Backtracking and error correction in DNA transcription

Margaritis Voliotis[1,2,*], Netta Cohen[1], Carmen Molina-París[2], & Tanniemola B. Liverpool[3]

[1] School of Computing, University of Leeds
[2] Department of Applied Mathematics, University of Leeds
[3] Department of Mathematics, University of Bristol

The survival of living cells crucially depends on the fidelity with which genetic information, stored in nucleotide sequence of DNA, is processed during cell division (DNA replication) and protein synthesis (DNA transcription and mRNA translation). However, thermodynamics introduces significant fluctuations which would incur massive error rates if efficient proofreading mechanisms were not in place [Hopfield (1974)].

Here, we review recent work on a putative mechanism for transcriptional error correction, which relies on backtracking of the RNA polymerase (RNAP). First, we present a detailed model of backtracking pauses as a first-passage process and study the distribution of their duration [Voliotis *et al.* (2008)]. We then present an error correction model that incorporates RNAP backtracking and mRNA cleavage [Voliotis *et al.* (submitted)]. We calculate the error rate as a function of the relevant rates (translocation, cleavage, backtracking and polymerisation) and show that the theoretical limit of the proposed model is reminiscent to that accomplished by a multiple-step kinetic proofreading mechanism [Hopfield (1974)].

Introduction

Recent single molecule experiments have shed light on the microscopic details of DNA transcription. In particular it has been observed that the RNA polymerase (RNAP) undergoes frequent pausing while transcribing a DNA template. A certain class of pauses seems to occur irrespective of the underlying DNA sequence and is associated with the backward translocation of the RNA polymerase on the DNA template, a phenomenon dubbed as backtracking [Shaevitz *et al.* (2003), *et al.* Forde (2002)]. Moreover it has been shown that pause lifetimes are significantly reduced with the addition of cleavage enzymes, whereas artificially enforcing nucleotide misincorporation increase the number of pauses [Shaevitz *et al.* (2003)]. These findings suggest that pausing due to backtracking is a crucial ingredient in the proofreading mechanism.

During backtracking the active site of the RNA polymerase disengages from the $3'$ end of the nascent mRNA, enabling the polymerase to translocate backwards on the DNA template without disrupting the mRNA [Greive and von Hippel (2005)]. The RNA polymerase shows weak exonuclease activity. However, specific proteins (Gre/TFIIS) can enhance the cleavage rate while the RNAP is backtracked [Borukhov *et al.* (1993), Fish and Kane (2002)], suggesting an error correction mechanism that relies on cleavage at the misincorporated nucleotides. Actually such a mechanism permits multiple attempts for error correction, since the RNAP can polymerise several nucleotides before backtracking and correcting an error (see Fig. 1). In general RNAP backtracking is restricted by hairpins and other mRNA structures that are formed as the nascent mRNA exits the transcription elongation complex. This restriction effectively suppresses the ability of the RNAP to backtrack beyond a certain point, so any errors further back on the RNA transcript would be fixated

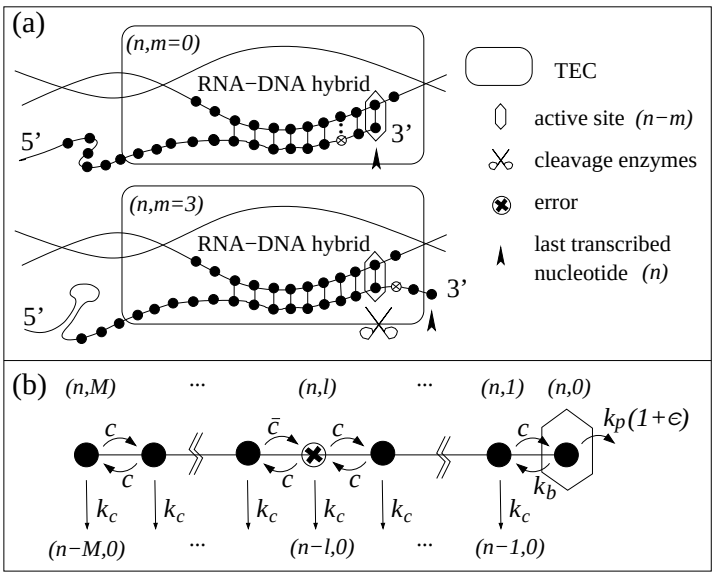

Figure 1: (a) Schematic illustration of the RNAP in the active $(n, m = 0)$ and in a backtracked state $(n, m > 0)$. The mRNA is marked by $3'$ and $5'$. Specific proteins (Gre/TFIIS) can enhance cleavage of the mRNA while the RNAP is backtracked. (b) Schematic illustration of backtracking at a specific template position. The RNAP will eventually polymerize forward $(n, m = 0) \rightarrow (n + 1, m = 0)$ or cleave at one of the backtracked states $(n, m = i) \rightarrow (n - i, m = 0)$. Figure adapted from [Voliotis *et al.* (submitted)]

