaturation and PCR in twenty three different trpE clones derived from the wild-type strain, GJ6504. Annotations in the figure are explained in legend to figure 4.6.

It was thus observed that there were vast amount of C-to-T changes on the template strand as inferred from the sequence analysis of the clones obtained. The significance of this observation was critical to be addressed and the implications of these results are elaborated upon in the discussion below. This discrepancy was later resolved when,

(i) DGGE in conjunction with bisulphite treatment was employed to test the nature of the amplified products as described in detail in Chapter 5 and evidence in favour of the hypothesis of R-loop formation in the rho and nusG mutants was obtained, and

(ii) By whole-genome analysis of bisulphite reactivity which showed that R-loops also occur genome-wide even in the wild-type strain, but at less frequency in the nusG mutant as described in Chapter 6.

**DISCUSSION**

a. **Standardization of the bisulphite mutagenesis protocol**

The results presented in this chapter describe the methodology adopted to standardize the assay of bisulphite mutagenesis in the absence of denaturation so as to employ this approach to test the
hypothesis that R-loops occur in the *rho* and *nusG* mutants of *E. coli*. Genetic evidence to this effect had been obtained in earlier studies from the laboratory (Harinarayanan and Gowrishankar, 2003) and it was proposed that one important function of Rho-dependent transcription termination is to act as a general surveillance system to prevent the occurrence of otherwise lethal R-loops on the bacterial chromosome (Gowrishankar and Harinarayanan, 2004). Accumulation of untranslated transcripts is prevented in wild-type strain proficient for Rho-dependent transcription termination in a gene that carries a premature nonsense mutation, due to induction of nonsense polarity immediately downstream of the mutation, thereby precluding (according to the model) the generation of R-loops from the nascent untranslated transcripts. This feature is compromised in a *rho* or *nusG* mutant, leading to relief of polarity at such loci, causing expression of downstream genes, as well as transcription of the mutant gene downstream of the premature termination codon that remains untranslated and therefore has the potential to form an R-loop with its complementary DNA strand. Thus it was hypothesized that R-loops occur downstream of the premature nonsense mutations at the *lacZ*118 and *trpE*9777 mutant loci, specifically in the *rho* and *nusG* mutant strains.

The results obtained so far indicate that bisulphite treatment in the absence of denaturation is a reliable method for detection of single-strandedness in the DNA samples being tested. This approach can hence be employed to detect single-strandedness due to R-loops in total nucleic acid preparations obtained from actively growing *E. coli* cultures of isogenic wild-type, *rho* and *nusG* strains carrying chain terminating mutations in the *lacZ* and *trpE* genes (construction described in detail in Chapter 3).

Various parameters of the bisulphite protocol were standardized as described in the main text. The protocol for treatment of nucleic acids with sodium bisulphite was standardized to maximize recovery of modified DNA. Native, converted nts-specific and converted ts-specific primers at various loci being tested were designed and their efficiencies of amplification tested. The PCR parameters for each of these primers were individually tested and optimized. Given the AT-richness of bisulphite modified DNA, various *T*$_{\text{ext}}$ values were tested and the *T*$_{\text{ext}}$ of 72°C was finally chosen for amplification of bisulphite treated DNA. Cloning of amplicons in TA cloning vectors and characterization of the cloned inserts was worked out in detail and the obtained sequences were finally depicted in a graphical format as explained (Appendix B).

The rate of C-to-T conversions on the non-template strand of transcription using this approach appears to be satisfactory and allows for extrapolation to assess the extent of R-loop formation at the test loci when pairs of specially designed primers at distinct genomic intervals are employed. The results conclusively establish that, in the *rho* and *nusG* mutant strains with constitutive transcription of the *lac* operon carrying an amber mutation in *lacZ* at codon 17 and the *trp* operon carrying a frameshift (Fr) mutation in *trpE* resulting in an opal mutation at codon 48, at least some proportion of the non-template strand of the 3’end of the *lacZ* and *trpE* genes is reactive to bisulphite. This implies that the non-template strand DNA at these loci is in single stranded form. The said results are consistent with the hypothesis of R-loop formation from nascent untranslated mRNA. However at this juncture these
results cannot be taken as strong evidence in favour of the model of R-loop formation in the rho and nusG mutants.

b. **Drawbacks of the current protocol**

A major drawback of the current protocol was that this approach is qualitative and the proportion of non-template strand molecules that were modified by bisulphite was not quantified and was probably very low. The use of primer pairs that were a mixture of native and converted primers complicated the analysis further. There was frequent detection of amplification in the wild-type strain with primer pairs having one converted and one native primer. Moreover, the occurrence of amplification with template-strand-specific primers even in the wild-type strain, called for further analysis to assess the biological significance of the observation. It remained to be dissected whether the observed amplifications in the wild-type strains with the nts-specific converted primers and also with the ts-specific converted primers in both the wild-type and mutant strains were a consequence of a biological phenomenon that required to be considered or whether it was merely an artifact of the approach, since PCR is a sensitive technique and can detect even low frequencies of ssDNA by this approach.

The latter possibility also gave rise to the speculation that the observed amplifications may simply be a consequence of an initial mispriming event of the converted primers onto an unconverted template. Alternatively, there could be another contributor to single-strandedness such as a moving replication fork. In LB-grown cells, there are, on average, two replication forks on each half of the chromosome. Assuming that each fork is associated with approximately 1 Kb of ssDNA, around 1:1000 cells are expected to have ssDNA at the region being probed, i.e., at the lacZ or trpE locus. It is thus possible that the single-stranded regions of DNA associated with the Okazaki fragment synthesis at the replication forks of asynchronously dividing cells (Kornberg and Baker, 1992) may also be targets for bisulphite-induced C-to-T conversions. One important consideration however is that such ssDNA is expected to be on the lagging strand of replication while the leading strand is base-paired with the nascent daughter strand. Since lacZ transcription is counterclockwise, the template strand at lacZ is the lagging strand of replication. Alternatively, the leading strand may also be single-stranded immediately behind the moving DnaB helicase. Other as yet unidentified sources of single-strandedness on the DNA may also be considered to contribute to bisulphite sensitivity. It could also be possible that R-loops occur pervasively across the genome in the rho and nusG mutants and even in the wild-type strain. This may be an important contributor to bisulphite sensitivity. However the current approach is not sensitive enough to dissect between these possibilities. Hence there was a need to further standardize the bisulphite protocol.

c. **Further standardization of the bisulphite protocol and results obtained – later chapters**

The work described in the subsequent chapters was undertaken to distinguish between these possibilities. Thus, another approach involving denaturing gradient gel electrophoresis (DGGE), in addition to
bisulphite mutagenesis was employed. DGGE is a sensitive technique that separates DNA molecules of the same size based on the differences in their melting points due to sequence variation. This approach was employed to test the composition of the amplification products obtained with each of the combinations of converted and native primer pairs. Also this technique would allow for direct quantification of the contribution of each population of amplicons (native and converted) to the final pool of amplification products. A detailed summary of the DGGE approach and the results obtained is outlined in Chapter 5. Having obtained suggestive evidence that R-loops occur at increased frequency in the rho and nusG mutants of E. coli, a detailed whole genome analysis was then undertaken of the bisulphite treated preparations and conclusive evidence for occurrence of R-loops in an E. coli strain defective for Rho-dependent transcription termination was obtained. The results also provided evidence that R-loops occur at low frequencies everywhere in the genome even in the wild-type bacterium (covered comprehensively in Chapter 6).