Chapter 1

Review of Literature
Organization of the literature review and rationale

The main focus of the work reported in this thesis is the investigation of the consequences of compromised factor-dependent transcription termination in the rho-4 (also known as rho-A243E) and nusG-G146D mutants (henceforth referred to as rho and nusG respectively, throughout the thesis, unless otherwise mentioned) of E. coli. Based on the hypothesis proposed from previous studies in the laboratory, the phenomenon of RNA-DNA hybrids (R-loops) occurrence on the chromosome in rho and nusG mutants was tested and evidence in support of the said hypothesis has been obtained. Genetic testing of synthetic interactions of rho and nusG with genes involved in recombination, replication fork resolution, transcription modulation etc., and their suppression by mutations in various rec genes has revealed novel insights into the roles of Rho and NusG in the maintenance of genomic integrity.

Accordingly, this introduction has been organized to provide a review of literature on the following topics:

i. The phenomenon of termination of transcription in E. coli (Part A)
ii. The connection between transcription and genomic integrity (namely, transcription-replication conflicts and their avoidance or resolution) (Part B) and finally,  
iii. The hypothesis investigated and the objectives of the present study (Part C).

1.1 Part A: The phenomenon of termination of transcription in E. coli

INTRODUCTION

The decoding of information encoded in the genetic material of the cell (DNA) to generate proteins is achieved by the two sequential steps of transcription and translation. The central dogma of molecular biology, first stated by Francis Crick in 1958, was introduced to indicate the flow of genetic information in organisms, as the formulation of general rules for information transfer from one polymer with a defined alphabet to another. According to this, the genetic information encoded in the DNA of an organism is transferred to RNA and from RNA to protein. This flow of genetic information is unidirectional and once it has passed into protein it cannot proceed further or be reversed (Crick, 1958; Crick, 1970).

1.1.1 Overview of transcription initiation, elongation and termination

The synthesis of RNA by the action of a DNA-dependent RNA polymerase (RNAP) is termed as transcription since the DNA sequence is copied into a corresponding RNA sequence. Transcription of DNA by RNAP is a sequential multistep process, which can be broadly divided into three distinct phases:

- Initiation,
- Elongation, and
• Termination.

A brief description of these three phases of transcription is given below, with emphasis on the third phase of transcription termination (which was the subject of the present study).

1.1.1.1 Initiation of transcription and its regulation

Gene expression in *E. coli* begins with the binding of RNAP specifically to the promoter region and initiation of transcription. The RNAP holoenzyme, which is a 1:1 complex of the core polymerase (core polymerase contains four polypeptides with the following composition: $\alpha_2\beta\beta'$ω) with the appropriate $\sigma$ factor for specificity, binds to the double-helical DNA of the promoter site, forming an initial “closed complex”. Double-helical DNA enters a long cleft in the surface of the enzyme and is held in place by a clamp domain, which is the flexible portion of the enzyme. Within the clamp domain the DNA is separated and RNA is paired to it. The $\sigma$ factor facilitates the recognition of the appropriate upstream transcription initiation site. The conversion of the “closed complex” to a binary “open complex” occurs at a specific 10-15 bp region of DNA containing the start site of transcription (+1). The nucleoside triphosphate (NTP) complementary to the template strand at the open start site (+1) binds to form a ternary initiation complex at the promoter. This complex advances by binding to the next ribonucleotide, as specified by the DNA template sequence and by catalysis of covalent bond formation to the previous ribonucleotide. The frequency of transcription initiation at the promoter by RNAP may be enhanced or retarded by transcriptional regulator proteins that are referred to as activators and repressors, respectively.

The elongation of an RNA molecule formally begins with the condensation of the first two nucleotides. However this step and the subsequent 10-12 nucleotide addition steps are considered part of the promoter clearance/promoter escape stage of initiation rather than elongation because RNAP readily releases nascent transcripts that are shorter than 13 residues and initiates new transcripts without dissociating from the promoter region. Thus, transcription initiation is a dynamic and complex process involving various types of $\sigma$ factors, activators and repressors and small molecule modulators of RNAP stability that respond to the physiological state and the external stimulus to the cell, thereby bringing about the expression of the appropriate set of genes in a particular environment.

1.1.1.2 Elongation of transcription and its regulation

The transition to the elongation phase occurs when the nascent RNA chain has 11-13 residues and is associated with three important changes in the transcription complex:

i. Movement of the core RNAP away from the promoter region of DNA,

ii. Increase in the stability of attachment of nascent RNA and

iii. Dissociation of the $\sigma$ factor from the core RNAP.

In the elongation phase, nucleotides are added to the 3’ end of the nascent RNA chain at an
average rate of about 45 nt/s until a terminator is reached. The trigger loop of RNAP makes direct substrate contacts and promotes nucleotide addition when folded as an α-helical hairpin structure. A short protein extension within the RNAP called as the rudder helps separate the RNA from the DNA and causes the two molecules to exit the RNAP molecule along different paths. The rudder and the lid domain of RNAP protrude from the mobile clamp of the enzyme. The GreA/B factors that prevent the formation of or reactivate arrested TECs due to misincorporation of ribonucleotides in the elongating chains, the Nus factors which act as antitermination factors and single molecule modulators such as ppGpp determine the rate and dynamics of RNAP TEC (the Gre, Nus factors and ppGpp are elaborated in later sections).

a. **Transcription elongation complex (TEC) stability and the definition of termination**

The TEC, a ternary complex of RNA, RNAP, and DNA is highly stable to dissociation. This stability is very important for the enzyme to catalyze polymerization of several hundreds to thousands of ribonucleotides during a single round of transcription without terminating prematurely. Once the RNAP has escaped the initiation phase of transcription, it was demonstrated that it maintains a set of contacts to ~18 bp of not-yet-transcribed DNA duplex downstream of the active site, and 8-10 bp RNA-DNA hybrid within a 12-14 nt transcription bubble, and ~5 nt of single stranded RNA at the RNA exit channel of the enzyme (Komissarova and Kashlev, 1998; Korzheva et al., 2000) However, because of this stability, the enzyme needs special signals and often assistance from special factors for termination of RNA synthesis when the end of a transcription unit is reached.

1.1.1.3 Termination of transcription

Termination occurs when the above mentioned contacts are sufficiently destabilized such that the rate of TEC inactivation and eventual dissociation becomes significant relative to the rate at which the next nucleotide is added to the growing RNA chain (von Hippel and Yager, 1992). The rate of transcript elongation defines a kinetic window within which termination occurs at a given DNA position. Transcriptional pausing is thus assumed to be the first step in a termination pathway and a prerequisite to efficient termination. However, true termination requires dissociation of the TEC, with the concomitant release of RNA and DNA from the RNAP. Thus, termination can be thought to occur in three steps (reviewed in Peters et al., 2011):

i. An initial pause,

ii. Formation of termination intermediate, and

iii. Dissociation of the TEC.

a. **Intrinsic termination and factor-dependent termination: two ways to terminate RNA synthesis**

Transcription termination occurs when a nascent RNA is released from its complex with RNAP and DNA template. The interactions between RNAP, DNA and the nascent RNA have evolved so that
almost all sequences are transcribed with a very low probability for release of the nascent chain. However, there are sequences where the probability of release is high enough to be competitive with continued elongation (Wilson and von Hippel, 1994). These sites function as intrinsic terminators and are defined by a small set of sequences. These terminators represent a major class of functionally active termination sites.

Many bacteria also use another mechanism that involves a different set of sequence signals and requires protein factors for termination. One of these factors is termed as Rho, and the factor-dependent termination is also known as Rho-dependent termination. All presently known terminators in *E. coli* can be classified either as intrinsic or Rho-dependent.

### 1.1.2 Intrinsic termination

Intrinsic termination, also known as Rho-independent or factor-independent termination relies on interactions of DNA and RNA with RNAP without the assistance of any transcription termination factors. Intrinsic terminators are found at the end of operons where they form mRNA 3’-ends and also between, within, or upstream of genes where they can regulate transcription via attenuation. Genomic analysis revealed that intrinsic terminators are present in nearly 50% of the annotated, protein-coding transcription units and approximately 70% of the non-coding RNA-transcription units (Lesnik *et al.*, 2001).

#### 1.1.2.1 Intrinsic termination signal

A large number of intrinsic terminators have been identified, and analyses of their sequences have revealed some very distinct motifs (Rosenberg and Court, 1979; Brendel *et al.*, 1986; Carafa *et al.*, 1990). They are defined by a GC-rich sequence of about 20 bp with an interrupted dyad symmetry upstream of a sequence of about 8 bp with a run of dA residues on the template strand.

The GC-rich sequence when transcribed by RNAP gives rise to an RNA segment that forms a very stable stem-loop structure. This is followed by a U-rich sequence on the RNA. The distance from the end of the stem-loop structure to the 3’ end of the RNA is about 7-9 nt (Rosenberg and Court, 1979; Carafa *et al.*, 1990; reviewed in Peters *et al.*, 2011). These U residues are paired with dA in the template DNA, and the ensuing hybrid occupies the entire site within the transcribing RNAP molecule, while the GC-rich stem-loop occupies the site engaged by the single stranded RNA that exits the polymerase. Steric hindrance in the RNAP molecule due to the stem-loop and RNA-DNA hybrid causes destabilization of the TEC owing to simultaneous disruption of interactions in the above mentioned sites in the RNAP.

Intrinsic termination process can be divided into at least four distinct steps whose features are fairly well understood:

- Transcriptional pausing at the intrinsic terminator,
• Nucleation of the terminator hairpin,
• TEC disruption, and
• TEC dissociation.

a. **Transcriptional pausing at the intrinsic terminator**

Pausing is the first step of termination where nucleotide addition stops, formation of the stem-loop structure occurs and the U-tract forms an RNA-DNA hybrid. The pause is primarily induced by the U-tract prior to stem-loop formation. It also appears likely that the sequences of the exiting RNA, the nucleotides in the active site, and the downstream DNA duplex also contribute to the duration of this pause (Lee et al., 1990; Chan et al., 1997; Kireeva and Kashlev, 2009; reviewed in Peters et al., 2011).

![Figure 1.1](image1.png)

**Figure 1.1**: Schematic representation of various steps in intrinsic termination: (a) Transcriptional pausing at the intrinsic terminator, (b) Nucleation of the terminator hairpin, and (c) TEC disruption (figure adapted from Peters et al., 2011).

b. **Nucleation of the terminator stem-loop**

Nucleation of the stem-loop occurs by closure of the stem-loop by one to several base pairs. The stem-loop secondary structure is thermodynamically very stable. Pausing induced by the U-tract seems to
favour the formation of the stem-loop. Stem-loop formation appears to be facilitated by the flap and zinc binding domains at the RNA exit channel. Upon nucleation, the stem is likely to extend quickly within 1-2 nt of the upstream end of the RNA-DNA heteroduplex where it encounters a potent barrier (reviewed in Nudler and Gottesman, 2002).

c. TEC disruption and dissociation

The stem-loop structure extends to less than 8 nt from the RNA 3’ end. This causes melting of ~3 bp of the RNA-DNA heteroduplex by extracting the RNA strand from the hybrid. Hybrid melting disrupts and destabilizes the TEC to the extent that dissociation becomes kinetically favourable (Gusarov and Nudler, 1999; Komissarova et al., 2002; Larson et al., 2008; reviewed in Peters et al., 2011).

d. Sequence conservation of the intrinsic terminators

A terminator can be defined as a point where the rate of release of an RNA transcript is greater than the rate of addition of the next nucleotide (Von Hippel and Yager, 1992). Intrinsic terminator sequences have characteristics that facilitate a significant increase in the rate of release of RNA and a significant decrease in the rate of nucleotide addition when compared with random sequences.

Three different aspects of a terminator sequence have properties that are likely to be conducive to such facilitations (figure 1.1):

i. GC-rich sequence with interrupted dyad symmetry: The sequence exhibits a strong bias for GC at the five positions nearest to the U-tract with a modest preference for G at –3’ end of the stem-loop. This is in agreement with the significance of the bottom of the stem in supplying the energy that destabilizes the TEC in the final step of termination.

ii. A run of dA residues in the template strand: This causes a run of U residues at the 3’ ends of the transcribed RNA. This stretch of residues aids in pausing of the TEC and forms a weak RNA-DNA hybrid (being comprised of A-U base pairs) which can be readily disrupted during dissociation of the transcription TEC.

