7

Epigenetics & Genetic Regulation in Cis

7.1 Nucleosomes and their enzymes

7.2 Nucleosome mediated epigenetics

7.2.1 *Simulate the A–U–M model from Fig. 7.5 without cell divisions, but with the modification that recruitments $U \to A$ and $U \to M$ are local, whereas the other recruitments are global. Use an $L = 60$ nucleosome system, ignore cell divisions and select a feedback strength that gives reasonable bistability. Examine behavior in a space-time plot of nucleosome states as a function of position and time measured in updates per nucleosome.*

Answer There is one parameter $\alpha \in [0; 1]$ with $F = \alpha/(1-\alpha)$. At each step select a random nucleosome $n_1 \in [1, L]$:

- A recruited conversion is attempted with probability α. Given this choice one selects, with probability $\beta = 0.5$, whether the recruitment should be local or global. **If global:** a second random nucleosome $n_2 \in [1, L]$ is selected ($n_2 \neq n_1$). If n_2 is in either the M or the A state, nucleosome n_1 is changed to U if and only if n_1 is in the opposite state of n_2. **If local:** If n_1 is in U state a second random nucleosome $n_2 = n_1 - 1$ or $n_2 = n_1 + 1$ is selected. If n_2 is either in the M or the A state, nucleosome n_1 is changed to the state of n_2.

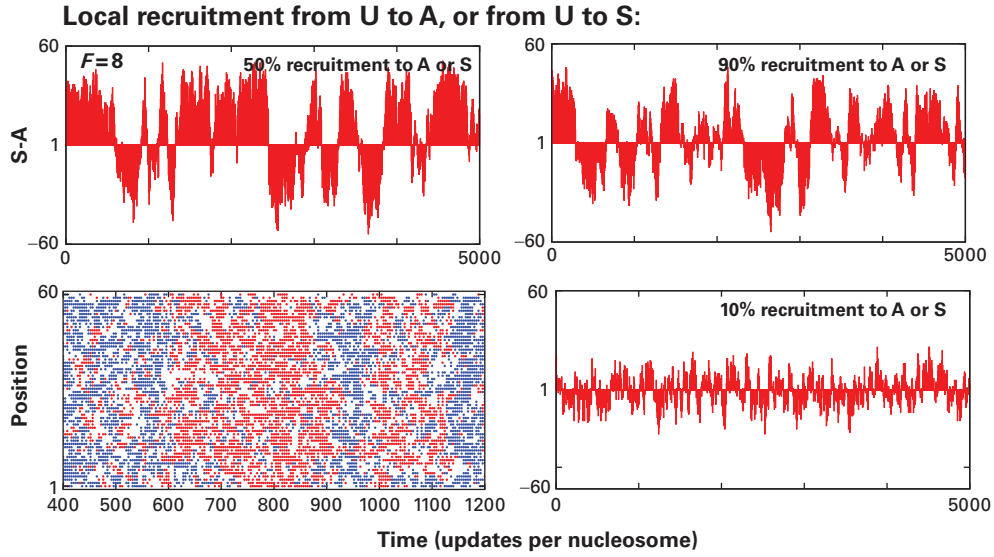

Figure 7.1 Nucleosome-mediated epigenetics, with mixed local and global recruitment. The value F sets overall recruitment to noise. Given the recruitment, one may still choose which of the two recruitment processes should be strongest (see right-hand panels).

- A noisy change is attempted with probability $1-\alpha$: a nucleosome is selected and converted to one of its neighbor states along the A–U–M axis.

Simulations for $F = 8$ are seen in Fig. 7.1, where we also explore variations in β in the relative strength β of the two recruitment processes. Notice that bistability now requires substantially larger F, reflecting the fact that we lose about $1/2$ the recruitment attempts by selecting the type of recruitment before selecting whether it actually is feasible.

