**Two promoter model**

**Generating model data for wild-type systems**

There are 5 binding sites on the gene of interest, three of which are 0A boxes (0A1, 0A2, and 0A3) and two of which are promoters (Pv and Ps). Assuming the binding at each site is binary, then there will be 25 possible configurations. We will use 1 to denote bound and 0 unbound, with the order of sites being [0A1, 0A2, 0A3, Ps, Pv]. An exemplary configuration would look like 10110, meaning 0A1, 0A3 and Ps are bound.

To calculate the total transcription rate for this 5-site system, we sum up the probabilities, , of each configuration , which are individually weighted by a transcription initiation rate[[1]](#footnote-1) unique to this configuration :

( 1 )

To get started on this equation, we can first solve for for each configuration. Assume strictly that each binding site can be bound by **only one type and one quantity** of their corresponding proteins: each 0A boxes is recognized and bound by one Spo0A~P molecule, Ps by one RNAp-sigmaH molecule, and Pv by one RNAp-sigmaA molecule. Then, the occupation probability of a particular configuration , with molecules of Spo0A~p, molecules of RNAp-sigmaH and molecules of RNAp-sigmaA in place can be computed as:

( 2 )

where the numerator is the partition function of a particular configuration, and the denominator is the sum of partition functions of all possible combinations resulting from different values. For configuration 10110, .

Note that is the standard free energy of a particular configuration. It is calculated by adding together all binding energies for individual sites that are occupied, , as well as interactive energies rising from the interactions between ligands when they bind, . These include interactions between Spo0A~P molecules,[[2]](#footnote-2) between Spo0A~P and RNAp-sigmaA,[[3]](#footnote-3) between Spo0A~P and RNAp-sigmaH, and between RNAp-sigmaA and RNAp-sigmaH. There are no tertiary interaction energies. A total of interactions are possible. We use to denote the occupancy of a site (1 = bound, 0 = unbound). Therefore,

( 3 )

**Generating model data for mutated systems**

When a site is mutated, no transcription factor or RNA polymerase can be bound. This eliminates some of the previously possible configurations, thus changing the total transcription rate. There are many ways to incorporate mutations into the model, and we demonstrate one way of doing it using the example of 12\*.

When 0A1 and 0A2 boxes are made infeasible for binding, we set the first two digits of all configurations to be 0 (i.e. 10110 🡪 00110, 11011🡪00011, etc.). The model will then calculate probability of each configuration using equation (2) and (3) and add them up using equation (1). However, some configurations would be considered multiple times; for instance, now 10110, 11110, 01110 and 00110 all are computed the same way as 00110. [[4]](#footnote-4)

Following this pattern, even though we still have 25 terms in equation (1), there are only 23 *unique* configuration in the current system; in other words, each configuration is counted four times. Therefore, to obtain the correct total transcription rate, we will divide the original result by 4, or 2number of mutations.

**Fitting to real data**

Overall, we demonstrated how to calculate the total transcription rate for a particular value of [Spo0A], for both wild-type and mutated strains. Now, if the concentration of Spo0A~P changes over time, the probability of each configuration goes up or down to affect the total transcription rate. For a set of [Spo0A~P] values , we will be able to compute a vector of total transcription rate, , as a function of [Spo0A~P], that is:

( 4 )

Dr. Fujita’s lab kindly provides the experiment data for wild-type strains and all mutated strains, including 1\*,2\*,3\*,12\*,13\*,23\*,123\*. Given a set of promoter activity for each type of strain

that varies with Spo0A~P concentration, we call them :

( 5 )

The goal of the computational model is to find parameters such that the distance between two vectors is minimized, that is

( 6 )

Specifically, variables to optimize include transcription initiation rate for each configuration, , binding energies of each site , and interactive energies between every possible pair of ligands, . We can choose to fit both wild-type and mutated types together, or separately if needed.

**Solve for time-dependent RNAp related terms**

We have not mentioned anything about where to get information on [RNAp]. Since concentration and binding energy of RNAP polymerase always appear together, we combine them into one parameter r for promoter Ps and Pv in our model, respectively:

( 7 )

For a two-promoter model, r is actually a known variable to us; we can obtain the value of r by fitting the one-promoter real data to a one-promoter model. We will explain this briefly here using the example of promoter Ps.

In a one-promoter model, the total transcription rate as a function of [Spo0A] is computed the same way as in the two-promoter model, except for i) there are 24 possible configurations, ii) there are no interactive energies involving Pv promoter, iii) equation (2) now becomes:

( 8 )

By taking away the Pv promoter, Dr. Fujita’s lab produced a Ps-only strain and tested out what effects different mutations have on promoter activity. When looking at the experimental data with three 0A boxes being mutated, , we saw that transcription rate still changes when no transcription factor Spo0A~P is bound. Theoretically, there are only two possible configurations in this case (0000 and 0001), and promoter activity is solely contributed by the configuration when RNAp binds:

( 9 )

From equation (9), we know that the only reason for transcription rate to vary with time is if [RNAp] varies with time. We then use a particle swarm algorithm to estimate how [RNAp] changes with time by minimizing the difference between and .

This way, we obtained time-dependent [RNAp-sigmaH] from Ps-only 123\* data and [RNAp-sigmaA] from Pv-only 123\* data. They will then be used as inputs into the two-promoter model.

1. We can reduce the number of by knowing that when none of the promoters is bound, transcription event will be very unlikely to happen. In other words, we assign all configurations ending with xxx00 with an initiation rate of 0, and reduce the number of ’s from 32 to 24. [↑](#footnote-ref-1)
2. We can potentially limit the interaction between 0A1- and 0A3-bound ligands to be 0 because they are not physically adjacent to each other. [↑](#footnote-ref-2)
3. Transcription machinery for Spo0A gene consists of a core enzyme and a sigma factor. [↑](#footnote-ref-3)
4. We used a map data structure in Matlab such that each configuration has a unique. Even if number of unique configurations now decreases and affects the order of how for each configuration is read into the model, this map made sure correct is assigned. [↑](#footnote-ref-4)