Among the terminators examined by Peters et al. (2009) using a subset of 100 intrinsic terminators listed in the RegulonDB database, approximately half had a perfect or near-perfect U-tract, whereas the remaining had at least two non-U residues with at least one C or G in the distal U-tract (imperfect U-tract).

iii. The sequence downstream of the termination stop point: The DNA sequence downstream of the terminator is not transcribed to become part of the RNA transcript and is within or just downstream of the contact point between the RNAP and DNA in the transcription TEC at the site of transcription termination. The downstream DNA sequence exhibited distinct differences in the AT content for the near-perfect and the imperfect U-tract containing terminator sequences. Imperfect U-tract terminators had a higher AT content at positions +10-12, while near-perfect U-tract terminators had a low AT content at the same positions. These two classes also exhibited opposite sequence preferences at positions +18-19 (reviewed in Peters et al., 2011). Thus, the downstream DNA sequence could possibly increase the rate of transcription
termination at the stop point.

1.1.2.2 Intrinsic termination mechanism

The process of intrinsic termination has two features:

• Changes in the structure of the nucleic acid, and
• Changes in the conformation of RNAP.

a. Changes in the structure of the nucleic acid during intrinsic termination

Chemical probing and cross-linking experiments suggest that the upstream 3 bp in the RNA-DNA heteroduplex melt upon the extension of the terminator stem-loop structure. Three alternative hypotheses have been proposed to explain this phenomenon of melting and release (reviewed in Peters et al., 2011):

• Hybrid shearing,
• Hypertranslocation, and
• Stem-loop invasion.

Hybrid shearing assumes that the extension of the stem-loop pulls the RNA out of the exit channel by transiently breaking and reforming base pairs in the hybrid as the RNA shifts out of register with the DNA strand. Thus, RNA is pulled out, while RNAP is static. In hypertranslocation, RNAP is pushed forward by the stem-loop extension but retains the register of the RNA-DNA heteroduplex. In both cases, the length of the RNA-DNA hybrid is reduced. The stem-loop invasion hypothesis assumes that the stem-loop extends into the main cleft of RNAP and causes RNA-DNA heteroduplex melting due to steric hindrance. Various lines of experimental evidence seem to support each of the hypotheses postulated above (Macdonald et al., 1993; Yarnell and Roberts, 1999; Komissarova et al., 2002; Touloukhonov and Landick, 2003; Santangelo and Roberts, 2004; Epshtein et al., 2007; Larson et al., 2008).

b. Changes in the conformation of RNAP

The TEC consists of RNAP, DNA and the nascent RNA. Although the structural and sequence characteristics of DNA and RNA are well documented, the nature of conformation changes in the RNAP have not yet been determined conclusively.

It is assumed that the RNAP clamp domain in the open conformation would favour transcription termination. The initial step of transcriptional pausing is proposed to involve clamp movement. The stem-loop may destabilize the TEC leading to the disruption of stabilizing interactions in the heteroduplex binding site and the RNA binding site thus causing instability of the DNA binding site, presumably by the opening up of the flexible clamp structure (reviewed in Nudler and Gottesman, 2002). The trigger loop of RNAP may be trapped in an inactive configuration with the nascent RNA stem-loops, the heteroduplex and downstream DNA and may facilitate transcription termination (Touloukhonov and Landick 2003; Touloukhonov et al., 2007).
Chapter 1

An “allosteric” model of intrinsic termination has been recently proposed to explain the conformation changes in RNAP. In this model, the terminator stem-loop dissociates the TEC, even as the RNA 3’ end remains in the active site by sweeping across the RNAP main cleft and disrupting the stabilizing contacts in the TEC. This causes trigger loop folding by direct contact on the trigger loop and dissociates the TEC (Epstein et al., 2007).

1.1.3 Rho-dependent transcription termination

In *E. coli*, transcription and translation are tightly coupled. Even as the nascent transcript is being synthesized, ribosomes occupy the emerging transcript and bring about translation. The translating first ribosome directly assists RNAP during elongation and thus facilitates a precise match of translational and transcriptional rates under various growth conditions (Proshkin et al., 2010; reviewed in Nudler, 2012). That such coupling is not an incidental consequence of the absence of a spatial barrier in bacteria is underscored by the phenomenon of nonsense polarity, first identified in the lac and trp operons of *E. coli* more than 40 years ago (Newton et al., 1965; Yanofsky and Ito, 1965). Nascent transcripts that are not simultaneously translated, such as at the ends of genes or operons or those in which there is stochastic uncoupling of transcription and translation, are terminated by Rho so that the occurrence of translation-uncoupled transcripts is minimized within the cells (Cardinale et al., 2008; Peters et al., 2009). Nonsense polarity refers to the abolition of expression of intact promoter-distal genes in an operon that bears a nonsense mutation which stops translation in a promoter-proximal gene, and is mediated by premature termination of transcripts in the region immediately downstream of the nonsense mutation (Adhya and Gottesman, 1978; Nudler and Gottesman, 2002).

In bacteria, a major mechanism of termination of transcription, apart from intrinsic termination, is via a protein called Rho, which functions as an RNA/DNA helicase. Rho binds to the nascent RNA and moves along the direction of transcription through energy derived from ATP hydrolysis (von Hippel and Delagoutte, 2001; Richardson, 2002; reviewed in Richardson, 2003). Rho is essential for viability in *E. coli*.

Rho-dependent terminator sequences are bipartite and spread over 150 - 200 bp of DNA (Richardson and Greenblatt, 1996). One part of the terminator region is a 40 nt long RUT (Rho-utilization site) element that lies upstream in the nascent RNA. It has little or no secondary structure and is enriched in C resides (Hart and Roberts 1991). The other part is the point where termination occurs, called the transcription stop point (tsp), and is generally spread over 100 bp of DNA in clusters of preferred points called subsites. Rho-dependent terminator sequences do not seem to conform to a simple consensus as they are very diverse. Rho-dependent terminators are widespread and cryptic within ORFs.

Studies on the general preference of Rho for RNA indicate that Rho has a very strong affinity for poly(C), a moderate affinity for single-stranded RNAs without C residues, and a much lower affinity for RNA molecules which are base paired among themselves or with other RNA molecules (Galluppi and
Richardson, 1980).

At least 1264 putative Rho-dependent terminator sites have been identified in the *E. coli* genome by Peters *et al.* (2012). They analyzed the variation in the distribution of RNAP in *E. coli* upon treatment with bicyclomycin (BCM), a Rho-specific inhibitor. By employing chromatin immunoprecipitation microarray (ChIP-on-chip) and RNA-seq they observed that BCM increased the proportion of antisense transcripts and proposed that this is caused by readthrough of Rho-dependent terminators. In an earlier study, they found that BCM seemed to shift RNAP downstream of the apparent termination site (Peters *et al.*, 2009; reviewed in Peters *et al.*, 2011).

### 1.1.3.1 Transcriptional polarity: Consequences of Rho-dependent termination

Termination of untranslated transcripts by Rho results in polarity. Polarity is defined as the decrease in expression of distal genes in an operon as a consequence of a premature stop codon or inefficient translation in an upstream gene within the same operon.

Peters *et al.* (2011) have described two prevailing models for polarity:

- **Rho-ribosome competition for RNA**
  
  According to this model, the presence of a ribosome prevents Rho from being associated with the nascent RNA (Adhya and Gottesman, 1978). Active translation causes the ribosomes to occupy the RUT sites, thereby preventing Rho activation and thus termination. Under conditions of optimal translation, Rho can only bind to RUT sites 5’ to the ribosome and hence cannot access the TEC due to the ribosome acting as a roadblock. In the event of slow or terminated translation, Rho can bind RUT near RNAP, causing termination to occur, thereby resulting in polarity.

- **Rho-ribosome competition for NusG**
  
  As per this model, Rho and the ribosome are two alternate targets for the C-terminal domain (CTD) of NusG. When transcription is coupled to translation, the S10 subunit (also called NusE) of the ribosome binds to the CTD of NusG, thereby blocking its access to Rho. However, in the event of dissociation of the ribosome, as in an encounter with a premature stop codon or in inefficient translation, NusG can bind to Rho at its CTD to cause polarity by enhancing transcription termination (Burmann *et al.*, 2010). Recent work by Peters *et al.* (2012) however, has implications on this model of transcriptional polarity. Since only a minority subset of Rho-dependent terminators was found to require NusG in this study, most polarity suppression is not explained by this model. Further work is needed to test this model.

#### a. Structure and biochemistry of Rho

Rho is a homohexamer of a single polypeptide with 419 amino acids encoded by the *rho* gene (Roberts, 1969; Richardson and Greenblatt, 1996). The subunits have a compact globular
shape with a diameter of approx. 4.2 nm and are organized in a flat hexameric ring (Skordalakes and Berger, 2006). Sequence and functional analyses have indicated that each Rho subunit has two distinct RNA binding and NTP binding domains. The first 125-130 residues of each monomer form the N-terminal RNA binding domain, which can bind to C-rich RNA (or DNA). Its structure consists of a three-helix bundle followed by a 5-stranded β-barrel with an OB-type fold (Allison et al., 1998; Bogden et al., 1999; reviewed in Richardson 2003). In the crystal structure of a functional Rho, the tertiary structures of the individual RNA binding domains are very similar to that determined for an isolated domain.

The C-terminal 131 through 419 residues contain the binding site for ATP and are involved in ATP binding and hydrolysis (Bear et al., 1985; Dombroski and Platt 1988; Dolan et al., 1990). The sequence of this domain is closely related to the sequences of the corresponding parts of the α and β subunits of the F₁ ATPase. This enabled modeling of the tertiary structure of Rho using the F₁ ATPase as the template. A single hexamer of Rho thus consists of a ring of ATP-binding subunits with three catalytic and three non-catalytic nucleotide binding sites (Miwa et al., 1995; Stitt and Xu 1998; Kim et al., 1999; reviewed in Richardson 2003).

Skordalakes and Berger (2006) have determined the functional form of Rho to be a closed ring “trimer of dimers”. They found that the Q-loop of Rho adopts two distinct configurations that alternate between adjacent protomers. RNA binds in a shallow channel formed by the interface of two adjacent ATPase domains, where it contacts Rho’s second translocation element, the R-loop.

Upon binding to RNA, a series of cycles of ATP binding, hydrolysis and release bring about conformational changes in the cavity of the hexamer that causes the transcript to pull through and induce Rho to translocate along the RNA in the direction of transcription. Functionally, the interactions between Rho and the nascent RNA that are coupled to ATP hydrolysis cause the dissociation of the RNA from its complex with RNAP and DNA (Richardson and Conaway, 1980). According to the “translocation model” Rho catches up with the TEC and pulls the transcript away from RNAP and the template DNA (reviewed in Nudler and Gottesman, 2002).

Rho appears to terminate RNA even from yeast RNAP II when the DNA has a strong RUT site (Wu and Platt, 1993). It also dissociates ssDNA base-paired to the 3’ end of an RNA molecule, thus acting as an RNA/DNA helicase. Rho does not seem to dissociate DNA molecules that are base paired at the 5’ end of the RNA (Brennan et al., 1987). It was also found that Rho can unwind RNA-RNA helices (Steinmetz et al., 1990).

b. Rho-dependent termination signal: The sequence elements

Rho-dependent transcription termination occurs in response to specific sequence elements in the RNA being terminated. One distinct characteristic feature of Rho-dependent terminators is the richness of rC residues in the mRNA. Rho-dependent terminators are bipartite, consisting of two distinct but partially overlapping parts that extend over 150 to 200 bp of DNA (Richardson, 1990; reviewed in Ciampi, 2006). These two fairly broad regions in the terminator signal are:
the RUT element, and

ii. The tsp region.

The minimal length of RNA required for termination is ~80 nt (Hart and Roberts 1991).

i. **The RUT element of Rho-dependent terminator signal**

Rho is an RNA binding protein that has a distinct preference for the type of RNA molecule to which it binds (Galluppi and Richardson, 1980). Its preference for binding to the RNA molecules correlates with the presence of C rich sites, designated as RUT sites. Binding of Rho was shown to be impaired when the nascent transcripts have not been extended to the end of the RUT region or when the full length transcripts were made in which the RUT site has been deleted (Faus and Richardson, 1989). Moreover DNA oligonucleotides that are complementary to the segment of RNA carrying the RUT site seemed to specifically block Rho-mediated transcription termination. The most likely explanation of this observation is that these oligonucleotides can compete for binding the RUT element with Rho and hence block access of Rho to the nascent RNA (Chen and Richardson, 1987).