7.2.2 *Simulate the A–U–M model above, but with the modification that recruitments $A \to U$ and $M \to U$ are local, whereas the other recruitments are global. Use the same overall approach as above.*

Answer Repeat the procedure from the question above, but switching the global and local recruitments, such that the global recruitment now makes $U \to M$, $U \to A$. Simulations for $F = 8$ is seen in Fig. 7.2.

7.2.3 *Repeat the above simulations, but now where global recruitments are acting across a distance x with probability $1/x$.*

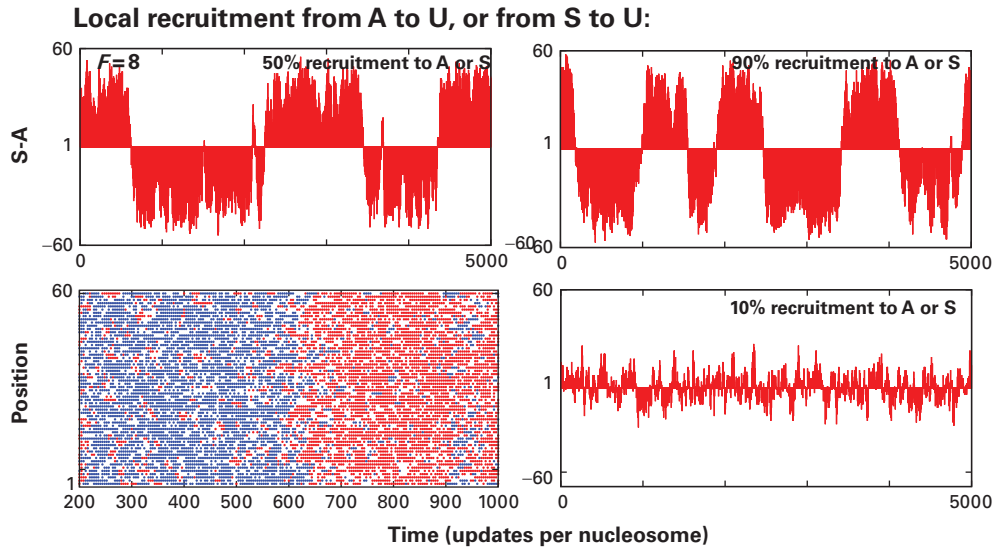

Figure 7.2 Nucleosome-mediated epigenetics, with mixed local and global recruitment. The value F sets overall recruitment to noise. Given recruitment one may still choose which of the two recruitment processes should be strongest (see right-hand panels).

Answer The algorithm is as for two previous questions, except that one global recruitment is replaced by selecting two nucleosomes in $[1, N]$ at a distance x with probability $1/x$. In practice, given the first position n_i, the distance x is chosen as $x = 60^{\text{random}}$ with random $\in [0, 1]$. Position n_j is then either $n_j = n_i + x$ or $n_j = n_i - x$. If $n_j < 1$ $n_j = n_i$ or $n_j > N$ is selected a new n_j is selected. The simulations for the two models are shown in Fig. 7.3.

7.2.4 *Reproduce the bifurcation plot in Fig. 7.7 by plotting fixed points for the model defined by Eqs. (7.1) and (7.2) as a function of F. Show analytically that bistability is only possible when $F > \frac{1}{\sqrt{2}-1} = 2.4142$.*

Answer Given $F = \alpha/(1-\alpha)$ and $\alpha = 1/(1+F)$, the steady-state values of the two equations give us two equations with two unknowns, a and m:

$$f \sim a \cdot (1-a) - 2a \cdot m + \frac{1-3a-m}{2F} = 0$$
$$g \sim m \cdot (1-m) - 2a \cdot m + \frac{1-a-3m}{2F} = 0$$

