An integrated mechanistic model for transcription-coupled nucleotide excision repair

Shwetal Patel\textsuperscript{a,\ast,1}, K.V. Venkatesh\textsuperscript{b}, Jeremy S. Edwards\textsuperscript{a}

\textsuperscript{a} Department of Chemical Engineering, University of Delaware, Newark, DE 19716, USA
\textsuperscript{b} Department of Chemical Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai-400076, India

Abstract

Preferential repair of the transcribed strand of active genes is usually attributed to a coupling protein that dislodges RNA polymerase stalled at a damage site and recruits repair enzymes. Experimental observations of the effect of transcription on preferential repair in \textit{Escherichia coli} are contradictory and inexplicable by this model. In this study, it is argued that the multiple conformations displayed by a stalled RNA polymerase result in two sub-pathways for repair: Mfd coupled and direct. Together with the fact that RNA polymerase recruits the repair enzymes in a promoter dependent manner, an integrated mechanistic model is proposed that is capable of explaining the effect of transcription on preferential repair reported in literature. The quantitative behavior of the model is illustrated by describing the various reactions using a biochemical network. The implications of the model on the mechanism for transcription-coupled repair in higher organisms are briefly discussed.

Keywords: Cockayne syndrome; In silico biology; Hypothesis; UvrABCD

1. Introduction

Transcription-coupled repair, the rapid and preferential mending of actively transcribing genes, is an important mode of DNA excision repair [1–7] defects in which are responsible for the symptoms of Cockayne syndrome [8,9]. Preferential repair of actively transcribing genes in \textit{Escherichia coli} is brought about by Mfd protein that dislodges a stalled RNAP and recruits repair enzymes (UvrA2B) [10]. The molecular mechanism of this pathway is remarkably conserved exhibiting mechanistic similarity across the living kingdom. Experiments in \textit{E. coli} have revealed two important properties of this pathway that are contradictory and inexplicable by the current model: (i) for the \textit{lacZ} gene, Mfd is required at low levels of transcription but not at elevated levels of transcription [11]; and (ii) for the \textit{tryT} gene, induction by the transcriptional activator Fis suppresses preferential repair [12]. In this report, an integrated mechanistic model that accounts for events at the promoter as well as at a stalled RNA polymerase is proposed that is capable of explaining the above observations. Because the model involves complex interactions between several factors, the quantitative behavior of the model is illustrated by using a biochemical network.

2. Model

Factors responsible for preferential repair of the transcribed strand arise during transcription initiation. DNA supercoiling and footprint analysis have revealed that the binding of RNAP to \textit{\lambda P\textsubscript{L}} promoter provides UvrA2B complex with a preferred binding site on the non-transcribed strand resulting in the activation of 5′–3′ helicase activity [13]. Moreover, crosslinked immunoprecipitates of isolated membrane-associated nucleoids from \textit{E. coli} show that UvrA is specifically associated with the \(\beta\) subunit of RNAP [14]. Based on these results and further studies on dimer excision, Ahn and Grossman suggested that upon recruitment of UvrA2B to the transcription bubble at the promoter region, a competent nucleoprotein complex for helicase action is formed that translocates along the non-transcribed strand.
Fig. 1. The presence of a dimer on the transcribed strand causes RNAP to stall. Subsequently, RNAP alternates between multiple backtracked conformations some of which are substrates for Mfd (stalled) while others allow direct binding of UvrA₂B to dimer (free), i.e., there are two sub-pathways for repair. Upon binding to DNA and stalled RNAP, Mfd protein pushes RNAP molecule off the DNA and recruits the UvrA₂B complex to initiate nucleotide excision repair. The free conformation allows UvrA₂B to bind directly to the dimer. The number of UvrA₂B molecules available at a dimer depends on its recruitment to the transcription bubble at the promoter through physical interaction with RNAP.