Analyses of well-characterized and putative RUT elements revealed that they consist of stretches of segments that are rich in C residues and are not likely to form a secondary structure or pair with other segments of RNA (Morgan et al., 1985). Another sequence characteristic of Rho-dependent terminators is a relatively low G and a high C content in the part just preceding the 3’ end of the terminated transcript (Alifano et al., 1991). The low level of G appears to be significant given the propensity of G residues for pairing with other residues and thus contributing to double-helical secondary structures in the RNA.

ii. **The tsp region of Rho-dependent terminator signal**

The sequence characteristics of the tsp region are far less understood than that of the RUT element and can be spread up to 120 bp distal to RUT. Several studies seem to suggest that Rho causes RNAP to terminate favourably at natural pause sites downstream of a RUT region. This behaviour of Rho also seems to be kinetically matched to the movement of RNAP along the DNA template (Jin et al., 1992). These observations also seem to suggest that Rho acts mainly as an RNA release factor and its specificity is determined by where it is able to bind to the RNA and also where the RNAP is on the DNA after the bound Rho has initiated its action. In general, the sequence elements that define a pause site also seem to determine a tsp region (reviewed in Nudler and Gottesman, 2002).

iii. **Distribution of targets of Rho**

Rho-dependent termination of transcription appears to be widespread in the prokaryotic world. However, most of the characterized Rho-dependent terminators occur in enteric bacteria. The phenomenon of Rho-dependent transcription termination is a complex process, which is subject to control in a variety of ways, enabling mechanisms to evolve in response to a variety of conditions.

Peters et al., (2012) performed genome-wide ChIP-chip analysis and RNA-seq analysis in the presence of bicyclomycin (BCM), which is a known inhibitor of Rho. They identified ~1264 putative Rho-dependent termination sites (termed as bicyclomycin sensitive targets or BSTs) in the genome of *E.
coli K-12. Three broad categories of Rho-dependent terminators were identified in this study:

i. Antisense BSTs were proposed to be caused by readthrough of Rho-dependent terminators either upstream of an oppositely oriented gene or within a gene.

ii. Sense BSTs were caused from readthrough of terminators at the ends of genes or unannotated transcripts or within genes.

iii. Intergenic BSTs resulted from readthrough of terminators at the ends of genes or where no gene is known to exist.

They found that a vast majority of Rho-dependent terminators controlled antisense transcription, which led to the proposal that the primary function of Rho is to suppress antisense transcription inside the cell.

1.1.3.2 Kinetic coupling model of Rho-dependent transcription termination

Rho translocates on RNA that is actively synthesized by RNAP. Termination is postulated to commence when there is little or no RNA between Rho and RNAP. Jin et al., (1992) proposed the kinetic coupling model after obtaining evidence for the relationship between the rate of translocation of Rho and the rate of elongation of RNAP. They analyzed the mechanism with which termination occurs in various allelic combinations of Rho and RNAP. They demonstrated directly that termination efficiency at Rho-dependent terminators is governed by a precise concurrence between the rates of action of RNAP and Rho. Even minor increases in the rate of elongation of RNAP or decreases in the rate of action of Rho led to considerable readthrough at several Rho-dependent terminators. Conversely, a small decrease in the elongation rate of RNAP significantly increased the termination efficiency of wild-type Rho. Thus, competing rates of RNA elongation and Rho translocation appear to dictate the efficiency of termination (Shashni et al., 2012).

1.1.3.3 Mechanism of Rho-dependent transcription termination

The key question as to how Rho dissociates the TEC remains yet to be answered. Different models have been proposed to explain how Rho terminates transcription. Transcription termination by Rho can be envisaged to occur in the following steps (Reviewed in Banerjee et al., 2006):

• Binding of Rho to RNA,
• RNA translocation through Rho and termination.

a. Binding of Rho to RNA

The structure of intact Rho enabled determination of how the RNA binding domains were connected to the C-terminal domains. These domains together form a cup-shaped structure with the RNA-binding clefts facing inwards towards the axis of the Rho hexameric ring. The clefts on each subunit are equidistant from each other around the inside and are oriented at an angle of 75° to a plane perpendicular to the ring axis (reviewed in Richardson 2003).
Electron microscopy images depict Rho hexamers in either a closed state or a “notched” state when a short (23 nt) RNA cofactor is present (Gogol et al., 1991). When longer RNAs (100 nt) are added that exceed the capacity of the primary site, Rho hexamers convert from the notched state to the closed state.

RNA seems to associate with a cleft in the RNA binding domain of Rho. This domain is large enough to fit only pyrimidines and exhibits interaction preferences with C residues. The six RNA binding domains become filled with nearly 60 nt to form a zigzag conformation from one subunit to the next. This is consistent with the minimum length of the sequence required for effective termination by Rho (>60 bp) (Hart and Roberts, 1994).

Based on the work done by Kim and Patel (2001), it has been proposed that RNA randomly collides with the Rho molecule to form a dynamic Rho-RNA complex. RNA then wraps around the primary binding site of each of the five protomers of Rho open complex to form a stable moiety. The RNA then passes towards the centre of the open ring of Rho and makes contact with the secondary binding site of Rho to form a steady complex. The Rho hexamer is completed by the addition of a sixth subunit after the RNA enters the ‘ring’ of Rho, leading to the formation of a highly stable ATPase competent moiety. The translocase/helicase activity of Rho is then fuelled by ATP hydrolysis to cause unwinding of the RNA-DNA heteroduplex (reviewed in Banerjee et al., 2006). Thomsen and Berger (2009) have determined four distinct ATP-binding states, representing potential catalytic intermediates within the Rho hexamer that are coupled to RNA positioning through a complex allosteric network.

Binding of Rho to C-rich nascent RNA triggers its RNA-dependent ATPase activity, which provides energy for translocation. While Rho can even bind to DNA, activation of ATP hydrolysis and helicase functions are specific for RNA. Crystallographic and electron micrographic studies suggest that Rho initially binds to RNA in an open, “lock-washer” conformation (Skordalakes and Berger, 2003). It then isomerizes into a closed ring structure as RNA moves to the central cavity. This transition from open to closed states is postulated to represent a shift from RNA loading to translocation-competent forms of Rho (reviewed in Peters et al., 2011).

Rho normally acts to terminate transcription at the ends of genes and can cause premature transcription termination in the absence of translation. As a consequence, the major deciding factor of Rho function in transcription termination is the absence of a ribosome upstream of the nascent RNA. In a typical operon, a ribosome loads onto the nascent RNA as soon as its attachment site is accessible and translates the RNA. The rate of translation is synchronous with the rate of transcription, such that the two processes are coupled. At the end of the operon, the ribosome does not progress along the RNA and the segment being transcribed becomes accessible to Rho. This situation is also consistent with the observed spacing of 90 nt between the translation termination codon of bacteriophage λ cro gene and the 3’ end of the RNA isolated from cells (Court et al., 1980). Of these 8 nt have been postulated to be protected by the ribosome, approximately 60 nt appear to occupy the primary binding site of Rho, 12 nt are expected to be needed to pass through the hexameric ring and another 14 nt are supposed to be buried in the
RNAP, thereby being conducive to Rho-dependent termination. The loading of Rho does not require a free 5' end of the target RNA (Burgess and Richardson, 2001; reviewed in Banerjee et al., 2006).

For RNAs that are not translated (tRNA and rRNA), the extensive secondary structure eliminates the possibility of an extended single-stranded region at the 3' end. Hence these molecules are not amenable to termination by Rho.

b. Translocation of RNA through Rho and termination

The actual mechanism by which Rho disrupts the elongating transcription complex is not very well understood. However, genetic, biochemical and structural data support the translocation model (Briercheck et al., 1998; Yu et al., 2000). Once Rho is bound to RNA, conformational changes in the cavity of the hexamer brought about by continuous cycles of ATP binding, hydrolysis and release cause the transcript to be pulled through, thereby translocating Rho along RNA in a 5' to 3' direction. This model postulates that Rho catches up with the TEC and pulls the transcript away from RNAP and the DNA template. Rho must therefore translocate faster than RNAP to cause efficient transcription termination.

Based on biochemical evidence obtained so far, two broad models of RNA translocation through Rho can be considered namely the “translocation model” and the “tethered tracking model”.

a. In the “translocation model”, RNA contacts the central cavity of Rho during translocation. Proposed by Geiselman et al. (1993), this model is based on the symmetry of the Rho hexamer and the detection of three strong and three weak ATP-binding sites as well as three strong and three weak RNA-binding sites per hexamer. These subunits are arranged with a pseudodyad symmetry in which one subunit is in a conformation with a strong RNA-binding site and the other is in a conformation that causes weak RNA-binding. The same arrangement is expected at the ATP binding site, although the strong ATP-binding site need not be on the same subunit as the strong RNA-binding site. The role of ATP is to switch conformations.

Thus, in a hexamer, a conformation switch in one dimer would cause the 5’ segment of an RNA bound to a strong site to be now bound in a weak conformation. This causes the 5’ segment to dissociate and be replaced by a 3’ segment such that it would have the right orientation so as to fit into the subunit with the strong site in the switched conformation. As this process continues along other dimers in the hexamer, Rho moves progressively towards the TEC.

b. The translocation model proposed by Geiselman et al., did not satisfactorily explain the continued contact of Rho with the RUT site on the RNA. The “tethered tracking model” (Steinmetz and Platt 1994) assumes that the RUT element of RNA remains bound to Rho during translocation at the strong RNA-binding sites. Dynamic interactions with the weaker sites would extend the contacts of Rho gradually towards the 3’ end. ATP hydrolysis provides the necessary energy for the helicase activity of Rho (Brennan et
The mechanism for ATP hydrolysis and hybrid unwinding is assumed to be similar to that of related DNA helicases (Bird et al., 1998).

In both models Rho attaches the RUT site and moves along the RNA. However the difference between the two models is that while in the former Rho leaves the RUT site, in the latter Rho remains attached at this site.

Rho dissociates the TEC, when halted at a pause site in the final step of termination. Thus, a pause site renders the TEC susceptible to termination by Rho. The distance from the RUT site, the stability of the TEC, sequence of the 3’ end of the RNA in the active site of RNAP, susceptibility to backtracking and the structure of the nascent RNA contribute to the efficiency of termination apart from the general determinants of pausing (reviewed in Peters et al., 2011).

1.1.3.4 Association of Rho with the TEC

Some workers have suggested recently that Rho associates with the TECs in vivo even when transcription termination does not occur. The evidence for Rho-TEC association in vivo was obtained by ChIP-chip experiments where similar genome-wide distributions of Rho and RNAP were observed (Mooney et al., 2009). In vitro Rho was observed to be retained by the beads on which the TECs were immobilized and was proposed to associate with RNAP prior to loading onto mRNA (Epshtein et al., 2010). On the other hand work done by Shashni et al., (2012) contradicted the view that Rho is associated with RNAP and provided evidence in support of kinetic coupling between Rho and RNAP.

It is also noteworthy that Rho is reported to be ~0.1% of total cellular protein in E. coli (Imai and Shigesada, 1978). This corresponds to ~1400 Rho hexamers per cell. Intriguingly, ~1250 of the 13,000 RNAP molecules are in TECs (Grigorova et al., 2006). Thus there appears to be a stoichiometric balance between the number of RNAP molecules in the TEC and the number of Rho hexamers.

The association of Rho with the TECs also lends support to the kinetic coupling model. Since the local concentration of Rho would be higher, it would allow Rho to engage RNA rapidly in termination when a RUT site becomes exposed.

1.1.3.5 Nucleic acid structure changes during Rho-dependent termination

Rho is a potent molecular motor that can apply force through RNA translocation (Schwartz et al., 2007) and is also known to be an RNA/DNA helicase. These activities of Rho substitute for the need of a stem-loop structure in a Rho-dependent terminator sequence. Consequently, (similar to what was discussed for the nucleic acid structure changes in intrinsic termination), Rho-dependent termination can be envisioned as occurring through one of the following mechanisms:

i. Hybrid shearing: The RNA is translocated through Rho until it becomes taut. This results in a pulling force that disrupts the RNA-DNA hybrid (Richardson, 2002).

ii. Hypertranslocation: Rho exerts a pushing force which causes RNAP to translocate forward on
the DNA template without the extension of the nascent RNA (Park and Roberts 2006).

iii. Invasion: The 3’ end of the RNA remains in the RNAP active site and the RNA-DNA hybrid is unwound by the helicase activity of Rho (Epshtein et al., 2010).

### 1.1.3.6 RNAP conformational changes during Rho-dependent termination

Epshtein et al., (2010) reported that Rho remains associated with RNAP throughout the transcription process. This binding is independent of the binding of Rho to the transcribed RNA, and is supposed to be crucial for the phenomenon of transcription termination, i.e., TEC inactivation and its dissociation. They proposed “conformational change model” which predicts that Rho pushes against the RNAP lid domain and causes clamp movement. This results in the unfolding of the trigger loop causing further clamp opening, loss of nucleic acid contacts and thereby, irreversible inactivation of the TEC. This allosteric mechanism of termination is general and likely to be preserved for all cellular RNAPs throughout evolution (reviewed in Peters et al., 2011). This proposal is substantiated by the observation that tagetitoxin, a chemical that binds directly to the trigger loop, inhibits Rho-dependent transcription termination. Additional evidence lending support to the model is yet to be obtained.

