BioBricks construction for Context Project 2011/2012
February 15, 2012

1 Defining gates

For simplicity of modeling and circuit design, two basic gates have been chosen:

- a NOT gate, see Figure 1(a).
- an AND gate, see Figure 1(b).

(a) NOT gate. (b) AND gate.

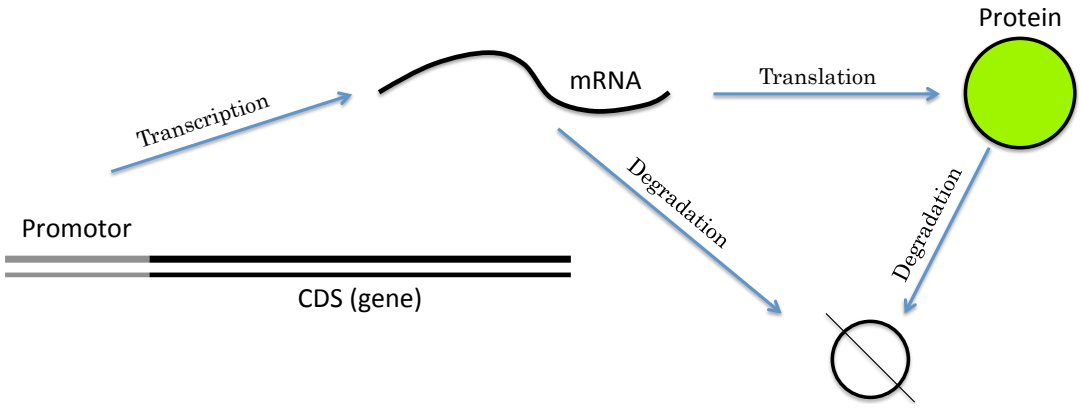
Figure 1: Symbols used for depicting local gates.

There are many approaches to modeling logic gates and logic gene networks They differ in complexity and level of abstraction from the underlying biological processes, such as cooperative or competitive *transcription factor* (TF) binding, RNA polymerase binding, transcription, ribosome binding, translation, and RNA and protein degradation. Typically, the more explicit a model is, the more parameters are required to define it and the more difficult is it to estimate these correctly.

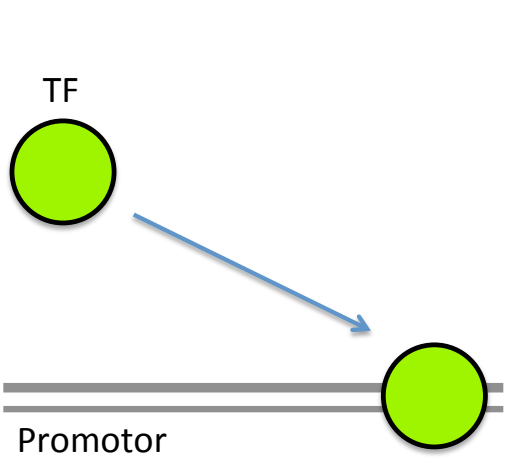
In this project a high-level abstraction leading to a simple model is proposed: gates are viewed as DNA sequences consisting of (a) a promotor region, binding one or more TFs with known affinity; (b) a gene *coding sequence* (CDS), which encodes for genes of known mRNA and protein degradation rates and known translation rate. Such an abstraction allows for defining gene coding sequences independently from each other and the gates in which they are used. Figure 2 shows the relevant steps considered in the model. Depending on the gate type (a NOT gate or an AND gate), either a promotor binding a single TF is used, or a promotor binding two TFs is used:

- NOT gates use promotor regions binding a single TF (see Figure 2(b)), which reduces expression (i.e. acts as a repressor).

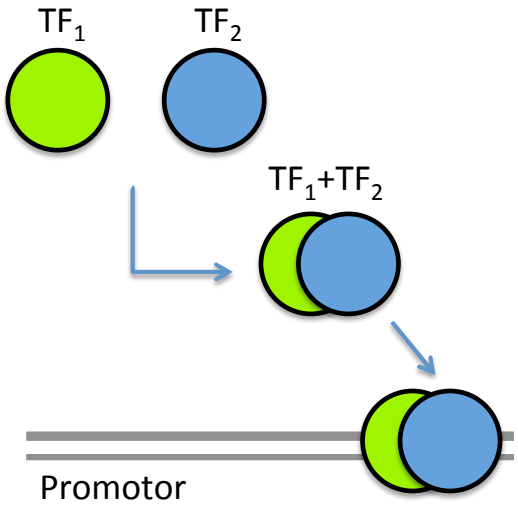
- AND gates use promotor regions binding a pair of TFs *in a cooperative fashion* (see Figure 2(c)) that induce expression (i.e. act as activators).



