Principles of Synthetic Biology Midterm

October 30, 2012

Instructions: This exam is an open book / open notes exam. You are not permitted to use any electronic devices during the course of this exam. You will have 90 minutes to complete the test. Please use exam booklets for your solutions. Answers included elsewhere will not be graded.

This exam is a bit lengthy so please exercise sound time management. Good Luck!

1. Circuit Parts and Compositors (10 Points)

Suppose we transformed the following transcriptional regulatory network into a cell:

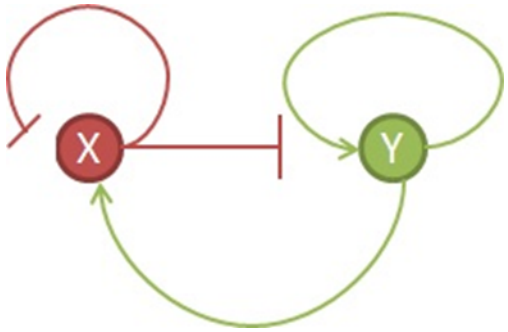


Fig. 1: A model transcriptional regulatory network

1.1. Qualitatively describe one or more possible behaviors of this system. Under what conditions would these behaviors manifest themselves? Furthermore, what dependencies on the host cell exist? (**4 points**)

This is a simple relaxation oscillator that consists of an activator Y and an inhibitor X. It is evident from the diagram that X uses a negative feedback motif to control its production while simultaneously repressing Y's production. Y, on the other hand, uses a positive feedback loop to catalyze its own production and activates the synthesis of X. If we started with a large amount of X (above self-repression threshold) and a small amount of Y (below self-activation threshold), then X effectively shuts down its production (not enough Y to catalyze X's production) because of negative feedback and actively continues to repress Y's production until the concentration of X falls below the self-repression threshold due to degradation. At this point, there is not enough X to keep repressing either itself and/or Y so we would expect to see an increase in Y as the positive feedback loop kicks in. However, since Y also activates its repressor, we would begin to see an uptick in X's concentration until it crosses the self-repression threshold upon which the cycle repeats. Consequently we have oscillations depending on parameter values that can be found by solving for the Hopf bifurcation point.

1.2. List the corresponding *PARTS* of this system and provide one possible mathematical representation of these parts. Draw a boundary around the respective parts on figure 1 itself. (**3 points**)

1) X represses its own production: $v_{rx} = V_{r-x}\left(\frac{K_{mx}^{i}}{K_{mx}^{i}+X^{i}}\right)$
2) X represses the production of Y: $v_{ry} = V_{r-y}\left(\frac{K_{nx}^{q}}{K_{nx}^{q}+X^{q}}\right)$
3) Y activates its own production: $v_{ay} = V_{a-y}\left(\frac{Y^{p}}{K_{ny}^{p}+Y^{p}}\right)$
4) Y activates the production of X: $v_{ax} = V_{a-x}\left(\frac{Y^{j}}{K_{my}^{j}+Y^{j}}\right)$
5) X is degraded: $v_{dx} = -\gamma_x X$
6) Y is degraded: $v_{dy} = -\gamma_y Y$

1.3. List the corresponding *COMPOSITORS* of this system and provide one possible mathematical representation of these compositors. (**3 points**)

$$\frac{dX}{dt} = v_{rx}v_{ax} + v_{dx}$$
$$\frac{dY}{dt} = v_{ry}v_{ay} + v_{dy}$$

X and Y maintain their state.

2. Circuit Analysis (30 Points)

Consider the circuit in figure 2. In this system, *A* is a transcriptional activator of gene *O* that binds to *O*'s promoter in a non-cooperative, Michaelis-Menton fashion. *B*, whose production is induced by the concentration of Arabinose presented, inactivates *A* by binding to it and forming a heterodimer complex *AB* that is no longer able to activate *O*. Assume that there is some basal production of *O* even when no activator is bound to its promoter. For simplicity, assume that all binding and unbinding reaction achieve equilibrium rapidly. Use the following sub-questions to guide you in this problem.