### 1.1.3.7 Physiological functions of Rho

Rho appears to have originated early in the evolutionary history of prokaryotes. Termination by Rho has been suggested to play various roles in cellular physiology. These (not mutually exclusive) functions include (reviewed in Peters et al., 2011):

i. generation of discrete transcripts and release of RNAP,

ii. silencing of horizontally transferred DNA (foreign DNA),

iii. suppression of antisense/pervasive transcription, enforcement of transcriptional polarity,

iv. prevention of R-loops (RNA-DNA hybrids) and

v. resolution of transcription-replication conflicts by eliminating stalled TECs that interfere with DNA replication.

Of these, function (i) is a well understood phenomenon, (ii) and (iii) are discussed in the following subsections, while (iv) and (v) are discussed separately in sections 1.1.3.9 and 1.2.3.1 respectively.

#### a. Silencing of foreign DNA

Genome-wide analyses by Cardinale et al., (2008) and Peters et al., (2009) indicated that Rho silences horizontally acquired DNA (reviewed in Peters et al., 2011). It is possible that foreign DNA will have a greater number of Rho-dependent terminators that are involved in regulation, such as the \( t_{\text{imm}} \) terminator which suppresses the induction of toxic genes of the rac prophage (de Crombugghe et al., 1973). Alternatively, insertion of foreign DNA into an active transcription unit may require Rho to terminate transcription in the foreign segment to compensate for the loss of a natural terminator. This
may occur if phage integrases remove intrinsic terminators from tRNA genes during recombination (Peters et al., 2009). Suboptimal codon usage in foreign DNA could inhibit translation, thus exposing RUT sites in the RNA, which could also be a possible mechanism of Rho action on foreign DNA to terminate transcription.

b. Suppression of antisense/pervasive transcription

Rho-dependent termination is thought to be better suited to terminate antisense or pervasive transcription than intrinsic termination because coding requirements on the sense strand make encoding an antisense intrinsic terminator difficult (Peters et al., 2009). Indeed, Dornenburg et al., (2010) reported the occurrence of ~1000 antisense RNA sequences by employing deep RNA sequencing. Antisense transcription may thus be a general target of Rho-dependent transcription termination in vivo. Pervasive transcription occurs on both genomic strands and has been discovered in E. coli as also in other bacteria (Selinger et al., 2000; Dornenburg et al., 2010; Thomason and Storz, 2010; Raghavan et al., 2012).

More recently Peters et al. (2012) have found that the antisense strands of nearly 34% of all genes of the E. coli genome were significantly upregulated upon BCM treatment that inhibits Rho-dependent termination. These results were proposed to indicate that a major function of Rho is to suppress antisense transcription.

1.1.3.8 Role of NusG in Rho dependent termination

Rho appears to be sufficient to terminate transcription in vitro at several Rho-dependent terminators tested. However, in vivo, an additional factor, NusG is required for efficient termination at several Rho-dependent terminators (reviewed in Ciampi, 2006).

NusG is a 21 kDa essential E. coli protein that enhances transcription termination by Rho. It is a monomeric protein and has two conserved domains – NTD (residues 1-118) and CTD (residues 126-181) with a flexible linker in between. The NTD has β’ clamp helices, interactions with which facilitate binding to RNAP (Mooney et al., 2009). The CTD serves as a binding site for Rho. Burmann et al., (2010) have found that the S10 protein (also known as NusE) also interacts with the CTD of NusG to form a stable complex during translation. Binding of Rho or S10 to the CTD of NusG is mutually exclusive. However, the affinity of NusG for Rho is much higher than that for S10 (reviewed in Peters et al., 2011).

NusG was originally isolated as a factor facilitating antitermination by N protein of bacteriophage λ (Li et al., 1992) It has long been considered as a transcription elongation factor that stimulates the rate of transcription (Burova et al., 1995) and enhances suppression of class II polymerase pause sites (Artsimovitch and Landick, 2000). Its requirement in Rho-dependent termination was investigated first by Sullivan and Gottesman (1992). They demonstrated that cells devoid of NusG had reduced efficiency of transcription termination at certain Rho-dependent terminators. Moreover, certain genes such as galE are totally dependent on NusG for Rho-dependent termination (reviewed in Ciampi,
Interactions with both Rho and RNAP are critical for enhancement of termination by NusG (Mooney et al., 2009). The interactions of NusG with RNAP may cause the enzyme to be more susceptible to Rho actions that can cause the release of the nascent transcript (Burns et al., 1999; reviewed in Ciampi, 2006).

It has been reported that NusG is essential in *E. coli* K-12 because it increases Rho-dependent termination upstream of the *kil* gene in the *rac* prophage (Mooney et al., 2009). However, NusG is dispensable in *Staphylococcus aureus* and *Bacillus subtilis* with even Rho being non-essential in the latter. Although Rho-dependent transcription termination was found to be stimulated by NusG the exact mechanism by which this occurs is yet to be elucidated (Sullivan and Gottesman 1992; Nehrke et al., 1993; Chalissery et al., 2011). Mooney et al. (2009) performed *in vivo* and *in vitro* studies to demonstrate that both NTD and CTD of NusG are required for enhancement of Rho-dependent transcription termination.

It is a matter of debate as to whether NusG stimulation occurs at all Rho-dependent terminators *in vivo* (Cardinale et al., 2008; Burmann et al., 2010) or a subset of Rho-dependent terminators (Sullivan and Gottesman, 1992; Peters et al., 2012).

### 1.1.3.9 Role of NusA in Rho-dependent transcription termination

Friedman and Baron (1974) identified NusA as a host factor that is required for N-mediated transcription antitermination during the lytic growth of phage λ. Bubunenko *et al.* (2007) found NusA to be essential for viability in *E. coli*. NusA is postulated to participate in termination of transcription by enhancing the pausing of RNAP, thus lowering the rate of transcription elongation (Zheng and Friedman 1994). By causing RNAP to pause at a point between the start point of translation and the first intragenic, Rho-dependent terminator, NusA would allow the ribosome to bind and initiate translation before the RNAP passes into the termination region with an unprotected nascent RNA. Gottesman and coworkers (Ward and Gottesman, 1981; Sullivan and Gottesman, 1992; Cardinale *et al.*, 2008) have suggested that both NusA and NusG act cooperatively to promote Rho-dependent termination.

Work in the current laboratory has identified a recessive substitution mutation in NusA (*nusA*-R258C) that confers relief of transcriptional polarity and was found to be lethal in combination with a missense mutation in *rho* (*rho*-A243E or *rho*-4) or a missense mutation in *nusG* (*nusG*-G146D) (Saxena and Gowrishankar, 2011b). The data presented in this study lend support to the findings of Cardinale *et al.* (2008) that NusA promotes factor-dependent transcription termination.

### 1.1.3.10 Avoidance of R-loops: A physiological function for Rho

In the event of uncoupling of transcription and translation, the untranslated nascent transcript is released from the transcription TEC by the action of Rho along with NusG (Adhya and Gottesman 1978;
Nudler and Gottesman 2002; Richardson 2002; Gowrishankar and Harinarayan, 2004).

The prevailing understanding has been that transcription-translation coupling is an adaptation by
the cell to prevent accumulation of non-functional transcripts in the cytoplasm (Richardson, 1991, 2002).
This notion was analogous to the phenomenon of nonsense mediated decay observed in eukaryotic cells
(Hilleren and Parker, 1999, Wilusz et al., 2001, Maquat, 2004). It was demonstrated that the stability of a
bacterial mRNA species is influenced by the efficiency of coupling of transcription and translation (Iost
and Dreyfus 1994, 1995; reviewed in Gowrishankar and Harinarayan, 2004).

However, recent findings in the current laboratory appear to support an alternative, but not
mutually exclusive, model that the purpose of coupling of translation to transcription is to prevent the
occurrence of otherwise lethal R-loops on the bacterial chromosome (Harinarayanan and Gowrishankar,
2003; reviewed in Gowrishankar and Harinarayan, 2004). In this manner, translation may be
envisioned to play the same role in prokaryotes as do other co-transcriptional processes in eukaryotic
cells, such as mRNA processing and export.

An R-loop is defined as a structure in which the RNA is heteroduplexed with one strand of the
double-stranded DNA, displacing the other DNA strand as a loop. R-loops were first reported in ColE1
family of plasmids to generate primers for replication initiation (Itoh and Tomizawa 1980).

In the event of uncoupling of transcription and translation, the RUT sites on the RNA get
exposed to which Rho binds and brings about termination of transcription. However, when there is
failure of active termination of transcription, an untranslated mRNA gives rise to R-loops. This condition
is unfavourable for cell viability and may be lethal when it occurs on a large scale on the chromosome.

Two well characterized examples of R-loop formation in bacterial cells are:

i. The physiological R-loop formed by plasmid encoded RNA-II in the process of replication of the
ColE1-like plasmids (Itoh and Tomizawa 1980, Selzer and Tomizawa 1982), and

ii. Transcription-associated pathological R-loop formation in topA mutants of E. coli, whose DNA
is hypernegatively supercoiled due to topoisomerase I deficiency (Drolet et al., 1995, Masse et
al., 1997, Masse and Drolet 1999, Broccoli et al., 2004; reviewed in Gowrishankar and
Harinarayanan, 2004).

Both R-loop formation and Rho-dependent transcription termination are kinetic phenomena that
are associated with the nascent untranslated transcript. This raises the possibility that Rho-mediated
termination of transcription has evolved to prevent the occurrence of R-loops on the chromosome.
Indeed, genetic evidence has been obtained which suggested that mutants deficient in either Rho or NusG
suffer increased R-loops on the chromosome (Harinarayanan and Gowrishankar, 2003).

The two known enzymes involved in resolution of R-loops in E. coli are RNase H1 and RecG,
which cause hydrolysis and unwinding of R-loops respectively. A combined deficiency of both enzymes
was found to be inviable by Kogoma (1997a). This suggests that R-loops occur in wild-type cells and can
be lethal if not removed. rho showed synthetic lethality with recG and nusG was found to be lethal with
rnhA. Transformation of the rho and nusG mutants with ColE1-family of plasmids (such as pACYC184
Chapter 1

and pUC19) resulted in inviability of the transformants owing to uncontrolled replication of these plasmids. Some of the phenotypes associated with these mutants were rescued by overexpression of RecG or RNase H1 (Harinarayanan and Gowrishankar, 2003). These results therefore seem to suggest that all bacterial transcription is R-loop prone and active mechanisms may exist to prevent the formation of R-loops.

It has been suggested that pathological R-loops may be avoided in any one of the following ways (Gowrishankar and Harinarayan, 2004):

i. Formation of RNA secondary structure, as in the case of stable RNAs (rRNA and tRNA),
ii. Coupling of translation to transcription and
iii. Rho and NusG mediated termination of untranslated mRNA, downstream of a translation stop codon or in the event of stochastic uncoupling of translation from transcription.

![Figure 1.2: Schematic depiction of R-loop formation by re-annealing of the nascent unstructured transcript to the template DNA strand upstream of the transcription TEC (A), and of its avoidance by either RNA secondary structure formation (B), coupling of translation with transcription (C), or Rho- and NusG-mediated termination of transcripts with premature stop codons (D). RNAP, RNA polymerase (figure adapted from Gowrishankar and Harinarayan, 2004).](image)

The uncontrolled replication of the CoIE1-plasmids in the rho and nusG mutants is explained as a consequence of titration of host factors such as RNase H1 or RecG by the chromosomal R-loops that may otherwise act to destabilize the R-loops at the plasmid replication origin (Harinarayanan and Gowrishankar, 2003).

Interestingly, increased plasmid replication is also observed in two other instances of transcription-translation uncoupling:

i. The classical method for amplification of plasmid content in cultures by the addition of chloramphenicol, which is an inhibitor of translation (Clewell and Helinski 1969, Clewell, 1972), and
Chapter 1

ii. Increased content of ColE1-like plasmids in relA mutant strains (ppGpp deficient) when subjected to amino acid limitation (Hecker et al., 1988; Riethdorf et al., 1989; reviewed in Gowrishankar and Harinarayanan, 2004).

Both perturbations are expected to lead to a global inhibition of translation without any effect on transcription efficiency. It is possible therefore that increase in the occurrence of chromosomal R-loops is the basis for increase in the ColE1 plasmid copy number in these instances (reviewed in Gowrishankar and Harinarayanan, 2004).

Both rho and nusG mutations have been extensively used in this study. These two alleles have also been employed in earlier studies in the laboratory (Harinarayanan and Gowrishankar, 2003, Saxena and Gowrishankar, 2011a, 2011b).

1.2. Part B: The connection between transcription and genomic integrity

Transcription is a crucial feature of DNA metabolism. DNA is also the template for replication repair and recombination apart from transcription. Thus, quite frequently, a given segment of DNA can be engaged in more than one of these processes. Therefore, there seems to be a physical and functional connection between these important cellular processes. This connection has emerged as a ubiquitous feature in all organisms from prokaryotes to eukaryotes and has an important impact on genome integrity. Recent research has shed considerable light on the impact of transcription on the preservation of genomic integrity and in the induction of genetic instability and diversity.