(a) The promotor region controls mRNA expression levels (transcription rates), whereas the CDS determines translation and degradation rates for the resulting mRNA and protein molecules.



(b) Promotor binding a single TF.



(c) Promotor cooperatively binding a pair of TFs.

Figure 2: Model overview.

2 Mathematical modeling

The processes depicted in Figure 2 can be modeled in a number of ways, ranging from the use of *ordinary differential equations* (ODEs) to simulated stochastic models. Modeling by means of differential equations is used here, for ease of understanding. Anything from a single gate to a complete gene network can be modeled by a systems of ODEs that describe how molecule (mRNA, protein, transcription factors) levels change over time. For a simple system with a single gene and a constitutive promotor with constant expression, the system can be modeled with two equations:

$$\begin{cases} \frac{d[\text{mRNA}]}{dt} & = k_1 - d_1\,[\text{mRNA}] \\ \frac{d[\text{Protein}]}{dt} & = k_2\,[\text{mRNA}] - d_2\,[\text{Protein}] \end{cases}, \tag{1}$$

where k_1 is the transcription rate (i.e. the number of mRNA molecules produces per gene and per unit of time), which in this example is constant due to use of a constitutive promotor with constant expression; d_1 is the mRNA degradation rate; k_2 is the translation rate (i.e. the number of protein molecules produces per mRNA molecule per unit of time); and d_2 is the protein degradation rate. Parameters d_1, d_2 and k_2 are constants describing a specific gene (CDS).

2.1 A NOT promotor

When a promotor with non-constant expression is used (i.e. a promotor binding TFs that activate or repress its activity), the transcription rate k_1 is modeled differently. The effect of repression through binding of a single transcription factor TF_1 can be modeled using a Hill function:

$$f([\text{TF}_1]) = \frac{k_1 K_m^n}{K_m^n + [\text{TF}_1]^n},$$

where [TF] is the repressing transcription factor concentration; k_1 is now the maximal transcription rate (when no repressor is present); K_m is the repression coefficient (the concentration of the repressor TF_1 required to repress the expression by 50%; and n is the so called Hill coefficient, which controls the steepness of the switch between no repression and full repression.

A promoter used in a NOT gate is then defined by (a) the TF that it binds and (b) the Hill function parameters k_1, K_m and n. Using (1), a NOT gate taking TF_1

as input can thus be modeled as:

$$\begin{cases} \frac{d[\text{mRNA}]}{dt} & = \frac{k_1 K_m^n}{K_m^n + [\text{TF}_1]^n} - d_1 [\text{mRNA}] \\ \frac{d[\text{Protein}]}{dt} & = k_2 [\text{mRNA}] - d_2 [\text{Protein}] \end{cases}, \tag{2}$$

where TF_1 can either be a logic network input signal, or an output of another gate.

2.2 An AND promotor

The use of a promotor that is activated through binding of a *single* transcription factor TF_1 requires modeling of transcription rate k_1 using a similar Hill function:

$$f([\text{TF}_1]) = \frac{k_1 [\text{TF}_1]^n}{K_m^n + [\text{TF}_1]^n},$$

where k_1 is the maximal transcription rate (in the presence of a large concentration of activating TF TF_1); K_m is the activation coefficient (the concentration of activator TF_1 required to activate the overall expression by 50%; and n is the Hill coefficient, which controls the steepness of the switch from no activation to full activation.

When the promotor requires cooperative binding of two TFs to activate expression, a different Hill function is used:

$$f([\text{TF}_1], [\text{TF}_2]) = \frac{k_1 ([\text{TF}_1] [\text{TF}_2])^n}{K_m^n + ([\text{TF}_1] [\text{TF}_2])^n}.$$

In this form, the Hill function considers the cooperative binding of TF_1 and TF_2 through the use of mass-action kinetics.