3. Results

To illustrate the quantitative behavior of the model, a mathematical description of the biochemical reactions in-
Fig. 2. The biochemical reaction network for the proposed model illustrated in Fig. 1. Excision of dimers proceeds along two parallel sub-pathways. The empirical parameter, $K^*$, determines the abundance of dimers with stalled RNAP as a function of the concentration of active RNAP ($P^*$) moving along the gene (see supplementary information). Free dimers are bound directly by UvrA2B, while UvrA2B binds dimers with stalled RNAP only after its recruitment following the dislodging of stalled RNAP (dislodged RNAP denoted by $P$) by Mfd. The irreversible binding of UvrB with the release of 2UvrA follows the binding of UvrA2B. Bound UvrB creates a kink in the DNA and its C-terminus recruits UvrC resulting in an active endonuclease which cleaves a phosphodiester bond four–seven phosphates $3'$ to the damage. This event activates a second nuclease in UvrC, which incises the eighth phosphate backbone $5'$ to the damaged residue. The oligonucleotide about 12bp long, containing the dimer, is excised by the helicase UvrD, the gap filled by the action of DNA Pol I and, finally, the nick stitched by DNA ligase [20,21]. $k$ and $k_d$ denote the Michaelis–Menten and first order rate constants, respectively. 

The biochemical reaction network for the proposed model illustrated in Fig. 1. Excision of dimers proceeds along two parallel sub-pathways. The empirical parameter, $K^*$, determines the abundance of dimers with stalled RNAP as a function of the concentration of active RNAP ($P^*$) moving along the gene (see supplementary information). Free dimers are bound directly by UvrA2B, while UvrA2B binds dimers with stalled RNAP only after its recruitment following the dislodging of stalled RNAP (dislodged RNAP denoted by $P$) by Mfd. The irreversible binding of UvrB with the release of 2UvrA follows the binding of UvrA2B. Bound UvrB creates a kink in the DNA and its C-terminus recruits UvrC resulting in an active endonuclease which cleaves a phosphodiester bond four–seven phosphates $3'$ to the damage. This event activates a second nuclease in UvrC, which incises the eighth phosphate backbone $5'$ to the damaged residue. The oligonucleotide about 12bp long, containing the dimer, is excised by the helicase UvrD, the gap filled by the action of DNA Pol I and, finally, the nick stitched by DNA ligase [20,21]. $k$ and $k_d$ denote the Michaelis–Menten and first order rate constants, respectively. 

In the parametric study reported above, the value of $\beta$ was a constant. In reality, $\beta$ is a function of the level of transcriptional activation. This notion opens the possibility of studying the process of recruitment of the repair enzymes to promoters with a variety of repressors and activators in detail using experimental as well as computational tools [19]. Furthermore, $\beta$ is likely to
Fig. 3. Performance of the biochemical network (Fig. 2) with the recruitment factor, \( \beta \), as a parameter. Four cases are considered comprising a combination of uninduced (Un) and induced (In) gene in the wild-type (\( mfd^+ \)) and mutant (\( mfd^- \)) strains. The fractional repair is computed as its mean value along the length of the transcribed strand. The fractional repair on the non-transcribed strand is 0.56 for all cases. There is a window of values of the recruitment parameter, \( \beta \), for which comparable repair of the wild-type and mutant strains may be achieved in the induced state but not in the uninduced state. For these values of \( \beta \), the fractional repair shows the strongest dependence on the level of transcriptional activation. In realistic promoters, the recruitment factor is a function of the level of transcriptional activation. For the \( lac \) promoter, induction removes bound repressor and, therefore, leads to an increase in the value of \( \beta \), resulting in faster repair by the direct pathway in the induced state. For the \( tyrT \) promoter, induction by the binding of Fis reduces the value of \( \beta \), resulting in slower repair in the induced state via the Mfd pathway. Initial dimer concentration: \([d]_0 = 0.005 \text{ nM/bp} \) (thin lines) and \([d]_0 = 0.05 \text{ nM/bp} \) (thick lines). (Solid lines) \( K^* = 10^{-2} \text{ nM}^{-1} \); (dashed lines) \( K^* = 10^{-4} \text{ nM}^{-1} \); \( t = 20 \text{ min} \); \( 0.05 \text{ nM/bp} \approx 1 \text{ dimer in every 80 nucleotides} \).

decrease as RNAP moves along the gene due to detachment. Together with the reduction in recruitment of the repair enzymes to the transcription bubble at the promoter upon binding of the transcription activator Fis, the observed slower repair in regions downstream of the transcription start site upon transcriptional activation of the \( tyrT \) gene [12] can be explained by the proposed model.