1.2.1 Consequences of replication fork and transcription complex collision

The rate of replication fork progression (~630 nt/s) is 12-30 times greater than the rate of transcription (~20-50 nt/s) in bacteria. Often, these two processes occur simultaneously on the same template. Moreover, specifically during chromosomal duplication, the pressure for heavily active transcription is assumed to be very high. Thus, one very obvious consequence of these two processes occurring on the same DNA template is the collision between the replication and transcription complexes. Transcription complexes pose as major impediments to replication fork progression due to their high stability and abundance on the DNA (Guy et al., 2009). Current evidence indicates that collisions between replication and transcription compromise genome integrity in prokaryotes and eukaryotes (reviewed in Pomerantz and O’Donnell 2010a and 2010b).

The high accuracy of genome duplication and efficiency of transcription therefore seem to suggest that mechanisms exist to prevent the collapse of the replication forks in the event of replisome-RNAP collisions, and maintain genomic integrity. Studies in bacteria and eukaryotes demonstrate that replisome-RNAP collisions cause DNA damage response, mutagenesis and chromosomal rearrangements. This indicates replications forks frequently collapse upon encounter with transcription elongation complexes (reviewed in Pomerantz and O’Donnell, 2010b).
1.2.1.1 Studies on replisome-RNAP collisions

Actively transcribing RNAP molecules could pose a problem for replisomes. In many genes RNAP pauses at regulatory sequences and at sites of DNA damage. Some paused or stalled RNAP complexes translocate backwards along the template. These backtracked complexes pose as barriers to approaching replisomes and succeeding RNAP complexes (reviewed in detail in McGlynn et al., 2012; Nudler, 2012).

Several studies of replisome-RNAP collisions have been performed in E. coli and Bacillus subtilis cells and it was concluded that replication is slower through head-on transcription units, which oppose the fork, as compared to co-directional transcription units, which face the same direction as the fork. Also there seems to be a correlation between the severity of replication fork arrest in head-on collisions and the level of gene expression (reviewed in Pomerantz and O’Donnell, 2010a).

Interestingly it has been observed that the orientation of highly transcribed genes, such as the rRNA genes, and highly expressed protein coding genes is codirectional with respect to replication fork movement (Guy and Roten, 2004). These observations suggest that the organization of chromosomes was selected in the course of evolution to minimize head-on collisions between the replication and transcription units. Indeed, Boubakri et al., (2010) found that rRNA operons when artificially oriented in a direction opposite to that of replication fork movement caused genomic instability and cell death.

Several observations indicate that replication forks pause after colliding with a single head-on RNAP, but collapse after colliding with a head-on RNAP array, a phenomenon thought to occur only in highly expressed genes. However, since highly transcribed regions are exclusively co-directional with replication in wild-type cells, replication forks probably remain mostly intact as they traverse the genome. (Maisnier-Patin et al., 2001; reviewed in Pomerantz and O’Donnell, 2010a).

Genomic integrity may also be compromised due to head-on collisions within transcription units that are not highly expressed (Vilette et al., 1996). Consequently, longer genes were also selected to be co-directional with replication which minimizes mutations that may occur due to head-on collisions (Omont and Kepes, 2004). Head-on collisions of replication forks and transcription complexes may lead to DNA recombination. Given the fact that replication is repeatedly initiated from the origin in bacteria, prolonged stalling of the replisome may result in double-strand breaks due to replication run-off of an upstream fork. Also, a stalled replication fork may regress which can also lead to double-strand breaks due to endonucleolytic cleavage of the DNA by the Holliday junction resolvase RuvABC (see figure 1.3).

Mutations at the site of fork collapse may be due to erroneous recombinational repair or from error-prone DNA polymerase activity since translesion DNA polymerases have been postulated to switch with the replicative DNA polymerase after replication fork arrest (Indiani et al., 2005). On the other hand, error-prone DNA polymerases may promote mutagenesis during recombinational repair of the fork since translesion DNA polymerases may act at recombination intermediates (McIlvraith et al., 2005).
The characteristics of transcription complexes seem to vary according to their phase of transcription and hence may have different effects on replication depending on their location along the gene. Transcription initiation complexes (especially closed complexes) are highly unstable. Hence they are not likely to have a significant effect on replication. As RNAP molecules leave the promoter to enter the elongation phase, the complexes become highly stable and processive and are thus likely to have the maximum impact on replication. RNAP undergoes a structural transition during termination to cause displacement of the RNAP molecule and transcript from the DNA template. A modest number of studies seem to indicate that replisome pausing can occur at the terminator, leading to fork arrest. Further investigation of this phase of transcription is needed to propose a final conclusion. While it is well accepted that replication is arrested within head-on transcription units, the fate of the replication fork following collisions with head-on transcription complexes is yet to be elucidated (reviewed in Pomerantz and O’Donnell, 2010a).

Figure 1.3: Different effects of transcription complexes on the progression of replication forks. (A) Replication slows upon collision with head-on transcription complexes (left), but is unaffected by co-directional transcription complexes (right). (B) Replication forks are arrested and collapse upon encountering a head-on RNAP array (left). Replication forks pause, but remain intact upon encountering a single head-on RNAP (right). (C) Replication fork
pausing due to a head-on collision may lead to double-strands breaks as a result of replication run-off of upstream forks. (D) Replication forks may regress following a head-on collision which leads to endonucleolytic cleavage of the DNA by RuvABC (figure adapted from Pomerantz and O’Donnell, 2010a).

1.2.2 Cellular mechanisms to resolve collisions

Evolution has favoured the selection of several mechanisms to resolve collisions between replisomes and transcription complexes. Such mechanisms include the action of auxiliary helicases and of various transcription elongation factors as discussed below.

Auxiliary helicases like Rep, UvrD and DinG in E. coli promote replication past protein roadblocks. Studies involving deletion of certain auxiliary helicases reported replication fork arrest, double-strand breaks and also cell death (Guy et al., 2009; Boubakri et al., 2010). Thus, DNA helicases appear to play an important role in preventing replication fork arrest due to encounters with protein roadblocks such as transcription. Furthermore, transcription factors and modulators such as Mfd, DksA, ppGpp and GreA/B, which play a role in elongation or dissociation of RNAP have also been demonstrated to play a significant role in reducing replication-transcription conflicts (McGlynn and Lloyd, 2000; Trautinger et al., 2005; Tehranchi et al., 2010).

These observations suggest that proper regulation of transcription elongation is necessary to reduce backed-up arrays of transcription complexes that may block the replication fork. Thus, transcription factors appear to contribute to genomic integrity by preventing conflict between replication and transcription (reviewed in Pomerantz and O’Donnell, 2010b; reviewed in McGlynn et al., 2012).

1.2.2.1 Role of Auxiliary helicases in resolving collisions

The function of auxiliary helicases in facilitating replication through protein barriers seems to be well conserved. However, the mechanism by which the DNA helicases facilitate replication fork movement through transcription complexes is not yet understood satisfactorily.

Rep and UvrD helicases are thought to facilitate replication through protein blocks such as transcription complexes (Heller and Marians, 2005). Rep and UvrD are SF1 superfamily helicases that translocate along the DNA in the 3’ – 5’ direction, and act on the leading strand to remove protein roadblocks. rep mutants replicate DNA at a slower rate and exhibit a greater number of replication forks along the chromosome, which may be a consequence of frequent fork stalling. uvrD mutants show reduced replication and the cells are filamented. rep uvrD double mutants are inviable, suggesting that these two proteins share a common function (reviewed in Pomerantz and O’Donnell, 2010a).

Rep and UvrD are involved in several replication repair processes such as replication restart and recombinational repair (Heller and Marians, 2005; Lestini and Michel, 2007). Recent observations have indicated that they promote replication through transcription complexes and repressors in E. coli (Boubakri et al., 2010; Guy et al., 2009). Rep was demonstrated to bind to the replicative DnaB helicase, and both helicases were shown to cooperate in unwinding the DNA duplex. This suggested that Rep may be a component of the replication fork and may act ahead of the DNA polymerase on the leading strand.
Chapter 1

DinG is an SF2 superfamily helicase and translocates along the ssDNA in the 5'–3' direction. It was proposed to act on the lagging strand and cooperate with either UvrD or Rep on the leading strand. DinG was also proposed to reduce conflicts between replication and head-on transcription complexes by removing R-loops that occur as a result of negative supercoils behind the moving RNAP (Boubakri et al., 2010).

UvrD, Rep and DinG were demonstrated to be required for cell growth when ribosomal genes were inverted. Replication fork accumulation was observed in the inverted ribosomal operons in the absence of one or more of these proteins (Boubakri et al., 2010).

1.2.2.2 Role of RNAP modulators in resolving collisions

Since RNAP is very abundant in the cell, transcription complexes are highly likely to become arrested at different types of DNA lesions apart from replication forks. The arrest of a single RNAP owing to DNA damage has been observed to lead to accumulation of an array of RNAPs. This may block the progress of a replication fork, especially when it is opposing the direction of the fork. Genetic analyses have revealed that RNAP modulators such as Mfd, ppGpp, DksA and GreA/B increased cell viability in strains deficient for DNA repair after UV exposure. Mutations in RNAP that decrease the stability of a TEC were observed to have similar effects of suppressing growth defects of these mutants (reviewed in Rudolph et al., 2007). These data seem to suggest that RNAP modulators help reduce conflicts between replication and transcription by dislodging stalled transcription complexes. Failure to do so may lead to RNAP arrays which block replication (reviewed in Pomerantz and O’Donnell, 2010a).

a. Control of bacterial metabolism by ppGpp

Guanosine tetraphosphate, ppGpp and guanosine pentaphosphate, pppGpp (collectively referred to below as ppGpp) were first described by Michael Cashel about 40 years ago. These nucleotides were found to accumulate rapidly in E. coli cells that are starved for amino acids, and cause inhibition of rRNA and tRNA synthesis (Cashel et al., 1996). This ppGpp induced response to starvation has been termed as the “stringent response” and the molecule itself is called “stringent factor”.

Two enzymes, RelA and SpoT have been recognized in E. coli that synthesize and degrade and maintain ppGpp levels in the cell through their synthesis and degradation activities (reviewed in Srivatsan and Wang 2008):

1. RelA is a ppGpp synthase that associates with the ribosomes in E. coli. During amino acid starvation, when the uncharged tRNAs bind to the ribosomal ‘A’ site due to stalled protein synthesis, an idling reaction is enabled in which pppGpp/ppGpp are synthesized by RelA from GTP/GDP, utilizing ATP in the process (Cashel et al., 1996).

2. A second protein, SpoT can synthesize ppGpp in stress conditions other than amino acid starvation, such as deprivation of phosphorus, iron, carbon source or fatty acids (Cashel et al., 1996). SpoT can both synthesize ppGpp and hydrolyze it to GTP/GDP and pyrophosphate.
Deletion of both relA and spoT leads to complete absence of cellular ppGpp (ppGpp\(^0\)) which renders E. coli multiauxotrophic and the strain displays a distinct set of amino acid requirements, which has been explained on the ground that ppGpp is required to activate transcription of the corresponding biosynthetic genes. The two major functions of ppGpp are activation of stable RNA (rRNA) synthesis and inhibition of amino acid operon activation.

It has been recently shown that physical interaction of the acyl carrier protein ACP, an essential cofactor in fatty acid metabolism, is required for the accumulation of ppGpp in response to fatty acid starvation by shifting the balance between the synthetic and hydrolytic activities of SpoT (Battesti and Bouveret, 2006). It was also speculated that this may also regulate ppGpp levels during carbon starvation, since carbon deprivation would affect glycolysis, leading to fatty acid starvation (reviewed in Srivatsan and Wang 2008). The unusually mobile structure of ACP enables it to interact with multiple partners, and hence ACP might convey diverse inputs to SpoT. Thereby, SpoT may monitor the physiological state of E. coli cells and tunes the levels of ppGpp accordingly.

i. **Control of genomic integrity by ppGpp**

Accurate genome duplication is a fundamental requirement for genomic integrity. This underscores the need for regulation of DNA replication in the event of nutritional or other stress conditions. It is therefore conceivable that the stringent response factor also affects DNA replication. In E. coli, replication initiation is inhibited by ppGpp in the event of the cell undergoing any form of stress. Replication is rapidly restored when favourable conditions are restored (Schreiber et al., 1995; Rudner et al., 1999).

