Transcription: Pausing and Backtracking: Error Correction

Mamata Sahoo and Stefan Klumpp

Theory and Bio-systems group, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany
Transcription is the efficient regulatory process in cells, organisms and tissues → Control the complex form of gene expression.

What happens?
- The genetic information → stored in DNA → RNA transcript.

How?
- Transcription → RNA polymerase moves along the length of a DNA template by a single base pair per stochastic nucleotide addition → creating a complementary RNA.
Transcriptional pausing

1. Heterogeneity in transcription rate → Transcription is not continuous ⇒ interrupted by pausing events.

2. Pauses: RNAP gets halt for times → forms inactive configuration.

2. Two general classes of Pauses → most frequent.

2. I. Artsimovitch et al., PNAS 97, 7090(2000)
Backtracking during transcription: Backtracking pauses

- Backtracking occurs in three phases.
  - Phase 1: Backtracking
  - Phase 2: Sliding → diffusional in nature.
  - Phase 3: Recovery of transcription

Questions Addressed?

- What happens to the transcription → pause and backtracking?
  - Pauses have negative effect on transcription → High transcription rate requires the pausing events to be suppressed.
  - Backtracking pauses → automatically suppressed by the trailing RNAP from behind. However, backtracking is required for the error correction and further recovery of transcription.
  - Making a pause → creating an error, Cleaving the transcript → Correcting the error.

- Questions??
  - What fraction of errors are corrected??
  - How the efficiency of error correction limited controlled??
  - How the accuracy can be improved??
Model studied for transcription
Both initiation and elongation limited.

4Low density and maximal current phase.

At high transcription initiation rate → transcription starts limiting by elongation.

Strongly affected by pausing events → elongation limited regime.

Suppresses⇒ with pausing and backtracking.

Efficiency of error correction, \( \text{fec} = \frac{\sum_{m=1}^{\infty} K_c P_m}{\sum_{m=1}^{\infty} K_c P_m + \epsilon_1 P_{m-1}} \)

For single RNAP transcription, \( \text{fec} = \frac{1}{1 + \sum_{m=1}^{\infty} \frac{\epsilon_1}{K_c P_m}} \)

Following the relation, \( \text{fec} = \frac{K_c a}{K_c a + \epsilon_1 (1-a)} \); 
\[
a = \left(1 + \frac{K_c}{2D}\right) - \frac{1}{2D} \sqrt{\left(4D^2 + K_c^2 + 4K_c D - 4DD_1\right)}.
\]
\[
= \left(1 + \frac{K_c}{2D}\right) - \frac{K_c}{2D} \sqrt{\left(1 + \frac{4D}{K_c}\right)} \quad \text{(for } D = D_1)\).
Fec is also both initiation and elongation limited.

- Increase of $D$ affect strongly in the elongation limited regime.
- Strong diffusivity suppresses the error correction $\Rightarrow$ RNAP spends much time in diffusive manner in any of the backtracked sites.
Fec with backward stepping rate($D_1$): Single RNAP and Multi-RNAP transcription

- Fec in multi-RNAP transcription is always reduced comparatively with single-RNAP transcription ⇒ Lack of free spaces that restricts diffusion of backtracked RNAP.
- The difference is strongly affected for higher $D_1$ regime.
- Further increase of $K_c$ reduces the difference between both cases ⇒ Push back effect of the trailing RNAP from behind in the multi-RNAP transcription.
Fec with the cleavage rate ($K_c$): Single-RNAP and Multi-RNAP transcription

- Fec for single-RNAP transcription is always above the fec for multi-RNAP transcription $\Rightarrow$ Available free spaces for error correction.
- Fec for multi-RNAP transcription is always reduced $\Rightarrow$ Dense traffic effect.
- Error correction in multi-RNAP case is improved for higher $K_c$. Further improvement is achieved with increase in $D_1$. 
Fec with both cleavage rate ($K_c$) and backward stepping rate ($D_1$)

- Fec is strongly controlled both by $D_1$ and $K_c$.
- Error correction $\rightarrow$ Strongly improved increasing both by backward stepping rate, $D_1$ and cleavage rate, $K_c$. 
Fec with distance \( L \) between an active RNAP and a paused RNAP

\[ \text{fec}(L) = \text{fec}_{\text{max}}^{\text{single}} \left\{ 1 - \exp\left(-\left(\frac{L}{L_0}\right)\right) \right\} \]

Approximation: \( L_0 = \frac{\epsilon}{K_c} \).

Gap distribution, \( P(L) = \left( \frac{\alpha}{\epsilon} \right) \left( \frac{\epsilon - \alpha}{\epsilon} \right)^L \).

\( \text{fec} \) increases with the distance: More free space available for error correction.

Larger gap size \( \Rightarrow \) Better error correction.
Efficiency of error correction: Multi-RNAP transcription

\[ D = D_1 = K_c = 0.07 \]

\[ \alpha_c = 0.08 \]

\[ \alpha_c = 0.04 \]

\[ D = D_1 = K_c = 0.07 \]

\[ \alpha_c = 0.04 \]

\[ \alpha_c = 0.08 \]

\[ \langle \text{fec}(\alpha) \rangle = \text{fec}_{\max} \sum_L \left\{ 1 - \exp\left( -\left( \frac{L}{L_0} \right) \right) \right\} \left( \frac{\alpha}{\epsilon} \right) \left( \frac{\epsilon - \alpha}{\epsilon} \right)^L \]

\[ = \text{fec}_{\max} \frac{(1 - \frac{\alpha}{\epsilon}) \exp\left( -\frac{1}{L_0} \right) \{\exp\left( \frac{1}{L_0} \right) - 1\}}{1 - (1 - \frac{\alpha}{\epsilon}) \exp\left( -\frac{1}{L_0} \right)} \]

- Analytical results valid for low value of \( \alpha \Rightarrow \) Semianalytical.
- The deviation starts from the critical value, \( \alpha_c = 0.04 \) where the density starts saturating.
- Beyond \( \alpha_c \), the error correction may depend on other parameters.
Transcription rate → suppressed both by pausing and backtracking (reduced saturated density effect).

We exactly calculate the efficiency of error correction for a single-RNAP and multi-RNAP transcription in a semi-analytical way.

Error correction can be strongly improved by increasing both the backward stepping rate and the transcript cleavage rate.
THANK YOU