AIV 2019/20120

Computational biology part

Gregory Batt

Guidelines: Provide a pdf report with written answers, including how you implemented things, and the plots you got together with their interpretations and possible comments you may have. Provide the matlab code in a zipped format. Name these two files LastName1_LastName2.pdf and LastName1_LastName2.zip. Send them to me by email (gregory.batt@inria.fr). Meet the deadline… and respect the guidelines!

The objective of this problem is to design an extension of the bacterial population control system developed by You *et al* that shows better performances. Indeed, the only demonstrated way to tune the cell density at steady state of the induced system is by changing the pH of the growth media. Firstly, this is not a very practical solution. Secondly, the dynamic range –defined here as the ratio of the highest to the lowest cell density that one can get at steady state– is rather limited this way.

To tentatively address these limitations, we investigate whether the additional use of an inducible promoter to control the intracellular levels of I and/or R proteins could lead to a system with a larger dynamic range for cell densities. We will do this in three steps. Firstly, we will model and simulate the existing system, and compute its predicted dynamic range. Secondly, we will propose various extensions of the existing system, and model them. Thirdly, we will simulate their behaviors and compute their predicted dynamic ranges. However, our initial knowledge of the functioning of the inducible promoter we would like to use is pretty bad, so we will need to characterize it before the third step. And to do this, you will need to propose experiments to perform and use the corresponding data for infer parameter values for the promoter.

Q1: *Propose a graphical representation of the system that represents the modeling viewpoint. That is, represent graphically the modeling assumptions. You can naturally get inspiration from Fig1a.*

Because we plan to play with the levels of the proteins I and R (seel below), we would like to make explicit the presence of these proteins in the model. We will assume that the cellular production rate of AHL is proportional to the I protein concentration, and that the production rate of the E protein is proportional to the AHL concentration and to the R protein concentration. To keep *kA* and *kE* parameter unchanged, we will set parameters for the I and R equations such that their steady state concentration is 1. Stated differently, we set the units of I and R as “relative

concentrations with respect to the original You system". We will also assume like the E protein, the apparent degradation of the I and R proteins comes from the dilution due to growth.

Q2: *Implement the model with the two additional differential equations and the additional parameters. Represent graphically and check that now the graphical representation of the model is very close to that of Fig1a. Which parameter should you set to zero to implement the OFF behavior? Simulate the ON and OFF behaviors at pH 7 as in Fig 3b. We will always assume that the system starts in the OFF state (ie, A, E, I and R are absent) and at a cell density of* 10^5 $CFU.mL^{-1}$*. For your representations, plot on the same graphs the variables that have comparable values, and use a log scale when appropriate.*

Q3: *Simulate the behavior of the system for the different pH provided in Table 1. What is the dynamic range of the existing system that one obtains by playing with the pH?*

The *luxR* and *luxI* genes are both driven by the $p_{lac\text{-}ara1}$ promoter, used in an ON/OFF manner by the authors. Using our inducible promoter to drive the *luxR* and/or *luxI* genes we could obtain better modulation capabilities without affecting the pH of the environment. From now on, we will always reason at pH 7.0.

Q4: *Like in question 2, represent graphically the different circuits that one can construct.*

To model these circuits, we need to have a model of the inducible promoter. We will assume that its activity is appropriately described by the following function $pa(m) = k_{basal} + k_{regulated} \frac{m^\eta}{\theta^\eta + m^\eta}$ where m represents the concentration of the inducer. We will assume that m ranges between 0 and 3, and will consider increments of 0.3 for subsequent computations.

Q5: *Write down the ODE models corresponding to the different circuits. Test your implementation using the following parameter values:* $k_{basal} = 0.2$, $k_{regulated} = 5$, $\theta = 1$, *and* $\eta = 2$. *That is, simulate the behavior of the three circuits in ON and OFF conditions and in presence or absence of the inducer (ie, m=0 or m=3). Check that everything is consistent with what you expect.*

Unfortunately, we do not know the values of the parameters appearing in the promoter activity function above. So you have to select a first circuit, called here a helper circuit, that will allow us to characterize these unknown parameter values, and once we will have characterized our promoter using this helper circuit, we will be able to compute which circuit among the ones we consider has the largest predicted dynamic range, and propose this one for final construction (if different from the helper circuit). The information we have is that $k_{regulated}$ and k_{basal} are in intervals centered respectively on the values 5 and 0.2, with ± 60% bounds. Moreover, we also assume that θ ranges between 0.4 and 1.6, and that η ranges between 1.8 and 2.3. We would like to calibrate our model using only lacZ tests (ie E readouts) as in the paper. Experimental data consists in the measurement of the steady state concentration of the E protein for many different

inducer concentrations. Since cloning is time consuming, you have to select just one helper circuit to construct. You should select the helper circuit for which the variance of the output is maximum when all parameters vary in their admissible ranges.

Q6: *Justify the above choice. Consider for each parameter 5 equally-spaced values in the provided intervals. Then, for each inducer concentration we consider, compute the steady state values of the measured quantity for all possible parameter combinations, and their variance. Identify the helper circuit that maximizes the overall variance, that is, the mean variance over all the considered inducer concentrations.*

Once you have selected your circuit to construct and hence the experiment to perform, send me an email with this information and I will provide you the corresponding (noisy) experimental data. The objective of the next question is to help you to carry out parameter fitting so that you can characterize your promoter.

Q7: *Use the function compute_cost and cmaes to search for parameter values of the inducible promoter that fit the experimental data. Simulate the behavior of the system with the parameters you found, and plot it on top of the experimental data.*

We now have characterized our promoter! Now we can decide which circuit to build to obtain the largest dynamic range.

Q8. *As in question 5, simulate and represent the behavior of the 3 circuits you can construct in the various conditions of interest, and compute their predicted dynamic range. Which circuit will you use? Is the result fully satisfying?*

References.

[1] Programmed population control by cell-cell communication and regulated killing, L. You, R. Sidney Cox III, R. Weiss & F.H. Arnold, Nature 428, 868–871.