In E. coli, ppGpp has been postulated to help ensure genomic integrity by resolving conflicts between replication and transcription. Backed-up arrays of RNAP due to stalling at DNA lesions, pose formidable blocks to replication. Evidence has been obtained for the model that ppGpp acts to destabilize TECs, thereby decreasing the likelihood that they impede replication fork progression (McGlynn and Lloyd, 2000; Trautinger et al., 2005).

Although further investigation is awaited, there seems to be compelling evidence to propose that by bringing replication, transcription, translation and DNA repair pathways together, ppGpp modulates the genomic stability and evolvability of microbes in response to stress (Galhardo et al., 2007; reviewed in Srivatsan and Wang, 2008).

b. **DksA protein: Role in preventing replication-transcription conflicts**

DksA is a 151 amino acid polypeptide which is absolutely required for regulation of rRNA expression (Paul et al., 2004). The structure of the DksA protein has been determined at 2A\(^o\) resolution. DksA is an RNAP-binding transcription factor that bears structural similarities to the Gre proteins and contains a zinc finger binding motif. It has been proposed that a coiled coil domain in DksA extends deeply into the secondary channel of RNAP (Perederina et al., 2004).

DksA acts in concert with ppGpp in modulating transcription initiation at certain promoters, such
as those of the rRNA operons (Mirkin et al., 2006). Rapid rise in the cellular levels of ppGpp causes inhibition of transcription initiation at ribosomal promoters in conjunction with DksA. Strains lacking DksA, like ppGpp\(^0\) strains, are multiauxotrophic but have different amino acid requirements that are a subset of those required by the latter (Brown et al., 2002). Perederina et al., (2004) have shown that DksA acts as a transcription elongation factor independently of ppGpp and in this capacity, can reduce RNAP pausing. In this connection, Tehranchi et al., (2010) have obtained evidence for a role for DksA in preventing conflicts between replication and transcription due to nutritional starvation.

Nutrient starvation causes a higher distribution of RNAP to genes other than ribosomal operons (Jin and Cabrera, 2006). Since a considerable fraction (~45%) of E. coli transcription units is oriented in a direction opposite to the direction of replication, this would raise the energetic barrier to fork progression along most of the chromosome. Although the detailed mechanism by which transcription factors reduce replisome-RNAP conflicts remains to be elucidated, the role of these factors appears to be important in preventing prolonged RNAP pausing (reviewed in Pomerantz and O’Donnell, 2010b).

c. Role of GreA/B in RNAP modulation

RNAP molecules pause during transcription at certain sequences. A functional role for pausing in the synthesis of an mRNA is to coordinate transcription with translation (Landick and Yanofsky, 1987). Recent studies argue that coupling of transcription and translation plays a key role in maintaining genome stability by preventing RNAP backtracking (Proshkin et al., 2010). One functional characteristic of a true pause site is that the paused RNAP molecule resumes elongation. However there are points on DNA where changes occur in the RNAP molecules that prevent them from resuming elongation (Arndt and Chamberlin, 1990; Krummel and Chamberlin, 1992). At these points the enzyme forms a dead-end complex. These sites differ from termination sites by not allowing the release of the RNA molecule.

The formation of such dead-end, arrested complexes has been observed during transcription of DNA in vitro with RNAP in the absence of the proteins GreA and GreB (Borukhov et al., 1993). One naturally occurring cause of formation of arrested complexes is the misincorporation of a noncomplementary residue during transcription (Erie et al., 1993). Gre proteins appear to prevent the damage caused by misincorporation, either by removing the misincorporated 3′-end residues before formation of the arrested state or by reactivating complexes that have become arrested because of misincorporation.

Cells deficient in GreA/B, exhibit a higher frequency of transcriptional pausing. This can lead to RNAP backtracking and further backed-up array of halted RNAPs with an extruded 3′ end of the RNA through the secondary channel. These can cause replication fork arrest and collapse (reviewed in Pomerantz and O’Donnell, 2010b). Gre factors suppress double strand breaks caused by replication fork arrest and collapse by restarting backtracked TECs (Borukhov et al., 2005).

d. Mutations in RNAP that prevent replication-transcription conflicts

A class of stringent mutations in RNAP, also known as rpoB* mutations, reduce replication-
transcription conflicts (McGlynn and Lloyd, 2000) and mimic ppGpp in their mode of action. Mahdi et al. (2006) proposed that the rpoB* mutations reduce replication-transcription conflicts owing to the reduced intrinsic stability of the mutant RNAP that can dislodge easily compared to wild-type RNAP which is more stable. Thus the rpoB* mutations restore viability by removing the roadblocks posed by otherwise stable RNAP complexes. On the other hand, Dutta et al., (2011) argue that backtracking of RNAP is the primary cause of replication-transcription conflicts and the rpoB* mutants help in resolution of these conflicts by reducing backtracking.

i. The omega subunit of RNA polymerase (RpoZ)

Omega (ω) is the smallest subunit of bacterial RNAP and also the least studied so far. Although ω was identified by Burgess (1969) at the same time as identification of the other subunits, the role of this protein in the enzyme complex remained unknown for almost two decades afterwards. Constitution of bacterial core RNAP was considered to be α2ββ′ till quite recently and ω was considered to be a tightly binding impurity of purified RNAP complex (Heil and Zillig, 1970; Berg et al., 1971; Burgess, 1971; Nusslein and Heyden, 1972; reviewed in Mathew and Chatterji, 2006).

Two studies conclusively established that ω is a part of RNAP complex (Gentry et al., 1993; Dove and Hochschild, 1998). Mukherjee and Chatterji (1997) provided the first indication that ω has a possible function in the RNAP complex. They demonstrated that RNAP purified from an rpoZ-null strain of E. coli had reduced activity compared to the wild-type enzyme. Further studies by Minakhin et al., (2001) revealed the finer details of the structure of ω and its structure-function relationship (reviewed in Mathew and Chatterji, 2006).

Although the ω subunit is a well conserved protein in the course of evolution, deletion of ω does not appear to affect the survival of the organism. The ω protein is encoded by the rpoZ gene which is upstream of spoT in the same operon. The proximity between rpoZ and spoT in the bacterial chromosome raised the speculation that ω might have a role in modulating the behaviour of RNAP in the stringent response. Igarashi et al., (1989) investigated the effect of ω on ppGpp regulated transcription in vitro and concluded that ω played a part in regulating the sensitivity of the enzyme to ppGpp. However, this was disputed by Gentry et al. (1991) who showed that a ΔrpoZ strain was completely proficient for the stringent response.

Recently, Vrentas et al., (2005) showed that RNAP purified in vivo by co-overproduction of all the core subunits except ω was insensitive to the inhibitory effect of ppGpp on transcription. They also showed that native RNAP purified from an rpoZ mutant strain of E. coli, that lacks ω behaved similarly. Supplementation of the RNAP complex with ω, either by co-overexpression in vivo or by in vitro addition, restored the response of the enzyme to ppGpp. This experiment clearly established a role for ω in the stringent response.

In the absence of ω, DksA appeared to mediate the effects of ppGpp on RNAP in vivo (Paul et al., 2004). This explains the absence of a phenotypic manifestation of deletion of rpoZ in E. coli.
Strains that lack both DksA and ω appear to have amino acid requirements different from strains lacking either DksA or ω alone (reviewed in Mathew and Chatterji, 2006).

The bonafide association of ω with RNAP and its role in stringent response makes it likely that ω might play a role in preservation of genomic integrity by resolving replication-transcription conflicts.

**1.2.2.3 The role of transcription-repair coupling factor Mfd in resolving conflicts between replication and transcription**

The concept of transcription coupled repair (TCR) was first proposed by Philip Hanawalt (1994). He postulated that the TCR pathway promotes elongation or dissociation of a halted RNAP and might therefore promote replication by removing frozen RNAP roadblocks from DNA. In bacteria, TCR specifically displaces a halted RNAP from DNA and recruits the nucleotide excision repair pathway to the site, thereby causing preferential repair of the transcribed strand.

In *E. coli*, TCR is performed by a single polypeptide protein called Mfd (reviewed in Roberts and Park, 2004). Mfd is an ATP-dependent DNA translocase which binds to the rear of a halted TEC and rewinds the transcription bubble behind the RNAP (Park and Roberts, 2006). Depending on whether the forward translocation of the molecule is facilitated or inhibited, Mfd binding results in the displacement or movement of the TEC, respectively (Park et al., 2002; reviewed in Pomerantz and O’Donnell, 2010b).

Pomerantz et al., (2010c), showed that the addition of Mfd in vitro following replisome-RNAP collision resulted in direct restart of the stalled replication fork, which was due to the ability of Mfd to facilitate displacement of the RNAP ahead of the fork. However, despite the important role of Mfd in the cellular milieu, mfd mutants appear to grow normally. This indicates that additional redundant factors exist that can also resolve or prevent conflicts between replication and transcription.

**1.2.3 Transcription termination and genomic integrity**

Efficient termination of transcription is an important cellular process in the maintenance of genomic integrity, to explain which two alternative modes have been proposed that are not mutually exclusive.

In cases where RNAP is stalled and forms backed-up arrays, either due to encounter with a lesion on the template DNA, or due to unavailability of the next nucleotide to be added, etc., it has been suggested that Rho along with its cofactor NusG terminates the stalled TEC and releases the RNAP from the DNA. This is critical for smooth replication fork progression and prevention of replisome-RNAP conflicts (reviewed in McGlynn et al., 2012).

Similarly, in the event of uncoupling of translation from transcription, termination factors
Rho and NusG have been found to act in concert to bring about termination of transcription. One catastrophic consequence of failure of termination in such an instance is continued transcription in the absence of translation which is proposed to lead to formation of R-loops on the chromosome. Thus, Rho mediated transcription termination is employed to prevent co-transcriptional R-loop formation (Harinarayanan and Gowrishankar, 2003).

R-loops form impediments to succeeding TECs thereby causing stalling and backed-up RNAP array formation. These may interfere with replication fork progression and lead to fork breakage or collapse (Drolet et al., 1995; 2003; Huertas and Aguilera, 2003; reviewed in Gowrishankar and Harinarayanan, 2004).

This possibility may explain the synthetic lethality of certain rho mutants with additional mutations in ssb or rep (Fassler et al., 1985). R-loops may also mediate the aberrant initiation of chromosomal DNA replication in bacteria, known as constitutive stable DNA replication (cSDR) (Kogoma 1997a).

1.2.3.1 Implication of R-loops in genomic integrity: A special case

As explained earlier, an R-loop occurs when an RNA molecule is partially or completely hybridized with its complementary DNA strand, leaving the other strand unpaired. Several studies have linked growth defects and cellular inviability to the phenomenon of R-loop formation in various genetic backgrounds (reviewed in Li and Manley 2006). Early evidence for R-loop formation during transcription came from studies in topA mutants of E. coli. R-loop formation during transcription in topA mutants was found to be a major problem linked to growth defects. Overexpression of RNase H stimulated the growth of these mutants, suggesting that R-loop formation is involved in the growth inhibition (Drlica, 1992; Drolet et al., 1994; Drolet et al., 1995; Masse and Drolet, 1999; Masse et al., 1997). Huertas and Aguilera (2003) through studies in yeast, were the first to provide evidence that R-loops formed from nascent mRNA is responsible for transcription elongation defect.

An important possible mechanism for R-loop mediated DNA damage is in the context of DNA replication. TECs halted or slowed down by R-loops potentially block the progress of replication forks and lead to DNA recombination or DNA DSBs in the newly replicated DNA (reviewed in Aguilera, 2002). In fact transcription-dependent replication fork collisions have been shown to induce recombination in bacteria and yeast (French, 1992; Krasilnikova et al., 1998; Takeuchi et al., 2003). Resolution of these recombination intermediates is crucial for viability.

It is noteworthy that R-loops have been observed to form preferentially when the non-template strand is G rich (Phoenix et al., 1997; Huertas and Aguilera 2003; Yu et al., 2003; Li and Manley 2005). Indeed Duquette et al., (2004) and Mizuta et al., (2005) have shown that guanine is indispensable for the formation of a stable RNA-DNA hybrid in the immunoglobulin S region in in
vitro transcription studies. This is consistent with the fact that the exceptional stability of the rG/dC base pairs (Sugimoto et al., 1995) facilitates R-loop formation by favouring RNA-DNA over DNA-DNA hybrids in regions where the non-template strand is G-rich, such as the S regions of immunoglobulin loci (Gritzmacher 1989, reviewed in Li and Manley 2006).

