

BE 150: Design Principles of Genetic Circuits

Justin Bois

Caltech

Spring, 2018

Caltech



Donna and Benjamin M.

Rosen Bioengineering Center

This document was prepared at Caltech with financial support from the Donna and Benjamin M. Rosen Bioengineering Center.

© 2018 Justin Bois.

All rights reserved except for figures taken from literature sources.

7 Kinetic proofreading

At the center of many of the genetic circuits we have discussed thus far are the processes of the Central Dogma, notably transcription and translation. The error rate in translation of mRNA to protein by ribosomes is about 10^{-4} errors per amino acid. In this lecture, we will work out whether or not this error rate is high or low, and then discuss mechanisms about how it is achieved.

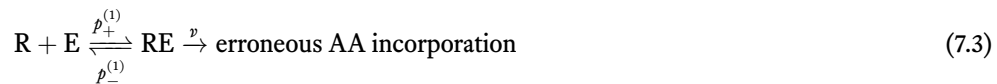
The design principle we will explore is more general: **Driven multistep pathways enable amplification of selectivity**. By the word “driven,” we mean energy consuming reactions are necessary for the multiple steps.

7.1 A simple model for incorporation of amino acids

We can think of the process of translation as consisting of two successive steps. First, a tRNA (T) reversibly associates with the ribosome/mRNA complex (R). When associated, an amino acid can be irreversibly added to the nascent peptide. This is represented by the following reaction scheme.



Unfortunately, there can also be incorporation of erroneous tRNAs and we can also get erroneous amino acid incorporation. So, we should really write two sets of chemical reactions, one for correct incorporation of an amino acid and one for erroneous.



Here, we have denoted chemical rate constants for erroneous incorporation with a p . Note that we have assumed that the irreversible incorporation rate is the same for correct and incorrect incorporation.

7.1.1 Dynamical equations

Using mass action kinetics, we can write the dynamical equations for amino acid incorporation. Importantly, we have

$$\text{rate of incorporation of correct AAs} \equiv r_c = v c, \quad (7.4)$$

$$\text{rate of incorporation of erroneous AAs} \equiv r_e = v e, \quad (7.5)$$

where c is the concentration of RC and e is the concentration of RE. We assume a separation of time scales, where incorporation of the amino acid is much slower than the other dynamics. Thus, the rate

of incorporation of amino acids is proportional to the steady state concentration of RC complexes (or RE complexes for erroneous addition). We can write differential equations for c and e .

$$\frac{dc}{dt} = k_+^{(1)} c_R c_C - k_-^{(1)} c, \quad (7.6)$$

$$\frac{de}{dt} = p_+^{(1)} c_R c_E - p_-^{(1)} e, \quad (7.7)$$

where c_R is the concentration of ribosomes, c_C is the concentration of correct tRNA and c_E is the concentration of erroneous tRNA. There are assumed to be constant, since they are in great excess and the cell is constantly producing them. So, in treating the dynamics, we only need the two above ODEs.

7.1.2 Error rates

We are interesting in computing the **error rate**, f_0 , which is the number of erroneous amino acids incorporated per correct amino acid. We will compute this at steady state, so the steady state error rate is

$$f_0 = \frac{\text{rate of incorporation of erroneous AAs}}{\text{rate of incorporation of correct AAs}} = \frac{e}{c}. \quad (7.8)$$

So, we just have to compute the steady state values for c and e . To do this, we take $dc/dt = de/dt = 0$ and solve. We get

$$c = \frac{k_+^{(1)}}{k_-^{(1)}} c_R c_C \quad (7.9)$$

$$e = \frac{p_+^{(1)}}{p_-^{(1)}} c_R c_E. \quad (7.10)$$

Thus, the error rate is

$$f_0 = \frac{e}{c} = \frac{p_+^{(1)} k_-^{(1)} c_E}{p_-^{(1)} k_+^{(1)} c_C}. \quad (7.11)$$