A promotor used in an AND gate can then be defined by (a) the two TFs that cooperatively bind to it; (b) the Hill function parameters k_1, K_m and n. Using (1), an AND gate taking TF_1 and TF_2 as inputs can be modeled as:

$$\begin{cases} \frac{d[\text{mRNA}]}{dt} & = \frac{k_1 ([\text{TF}_1][\text{TF}_2])^n}{K_m^n + ([\text{TF}_1][\text{TF}_2])^n} - d_1 [\text{mRNA}] \\ \frac{d[\text{Protein}]}{dt} & = k_2 [\text{mRNA}] - d_2 [\text{Protein}] \end{cases}. \tag{3}$$

3 BioBricks

Given the proposed abstraction, it is reasonable to choose genes (coding sequences) and promotors as the smallest building blocks (BioBricks) for this projects. A CDS

is then defined by its translation and degradation rates k_2, d_1 and d_2; a promotor is defined by a list of TFs that it binds, the maximum transcription rate k_1 and the Hill function parameters.

3.1 Coding sequences

Ten coding sequences have been generated for the first part of the project: Gene "A" to "J" (Table 1).

Gene	Parameters	Gene	Parameters	Gene	Parameters
Gene A	$k_2 = 4.6337$ $d_1 = 0.0240$ $d_2 = 0.8466$	Gene B	$k_2 = 4.6122$ $d_1 = 0.0205$ $d_2 = 0.8627$	Gene C	$k_2 = 4.1585$ $d_1 = 0.0235$ $d_2 = 0.8338$
Gene D	$k_2 = 3.0938$ $d_1 = 0.0197$ $d_2 = 0.8101$	Gene E	$k_2 = 2.0315$ $d_1 = 0.0163$ $d_2 = 0.8369$	Gene F	$k_2 = 3.0805$ $d_1 = 0.0157$ $d_2 = 0.8878$
Gene G	$k_2 = 3.5894$ $d_1 = 0.0171$ $d_2 = 0.7222$	Gene H	$k_2 = 2.5034$ $d_1 = 0.0164$ $d_2 = 0.8595$	Gene I	$k_2 = 4.3378$ $d_1 = 0.0243$ $d_2 = 0.9330$
Gene J	$k_2 = 1.8689$ $d_1 = 0.0237$ $d_2 = 0.7407$				

Table 1: A set of generated NOT promotors and parameters describing them.

3.2 NOT promotors

For each coding sequence, a NOT promotor binding it has been generated (Table 2), allowing for construction of NOT gates that take proteins produced by the coding sequences (through transcription and translation) as TFs.

TF	Parameters		TF	Parameters		TF	Parameters	
	k_1	$= 4.7313$		k_1	$= 2.8753$		k_1	$= 3.2155$
A	K_m	$= 224.0227$	B	K_m	$= 281.3545$	C	K_m	$= 213.2011$
	n	$= 1$		n	$= 1$		n	$= 1$
	k_1	$= 1.3884$		k_1	$= 4.8549$		k_1	$= 4.1169$
D	K_m	$= 270.1843$	E	K_m	$= 217.3772$	F	K_m	$= 228.4511$
	n	$= 1$		n	$= 4$		n	$= 5$
	k_1	$= 6.1147$		k_1	$= 5.0024$		k_1	$= 1.6675$
G	K_m	$= 293.8169$	H	K_m	$= 200.7356$	I	K_m	$= 236.7109$
	n	$= 1$		n	$= 1$		n	$= 2$
	k_1	$= 3.1634$						
J	K_m	$= 224.9620$						
	n	$= 2$						

Table 2: A set of generated NOT promotors and parameters describing them.

3.3 AND promotors

For every pair of coding sequences an AND promotor has been generated (Table 3), to which proteins produced by these sequences collectively bind as TFs. For the 10 CDSs from Section 3.1 a total of 45 different promotor regions have been designed, leading to $45 \times 8 = 360$ different possible AND gates (8 possible coding sequences for every pair of input signals).

TF_1	TF_2	Parameters		TF_1	TF_2	Parameters	
		k_1	$= 4.5272$			k_1	$= 6.5432$
A	B	K_m	$= 238.9569$	A	C	K_m	$= 290.5048$
		n	$= 3$			n	$= 3$
		k_1	$= 4.3077$			k_1	$= 5.8487$
A	D	K_m	$= 290.9701$	A	E	K_m	$= 266.1395$
		n	$= 3$			n	$= 3$

