

CONTENTS

| | |
|---------------------------|---|
| 1. Introduction: | 2 |
| 2. Choice of Model | 2 |
| 3. Mass Action Law | 2 |
| 4. Degradation | 3 |
| 5. Transcription | 4 |
| 6. Translation | 5 |
| 7. Protein Activity | 6 |
| 8. Initial Data | 6 |
| 9. Simplification | 6 |
| 10. MATLAB Implementation | 7 |
| References | 7 |

1. INTRODUCTION:

In the process of developing our model, we had certain difficulties to find reference or guidelines to model genetic networks in prokaryotic systems. The only book that covered differential equations for gene expression in our library was [1]. It covers only very rudimentary basics but if you have no prior experience in the topic it will certainly be helpful. To ease the modeling for other teams, we put together a guide that explains how we developed the model for our system. The guide will not give minute details on what the biological/chemical/physical backgrounds are, for this we linked some of the titles to the corresponding wikipedia page that explains the processes in detail.

2. CHOICE OF MODEL

The first step in modeling is always to choose whether you want to model your system in a stochastic or deterministic and time/state-discrete or time/state-continuous manner. The most convenient way to select between a discrete and continuous model is to look at the behavior of the variables in the real world, since it makes no sense to model single molecule interactions with quantities that include fractions of molecules. A model where the state variable is discrete in most cases limits you to a stochastic model. If you pick a continuous model the choice to model the system with a deterministic or stochastic ansatz is highly dependent on what characteristics of the system you want to analyze in your model. For quantitative data, like what amount of Protein XY is produced, a deterministic model fits your needs. For qualitative data, like the variance in the production of Protein XY. The decision between a stochastic and deterministic models is sometimes also dependent on which parameters for the system are available or rather attainable.

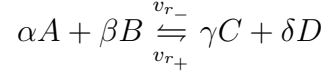
The next step is to pick a level of detail. Here your decision should be based on what level of detail is feasible and what level of detail is needed and you should always choose the lower level of those two. The feasibility again depends on the available parameters and how well the considered system is known. In your implementation the level of detail should be reflected by the tolerances that you pass to your solver and the amount of simplifications that you apply to your equations.

In biological systems negative concentrations or populations do not make any sense. This should be reflected by your equations and if you experience any negative data, you should probably check whether there is something wrong with your modeling.

The following sections will cover most of the basics that are needed for the development of a deterministic continuous model.

3. MASS ACTION LAW

The fundamental mathematical model for chemical reactions is the Mass Action Law, It provides a mathematical model for reactions of the type

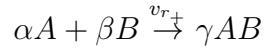


The resulting differential equations are

$$\begin{aligned} [\dot{A}] &= \alpha v_{r+} [C]^\gamma [D]^\delta - \alpha \gamma v_{r-} [A]^\alpha [B]^\beta \\ [\dot{B}] &= \beta v_{r+} [C]^\gamma [D]^\delta - \beta v_{r-} [A]^\alpha [B]^\beta \\ [\dot{C}] &= \gamma v_{r-} [A]^\alpha [B]^\beta - \gamma v_{r+} [C]^\gamma [D]^\delta \\ [\dot{D}] &= \delta v_{r-} [A]^\alpha [B]^\beta - \delta v_{r+} [C]^\gamma [D]^\delta \end{aligned}$$

Where the $[]$ brackets denote that the concentration is used

This method can be used for arbitrarily complex reactions that only involve one mechanistic step. In most cases the mass action law is used to model the boning of two elements



which is modeled as described above

$$\begin{aligned} [\dot{A}] &= -\alpha v_{r+} [A]^\alpha [B]^\beta \\ [\dot{B}] &= -\beta v_{r+} [A]^\alpha [B]^\beta \\ [\dot{AB}] &= \gamma v_{r+} [A]^\alpha [B]^\beta \end{aligned}$$

Notice that in this case we did not include the backwards reaction. If the compound AB can dissociate in its products the backwards reaction of course should be included.

4. DEGRADATION

A compound does not always dissociate into its process and there are certain chemical processes that rob RNA and Proteins of their initial functionality. To take account for those degradation processes where the substance X decays, such that the decay-products are of no further importance to the system, we impose a degradation term on all degradable substances:

$$[\dot{X}] = \dots - k[X]$$

where k is typically the the inverse of the half-value period of the degradation process. Sometimes this k also takes account of cell division, then the inverse of the half-value period of the cell cycle is added.

5. TRANSCRIPTION

Regulation of gene transcription is the essential to the behavior of genetic networks so special attention should be given to this part of the modeling. The differential equation for the transcription consists of up to three terms. First an optional term for the constituent transcription of mRNA

$$[mRNA] = k_l + \dots \quad \text{with constituent transcription rate } k_l$$

then the term for the degradation of mRNA

$$\dots + \frac{1}{\tau_{\frac{1}{2}}} [mRNA] + \dots \quad \text{with half-value period } \tau_{\frac{1}{2}}$$

and finally the most important term that models the influence of activators and repressors binding to the regulatory region of the promoter:

$$\dots + v_{max} g([A], [R])$$

with v_{max} maximum transcription rate of the gene and some function $g([A], [R])$ that can depend on the activator concentration $[A]$ and the repressor concentration $[R]$.