Note that the two ratios of rate constants are the dissociation constants for the interaction of the tRNA with the ribosome/mRNA complex; $K_{d,c} = k_-^{(1)}/k_+^{(1)}$ and $K_{d,e} = p_-^{(1)}/p_+^{(1)}$. Finally, to ease notation, we define the ratio of concentrations of erroneous and correct tRNAs to be q . Thus, we have

$$f_0 = q \frac{K_{d,c}}{K_{d,e}}. \quad (7.12)$$

From thermodynamic considerations, the dissociation constant is given by

$$K_d = \exp \left[-\frac{E_R + E_T - E_{RT}}{k_B T} \right], \quad (7.13)$$

where T is either E or C. Since $E_C \approx E_E$, the ratio of the dissociation constants is

$$\frac{K_{d,c}}{K_{d,e}} = \exp \left[-\frac{E_{RE} - E_{RC}}{k_B T} \right]. \quad (7.14)$$

So, the error rate for this simple model is

$$f_0 = q \exp \left[-\frac{E_{RE} - E_{RC}}{k_B T} \right]. \quad (7.15)$$

7.1.3 Does this work?

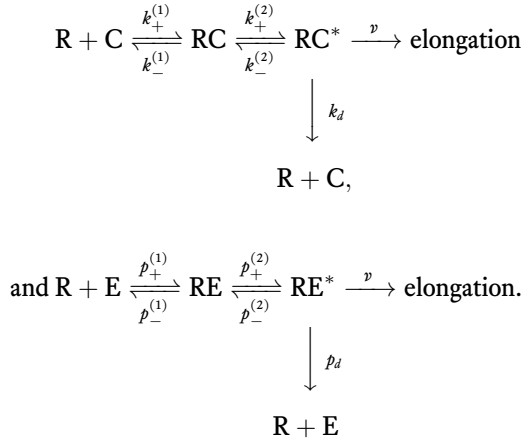
As mentioned before, about 10^{-4} erroneous amino acids per residue are incorporated. We will take $q \approx 20$, though this can vary significantly depending on the amino acid, since tRNA abundance is variable. So, to get an error rate of 10^{-4} , we need

$$\frac{E_{RE} - E_{RC}}{k_B T} \approx -\ln(5 \times 10^{-6}) \approx 12. \quad (7.16)$$

So, the difference in the interaction between an erroneous and correct tRNA and the ribosome/mRNA complex would have to be at least $12k_B T$ (maybe more because some tRNAs are in higher abundance). So, the question is what is the energy difference between correctly and incorrectly bound tRNA-mRNA pairs and how does this compare to $12k_B T$? Being off by a single base pair could result in an erroneous amino acid being incorporated. The misalignment of the base pairs would result in disrupted hydrogen bonds. Base pairs have two or three hydrogen bonds. Presumably, at least one could form if the bases are misaligned. Thus, we could get an error with a single hydrogen bond being off. The stacking interactions are not as important here, since the bases are confined within the geometry of the ribosome. So, we will estimate that the energy difference is about one hydrogen bond, or $2k_B T$. So, there is just not enough energy in hydrogen bonding to give the necessary energy difference to give the low error rate.

7.2 Kinetic proofreading

Experimentally, it is known that there are intermediate steps in translation. In 1974, John Hopfield's big insight was that these steps require energy (e.g., GTP hydrolysis involving the elongation factor EF-Tu), and can therefore serve to decrease the error rate since the energy serves to overcome the limit set by thermodynamics. These hydrolysis events can cause a conformational change in the tRNA-codon complex that allow elongation and also for dissociation of tRNA from the codon. For this scheme, the chemical reactions are



7.2.1 Error rate by kinetic proofreading

We will now compute the error rate for this new scheme. To begin, we compute the steady rate of elongation with correct amino acids, which we will call r_c . Our goal is to compute the steady state concentration of RC^* , c^* , since $r_c = \nu c^*$. Some differential equations describing the dynamics of the system under mass action kinetics are

$$\frac{dc_R}{dt} = -k_+^{(1)} c_R c_C + k_-^{(1)} c + k_d c^* \quad (7.17)$$

$$\frac{dc}{dt} = k_+^{(1)} c_R c_C - k_-^{(1)} c - k_+^{(2)} c + k_-^{(2)} c^*. \quad (7.18)$$

Setting the time derivatives to zero, equation (7.17) gives

$$c = \frac{1}{k_-^{(1)}} \left(k_+^{(1)} c_R c_C - k_d c^* \right). \quad (7.19)$$

Substituting this expression into equation (7.18) (with the time derivative set to zero) yields

$$k_d c^* - \frac{k_+^{(2)}}{k_-^{(1)}} \left(k_+^{(1)} c_R c_E - k_d c^* \right) + k_-^{(2)} c^* = 0 \quad (7.20)$$