		k_1	$= 2.3735$			k_1	$= 2.2344$
A	F	K_m	$= 260.2590$	A	G	K_m	$= 286.1992$
		n	$= 1$			n	$= 1$
		k_1	$= 5.0349$			k_1	$= 5.6093$
A	H	K_m	$= 243.7318$	A	I	K_m	$= 230.3279$
		n	$= 4$			n	$= 1$
		k_1	$= 6.2219$			k_1	$= 6.5982$
A	J	K_m	$= 236.9308$	B	C	K_m	$= 227.4099$
		n	$= 4$			n	$= 5$
		k_1	$= 6.2147$			k_1	$= 4.3980$
B	D	K_m	$= 279.1344$	B	E	K_m	$= 238.6681$
		n	$= 4$			n	$= 5$
		k_1	$= 4.7208$			k_1	$= 5.9266$
B	F	K_m	$= 278.1641$	B	G	K_m	$= 285.2921$
		n	$= 5$			n	$= 3$
		k_1	$= 6.4610$			k_1	$= 4.0897$
B	H	K_m	$= 233.3847$	B	I	K_m	$= 203.1209$
		n	$= 2$			n	$= 5$
		k_1	$= 2.6516$			k_1	$= 5.2615$
B	J	K_m	$= 267.7499$	C	D	K_m	$= 265.0759$
		n	$= 4$			n	$= 2$
		k_1	$= 3.9838$			k_1	$= 2.3070$
C	E	K_m	$= 293.6419$	C	F	K_m	$= 227.2472$
		n	$= 2$			n	$= 4$
		k_1	$= 3.0713$			k_1	$= 1.5466$
C	G	K_m	$= 260.3075$	C	H	K_m	$= 230.8915$
		n	$= 1$			n	$= 2$

		k_1	$= 6.6824$			k_1	$= 4.7103$
C	I	K_m	$= 293.6884$	C	J	K_m	$= 204.3818$
		n	$= 1$			n	$= 3$
		k_1	$= 4.2599$			k_1	$= 2.5950$
D	E	K_m	$= 284.0339$	D	F	K_m	$= 290.4691$
		n	$= 4$			n	$= 4$
		k_1	$= 4.5167$			k_1	$= 3.1108$
D	G	K_m	$= 289.7154$	D	H	K_m	$= 261.8855$
		n	$= 2$			n	$= 1$
		k_1	$= 3.3469$			k_1	$= 5.5203$
D	I	K_m	$= 218.7381$	D	J	K_m	$= 267.4674$
		n	$= 5$			n	$= 5$
		k_1	$= 6.1310$			k_1	$= 1.5918$
E	F	K_m	$= 254.8725$	E	G	K_m	$= 251.4758$
		n	$= 1$			n	$= 5$
		k_1	$= 6.2691$			k_1	$= 3.3283$
E	H	K_m	$= 293.5012$	E	I	K_m	$= 232.0901$
		n	$= 5$			n	$= 4$
		k_1	$= 2.3264$			k_1	$= 3.2858$
E	J	K_m	$= 295.9152$	F	G	K_m	$= 234.4912$
		n	$= 5$			n	$= 4$
		k_1	$= 5.4788$			k_1	$= 1.6073$
F	H	K_m	$= 280.0343$	F	I	K_m	$= 255.9509$
		n	$= 3$			n	$= 1$
		k_1	$= 5.7352$			k_1	$= 2.4187$
F	J	K_m	$= 265.9648$	G	H	K_m	$= 271.3714$
		n	$= 3$			n	$= 4$

		k_1	$= 3.1514$			k_1	$= 5.8144$
G	I	K_m	$= 203.2751$	G	J	K_m	$= 250.9647$
		n	$= 5$			n	$= 4$
		k_1	$= 5.9448$			k_1	$= 4.6044$
H	I	K_m	$= 233.0669$	H	J	K_m	$= 228.7909$
		n	$= 5$			n	$= 2$
		k_1	$= 6.5636$				
I	J	K_m	$= 236.6101$				
		n	$= 1$				

Table 3: A set of generated AND promotors and parameters describing them.

3.4 Choosing rate constants

The rate constants involved in transcription, translation, activation and repression were chosen randomly from Guassian $N(\mu, \sigma)$ and uniform integer $U(a, b)$ distributions:

- For coding sequences: $k_2 \approx N(1.5, 2)$, $d_1 \approx N(0.01, 0.12)$ and $d_2 \approx N(0.7, 0.5)$.
- For NOT promotors: $k_1 \approx N(1, 2.5)$, $K_m \approx N(200, 10)$ and $n \approx U(1, 5)$.
- For AND promotors: $k_1 \approx N(1, 2.5)$, $K_m \approx N(200, 10)$ and $n \approx U(1, 5)$.

The resulting values are most likely not biologically plausible.