Additionally, the single-stranded G-rich regions are prone to form stable parallel four-stranded DNA structures known as G-quartets or G4 DNAs (reviewed in Gilbert and Feigon 1999). This may further help stabilize the single-strand nature of the non-template strand, thus facilitating RNA-DNA hybrid formation during transcription. Such RNA-DNA hybrids spanning G-quartets are known as G-loops and possess extraordinary stability. Negative supercoiling and high G-density have been proposed to facilitate opening of the bubble between two DNA strands (reviewed in Aguilera and Garcia-Muse, 2012). Indeed, G-quartets have been observed during in-vitro transcription of S regions by electron microscopy and also on S-region containing plasmids that were propagated in E. coli (Duquette et al., 2004). Several nucleases have been identified in yeast and mammalian systems that can cleave the single-stranded region 5' to the G-quartets with high specificity (Liu et al., 1993; Liu and Gilbert 1994; Sun et al., 2001). These observations suggest that structures such as G-quartets help drive the formation of co-transcriptional R-loop formation and also generate susceptible sites for specific nucleases. However further analysis is needed to conclusively establish the role of G-quartets in genomic instability (reviewed in Li and Manley 2006; Aguilera and Garcia-Muse, 2012).

The notion that co-transcriptional R-loop formation interferes with replication fork progression to induce DNA damage and compromise genomic integrity is an interesting possibility that needs further investigation. Thus, R-loop formation is a serious threat to genomic integrity and different species have employed various co-transcriptional processes to prevent the formation of R-loops. A very important adaptation in bacteria is coupling of translation to transcription such that translation commences even as the nascent protein coding transcript is being synthesized. In the event of failure of such a coupling, immediate termination of transcription by Rho and NusG prevent formation of untranslated mRNA, thereby precluding the occurrence of R-loops. Stable RNAs (rRNA and tRNA) do not get translated but adopt secondary conformations immediately even as they are transcribed that prevents their annealing to the template DNA, thereby preventing formation of R-loops.

1.2.4 Consequences of replication fork arrest following replication-transcription conflict

Encounter of replication forks with a roadblock such as a stalled RNAP complex results in single-strand gaps. A block encountered in the leading strand template leads to co-directional conflict between replisome and RNAP. Although the potential for such conflicts is well
understood (Rudolph et al., 2007; Mirkin and Mirkin, 2007; Birch and Zomerdijk, 2008), they had not been detected for considerable time in vivo (French, 1992). The exception to this observation is the co-directionally placed transcription terminators that can inhibit replication fork progression (Zeigler et al., 2008). Also, co-directional encounters designed in vitro caused the replisome complex to displace RNAP translocating along the leading strand, a process facilitated by the replicative helicase translocating along the lagging strand of the replication complex (Mirkin and Mirkin, 2007; Klumpp and Hwa, 2008). Thus codirectional encounters were long thought to have little or no effect on replication (reviewed in Soultanas, 2011). However, recently Dutta et al., (2011) have demonstrated that codirectional collisions with backtracked (arrested) RNAP complexes lead to DNA double-strand breaks (DSBs). Merrikh et al. (2011) observed that highly transcribed rRNA genes are hotspots for co-directional collisions in rapidly growing B. subtilis cells.

Co-directional replisome-RNAP encounters can result in single-strand gaps if the lagging strand becomes uncoupled and continues to extend and the leading strand is resumed beyond the point of collision by de novo replisome assembly. During the second round of replication the strand with the gap acts as a lagging strand template. Discontinuous synthesis of DNA occurs on this template and when it approaches the site of the gap DNA synthesis stops due to unavailability of a continuous template further downstream. This leads to a double-strand break with a 3’ ssDNA tail, the length of which depends on where the last lagging strand Okazaki fragment initiated synthesis.

Head-on collisions between the replication and transcription machineries occur when the transcription complex is on the lagging strand. These can lead to replication fork arrest and genomic instability. Therefore to avoid head-on collisions, most genes, especially the essential and highly transcribed genes, are encoded on the leading strand such that transcription and replication are co-directional. Biochemical analyses indicate that head-on collisions are more deleterious than co-directional collisions (Klumpp and Hwa, 2008; Azvolinsky et al., 2009).

Encounter of a stalled RNAP complex on the lagging strand by an approaching replication fork can cause the leading strand synthesis to continue past the point of lesion. The Okazaki fragment that encounters the lesion gets stalled, but the lagging strand synthesis resumes beyond the point of collision when new Okazaki fragments are synthesized. A consequence of this is also a single strand gap at the point of encounter. However, in this case the gapped strand acts as a leading strand template in the succeeding round of replication. When the replisome reaches the gap on the leading strand template, run off replication occurs resulting in a blunt double strand break at the site of collision.

Another possible consequence of the inability of a replication fork to proceed because of a block is that it migrates backward so as to reanneal the original template strands while extruding the newly synthesized DNA strands as a short duplex branch. This phenomenon is referred to as
“Replication fork regression”. This creates a special type of Holliday junction which is commonly referred to as the “Chickenfoot structure” (Postow et al., 2001). Such chickenfoot structures have been observed by electron microscopy in various studies (Higgins et al., 1976; Robu et al., 2004; Viguera et al., 2000). Superhelical stress caused by replication fork movement may cause even normal replication forks to regress spontaneously, resulting in chickenfoot structure formation (Olavarrieta et al., 2002; Postow et al., 2001).

A double strand break and chickenfoot structure can serve as intermediates in homologous recombination. This highlights the importance of homologous recombination and recombination proteins in maintenance of genomic integrity.

1.2.5 Role of recombination proteins in rescue of stalled replication forks

Recombination proteins play a key role in the rescue of stalled replication forks in E. coli. In cases of replication fork arrest, often the repair or remodeling of inactivated replication forks is first needed which is followed by restart of replication at the site of arrest. This repair activity requires recombination proteins.

Numerous obstacles are encountered by replication forks during their progression which can cause replication forks to arrest. The in vivo structure of the arrested replication forks is still unknown, as even the length of the leading and lagging strand ends remain to be elucidated. Increasing body of work seems to acknowledge the role of homologous recombination in DNA repair and damage tolerance, and they seem to play key roles in stabilizing and restoring blocked replication forks (Cox, 1994; Kogoma, 1996; Kuzminov, 1996; Haber, 1999; Marians, 1999; Kowalczykowski, 2000; Kreuzer, 2000; Michel et al., 2001; Grompone et al., 2002; McGlynn and Lloyd, 2002; Courcelle et al., 2003; Friedberg et al., 2006).

Homologous recombination can occur by two main pathways, depending on the nature of DNA at the point of recombination (reviewed in Michel et al., 2007) (figure 1.4):

i. Recombination at gaps, (also called gap repair) and

ii. Recombination at a DNA double-strand end (also called double-strand break repair).

Both the pathways require the presence of RecA as a key player. RecA binds ssDNA and forms a RecA-ssDNA filament. The RecA loaded filament then invades a homologous dsDNA, pairs with its complementary strand, and strand exchange by RecA-facilitated branch migration occurs (reviewed in Kuzminov, 1999; Cox, 2003; Michel et al., 2007).

RecA mediated strand exchange reaction results in the formation of a Holliday junction, which is a four-arm dsDNA structure. RuvAB and RecG helicases recognize Holliday junctions and catalyze branch migration. RuvC is an endonuclease that associates with RuvAB, forming a RuvABC complex. This complex resolves the junction resulting in strand exchange (reviewed in Michel et al., 2007).
Figure 1.4: Two possible Recombination pathways: (A) Recombination at gaps. A gap is enlarged by the 5’–3’ exonuclease RecJ (helicase activity may be provided by redundant helicases or may not be needed, as no specific helicase is essential for gap repair). The ssDNA region is bound by SSB (not shown) and RecFOR allows RecA binding. RecA promotes strand exchange. The Holliday junctions are resolved by RuvABC (or by a RecG-dependent branch migration mechanism in a ruvAB mutant). (B) Recombination at a DNA double-strand end. RecBCD binds the dsDNA end and degrades it up to a χ site. At χ, RecBCD switches from its exonuclease V to its recombinase activity and RecA is loaded onto the 3’-ended single-strand. RecA promotes strand-exchange and the Holliday junction is resolved by RuvABC (or by the action of RecG in a ruvABC mutant). The D-loop is acted upon by PriA and a replisome is assembled. Blue and red lines represent the DNA strands of two homologous molecules, arrows are 3’-ends. Green indented circle RecJ, green circles RecFOR, magenta indented circle RecBCD, yellow circles RecA, purple triangle PriA (figure adapted from Michel et al., 2007).
1.2.5.1 *RecA* protein: Role in homologous recombination, DNA repair and maintenance of genomic integrity

RecA is a 352 amino-acid protein with DNA-dependent ATPase and ATP-dependent DNA-binding activities. It is the most crucial protein implicated in all homologous recombination reactions in *E. coli*. Bacterial RecA protein is unique in that it is able to use ATP hydrolysis to drive strand exchange, a property not reported in the homologs of RecA among other classes of organisms (Cox and Lehman, 1982; Kowalczykowski and Krupp, 1987; Menetski *et al*., 1990; Cox, 2003).

RecA can pair two homologous DNA molecules if one of them is single stranded or partially single stranded (Radding, 1991; West, 1992; Kowalczykowski, 2000; Kuzminov, 2001; McGrew and Knight, 2003; Cox, 2003). The active species in RecA-mediated DNA strand exchange is a nucleoprotein filament composed of DNA and RecA. It can be envisioned as a scaffold to facilitate DNA pairing and strand exchange with a naked DNA molecule. RecA can form filaments on both ssDNA and dsDNA. When coupled to ATP hydrolysis, RecA promoted DNA strand exchange reactions are unidirectional, with branch movement occurring in 5’ to 3’ direction relative to the bound ssDNA in the gap (Jain *et al*., 1994).

RecA can facilitate replication fork movement if the stalled fork contains a substantial single-strand gap. RecA can nucleate onto the single stranded DNA in the gap and promote fork regression (Robu *et al*., 2001). The outcome of RecA-facilitated fork movement depends on the strand which contains the single-strand gap. If a RecA filament loads onto a leading strand gap, movement of the nucleoprotein filament in the 5’ to 3’ direction results in fork regression. However, RecA loading onto a lagging strand gap causes movement of the nucleoprotein filament in the direction opposite to that required for fork regression (reviewed in Cox, 2001).

Thus, RecA appears to function in both facilitating fork regression resulting in genomic as well as in reversal of fork regression thereby contributing to maintenance of genomic integrity.

1.2.5.2 The RecFOR complex

The RecFOR complex promotes the formation of RecA-DNA nucleoprotein complex on SSB-coated ssDNA (Morimatsu and Kowalczykowski, 2003). This property of RecFOR is specific for gapped DNA and does appear to happen with free ssDNA. Additionally, the reaction requires a base-paired 5’ terminus at the ssDNA/dsDNA junction (reviewed in Kogoma, 1997b).

RecO interacts with both RecF and RecR to form a three-protein complex and facilitates RecA loading in a two-step process (Umezu and Kolodner, 1994; Hegde *et al*., 1996). The RecFR complex has a high affinity for ssDNA/dsDNA junction, nucleates at this junction and progressively binds to the gapped DNA, thus removing SSB bound to the ssDNA. The RecOR complex recognizes ssDNA-RecFR complex and facilitates RecA loading by displacing RecFR. Thus, binding of RecFR to ssDNA prevents RecA loading beyond the junction into dsDNA (Webb *et al*., 1997). The RecOR complex also has the effect of preventing the dissociation of RecA from linear ssDNA by likely stabilization of RecA-ssDNA nucleoprotein filament (Shan *et al*., 1997).
1.2.5.3 RecQ helicase and RecJ nuclease

The RecFOR pathway can process DNA double strand breaks to produce 3’ tails. DNA helicase activity for this to occur is provided by RecQ helicase. RecQ lacks nuclease activity, and this is provided by RecJ, which is a 5’–3’ single-strand exonuclease (Chase and Richardson, 1974; Lovett and Kolodner, 1989). RecJ can degrade the 5’ DNA strand that was displaced by RecQ to generate 3’ single stranded tail DNA. RecQ and RecJ together degrade the nascent lagging strand at UV-induced stalled DNA replication forks to promote fork stabilization and suppress inappropriate recombination (Courcelle et al., 2003). This is acted upon by the RecFOR complex to load RecA and facilitate homologous recombination functions.

The RecQ family of helicases is a conserved group of enzymes that are required for maintaining genomic integrity by functioning as suppressors of inappropriate recombination (Magner et al., 2007; Fonville et al., 2010). *E. coli* RecQ has broad DNA substrate specificity and acts on DNA duplexes containing blunt or forked termini, duplexes with 3’ or 5’ single strand tails, D-loops and three-or four-way Holliday junctions (Hermon and Kowalczykowski, 1998). RecQ functions as a monomeric protein (Xu et al., 2003) (reviewed in detail in Opresko et al., 2004).

1.2.5.4 Branch migration and resolution: The RuvABC complex, RecG and PriA helicases

Three enzymes catalyze the process of Holliday junction branch migration – RuvAB, RecG and PriA.