This is readily solved for c^* to give

$$c^* = \frac{k_+^{(1)} k_+^{(2)} c_R c_E}{k_+^{(2)} k_d + k_-^{(1)} \left(k_-^{(2)} + k_d \right)}. \quad (7.21)$$

Similarly, we have for e^* ,

$$e^* = \frac{p_+^{(1)} p_+^{(2)} c_R c_E}{p_+^{(2)} p_d + p_-^{(1)} \left(p_-^{(2)} + p_d \right)}, \quad (7.22)$$

giving an error rate of

$$f = \frac{r_e}{r_c} = \frac{e^*}{c^*} = q \left(\frac{p_+^{(1)} p_+^{(2)}}{k_+^{(1)} k_+^{(2)}} \right) \left(\frac{k_+^{(2)} k_d + k_-^{(1)} \left(k_-^{(2)} + k_d \right)}{p_+^{(2)} p_d + p_-^{(1)} \left(p_-^{(2)} + p_d \right)} \right). \quad (7.23)$$

7.2.2 Approximate error rate

For ease of comparison and to build intuition, we will consider the following approximations.

- i) The forward reactions of tRNA-codon binding have the same rate. That is to say, $k_+^{(1)} \approx p_+^{(1)}$. This would be expected if these reactions are diffusion limited, since both correct and erroneous tRNAs have the same diffusion coefficient.
- ii) The dissociation of the activated complexes is much faster than their deactivation. This is essentially a statement of the irreversibility of the hydrolysis reaction. Mathematically, this means $k_-^{(2)} \ll k_d$ and $p_-^{(2)} \ll p_d$.

- iii) The equilibrium prior to hydrolysis is fast; hydrolysis and conformational change are slow. Hence, $k_+^{(2)} \ll k_-^{(1)}$, and $p_+^{(2)} \ll p_-^{(1)}$.
- iv) The hydrolysis reaction is nonspecific, such that $k_+^{(2)} \approx p_+^{(2)}$.

Applying approximation (ii), we have

$$k_+^{(2)}k_d + k_-^{(1)}(k_-^{(2)} + k_d) \approx (k_-^{(1)} + k_+^{(2)})k_d. \quad (7.24)$$

Applying approximation (iii) gives

$$(k_-^{(1)} + k_+^{(2)})k_d \approx k_-^{(1)}k_d. \quad (7.25)$$

Similar results hold for the terms describing erroneous binding. Thus, we have

$$f \approx q \frac{p_+^{(2)}}{k_+^{(2)}} \frac{k_-^{(1)}k_d}{p_-^{(1)}p_d}. \quad (7.26)$$

Applying approximation (i) allows us to write

$$f \approx q \frac{p_+^{(2)}}{k_+^{(2)}} \frac{K_{d,c}k_d}{K_{c,e}p_d}, \quad (7.27)$$

since $K_{d,e} = p_-^{(1)}/p_+^{(1)} \approx p_-^{(1)}/k_+^{(1)}$. We use the approximation that the hydrolysis reaction is nonspecific, such that $k_+^{(2)} \approx p_+^{(2)}$, giving

$$f \approx q \frac{k_d}{p_d} \frac{K_{d,c}}{K_{d,e}} = \frac{k_d}{p_d} f_0, \quad (7.28)$$

which is a k_d/p_d -fold improvement. So, to get improved fidelity, we need $k_d < p_d$. I.e., incorrectly paired complexes dissociate faster than correctly paired ones. To bring the error rate down four more orders of magnitude to the observed error rate, we would need this ratio to be about 1/10000.

7.3 Multiple steps enable greater selectivity

We have shown that adding the extra activation step allows for greater selectivity of correct amino acid incorporation in the context of translation. This extra step is a “proofreading” step that requires energy, hence the name kinetic proofreading. The kinetic proofreading principle is, in fact, more general. We have studied a *recognition problem* where there are two steps: bind a substrate and trigger a response. In this case, the substrate is tRNA and the response is addition of a new amino acid.

Kinetic proofreading has been applied extensively to recognition of major histocompatibility complex by T cell antigen receptors. The result of this recognition is T cell activation. This system exhibits extraordinary sensitivity, with the ability to detect a single ligand in a sea of thousands. This is better even than electronic circuits. In your homework, you will work out that this can be accomplished with many added steps of kinetic proofreading.