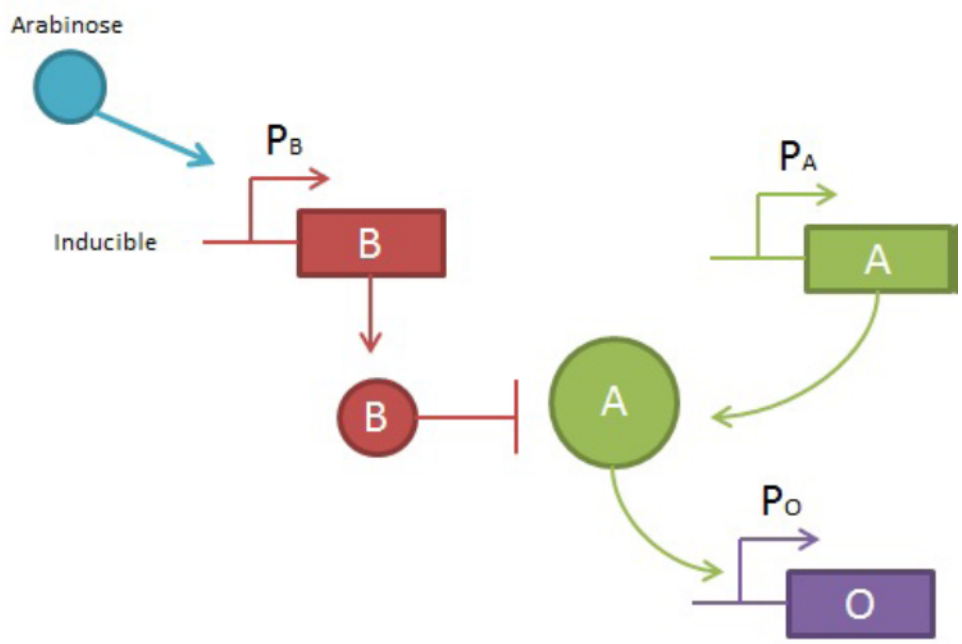
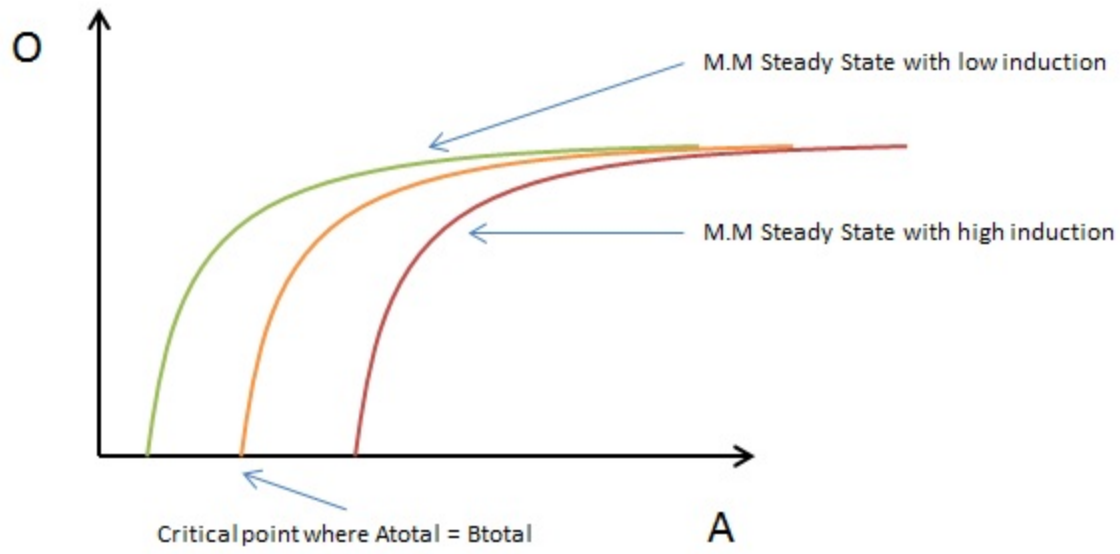


Fig. 2: A model gene circuit

2.1. If the formation of the heterodimer complex *AB* is vastly favored over free *A*, what would you expect to happen at *low* induction, in which the total amount of *B* in the cell is much less than that of *A*? What would you expect to happen at *high* induction? (**6 points**)

At low induction, A is essentially free to function as an activator of O; therefore we would expect that the system reaches some steady state which is a combination of the basal production and Michealis-Menton production that saturates at V max. At high induction limit, there are few free A molecules left in the system to actively catalyze O's transcription, so we would get some steady state production of O for all time which is a ratio of the basal production and decay terms. We would expect this steady state level to be lower than in the activated case.

