Exercises

8.11 **The repressilator.**[1] [2] (Biology, Computation) ④

The 'central dogma' of molecular biology is that the flow of information is from DNA to RNA to proteins; DNA is *transcribed* into RNA, which then is *translated* into protein.

Now that the genome is sequenced, it is thought that we have the parts list for the cell. All that remains is to figure out how they work together! The proteins, RNA, and DNA form a complex network of interacting chemical reactions, which governs metabolism, responses to external stimuli, reproduction (*proliferation*), *differentiation* into different cell types, and (when the cell perceives itself to be breaking down in dangerous ways) programmed cell death, or *apoptosis*.

Our understanding of the structure of these interacting networks is growing rapidly, but our understanding of the dynamics is still rather primitive. Part of the difficulty is that the cellular networks are not neatly separated into different modules; a given protein may participate in what would seem to be several separate regulatory pathways. In this exercise, we will study a model gene regulatory network, the *repressilator*. This experimental system involves three proteins, each of which inhibits the formation of the next. They were added to the bacterium *E. coli*, with hopefully minimal interactions with the rest of the biological machinery of the cell. We will implement the stochastic model that the authors used to describe their experimental system [37]. In doing so, we will

- implement in a tangible system an example both of the central dogma and of *transcriptional regulation*: the control by proteins of DNA expression into RNA,
- introduce sophisticated Monte Carlo techniques for simulations of stochastic reactions,
- introduce methods for automatically generating continuum descriptions from reaction rates, and
- illustrate the *shot noise* fluctuations due to small numbers of molecules and the *telegraph noise* fluctuations due to finite rates of binding and unbinding of the regulating proteins.

Figure 8.12 shows the biologist's view of the repressilator network. Three proteins (TetR, λCI, and LacI) each repress the formation of the next. We shall see that, under appropriate circumstances, this can lead to spontaneous oscillations; each protein peaks in turn, suppressing the suppressor of its suppressor, leading to its own later decrease.

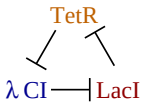

Fig. 8.12 Biology repressilator. The biologist's view of the repressilator network. The T-shapes are blunt arrows, signifying that the protein at the tail (bottom of the T) suppresses the production of the protein at the head. Thus LacI (pronounced lack-eye) suppresses TetR (tet-are), which suppresses λ CI (lambda-see-one). This condensed description summarizes a complex series of interactions (see Fig. 8.13).

The biologist's notation summarizes a much more complex picture. The LacI protein, for example, can bind to one or both of the *transcriptional regulation* or *operator* sites ahead of the gene that codes for the tetR mRNA.[3] When bound, it largely blocks

[1] From *Statistical Mechanics: Entropy, Order Parameters, and Complexity* by James P. Sethna, copyright Oxford University Press, 2007, page 179. A pdf of the text is available at pages.physics.cornell.edu/sethna/StatMech/ (select the picture of the text). Hyperlinks from this exercise into the text will work if the latter PDF is downloaded into the same directory/folder as this PDF.

[2] This exercise draws heavily on Elowitz and Leibler [37]; it and the associated software were developed in collaboration with Christopher Myers.

[3] Messenger RNA (mRNA) codes for proteins. Other forms of RNA can serve as enzymes or parts of the machinery of the cell. Proteins in *E. coli* by convention have the same names as their mRNA, but start with capitals where the mRNA start with small letters.

[4] *RNA polymerase*, the molecular motor responsible for transcribing DNA into RNA, needs to attach to the DNA at a *promoter site*. By binding to the adjacent operator sites, our repressor protein inhibits this attachment and hence partly blocks transcription. The residual transcription is called 'leakiness'.

the translation of DNA into tetR.[4] The level of tetR will gradually decrease as it degrades; hence less TetR protein will be translated from the tetR mRNA. The resulting network of ten reactions is depicted in Fig. 8.13, showing one-third of the total repressilator network. The biologist's shorthand (Fig. 8.12) does not specify the details of how one protein represses the production of the next. The larger diagram, for example, includes two operator sites for the repressor molecule to bind to, leading to three states (P_0, P_1, and P_2) of the promoter region depending upon how many LacI proteins are bound.

You may retrieve a simulation package for the repressilator from the computational exercises portion of the book web site [129].