The *ruvAB* operon is regulated as a part of the SOS response (Shurvinton and Lloyd, 1982; Benson et al., 1988; Shinagawa et al., 1988). RuvA and RuvB proteins interact to form the RuvAB complex that binds specifically to Holliday junctions and acts as a motor to cause branch migration to occur. The direction of branch migration depends on the way RuvAB loads onto the DNA (Parsons et al., 1992; Shiba et al., 1993; Parsons and West, 1993). However, in the event of event of RuvAB binding an arrested fork to cause repair, there is a need for directed branch migration. Iype et al., (1995) have observed that in the event of RecA facilitated strand exchange, RuvAB loading occurs preferentially to reverse the reaction. The branch migration promoted by RuvAB is much faster and more energy efficient than that promoted by RecA (Tsaneva et al., 1992).

A third gene at the *ruv* locus, which is not a part of the *ruvAB* operon encodes the RuvC protein which is a 19 kDa endonuclease. RuvC can resolve Holliday junctions into recombinant molecules by specific endonucleolytic cleavage across the point of strand exchange (Connelly et al., 1991; Iwasaki et al., 1991; Bennett et al., 1993; Friedberg et al., 2006). The *ruvC* gene is not SOS regulated.

RecG is another *E. coli* protein that can cause the migration of Holliday junction, apart from its several other attributed functions. It is a monomeric 76 kDa DNA-dependent ATPase. Whitby et al. (1993) have demonstrated that it can counter RecA catalyzed branch migration by causing the strand exchange to occur in the reverse direction.
RecG consists of three domains: The N terminal domain (residues 1-350) is the largest and contains a “wedge” that binds to the single stranded region between two double stranded DNA arms in a DNA junction (Singleton et al., 2001; Dillingham and Kowalczykowski, 2001). The central (residues 351-549) and C terminal domain (residues 550-780) comprise of the superfamily II helicase-like motor, and are called as helicase domains. The helicase domains are assumed to form a dsDNA translocation motor and drag DNA across the wedge domain. The result of this action is the unwinding of the leading and lagging strands from their respective parental template strands, allowing the parental strands to reanneal, and the daughter strands to anneal with each other. This action of RecG brings about replication fork reversal, creating a chicken-foot structure at the wedge domain (reviewed in McGlynn and Lloyd, 2002).

Deficiency of either RecG or RuvAB causes modest deficiency in recombination, while the deficiency of both causes severe deficiency in recombination (Sharples and Lloyd, 1993). This observation led to the assumption that apart from RuvC, *E. coli* may also have a second resolvase that functions in conjunction with RecG (Mandal et al., 1993; West, 1994). A search for such an enzyme using suppressor studies led to the discovery of RusA, a small metal ion-dependent enzyme that can cleave Holliday junctions. However, rusA is a cryptic enzyme and is not normally expressed in *E. coli* (Bolt and Lloyd, 2002; Zhang et al., 2010). Thus so far, no endonuclease normally expressed in *E. coli* that works in concert with RecG has been identified. RecG branch migration activity helps in restoration of replication forks from migration of a chickenfoot structure thereby preventing unnecessary recombination (Mahdi et al., 2006). RecG also limits damage-inducible stable DNA replication (SDR) which has been shown to be dependent on recombination (Kogoma, 1997a). SDR is dramatically exacerbated in the absence of RecG (Hong et al., 1995; Rudolph et al., 2009; reviewed in Rudolph et al., 2010).

RecG might also help in avoidance of replication fork collapse by unwinding R-loops on the chromosome. Persistence of R-loops may allow priming of DNA replication by PriA, leading to replication initiation when it is not needed. R-loops are generally removed by digestion of the hybridized RNA via RNase H. However, mutants lacking RNase H were found to be viable. Instead deprivation of both RNase H and RecG rendered *E. coli* inviable (Hong et al., 1995; reviewed in McGlynn and Lloyd, 2002).

The PriA protein also possesses 3’- 5’ helicase activity apart from its primary function of primosome assembly and can participate in branch migration. PriA specifically recognizes branched DNA structures and catalyzes the loading of the primosome onto the lagging strand template for DnaB loading. A search for suppressors of *recG* led to the recovery of mutations within or near the helicase domain of PriA (Al-Deib et al., 1996). This implicates the helicase function of PriA in RecG-dependent processing of replication forks and suggests that PriA and RecG helicase activities might be in balance within the cell (reviewed in McGlynn and Lloyd, 2002).

PriA, along with PriB causes replication restart at D-loops (which are triple helical structures
with one strand of a DNA duplex base-paired with its complementary newly synthesized daughter DNA while the other strand is unpaired). This pathway is called as the PriA-PriB pathway. Apart from this pathway, two additional pathways exist for replication restart that require another protein called PriC. PriC can act either with PriA (PriA-PriC pathway) or the 3′–5′ DNA helicase, Rep (Rep-PriC pathway) (Sandler et al., 1999, 2001; Sandler, 2000). Xu and Marians (2003) provided supporting biochemical evidence for PriA-PriB pathway. Heller and Marians (2005) showed that PriC interacts with Rep and coordinates the action of Rep and DnaB for replication restart.

### 1.2.5.5 The RecBCD complex

RecBCD complex, also known as exonuclease V is a heterotrimeric helicase-nuclease that processes the double strand ends of DNA molecule to generate a 3′ tail suitable for homologous recombination (reviewed in Smith, 1991; Myers and Stahl, 1994; Kowalczykowski, 2000; Friedberg et al., 2006). The complex unwinds the dsDNA at a speed of approx. 1000 bp/s, for about 30 kb before dissociating. As the helicase activity of the enzyme unwinds the DNA, the nuclease component simultaneously degrades the strand that was 3′ terminal at the enzyme entry point.

RecBCD is an unusual enzyme in that it has two DNA helicase activities of opposite polarities (Dillingham et al., 2003). The RecD component of the enzyme is a fast moving helicase with 5′–3′ polarity whereas the RecB component is a slow helicase with 3′–5′ polarity. Due to this disparity in the speed of helicase activities, a single stranded loop frequently accumulates on the 3′ ended strand due to RecD activity on the opposite strand, until it is released by a next round of RecB activity (Taylor and Smith, 2003). Additionally, the RecBCD complex also possesses 3′–5′ exonuclease activity.

The 3′–5′ exonuclease activity of RecBCD predominates till the enzyme reaches a χ site on the DNA. The χ site is characterized by an octameric sequence, 5′GCTGGTGG3′ in *E. coli*. Upon encounter with χ, RecD is uncoupled from the enzyme complex, the translocase activity is slowed to nearly half the original rate, the stronger 3′–5′ nuclease activity is attenuated and a weaker 5′–3′ nuclease activity is activated (Dixon and Kowalczykowski, 1993; Anderson and Kowalczykowski, 1997; Spies et al., 2003; Friedberg et al., 2006). A 3′ ssDNA tail is generated as a result of this change in nuclease activity, with χ at its extreme end. This is a substrate for binding by RecA, thereby facilitating RecA mediated strand invasion and homologous recombination.

RecBC enzyme was constitutively shown to load RecA onto the 3′ end of DNA strand, even when the χ sequence is absent. This indicates that RecA loading is a constitutive property of the RecBCD holoenzyme that is normally blocked by the RecD subunit and is revealed upon interaction with χ (Churchill et al., 1999).

### 1.2.6 Recombinational repair of DNA double-strand breaks

A double strand break can occur on the DNA either directly by agents such as ionizing radiation,
or by a replication fork encountering a single-stranded break on the template DNA, or by a resolvase processing a chickenfoot structure created by replication fork regression.

![Diagram of homologous recombination](image)

**Figure 1.5:** General model for homologous recombination as suggested by Szostak et al. (1983) (Figure adapted from Friedberg et al, DNA repair and mutagenesis, 2nd Edition, 2006).

Recombinational repair of DNA a double strand break requires the presence of another homologous DNA duplex (Krasin and Hutchinson, 1981). Michael Resnick (1976) first proposed the mechanism of double strand break repair. Several modifications have since been proposed by others (Kobayashi and Ikeda, 1983; Szostak et al., 1983; Wang and Smith, 1986; Smith, 1987; Friedberg et al., 2006).

The general model for double strand break repair as proposed by Szostak et al., (1983) is depicted in the figure (figure 1.5). According to this model, the initial DSB is processed via exonuclease action to generate exposed 3’ ends. The 3’ ends can then invade the intact homologous duplex to prime DNA repair
synthesis. This invasion is catalyzed by RecA. This eventually leads to the formation of two Holliday junctions that are then resolved by the RuvABC complex and RecG. Depending on how the Holliday junctions are resolved, the resulting DNA molecules would be either recombinant or non-recombinant.

1.2.7 Role of recombinational repair in recovery of arrested or collapsed replication forks: A perspective

For several years it was assumed that the primary purpose of homologous recombination was to promote genetic diversity. However, in recent times a new role of this cellular process has emerged, and that is to enable the stabilization of arrested or collapsed replication forks, which have occurred as a result of endogenous or exogenous DNA damage, and to ensure their nonmutagenic recovery (Bernstein, et al., 1985; Clark and Sandler, 1994; Cox, 1994; Kuzminov, 1996; Galitski and Roth, 1997; Michel et al., 2004; Friedberg et al., 2006). As discussed in the previous sections, recombination proteins such as RecA, the RecBCD and RecFOR complexes, the RuvABC complex, RecG act in concert to process blocked replication forks by bringing about homologous recombination and providing a means to restart such forks (reviewed in detail in McGlynn and Lloyd, 2002).

When analyzed from this perspective, the role of recombination proteins in processes such as daughter strand gap repair is to promote survival by helping cells restore genomic integrity and prevent generation of anomalous DNA structures that can arise due to replication involving a damaged DNA template.

1.2.8 The link between transcription and genome stability

It is well understood that mutations in a particular gene can be induced by transcription (Brock, 1971; Herman and Dworkin, 1971; Datta and Jinks-Robertson, 1995; Beletskii and Bhagwat, 1996; Wright et al., 1999; Beletskii et al., 2000). For instance, spontaneous deamination of cytosine is 140 fold more efficient on ssDNA than on dsDNA (Frederico et al., 1990). It has also been observed in the E. coli lacI and human hprt genes that the non-transcribed ssDNA is more susceptible to mutations than the transcribed strand (Fix and Glickman, 1987; Skandalis et al., 1994). In the E. coli tac region, a 4 to 5 fold increase in C-to-T mutations was observed due to transcription (Beletskii and Bhagwat, 1996). Moreover, transcription associated mutation is higher in a mutant of T7 RNAP that has a slower elongation rate (Beletskii et al., 2000). These observations indicate that C deamination in the non-transcribed strand may be dependent on the length of time that the DNA remains open during transcript elongation (reviewed in Aguilera, 2002).

Lindahl and Nyberg (1974) had demonstrated that ssDNA is more vulnerable to mutations than dsDNA. Extensive R-loop formation may make certain transcribed regions of the genome more susceptible to mutations by exposing ssDNA to DNA damaging agents at increased frequency. This is supported by observations that transcription synergistically increases the mutagenic effects of DNA-damaging agents in bacteria (Brock 1971; Herman and Dworkin 1971; Garcia-Rubio et al., 2003).
Protein factors that specifically recognize R-loops, such as AID (Muramatsu et al., 2000; Okazaki et al., 2002; Longerich et al., 2006) may cause generation of mutagenic/recombinant DNA lesions at these structures (reviewed in Li and Manley 2006).

1.3 Part C: Objectives of the present study

A previous study reported from this laboratory had identified a novel nusG mutation that phenocopied a rho mutation in conferring global deficiency in transcription termination. Both rho and nusG mutants were killed following transformation with plasmids such as pACYC184 or pUC19, which have R-loop replication intermediates and the lethality was correlated with greatly increased content of these plasmids. Following these observations, it was proposed that an increased occurrence of chromosomal R-loops in the rho and nusG mutants leads to titration of cytoplasmic host factor(s) that negatively modulate the stability of plasmid R-loop replication intermediates and consequently to runaway plasmid replication (Harinarayanan and Gowrishankar, 2003).

The present studies were aimed to identify and understand the cytoplasmic host factor(s) involved in processing R-loops and to biochemically test for the presence of R-loops on the chromosome of rho and nusG mutants.

Following were the strategies employed to address each of the objectives:

i. A biochemical and molecular assay involving in vitro bisulphite mutagenesis was designed, standardized and employed to test for R-loops and evidence for formation of these structures in the rho and nusG mutants was obtained (Chapters 3, 4 and 5).

ii. Whole genome analysis of the bisulphite treated total nucleic acid preparations by employing whole genome sequencing approach was undertaken to gain further insight into the dynamics of R-loop formation on the chromosome (Chapter 6).

iii. A candidate gene approach was employed to search for synthetic lethal phenotypes with rho and nusG and their contribution to cell viability was studied (Chapter 7)