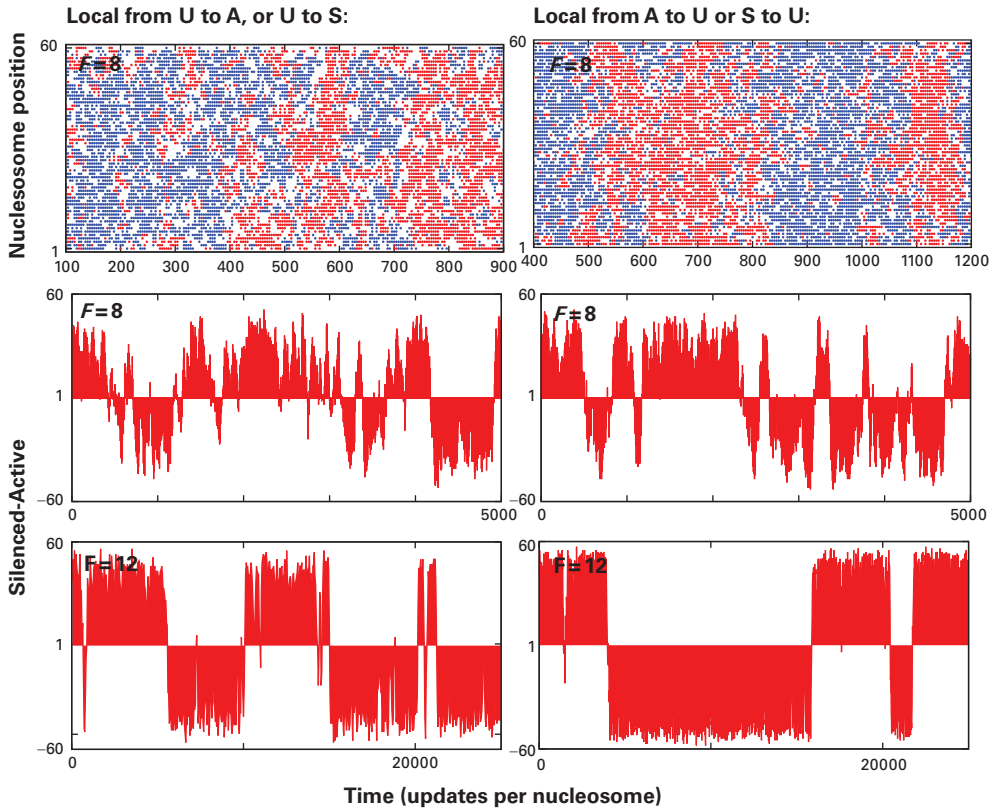

Figure 7.3 Nucleosome-mediated epigenetics, with mixed local and global recruitment, and global recruitment selected at distance x with probability $1/x$.

giving

$$f - g = (a - m)\left(1 - (a + m) - \frac{1}{F}\right) = 0$$

which is equal to zero when $a = m$ or when $a+m = 1-1/F$. These two types of solution can be inserted into the $f = 0$ equation. The $-m = a - 1 + 1/F$ solution gives

$$\begin{aligned} f &\sim a - a^2 - 2a \cdot \left(1 - a - \frac{1}{F}\right) + \frac{1}{2F}\left(-2a + \frac{1}{F}\right) \\ &= -a + a^2 + \frac{2a}{F} - \frac{a}{F} + \frac{1}{2F^2} = \end{aligned}$$

$$f \sim -\left(1-\frac{1}{F}\right)a + a^2 + \frac{1}{2F^2} = 0 \Rightarrow$$
$$a = \frac{1}{2}\left(\left(1-\frac{1}{F}\right) \pm \sqrt{1-\frac{2}{F}-\frac{1}{F^2}}\right)$$

expressing that, for $F \to \infty$, the fixed point for a approaches respectively 1 or 0. When a takes one of these solutions for finite F, then the methylated fraction m takes the other, and the unmodified nucleosome fraction:

$$u = 1 - m - a = 1 - \left(1-\frac{1}{F}\right) = \frac{1}{F}$$

The two solutions for the fixed point are possible when:

$$1 - \frac{2}{F^2} - \frac{1}{F} > 0 \Rightarrow F > \frac{1}{\sqrt{2}-1} = 2.4142$$

which is thus the critical minimal F for obtaining bistability.

