Chemical equation for RNAP movement

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

$$\mathrm{R_i^e} \rightleftharpoons \mathrm{R_i^s} \rightleftharpoons \mathrm{R_i^sN_t} \rightleftharpoons \mathrm{R_{i+1}^e} + \mathrm{PP_i} \tag{1}$$

The following terms are used in equation (1):

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

See also Figure 1 adopted from the Landick paper (Hein *et al.*) for what these symbols represent.

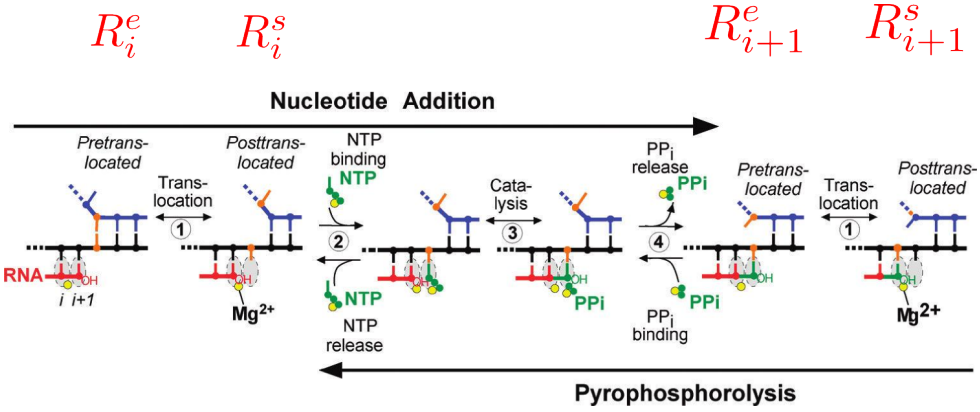

Figure 1: Polymerase at pre-translocated and post-translocated states for nucleotides i and $i+1$

i refers to the length of the growing RNA. Since backtracking is not included, this is equivalent to the template position at the active site. All expressions (like $\mathrm{R}_i^{\mathrm{e}}$) should be interpreted as concentrations ($mmol/L$ for example).

In the reaction equation (1) RNAP moves 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 may then be incorporated, bringing RNAP to position $i+1$ and at the same time releasing PP_i. This reaction ignores all pathways related to backtracking and abortion.

To make a model fit for simulation, we want to simplify the reaction kinetics. The first simplifying assumption is to make the nucleotide binding and synthesis step implicit:

$$\mathrm{R_i^e} \rightleftharpoons \mathrm{R_i^s} \rightleftharpoons \mathrm{R_{i+1}^e}\,. \tag{2}$$

Next, we assume that PPi binding is negligible. Then we reach the following simplified reaction

$$\mathrm{R_i^e} \rightleftharpoons \mathrm{R_i^s} \xrightarrow{\mathrm{k_{1i}}} \mathrm{R_{i+1}^e}\,. \tag{3}$$

Here, the reaction rate k_1 is the rate of the nucleotide incorporation step.

To simplify further, we assume that the step between pre-translocation and post-translocation reaches equilibrium. This assumption is not likely to be true, but we perform it in order to be able to introduce the Keq parameter into our equation and to simplify the model. Even though the assumption is not be entirely true, we only need that it is approximately true. We sacrifice some accuracy in order to have a more simplified model. The reaction we end up with is then:

$$\mathrm{R_i^e} \xrightleftharpoons{\mathrm{Keq_i}} \mathrm{R_i^s} \xrightarrow{\mathrm{k_{1i}}} \mathrm{R_{i+1}^e}\,. \tag{4}$$

Here Keq_i is the assumed equilibrium constant for the translocation step:

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

We will use only the last reaction step in our model:

$$\mathrm{R_i^s} \xrightarrow{\mathrm{k_{1i}}} \mathrm{R_{i+1}^e} \tag{6}$$

From chemical equation to differential equation

The rate equation for equation (6) is by definition:

$$r = k_{1_i} R_i^{\mathrm{s}}$$

Since the rate is the change of R_{i+1}^{e} with time, we form the differential equation

$$\frac{\mathrm{d}R_{i+1}^{\mathrm{e}}}{\mathrm{d}t} = k_{1_i} R_i^{\mathrm{s}}$$

Now we can make use of the relationship in equation (5) and write

$$\frac{\mathrm{d}R_{i+1}^{\mathrm{e}}}{\mathrm{d}t} = k_{1_i} \frac{[\mathrm{R_i^e}]}{Keq_i} \tag{7}$$

So far, by making a series of simplifying steps, we have arrived at an equation that relates the concentration of RNAP at position $i+1$ to the concentration of RNAP at position i. The next question is, what does k_1 look like? A general expression for the rate of a reaction is the following:

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

where ΔG is the Gibbs free energy change associated with the reaction, R is the universal gas constant and T is the temperature in Kelvin. By inserting this expression into equation (7), we get

$$\frac{\mathrm{dR}_{i+1}^{\mathrm{e}}}{\mathrm{d}t} = Ke^{-\frac{\Delta G}{RT}} \frac{\mathrm{R}_i^{\mathrm{e}}}{Keq_i}$$

Now we have exponential we initially sought. We need to get Keq into that exponential. By using

$$Keq_i = e^{\ln(Keq_i)} \text{ (from } x = e^{ln(x)})$$

we obtain

$$\begin{aligned}\frac{\mathrm{dR}_{i+1}^{\mathrm{e}}}{\mathrm{d}t} &= \frac{Ke^{-\frac{\Delta G}{RT}}\mathrm{R}_i^{\mathrm{e}}}{e^{ln(Keq_i)}} \\ &= Ke^{-\frac{\Delta G}{RT} - ln(Keq_i)}\mathrm{R}_i^{\mathrm{e}}\end{aligned}$$

The term RT evaluates to 0.62. Thus we have

$$\frac{\mathrm{dR}_{i+1}^{\mathrm{e}}}{\mathrm{d}t} \sim e^{-1.6\Delta G - \ln(Keq)} \tag{8}$$

which is a term very similar to what we started out with, which was:

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

Intermission

By performing some simplifying assumptions on the reaction (1), we have arrived at a term for the rate of RNAP elongation that is similar to the term we originally started with.