4. Discussion

The proposed model can be rigorously validated through well-designed experiments. For example, controlled expression of UvrA using plasmid constructs can increase the concentration of UvrA2B (UvrB is in excess of UvrA) leading to an increase in the recruitment of this complex to the transcription bubble at the promoter as well as availability of this complex to dimers through non-specific DNA binding. Another interesting possibility is to mutate RNAP or UvrA so as to alter their affinity for each other and thereby affect the efficiency with which the complex is recruited to the transcription bubble. The above model also suggests that the mechanism of transcription-coupled repair may be determined by the properties of the promoter and the manner in which transcription is regulated. Two promoters, resulting in different mechanisms of repair were mentioned above, namely, \( lac \) and \( tyrT \). Further studies in other promoters will help elucidate the role of the promoter in determining the mechanism of transcription-coupled repair.

The existence of an alternative pathway for preferential repair is mentioned in the works of Ahn and Grossman [13] and Grossman and co-workers [14]. This pathway relies on
Fig. 4. Comparison of model calculations with experimental data for transcription-coupled repair of the *lacZ* gene reported in literature. Model calculations: fractional repair vs. time (min) on the non-transcribed (thin lines) and transcribed (thick lines) strands. Experimental data (•) non-transcribed strand; (+) transcribed strand [11]. Region B, extending 150 bp after the mRNA start site. A value of \( \beta \) for which the fractional repair is most sensitive to transcriptional activation was chosen, i.e., \( \beta = 10 \text{nM}^{-1} \). The values of other parameters are: \( k^* = 10^4 \text{nM}^{-1} \); \( d_0 = 0.025 \text{nM/bp} \). The model clearly captures the fact that Mfd is required in the uninduced state (left two panels) but not required for preferential repair in the induced state (right two panels).

the enhanced delivery of UvrA2B due to interaction and translocation with RNAP but still requires Mfd to dislodge the stalled RNAP.

The recent observation that conformation modulating sub-units (Rpb4 and Rpb9) of RNA Pol II in *Saccharomyces cerevisiae* are capable of inducing a switch between Rad26 (Mfd pathway) and direct repair pathways [22] can be explained by the above model and unknown proteins need not be invoked, i.e., it is likely that these conformation modulating sub-units alter the distribution of the stalled and free states. The recruitment of the repair enzyme complex to the transcription bubble at the promoter region through physical association with RNAP appears to have a strong analogy with the loading of TFIIH (a basal transcription factor and an important NER protein [23-24]). TFIIH is usually jetisoned after about 30-60 nucleotides downstream from the transcription start site [6] and, is therefore, implicated in preferential repair in the immediate vicinity of the start site that does not require Rad26 (Mfd) [23]. A similar phenomenon has been noted in vitro using human nucleotide excision repair proteins [26] although no preferential repair was observed in the absence of CSB. Moreover, footprint analysis revealed that RNA Pol II stalled at a thymine dimer covers \( \sim 40 \text{nt} \) nearly symmetrical around the dimer. As static footprints capture only the average of many dynamic conformations, they do not provide adequate insights into the extent of backtracking and the underlying mechanism of repair. These examples provide ample motivation to investigate the interplay between RNAP conformations, delivery of repair enzymes, and the choice between the sub-pathways for preferential repair in eukaryotes within the framework of the model proposed above.

Methods

The mathematical representation of the biochemical network (Fig. 2) is described in detail in supplementary information. The parameter values were taken from literature and reasonable values were assigned to those whose values are not reported. The model equations were integrated in time using the function ode15s available in MATLAB® (The Mathworks, Natick, MA).
Acknowledgements

The generous support from the US Department of Energy’s Office of Biological and Environmental Research to Dr. Jeremy S. Edwards is greatly appreciated.

References