The $m = a$ solution:

$$f \sim -3a^2 + \left(1-\frac{2}{F}\right)a + \frac{1}{2F}$$
$$f = 0 \Rightarrow a^2 - \frac{1}{3}\left(1-\frac{2}{F}\right)a - \frac{1}{6F} = 0 \Rightarrow$$
$$a = \frac{1}{6}\left(1-\frac{2}{F} \pm \sqrt{(1-\frac{2}{F})^2 + \frac{6}{F}}\right)$$

where only the largest solution is above 0. These equation have solutions $a = m = u \sim 1/3$ for $F \to \infty$. The solutions are shown in Fig. 7.4.

7.2.5 *The linear stability analysis of a set of equations* $\mathrm{d}x/\mathrm{d}t = f$ *and* $\mathrm{d}y/\mathrm{d}t = g$ *consider the dynamics of a small deviation* $(\delta x, \delta y)$*, around a fixed point* (x_0, y_0)*:*

$$\frac{\mathrm{d}(\delta x)}{\mathrm{d}t} = f(x,y) - f(x_0,y_0) = \frac{\mathrm{d}f}{\mathrm{d}x}\delta x + \frac{\mathrm{d}f}{\mathrm{d}y}\delta y = f_x\delta x + f_y\delta y$$
$$\frac{d(\delta y)}{\mathrm{d}t} = g(x,y) - g(x_0,y_0) = \frac{\mathrm{d}g}{\mathrm{d}x}\delta x + \frac{\mathrm{d}g}{\mathrm{d}y}\delta y = g_x\delta x + g_y\delta y$$

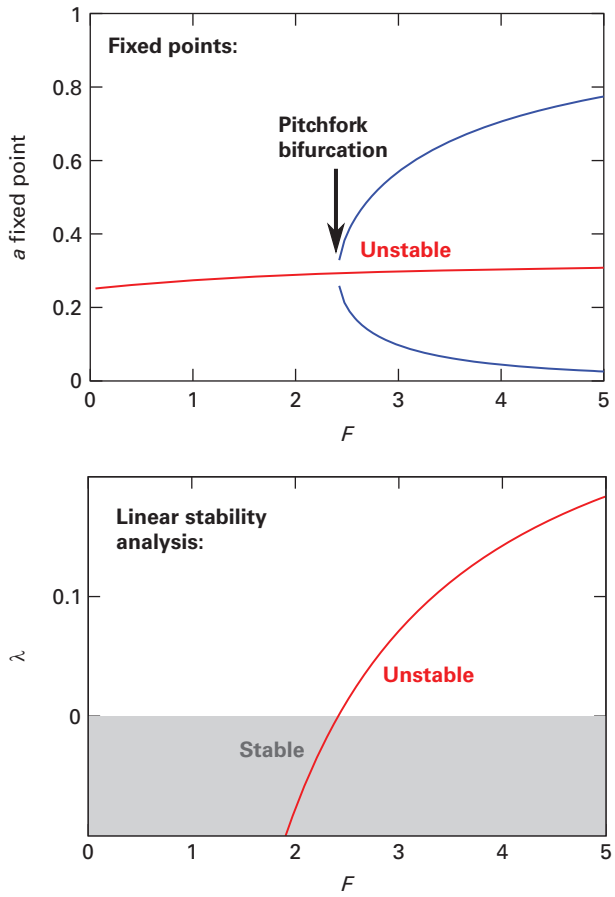

Figure 7.4 Upper panel: fixed points as a function of F for the standard nucleosome model from the text. Lower panel. The largest eigenvalue for the m = a solution using linear stability analysis (from Question 8.2.5).

The stability of a fixed point is given by the sign of the largest eigenvalue:

$$\lambda = \frac{f_x + g_y}{2} + \sqrt{\frac{(f_x + g_y)^2}{4} - f_x g_y + f_y g_x}$$

Insert Eqs. (7.1) and (7.2) into this equation and plot the largest eigenvalue as function of F for the $m = a$ fixed-point solution.