The biggest difference between the two expressions is in the sign of $\Delta G_{\text{RNA-DNA}}$. The difference between $log(Keq)$ and Keq is not large since the values for Keq are around 1 ($log(x)$ is approximately linear around $x = 1$).

To test if this simplified model would work at all, I simulated a set of ordinary differential equations based on equation (8), using $\Delta G_{\text{RNA-DNA}}$ in place of ΔG. I get a nice correlation ($r = 0.7$) – except that I must change the sign of $\Delta G_{\text{RNA-DNA}}$ from negative to positive. Otherwise the correlation breaks down.

The fact that the ΔG term from equation (8) has a positive instead of a negative sign, indicates that this term is working to 'hold back' the reaction. A candidate for the ΔG term other than $\Delta G_{\text{RNA-DNA}}$ is the one involved with scrunching of DNA during initiation. The $\Delta G_{\text{DNA-DNA}}$ energies are similar to the $\Delta G_{\text{RNA-DNA}}$ energies but not identical. The next challenge is to try to incorporate DNA scrunching into the model.

Scrunching

To introduce scrunching into the model, the following assumptions are made about the physics of transcription initiation

1. the RNA-DNA hybrid acts to stabilize the initiation complex

2. the scrunched DNA-DNA bubble de-stabilizes the initiation complex
3. the $\Delta G_{\text{RNA-DNA}}$ and $\Delta G_{\text{DNA-DNA}}$ energies were obtained for helical conformations; how these parameters are applicable to the RNA-DNA hybrid and the scrunched DNA bubble is not straightforward
4. modeling the non-abortive pathway of transcription initiation is a good enough approximation to describe the system

Let us now revisit equation (8):

$$\frac{\mathrm{dR}^{\mathrm{e}}_{i+1}}{\mathrm{d}t} \sim e^{-\frac{\Delta G}{RT} - \ln(Keq)}$$

The ΔG term is the free energy term that (together with the value of Keq) determines if the reaction

$$\mathrm{R}^{\mathrm{s}}_{\mathrm{i}} \xrightarrow{\mathrm{k_{1i}}} \mathrm{R}^{\mathrm{e}}_{\mathrm{i+1}}$$

will take place. Based on assumptions 1 and 2 in the above list, we propose that this energy term is on the following form:

$$\Delta G = \Delta G_{\text{Stabilizing}} - \Delta G_{\text{Destabilizing}}$$

where

- $\Delta G_{\text{Stabilizing}}$ is the $\Delta G_{\text{RNA-DNA}}$ energy of the growing RNA-DNA hybrid
- $\Delta G_{\text{Destabilizing}}$ is the $\Delta G_{\text{DNA-DNA}}$ energy of the scrunched DNA bubble

However, as stated in assumption 3, we do not know the relationship between the measured values of $\Delta G_{\text{RNA-DNA}}$ and $\Delta G_{\text{DNA-DNA}}$ from the literature, and the role these values play in transcription initiation. As a first approximation, we assume that the values in transcription initiation correspond to a simple scaling of the measured values by multiplying factor. Thus, we rewrite equation (8) to obtain:

$$\frac{\mathrm{dR}^{\mathrm{e}}_{i+1}}{\mathrm{d}t} \sim e^{\frac{-c1\Delta G_{\text{Stabilizing}} - c2\Delta G_{\text{Destabilizing}}}{RT} - \ln(Keq)} \tag{9}$$

In this model, the term $c1\Delta G_{\text{Stabilizing}} - c2\Delta G_{\text{Destabilizing}}$ will be positive during transcription initiation, at least for $c1$ and $c2$ equal to 1. We have thus found a way of including 'scrunching' in the model in such a way that the sign of the ΔG term is positive.

Interpretation of ΔG in transcription initiation

The starting value for the $\Delta G_{\text{Stabilizing}}$ term will be 0 (hybrid of length 0), and the starting value for $\Delta G_{\text{Destabilizing}}$ will be equal to that of the 13 bp initial bubble. As initiation proceeds, the $\Delta G_{\text{Stabilizing}}$ will grow (hybrid increases in size), but the term will be more or less equalized by the growth of the $\Delta G_{\text{Destabilizing}}$ term (the growth of the bubble). Only after the RNA-DNA hybrid has reached 8-9 basepairs will the $\Delta G_{\text{Stabilizing}}$ term stop to grow. From here on, ΔG will be come increasingly positive as the DNA-DNA bubble continues the grow but the RNA-DNA hybrid remains constant.

It is possible that the σ barriers at 9 and 15 nt should be implemented in the model, but it's not clear how this should be done.

Conclusion

I will now try to model equation (9) and see if this model fits the data.