Background

We have found a strong and consistent correlation (r $0.8, p < 0.0001$) between the productive yield (PY) and the parameters $\Delta G_{\text{RNA-DNA}}$ and Keq. These are respectively the free energy of RNA-DNA hybrid and the equilibrium constant for the movement of RNAP between the post-translocated and pre-translocated states:

$$\mathrm{R^e} \rightleftharpoons \mathrm{R^s} \tag{1}$$

The correlation we have found is on this form:

$$PY \sim e^{\Delta G_{\text{RNA-DNA}} - Keq}$$

In plain text, the productive yield of a promoter correlates with the exponent of the RNA-DNA free energy minus the equilibrium constant for translocation. In the actual calculation, the two terms are calculated for each dinucleotide of the initially transcribed sequence (ITS) of a promoter.

RNA-DNA and Keq are known, or assumed to be involved in RNAP translocation, and therefore this correlation makes sense. We would like to say that a combination of RNA-DNA free energy and translocation explain the ITS-specific abortive trascription rates. The challenge in this report is to find a physical motivation for the term

$$e^{\Delta G_{\text{RNA-DNA}} - Keq}$$

Chemical equation for RNAP movement

Below is a reaction equation for the RNA polymerase as it is translocating on DNA.

- $\mathrm{R}^{\mathrm{e}}_i$ is the pre-translocated polymerase at nucleotide i
- $\mathrm{R}^{\mathrm{e}}_{i+1}$ is the pre-translocated polymerase at nucleotide $i+1$
- $\mathrm{R}^{\mathrm{s}}_i$ is the post-translocated polymerase at nucleotide i
- $\mathrm{N_t}$ is the incoming NTP
- PP_i is the released PP_i from nucleotide incorporation

In the reaction equation RNAP osscillates between the pre-translocated and the post-translocated step at some nucleotide position i. In the post-translocated step, a nucleotide $\mathrm{N_t}$ may bind in the active site. This nucleotide might then be incorporated, bringing RNAP to position $i+1$ and at the same time releasing PP_i.

$$\ldots \mathrm{R^e_i} \rightleftharpoons \mathrm{R^s_i} \rightleftharpoons \mathrm{R^s_i N_t} \rightleftharpoons \mathrm{R^e_{i+1}} + \mathrm{PP_i} \ldots \tag{2}$$

I wish to simplity equation 2 in three steps. In the first step, I ignore the concentration of PP_i and assume that the nucleotide incorporation step is irreversible.

$$\ldots \mathrm{R_i^e} \rightleftharpoons \mathrm{R_i^s} \rightleftharpoons \mathrm{R_i^s N_t} \longrightarrow \mathrm{R_{i+1}^e} \ldots \tag{3}$$

In the second step, we skip the nucleotide binding step, assuming this step not to be rate-limiting or sequence dependent.

$$\ldots \mathrm{R_i^e} \rightleftharpoons \mathrm{R_i^s} \longrightarrow \mathrm{R_{i+1}^e} \ldots \tag{4}$$

In the third step, we assume that the osscillation between post-translocation and pre-translocation steps is rapid compared to nucleotide binding and catalysis. I can then realte the relationship between the pre- and post-translocated states by the equilibrium constant for that reaction (Keq_i):

$$\mathrm{R}_i^{\mathrm{s}} = \frac{\mathrm{R}_i^{\mathrm{e}}}{Keq_i} \tag{5}$$

I also assign the reaction rate k_1 to the nucleotide incorporation step:

$$\ldots \mathrm{R_i^e} \overset{\mathrm{Keq_i}}{\rightleftharpoons} \mathrm{R_i^s} \overset{\mathrm{k_{1i}}}{\longrightarrow} \mathrm{R_{i+1}^e} \ldots \tag{6}$$

From chemical equation to differential equation

At this stage, we would like to transfer the reaction equation into differential equations which represent rate of change of RNAP at the different positions. The expression for $\mathrm{R}_{i+1}^{\mathrm{e}}$ is according to mass action kinetics

$$\frac{\mathrm{dR}_{i+1}^{\mathrm{e}}}{\mathrm{d}t} = k_1 \mathrm{R}_i^{\mathrm{s}}$$

By inserting the relationship from equation 5, we get the following final expression

$$\frac{\mathrm{dR}_{i+1}^{\mathrm{e}}}{\mathrm{d}t} = \frac{k_1}{Keq_i} \mathrm{R}_i^{\mathrm{e}} \tag{7}$$

This equation relates the rate of change of RNAP at position $i+1$ to the amount of RNAP at position i.

The reaction constant k_1

We write the reaction constant k_1 as a special form of the Arrhenius equation:

$$k_1 = Ke^{-\frac{\Delta G}{RT}}$$

Here, ΔG is the Gibbs free energy for the reaction, R is the universal gas constant and T is the temperature in Kelvin.

To evaluate the term

$$\frac{k_1}{Keq}$$

from equation 7 (without the index i for simplicity), we write

$$Keq = e^{\ln(Keq)},$$

where ln is the natural logarithm. We insert this into $\frac{k_1}{Keq}$ to get

$$\begin{aligned}\frac{k_1}{Keq} &= \frac{Ke^{-\frac{\Delta G}{RT}}}{e^{\ln(Keq)}} \\ &= Ke^{-\frac{\Delta G}{RT} - \ln(Keq)}\end{aligned}$$