Answer In general, the eigenvalues for the matrix that describes the dynamics of a small deviation $(\delta x, \delta y)$, around fixed point (x_0, y_0):

$$\frac{\mathrm{d}(\delta x)}{\mathrm{d}t} = f(x,y) - f(x_0,y_0) = \frac{\mathrm{d}f}{\mathrm{d}x}\delta x + \frac{\mathrm{d}f}{\mathrm{d}y}\delta y = f_x\delta x + f_y\delta y$$
$$\frac{\mathrm{d}(\delta y)}{\mathrm{d}t} = g(x,y) - g(x_0,y_0) = \frac{\mathrm{d}g}{\mathrm{d}x}\delta x + \frac{\mathrm{d}g}{\mathrm{d}y}\delta y = g_x\delta x + g_y\delta y$$

where the last step uses the fixed point condition $f(x_0,y_0) = g(x_0,y_0) = 0$. With $\mathcal{M}$ as the matrix for linear variations, the eigenvalues λ for linear stability are given by:

$$\frac{\mathrm{d}\delta\mathbf{r}}{\mathrm{d}t} = \mathcal{M}\delta\mathbf{r} = \lambda\delta\mathbf{r}$$

Solving this:

$$\begin{aligned}
\det(\mathcal{M} - \lambda\cdot\mathbf{1}) &= 0 \\
(f_x-\lambda)\cdot(g_y-\lambda) - g_xf_y &= 0 \Rightarrow \\
\lambda^2 - (f_x+g_y)\cdot\lambda + f_xg_y - g_xf_y &= 0 \Rightarrow \\
\lambda &= \frac{f_x+g_y}{2} \pm \sqrt{\frac{(f_x+g_y)^2}{4} - f_xg_y + f_yg_x}
\end{aligned}$$

For our model:

$$\begin{aligned}
f &\sim a\cdot(1-a) - 2a\cdot m + \frac{1-3a-m}{2F} = 0 \\
g &\sim m\cdot(1-m) - 2a\cdot m + \frac{1-a-3m}{2F} = 0
\end{aligned}$$

Giving:

$$\begin{aligned}
f_a &\sim 1 - 2a - 2m - \frac{3}{2F} \\
f_m &\sim -2a - \frac{1}{2F} \\
g_a &\sim -2m - \frac{1}{2F} \\
g_m &\sim 1 - 2m - 2a - \frac{3}{2F}
\end{aligned}$$

giving the largest eigenvalue:

$$\begin{aligned}
\lambda &= \frac{f_a+g_m}{2} \pm \sqrt{\frac{(f_a+g_m)^2}{4} - f_ag_m + f_mg_a} \\
&= 1 - 2a - 2m - \frac{3}{2F}
\end{aligned}$$

$$
\begin{aligned}
&+\sqrt{\left(1-2a-2m-\frac{3}{2F}\right)^2-\left(1-2a-2m-\frac{3}{2F}\right)^2+\left(2a+\frac{1}{2F}\right)\cdot\left(2m+\frac{1}{2F}\right)}\\
&=1-2a-2m-\frac{3}{2F}+\sqrt{\left(2a+\frac{1}{2F}\right)\cdot\left(2m+\frac{1}{2F}\right)}\\
&=1-4a-\frac{3}{2F}+2a+\frac{1}{2F}=1-2a-\frac{1}{F}\\
&=1-\frac{1}{F}-\frac{1}{3}\left(1-\frac{2}{F}+\sqrt{\left(1-\frac{2}{F}\right)^2+\frac{6}{F}}\right)\\
&=\frac{2}{3}-\frac{1}{3F}-\frac{1}{3}\sqrt{\left(1-\frac{2}{F}\right)^2+\frac{6}{F}}
\end{aligned}
$$

which is plotted in the lower panel of Fig. 7.4. The $m = a$ solution is unstable when $\lambda > 0$. Notice that the $m = a$ solution becomes unstable when F exceeds 2.4.

