
Problem Set 1

- You may use your course materials and/or any literature resources (as well as the internet) to formulate your solutions.
- You may work in teams. However, each student must submit their individual work. *Solutions must be typed*. All model/analysis code must be submitted to GitHub and the link provided to the teaching staff for *each* student. Solutions should be submitted electronically to the teaching staff.
- Problem Set 1 is due on **Thursday, March 16, 2017** by **4:59 PM**. Problem Set 1 is worth 100 points. A 50% penalty will be charged for each late day.

---

1. (20 pts). **Open complex formation.** The effective transcription model we developed in class to describe the transcription of gene j implicitly lumped the time required to form the open complex (initiation time) into the control variable u_j. The study of McClure (PNAS, **77**:5634-5638, 1980) developed an *in-vitro* technique to measure the initiation time using *E.coli* RNA polymerase (RNAP).

    a) Modify the effective transcription model presented in class to explicitly account for open complex formation.

    b) *Qualitatively* reproduce the difference between the cases: (i) RNAP pre-incubated with the DNA template (closes symbols) and (ii) RNAP pre-incubated without the DNA template (open symbols) in Fig. 1 of the McClure study.

    c) What is the impact of gene read length (consider $\mathcal{L}_{T,j}$ = 10bp, 100bp and 1000bp) with a characteristic length of $\mathcal{L}_T$ = 100bp)?.

c) (**Optional**) *Quantitatively* reproduce the difference between the cases: (i) RNAP pre-incubated with the DNA template (closes symbols) and (ii) RNAP pre-incubated without the DNA template (open symbols) in Fig. 1 of the McClure study.

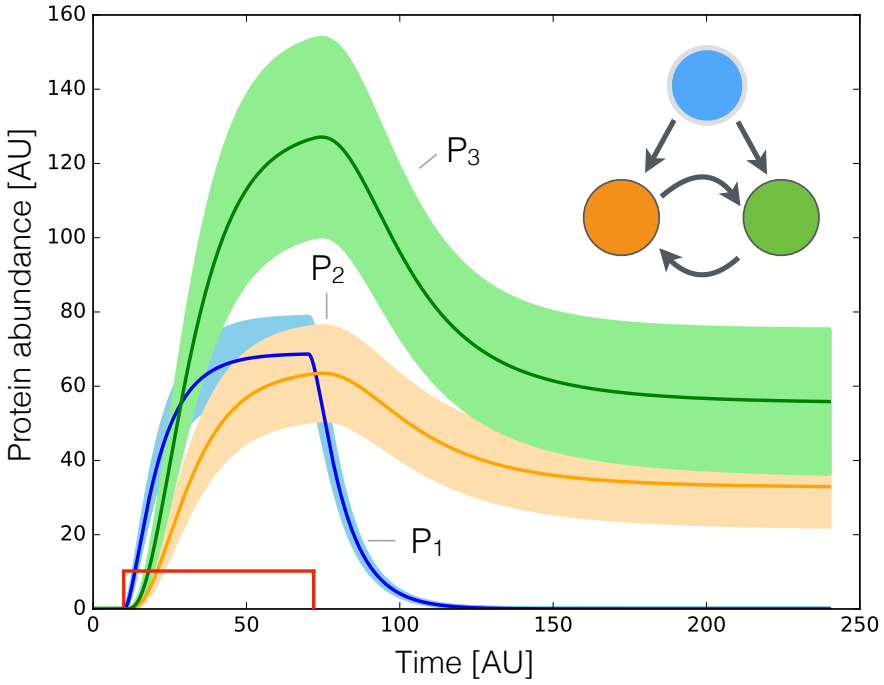

Figure 1: Schematic and simulation of *prototype* three gene memory motif. Inducer $I(t)$ (shown in red) drives the expression of P_1.

2. (30 pts). **Sensitivity analysis.** We developed sensitivity analysis tools in class to understand which model parameters controlled both the dynamic and steady-state performance of a biochemical model. Use these tools to understand the performance of the three-gene memory motif shown in Fig. 1. Assume this motif is being analyzed in a growing population of *E. coli* cells and $\mathcal{L}_{T,1}$ = 1080bp, $\mathcal{L}_{T,2}$=1251bp and $\mathcal{L}_{T,3}$ = 3075bp.

a) Assume all mRNA and protein species are measurable at any frequency (as fast as you want). Use singular value decomposition (SVD) to decompose the time-averaged sensitivity array to estimate which model parameters (and pa-

rameter combinations) control the dynamics of P_j induction. Which parameters control the steady-state abundance of P_j?

b) Which parameters are identifiable from measurements of mRNA$_1$ and P_3 versus all species being measured as a function of sampling frequency?

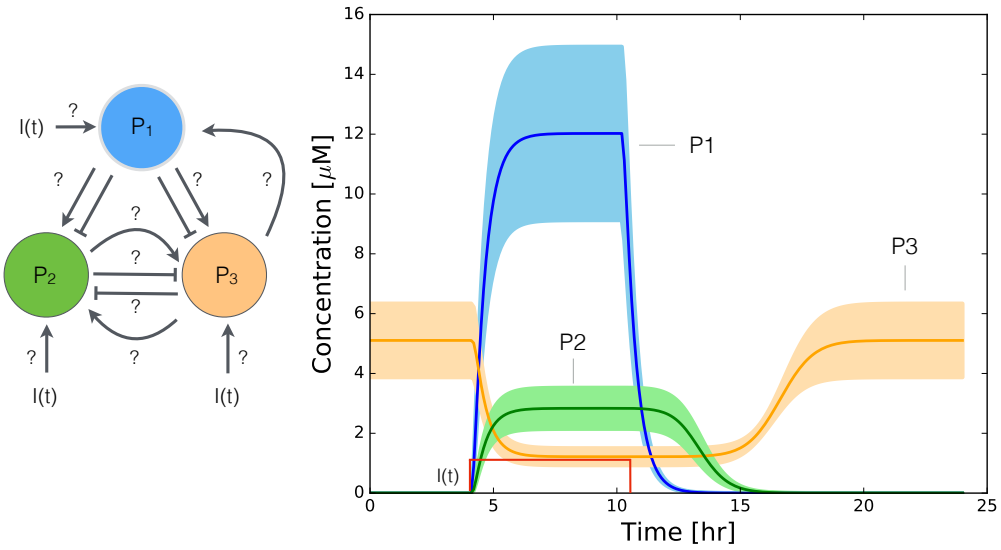

Figure 2: Model discrimination problem. Hypothetical three-gene model connectivity (left). True model behavior following the addition and washout of inducer $I(t)$ (right).

3. (50 pts). **Model discrimination**. Use model discrimination, sensitivity analysis and code generation tools to estimate the biochemical connectivity of the *true* model (Fig. 2, right) given the probable model structures (Fig. 2, left). Assume the expression of only one gene can be induced by $I(t)$. Lastly, assume your experiments are being conducted in a growing population of *E.coli* cells and $\mathcal{L}_{T,1}$ = 1080bp, $\mathcal{L}_{T,2}$=1251bp and $\mathcal{L}_{T,3}$ = 3075bp.

a) Develop a strategy to discriminate between competing model structure hypotheses. This can involve any of the tools we discussed in class (or any studies in the literature).

b) Implement your strategy to rank-order probable model structures. Toward your strategy, the teaching staff will run *synthetic experiments* for using the true model, and will return the synthetic measurements as text file.