2.2. Qualitatively describe the effect of increasing the total amount of *B* in this system. Sketch the output-input (*O* vs. *A*) curve in the graph below as you turn on induction of *B* with arabinose. Speculate on any changes to the steady state levels of *O*. Sketch three curves representing increasing induction and clearly label them as either *low*, *medium*, or *high* induction. (**6 points**)



As we ramp up induction of B's promoter, we will slowly begin to accumulate B in the system. In doing so, the commensurate amount of A is sequestered by tight binding and is therefore unable to catalytically drive O's transcription. Consequently, the production of O remains slow until the concentration of A approaches the critical point where it is roughly equivalent to that of B. To the right of this point, we would expect to see a rapid increase in O since there are now enough activator molecules present to catalytically drive O's production.

2.3. Qualitatively describe what happens in the regime where the total concentrations of *A* and *B* in the system are close to each other. (**4 points**)

To the right of this point, we would expect to see a rapid increase in O since there are now enough activator molecules present to catalytically drive O's production. Consequently, this system is ultrasensitive to relative changes in A's concentration in a neighborhood of this critical point.

2.4. Now consider the same scenario with a positive feedback loop in which *A* can catalyze its own production (see figure 3). Suppose you transformed this system into *E. coli*. Bacteria with high Yellow Fluorescent Protein (YFP) expression (above some basal threshold) are

considered ON, whereas those that do not are considered OFF. Why is the gene for Cyan Fluorescent Protein (CFP) also included in this system? (**4 points**)

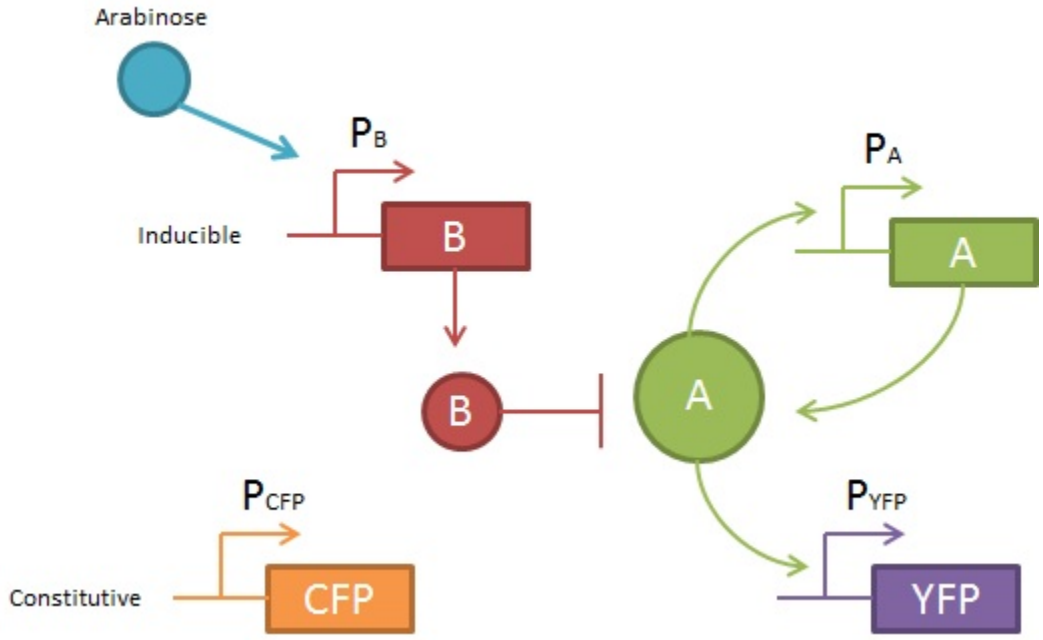


Fig 3: **A model gene circuit with a positive feedback element**

It is a negative control. Bacteria that are alive but OFF would still express CFP. In addition, since there are stochastic fluctuations in metabolic activity between cells, this gives us a measurable baseline which we can use to normalize YFP output.