7.3 A Regulated 2-State Model

7.3.1 *Simulate the two-state model for different F values for an $L = 30$ system.*

Answer Consider the $L = 30$ system, where each position can be 0 or 1. Define $\alpha = 1/(F + 1)$. Algorithm reads: choose random number $r \in [0, 1]$ uniformly. If $r < \alpha$ select two nucleosomes. If in the different state, nothing is done. If in same state, a third nucleosome is selected. This nucleosome is then set to the majority state. If $r > \alpha$, a nucleosome at a random position is switching state, $s(i) = 1 - s(i)$. Result is shown in Fig. 7.5 for various F values, pinpointing $F = 4$ as the critical F for this model.

7.3.2 *Discuss the two-state model in terms of movements in an epigenetic landscape of the form:*

$$
\begin{aligned}
V &\sim -\int \langle \mathrm{d}m/\mathrm{d}t\rangle \mathrm{d}m\\
&\propto \int \left(m(1-m)-\frac{1}{F}\right)(1-2m)\mathrm{d}m\\
&= m(1-m)\cdot\left(\frac{1}{2}m(1-m)-\frac{1}{F}\right)
\end{aligned}
\tag{7.1}
$$

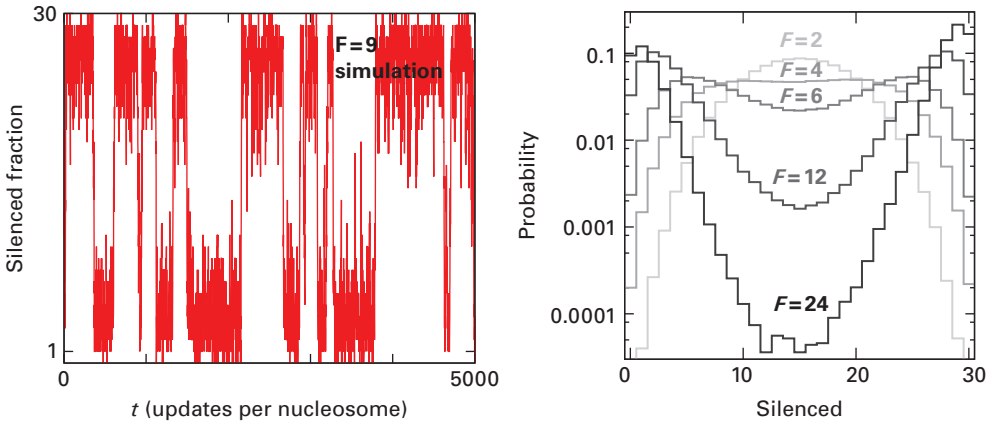

Figure 7.5 Two-state model simulated for various ratios of feedback to passive conversions.

Plot the potential, and discuss how residence time in one of the "potential wells" depends on the system size L.

Answer The assumed potential in the question would be correct if the diffusion and mobility of the m-particle was independent of its position. We therefore compare the real system simulation with this approximate potential. Results shown in the left-hand panels of Fig. 7.6. A more careful calculation that takes into account that the diffusion depends on position $m \in [0,1]$ gives a potential:

$$\frac{V}{L} = 2 \cdot m \cdot (1-m) + \left(\frac{1}{L} - \frac{4}{F}\right) \ln(Fm(1-m)+1)$$

$$D = \frac{1}{2L^2} \cdot \left(\frac{m(1-m)\cdot F}{F+1} + \frac{1}{F+1}\right)$$

where the last equation expresses the diffusion constant. D has a maximum at $m = 1/2$, reflecting the fact that dynamics is most active when passing the barrier. Notice that the full potential above is more pronounced than the approximate potential used in the question, although the two potentials are qualitatively similar. For more details on how to derive the real potential see Micheelsen et al., q-bio 1002.1600v2, 23 Sep. 2010.