The most basic model for g is a Hill-Function

$$g = \begin{cases} 1 - \frac{\left[\frac{R}{k_R}\right]^n}{\left(1 + \left[\frac{R}{k_R}\right]^n\right)} & \text{for repression} \\ \frac{\left[\frac{A}{k_A}\right]^n}{\left(1 + \left[\frac{A}{k_A}\right]^n\right)} & \text{for activation} \\ \frac{\left[\frac{A}{k_A}\right]^n}{\left(1 + \left[\frac{A}{k_A}\right]^n\right)} * \left(1 - \frac{\left[\frac{R}{k_R}\right]^n}{\left(1 + \left[\frac{R}{k_R}\right]^n\right)}\right) & \text{for both} \end{cases}$$

where the parameter n is used to fit the shape of the function to gathered data. If no data is available a general rule of thumb is that n is the number of molecules involved in the binding process. k_A and k_R are the binding rates of activator and repressor.

This is of course a very basic model, but in our case we lacked some parameters and decided that a more sophisticated model would only give the illusion of a higher level of detail. However if parameters and exact behavior of the regulatory region are known the following prototypes for g in case of activation only can be used

$$g = \left\{ \begin{array}{ll} \left\{ \begin{array}{l} \frac{\left[\frac{[A]}{k_A}\right]^n}{\sum_{k=0}^n \left(\left[\frac{A}{k_A}\right]\right)^k} \\ \frac{\left[\frac{A}{k_A}\right]^n}{\sum_{k=0}^n \binom{n}{k} \left(\left[\frac{A}{k_A}\right]\right)^k} \end{array} \right. & \begin{array}{l} \text{sequential binding} \\ \text{independent binding} \end{array} \\ \left\{ \begin{array}{l} \frac{\left[\frac{A}{k_A}\right]^n}{\prod_{i=1}^n k_{A_i}} \\ \frac{\left[\frac{A}{k_A}\right]^n}{\sum_{j=0}^n \frac{[A]^j}{\prod_{i=1}^j k_{A_i}}} \\ \frac{\left[\frac{A}{k_A}\right]^n}{1 + \sum_{j=0}^n \frac{\binom{n}{j} [A]^j}{\prod_{i=1}^j k_{A_i}}} \end{array} \right. & \begin{array}{l} \text{sequential binding} \\ \text{independent binding} \end{array} \end{array} \right. \quad \begin{array}{l} \text{for no cooperativity} \\ \text{for cooperative binding} \end{array}$$

where cooperative binding means that the binding rate of the other binding sites is affected if an activator binds to a specific binding site and sequential binding means that an activator can only bind if the previous binding site is occupied. For cooperative binding the different binding rates are denoted by k_{A_i} whereas in case of no cooperativity all k_{A_i} are equal and k_A is used respectively.

The formulas given above can be adapted for repression only and activation and repression in the same manner as in the basic version.

(ref: lecture notes system biophysics)

6. TRANSLATION

The next step after transcription of the DNA is the translation of the mRNA into proteins. Since the binding rates of the ribosome and folding process are widely unknown it is advisable to use a very simplified model for the translation process which only involves a translation rate v_{trans} linearly scaling with availability of mRNA

$$\dot{[Protein]} = v_{trans} * [mRNA] - \frac{1}{\tau_{\frac{1}{2}}} [Protein]$$

If the time τ_{fold} needed for the folding of the protein is known it could be implemented by using Delay Differential Equations with the folding time as constant delay

$$\dot{[Protein]} = v_{trans} * [mRNA](t - \tau_{folding}) - \frac{1}{\tau_{\frac{1}{2}}} [Protein]$$

Since solving DDEs requires a special much slower solver and the impact of the change can be expected to be very small, this possible improvement was omitted in our model.

If there are other regulatory processes involved in the mRNA translation as in the case of our AND-Gate where the translation of the T7 RNA Polymerase depends on the availability of tRNA a combination of the techniques for Mass Action Law and Transcription can be used to model their influence.

7. PROTEIN ACTIVITY

As protein activity can be regulated by activators and repressors in a similar manner as the transcription of DNA, analogous methods can be used. The the constituent as well as the regulated activity rate will of course additionally depend on the concentration of the substrate and of the protein. In general Protein activity is modeled with the Michaelis-Menten principle

$$[Product] = v_{max}[Protein] \frac{[Substrate]}{K + [Substrate]}$$

where v_{max} is the maximum protein activity and K is the concentration of substrate at which half the maximum activity is achieved. If the initial substrate concentration is much larger than K, we can simplify this equation to

$$[Product] = v_{max}[Protein]$$

(Ref: Chaos Lac)

8. INITIAL DATA

If you design a bacterial application that reacts to a specific signal, the reactions occur independent of the signal will normally have reached a steady state. Thus it is advisable to set their initial values in your simulation to their respective steady state.

9. SIMPLIFICATION

In some cases your model will involve multiple subsystems that work on different timescales which will most probably lead to ill-posed problems which cause numerical problems. These timescales can be identified by dimensional analysis and sufficiently fast subsystems can be assumed to reach their steady state instantly. This means that if the reaction for x is fast its differential equation

$$\dot{x} = f(x, y, \dots)$$

simplifies to an algebraic equation

$$0 = f(\bar{x}, y, \dots)$$

whose solution can be substituted into the other equations.

10. MATLAB IMPLEMENTATION

If no simplifications regarding fast/slow subsystems are made or if you experience your code to be unexpected slow, it is advisable to use a solver for stiff differential equations as ode23s.

REFERENCES

- [1] S.P. Ellner and J. Guckenheimer. *Dynamic models in biology*. Princeton Paperbacks. Princeton University Press, 2006.