2.5. Speculate as to what effect(s) the positive feedback loop might have on circuit dynamics as opposed to the open loop case (figure 2). Consider the following experiment: Grow bacteria in arabinose deficient media and then inoculate them into media containing intermediate concentrations of arabinose. Measure YFP levels for each cell. Separately, grow bacteria at full induction of *B* for the same length of time and then inoculate a sample into the same intermediate concentrations of arabinose. Measure YFP levels for each cell. How would you expect these two sets of measurements to compare? Would they be the same or different and why? What useful property does this circuit have? (**10 points**)

Under a positive feedback loop, production rates are initially small when the concentration of free activator is low (high induction); as we move towards lower induction, production will gradually begin to accelerate as more activator molecules are free from the sequestering agent. On the other hand, production rates will continue to remain high as the concentration of free activator molecules is sequentially reduced by sequestration (increasing induction) so long as the concentration of the activator molecules is above the activation threshold. The net effect is that the path traversed from low induction to high induction is not the same as that from high induction to low induction; in effect, we have a hysteretic switch which remembers the previous state. CFP expression should not vary in either state and serves as a baseline

3. Cell Classifier (30 Points)

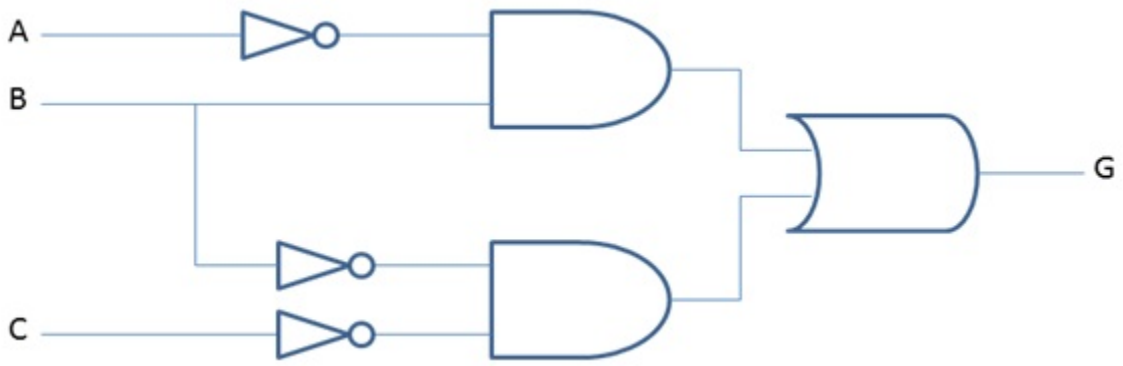
You find three markers *A, B,* and *C* which you can use to identify a cell line. You would like to build a circuit that implements the appropriate logic (given below) and only produces GFP during specific instances of marker expression that match this profile.

A	*B*	*C*	**G (GFP)**
0	0	0	**1**
1	0	0	**1**
0	1	0	**1**
1	1	0	**0**
0	0	1	**0**
1	0	1	**0**
0	1	1	**1**
1	1	1	**0**

3.1. Find the simplest logic expression, $G = G(A, B, C)$, that captures the input-output relationship demonstrated in the truth table above. Write this as a logic function. (**5 points**)