In any case, the potential scale with N or the diffusion constant decreases as $1/L$, and therefore the barrier-passing time:

$$\tau \sim \exp(\text{constant} \cdot L) \tag{7.2}$$

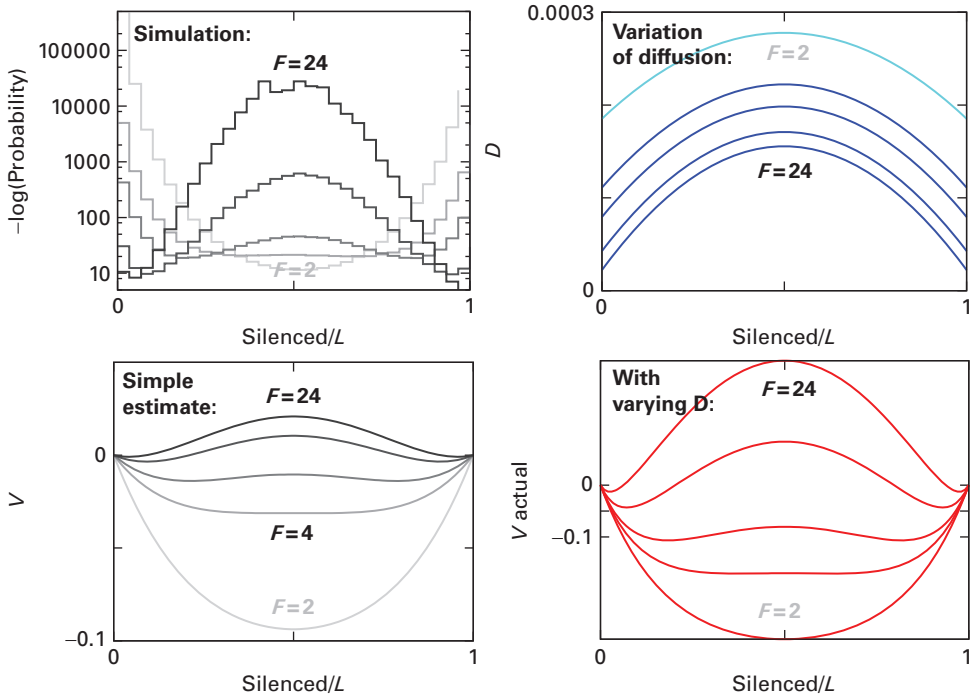

Figure 7.6 Two-state model: Effective potential, and potential associated to deterministic part of simulation (V).

and thus a doubling of system size should leads to a lifetime raised to the power 2 (in units of natural timescale).

7.3.3 *Implement and simulate gene regulation in a two-state version of the nucleosome model where the co-operative recruitment only acts in one direction. Thus the co-operative recruitment acts by converting S to A, when two random nucleosomes are in the A state, and the effect of the transcription factor is to convert the A state to S with some adjustable rate R.*

Answer Consider the $L = 30$ system, where each position can be 0 or 1. Define $\alpha = F/(F+1)$, $R \in [0,1]$ as representative of regulator concentration. For illustration we here provide the algorithm in a Gillespie version: Select three times:

(1) Set $t_1 = -\ln(\text{ran})/\alpha$
(2) Set $t_2 = -\ln(\text{ran})/(1-\alpha)$
(3) set $t_3 = -\ln(\text{ran})/R$

select the process with minimal t_i. If this is process $i = 1$ then select two different nucleosomes and check whether both are in state 1. If so then select a third nucleosome and set this to state 1.

If the chosen process is $i = 2$, a nucleosome at a random position k switches state, $s(k) = 1 - s(k)$.

