

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 . The correlation we found was on the following form:

$$PY \sim \Delta G_{\text{RNA-DNA}} - Keq \quad (1)$$

We found this relationship almost by accident. Now we need to ascertain the physical meaning of this relationship.

A clue can be found in the scatterplot: there is an exponential relationship between the right and left side of equation 1. Thus, we are looking for a term on the form

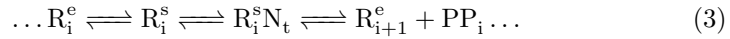
$$PY \sim e^{\Delta G_{\text{RNA-DNA}} - Keq} \quad (2)$$

Chemical equation for RNAP movement

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

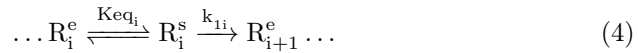
- R_i^e is the pre-translocated polymerase at nucleotide i
- R_{i+1}^e is the pre-translocated polymerase at nucleotide $i + 1$
- R_i^s is the post-translocated polymerase at nucleotide i
- N_t is the incoming NTP
- PP_i is the released PP_i from nucleotide incorporation

In the reaction equation RNAP oscillates between the pre-translocated and the post-translocated step at some nucleotide position i . In the post-translocated step, a nucleotide 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 .



This reaction ignores all pathways related to backtracking and abortion.

By skipping the nucleotide binding step and assuming that translocation is at equilibrium, we obtain the following simplified equation:



Here (Keq_i) is the equilibrium constant for the translocation step:

$$R_i^s = \frac{R_i^e}{Keq_i} \quad (5)$$

Also the reaction rate k_1 is the rate of the nucleotide incorporation step:

From chemical equation to differential equation

By assuming mass action kinetics, and using

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

we arrive at the following expression:

$$\frac{dR_{i+1}^e}{dt} = Ke^{-\frac{\Delta G}{RT}} R_i^s.$$

Inserting for equation 5, we get

$$\frac{dR_{i+1}^e}{dt} = Ke^{-\frac{\Delta G}{RT}} \frac{R_i^e}{Keq_i}$$

By using

$$Keq_i = e^{\ln(Keq_i)}$$

we obtain

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

Thus

$$\frac{dR_{i+1}^e}{dt} \sim e^{-\frac{\Delta G}{RT} - \ln(Keq)} \quad (6)$$

If we assume that ΔG from the rate of nucleotide incorporation is similar to $\Delta G_{\text{RNA-DNA}}$. Further the term RT evaluates to 0.62. Thus we have

$$\frac{dR_{i+1}^e}{dt} \sim e^{-1.6\Delta G_{\text{RNA-DNA}} - \ln(Keq)}$$

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

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

Results

Conclusion

By performing some simplifying assumptions on the reaction 3, 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 $\Delta G_{\text{RNA-DNA}}$. Since values for Keq are around 1, taking the logarithm does not change these values much in magnitude ($\log(x)$ is near linear around $x = 1$).

As a result, I simulated a set of ordinary differential equations based on equation 6. 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.

What is the cause of the sign-mismatch? I have gone over the maths several times and find no mistakes. Is there a way around this which makes sense physically?