$$G = \bar{A}B + \bar{B}\bar{C}$$

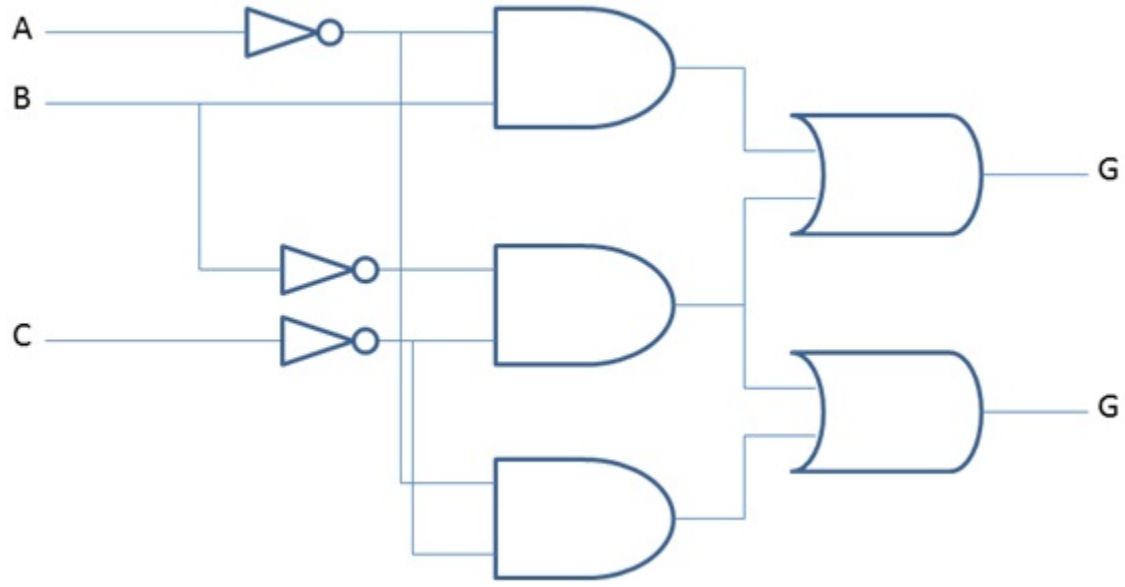
3.2. Draw the simplest logic circuit, using as few logic gates as possible, that captures the input-output relationship demonstrated in the truth table above. You may only use NOT, AND, and OR logic gates. (**5 points**)



3.3. List any hazards in your circuit. Again using as few logic gates as possible, draw the simplest hazard-free circuit. If your design does not have any hazards, write NO HAZARDS, and proceed to the next question. (**10 points**)

A hazard occurs when the following conditions are met:

Time	**A**	**B**	**C**
0	0	1	0
1	0	0	0



3.4. Implement your hazard free circuit using only the following biological parts: Repressor protein (Rep), protein comprising an Activation domain (ActD) coupled with a Leucine zipper (Lx), and a protein comprising a DNA binding domain (DBD) coupled with a Leucine zipper (Ly). The corresponding singly-activated and singly-repressed promoters are also available. For this question, assume that Leucine zippers can form heterodimers specifically. In your answer, draw the biological circuit, and indicate the specific Leucine zipper pairs (i.e L*i* binds to L*j*). (**10 Points**)

A(RepA) –inverter: expresses ActD(A)
B(RepB)-inverter: expresses ActD(B1)
B also ActD-DBD(B0)-delay: expresses BDB(B2)

C(RepC)-inverter: expresses DBD(C1) and DBD(C2)
ActD(A)-DBD(C1): expresses G
ActD(A)-DBD(B2): expresses G
ActD(B1)-DBD(C2): expresses G

4. Noise Generator (30 Points)

Galenman is interested in understanding bacterial population dynamics. He proposes to construct a gene circuit that can drive a homogeneous bacterial population into two distinct subpopulations in such a manner that the ratio of the two subpopulations can be controlled by a chemical. To accomplish this feat, he intends to design a noise generator whose noise level can be modulated by an exogenous chemical *C*, and then connect this noise generator to a bi-stable switch. The population type is determined by whether or not the output of the bi-stable switch, which is based on a positive feedback loop, is high. Galenman finds a tetrameric activator *A* and promoter *P* pair. The promoter can be turned off by adding chemical *B*. He also finds a series of degradation tags that he can use to modulate protein concentration in the cell. The protein production curve from this promoter is shown below (figure 4).