Backtracking Model Results

We model the transcription elongation process in terms of two discrete variables n and m (Fig. 1(a)). Variable $n = 0, \ldots, N$ denotes the position of the last transcribed nucleotide or equivalently the current length of the mRNA transcript. The second variable $m = 0, \ldots, M$ denotes the position of the RNAP active site relative to n. Within this elongation process, we describe backtracking as a symmetric hopping process among states $0 < m \leq M$ (for any template position n) with hopping rate c and a reflecting boundary at $m = M$ (see Fig. 1(b)). A backtracking pause starts with the RNAP at position $(n, m = 1)$ and terminates once the RNAP returns to the so-called active state $(n, m = 0)$, from which further polymerisation is possible. Using the Laplace transform we solve this first passage problem and obtain an analytic form for the probability distribution of pause durations $\mathcal{P}(t)$ (see Fig. 2) [Voliotis et al. (2008)]. The pause durations can be classified into different regimes, for which the asymptotic behaviour is given by

$$\mathcal{P}(t) \sim \begin{cases} \dfrac{A}{t^{3/2}} & , \dfrac{1}{c} \ll t \ll \dfrac{M^2}{c}, \\ Be^{-Ct} & , t \gg \dfrac{M^2}{c}. \end{cases} \tag{1}$$

For times short compared to the time scale of diffusion to the reflecting state $m = M$ ($t \ll M^2/c$), but still longer than the time for the active site to diffuse by one nucleotide ($t \gg 1/c$), $\mathcal{P}(t)$ scales as $t^{-3/2}$. Interestingly, the power law behaviour characteristic of this regime is consistent with the heavily skewed and heavy-tailed distribution observed by Shaevitz *et al.* (2003). Conversely, for times much longer than M^2/c, which ensures reflection, the asymptotics are altered and $\mathcal{P}(t)$ exhibits a rapid exponential decay. The two different asymptotic behaviours are illustrated in Fig. 2, where the analytic results have been plotted together with the data obtained from stochastic simulations of the model.

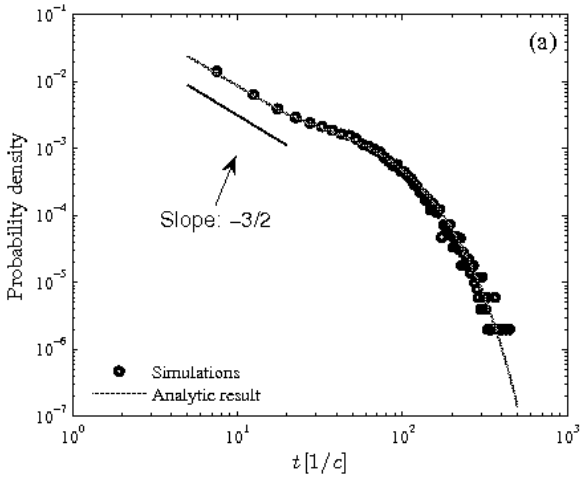

Figure 2: Distributions of pause duration $P(t)$ for $M = 10$. Plotted are the analytic results as solid lines and the results of stochastic simulations as circles. $P(t)$ and a power law decay for $1/c \ll t \ll M^2/c$, followed by an exponential cutoff in long time limit ($t \gg M^2/c$). Figure adapted from [Voliotis *et al.* (2008)]

Proofreading Model Results

To assess the effectiveness of the transcriptional proofreading we have extended the above model to include the polymerisation of errors and cleavage from backtracked states [Voliotis *et al.* (submitted)]. A schematic illustration of the model is given in Fig. 1(b). We calculate the error fraction at a given position of the transcript, $\mathcal{E}n$, defined as the ratio of the probabilities of incorporating at that position an incorrect as compared to a correct nucleotide. In the limit of slow polymerization and frequent backtracking $\mathcal{E}$ is given by

$$\mathcal{E} \approx \left(\frac{c_e}{e}\right)^{M+1} M^M/M!, \tag{2}$$

Thus, in this limit the error fraction depends only on the backtracking limit (M) and the ratio of the hopping rates. This behaviour is reminiscent of kinetic proofreading with M intermediate steps [Hopfield (1974)].

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council under grants No. EP/D003105 and EP/C011953/1 (N.C.), the Medical Research Council under grant No. G0300556 (N.C., C.M.P., T.B.L.), the Royal Society (T.B.L.), and the University of Leeds (M.V.). T.B.L. acknowledges the hospitality of the Curie Institute in Paris.

References

Borukhov, S., Sagitov, V. and Goldfarb, A. (1993). Transcript Cleavage Factor from *E. coli*. *Cell*, **72**, 459-466

Fish, R. N. and Kane C. M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta*, **1577** , 287-307.

Forde N.R, Izhaky D., Woodcock, G. R., Wuite, G. J. L. and Bustamante, C. (2002). Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Science*, **99**, 11682-11687.

Galburt, E. A. Grill, S. W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E.,Kashlev, M. and Bustamante, C. (2007). Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*, **446**, 820-823

Greive, S. J., and von Hippel, P. H. (2005). Thinking Quantitatively about Transcripptional Regulation. *Nat. Rev. Mol. Cell Biol.* , **6**, 221-232.

Hopfield, J. J. (1974). Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity . *Proceedings of the National Academy of Science*, **7**, 4135-4139.

Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. and Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near base pair resolution. *Nature*, **426**, 684-687.

Voliotis, M., Cohen, N., Molina-París, C. and Liverpool T. B. (2008). Fluctuations, Pauses, and Backtracking in DNA Transcription. *Biophys. J.*, **94**, 334-348.

Voliotis, M., Cohen, N., Molina-París, C. and Liverpool T. B. (2008). Proofreading in DNA Transcription. Submitted.