(a) *Run the simulation for at least 6000 seconds and plot the protein, RNA, and promoter states as a function of time. Notice that*

- *the protein levels do oscillate, as in [37, figure 1(c)],*
- *there are significant noisy-looking fluctuations,*
- *there are many more proteins than RNA.*

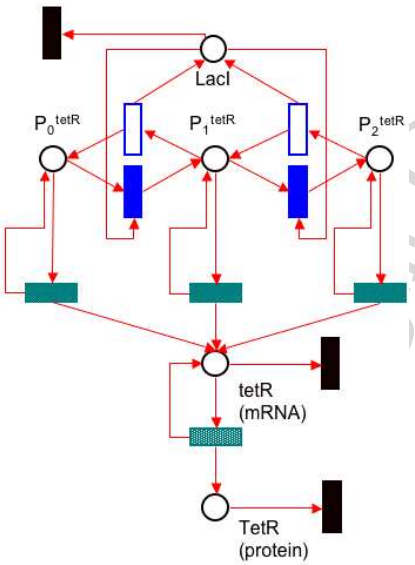

Fig. 8.13 Computational repressilator. The Petri net version [50] of *one-third* of the repressilator network (the LacI repression of TetR). The biologist's shorthand (Fig. 8.12) hides a lot of complexity! We have implemented these equations for you, so studying this figure is optional. The solid lighter vertical rectangles represent binding reactions $A + B \to C$, with rate $k_b[A][B]$; the open vertical rectangles represent unbinding $C \to A + B$, with rate $k_u[C]$. The horizontal rectangles represent catalyzed synthesis reactions $C \to C+P$, with rate $\gamma[C]$; the darker ones represent transcription (formation of mRNA), and the lighter one represent translation (formation of protein). The black vertical rectangles represent degradation reactions, $A \to$ nothing with rate $k_d[A]$. The LacI protein (top) can bind to the DNA in two *promoter sites* ahead of the gene coding for tetR; when bound, it largely blocks the transcription (formation) of tetR mRNA. P_0 represents the promoter without any LacI bound; P_1 represents the promoter with one site blocked, and P_2 represents the doubly-bound promoter. LacI can bind to one or both of the promoter sites, changing P_i to P_{i+1}, or correspondingly unbind. The unbound P_0 state transcribes tetR mRNA quickly, and the bound states transcribe it slowly (leaky repression). The tetR mRNA then catalyzes the formation of the TetR protein.

To see how important the fluctuations are, we should compare the stochastic simulation to the solution of the continuum reaction rate equations (as we did in Exercise 8.10). In [37], the authors write a set of six differential equations giving a continuum version of the stochastic simulation. These equations are simplified; they both 'integrate out' or coarse-grain away the promoter states from the system, deriving a Hill equation (Exercise 6.12) for the mRNA production, and they also rescale their variables in various ways. Rather than typing in their equations and sorting out these rescalings, it is convenient and illuminating to write a routine to generate the continuum differential equations directly from our reaction rates.

(b) *Write a* `DeterministicRepressilator`*, derived from* `Repressilator` *just as* `StochasticRepressilator` *was. Write a routine* `dcdt(c,t)` *that does the following.*

- *Sets the chemical amounts in the reaction network to the values in the array* `c`.
- *Sets a vector* `dcdt` *(of length the number of chemicals) to zero.*
- *For each reaction:*
  - *compute its rate;*
  - *for each chemical whose stoichiometry is*

changed by the reaction, add the stoichiometry change times the rate to the corresponding entry of `dcdt`.

Call a routine to integrate the resulting differential equation (as described in the last part of Exercise 3.12, for example), and compare your results to those of the stochastic simulation.

The stochastic simulation has significant fluctuations away from the continuum equation. Part of these fluctuations are due to the fact that the numbers of proteins and mRNAs are small; in particular, the mRNA numbers are significantly smaller than the protein numbers.

(c) *Write a routine that creates a stochastic repressilator network that multiplies the mRNA concentrations by* `RNAFactor` *without otherwise affecting the continuum equations. (That is, multiply the initial concentrations and the transcription rates by* `RNAFactor`*, and divide the translation rate by* `RNAFactor`*.) Try boosting the* `RNAFactor` *by ten and one hundred. Do the RNA and protein fluctuations become significantly smaller?* This noise, due to the discrete, integer values of chemicals in the cell, is analogous to the *shot noise* seen in electrical circuits due to the discrete quantum of electric charge. It scales, as do most fluctuations, as the square root of the number of molecules.

A continuum description of the binding of the proteins to the operator sites on the DNA seems particularly dubious; a variable that must be zero or one is replaced by a continuous evolution between these extremes. (Such noise in other contexts is called *telegraph noise*—in analogy to the telegraph, which is either silent or sending as the operator taps the key.) The continuum description is accurate in the limit where the binding and unbinding rates are fast compared to all of the other changes in the system; the protein and mRNA variations then see the average, local equilibrium concentration. On the other hand, if the rates are slow compared to the response of the mRNA and protein, the latter can have a switching appearance.

(d) *Incorporate a* `telegraphFactor` *into your stochastic repressilator routine, that multiplies the binding and unbinding rates. Run for 1000 seconds with* `RNAFactor` $= 10$ *(to suppress the shot noise) and* `telegraphFactor` $= 0.001$. *Do you observe features in the mRNA curves that appear to switch as the relevant proteins unbind and bind?*