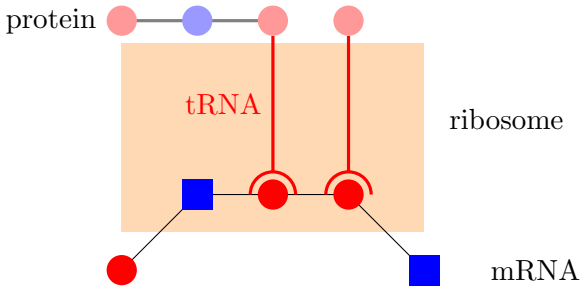
Kinetic Proofreading

DNA can be copied remarkably accurately, with an error occurring with probability of about 10^{-8} per base at the initial replication. Proteins can also be coded quite accurately, with error probabilities of about 10^{-4} per base. It is actually not trivial to achieve this task, as we will demonstrate in this problem. The mechanism which seems to be often used is called **kinetic proofreading**.

Before moving to the actual mechanism of replication, let us start by looking at a toy example, which will highlight for us the reason why we need the kinetic proofreading mechanism. Consider the formation of a protein, which roughly occurs as follows: a polymer called mRNA feeds through another protein called a ribosome. The ribosome itself attracts tRNA: proteins which, on one end, binds favorably to specific sites on the mRNA, and on the other end, carries an amino acid. The ribosome then pulls off the amino acid, and attaches it to the growing strand of a protein.



In the simple sketch of a ribosome shown above, we have two types of bases on the mRNA: red and blue, and two associated amino acids and types of tRNA. Let us assume that, as in the sketch, we have a red base on the mRNA. We have the following reaction possible, with rates shown on the reaction:

$$\text{ribosome} \underset{\mu}{\overset{\alpha}{\rightleftharpoons}} \text{correct RNA:ribosome} \xrightarrow{\omega} \text{correct protein}$$

This is the reaction we want to occur. However, we have also the following bad reaction:

$$\text{ribosome} \underset{\nu}{\overset{\alpha}{\rightleftharpoons}} \text{incorrect RNA:ribosome} \xrightarrow{\omega} \text{incorrect protein}$$

Note that we might expect $\alpha \sim \nu$, since there should be little attraction between the wrong tRNA and the mRNA, but we do expect α to be the fastest time scale in the problem, and to have $\alpha \gg \mu$. It is also reasonable to expect ω to be the slowest time scale in the problem – this will help simplify your calculations a bit. Model the ribosome as a continuous time Markov chain in its stationary distribution, neglecting the ω reaction when finding the stationary distribution. In this problem, you will also want to relate free energy differences to the reaction rates: assume that there are no important intermediate steps in the reactions.

Let us denote the error rate f with

$$f \equiv \frac{\text{rate of production of incorrect protein}}{\text{rate of production of correct protein}}.$$

(a) Within the above approximations, what is the rate of creation of incorrect proteins, and the rate for correct proteins?

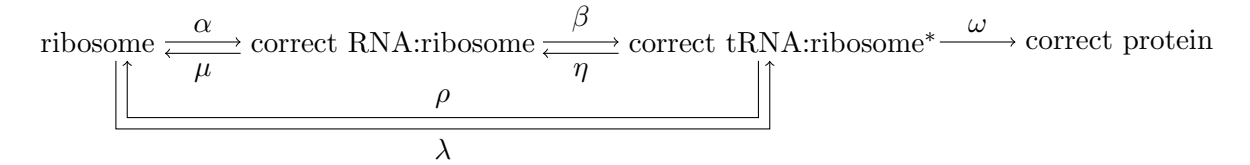
(b) Show that the error rate $f = f_0$, where

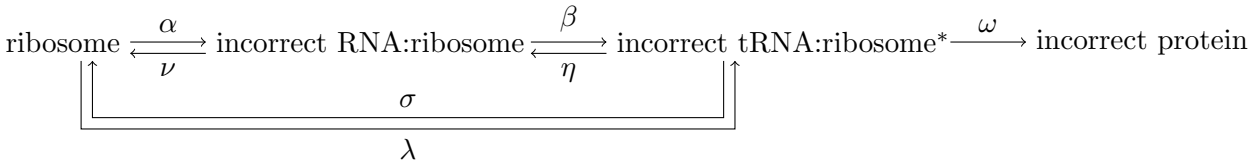
$$f_0 \equiv \mathrm{e}^{-(G_{\mathrm{err}} - G_{\mathrm{corr}})/k_{\mathrm{B}}T}$$

where G_{err} is the free energy of the incorrect tRNA:ribosome complex, and G_{corr} is the free energy of the correct tRNA:ribosome complex.

(c) It is reasonable to assume that $\alpha \sim 10^3\ \mathrm{s}^{-1}$. A ribosome in vivo can produce about 10 base pairs on a protein per second. Is it possible to achieve this *and* have the low error rate?

Before moving on to kinetic proofreading, let us consider a more naïve guess: that we can improve the error rate by simply adding a *second* internal configuration for the tRNA:ribosome complex. Roughly speaking, this means that we "check twice":





(d) Show that, in general, the error rate is now given by

$$f = \frac{(\rho(\mu+\beta)+\mu\eta)(\lambda(\nu+\beta)+\alpha\beta)}{(\sigma(\nu+\beta)+\nu\eta)(\lambda(\mu+\beta)+\alpha\beta)}$$

(e) Suppose that we do not expend any free energy during this process. Show that the constraints on the reaction rates added by this constrain $f = f_0$. So the extra step does not do any good.

Now, the idea behind kinetic proofreading is that the β reaction, which converts between the two internal ribosome states, is driven by expending free energy. In biology, this is driven by ATP phosphorylation and this means that the second internal state may be at a very high free energy, relative to the other configurations. This renders some of the conclusions of part (e) moot.

(f) Show that by expending free energy, we now have the bound $f \geq f_0^2$. How do we saturate this bound in practice?

(g) Give an intuitive explanation for why the bound is f_0^2. Guess at how we obtain bounds of f_0^n for any positive integer n.

(h) Earlier in the problem, we said to assume that ω is the smallest time scale in the problem. How much can we relax this assumption without qualitatively changing the answer to part (f)?

(i) Return to the biological numbers given in part (c). Can you now explain how the cell could transcribe proteins with the given speed and error rate? Be sure to consider the rate of production of proteins given the kinetic proofreading scheme, and not just the relative rate f, although feel free to use any approximations you found in part (f) and (h) make the kinetic proofreading scheme most efficient.