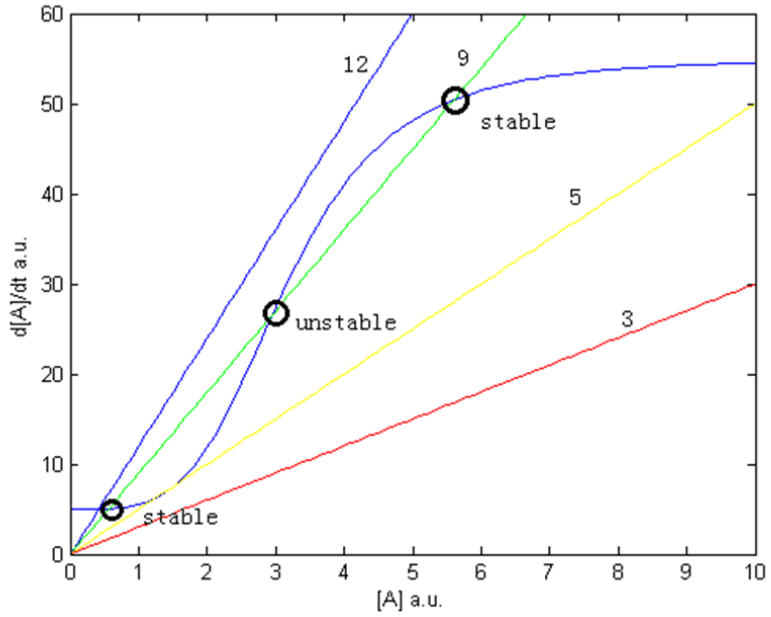


Fig. 4: Production curve of A from P

4.1. From the table below, choose a degradation tag that can form a bi-stable switch. On figure 4 itself:

Tag Name	Degradation rate (a.u. per unit time)
Slowtag	3
Medtag	5
Fastag	9
Spedtag	12

4.1.1. Sketch the degradation curve for the chosen degradation tag (**3 points**)

4.1.2. Circle any and all stationary points (**3 points**)

4.1.3. Label the stability of each stationary point as either stable or unstable (**4 points**)

4.1.4. Assuming that the tetramer formation is highly cooperative and fast relative to the transcription and translation of *A*, estimate the numerical value of the three parameters in the ODE provided below such that the model accurately fits the production curve in figure 4. (**4 Points**)

$$\frac{d[A]}{dt} = a + b\frac{k \cdot [A]^4}{1 + k \cdot [A]^4}$$

The parameter values that best fit these curves are

$$a = 5; b = 50; k = 0.01$$

4.2. To make the noise generator, Galenman finds a low noise promoter that can reliably maintain its mRNA copy level at *m* copies per cell. Assuming that the translation rate, k_{tr} (events per unit time), is given by $k_{tr} = 2C$, where *C* is the concentration of chemical *C* and that translation events occur independently, compute the mean and variance of protein production per unit time in terms of *m* and *C*. Justify your result. (**6 points**)

$$Mean = 2mC$$

$$Variance = 2mC$$

4.3. The output from the noise generator will be the same tetrameric activator. In order to control the population ratio, Galenman needs to determine the appropriate amount of *C*. To simplify the problem, assume that the distribution you computed in question 3.2 can be approximated

by a continuous triangle function that peaks at the mean and drops to zero at one standard deviation of the mean, i.e at $mean \pm \sqrt{variance}$. Assume that both the bi-stable switch and the degradation processes are strictly deterministic and that the noise generator alone is stochastic. Calculate the concentration of *C* needed in order to set the *A-HIGH* population to 20% of the total population when $m = 18$. Suggest an experimental procedure that could be used to recreate this scenario. (**10 points**)

The unstable point is around A = 3 a.u. Mean + (1-20%*2)*Variance = 3

2mC/9 + (1-20%*2)*2mC/9 = 3

if using square root of variance

$$Mean + (1 - 20\% \times 2) \times \sqrt{Variance} = 3$$

$$\frac{2 \times 18C}{9} + 0.6 \times \sqrt{\frac{2 \times 18C}{9}} = 3$$

$$C \cong 0.73$$

Add B and C, wait until steady state, remove B and C wait until steady state