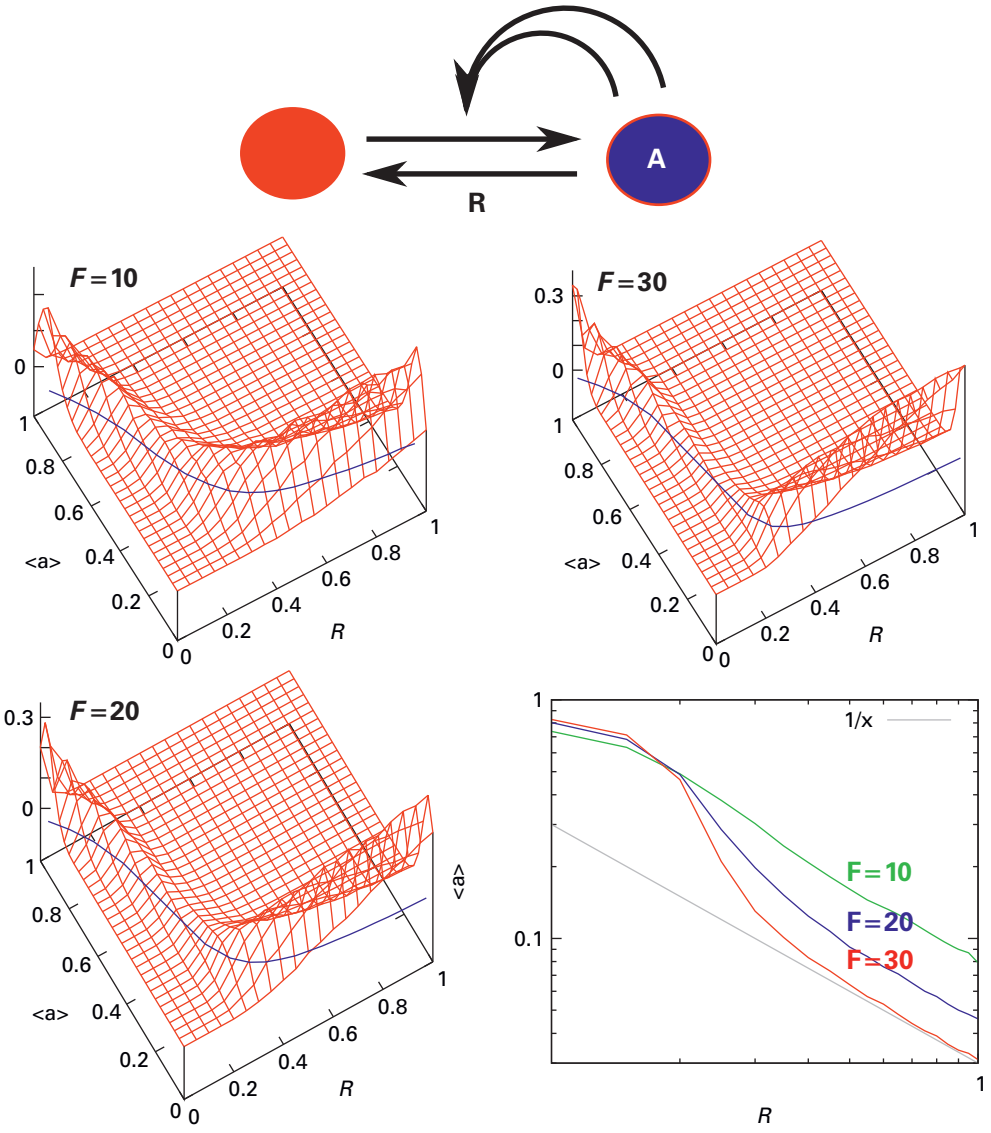

Figure 7.7 Two-state model with regulator, simulated for various values of feedback F and concentrations of regulator.

If the chosen process is $i = 3$ then set a random nucleosome to state = 0.

Notice that the Gillespie-like formulation allows for easy extension to cases where some rates exceed 1.

Figure 7.7 shows distribution of nucleosomes in state 1 as a function of strength of regulator. Notice that at a critical value of R, the system is bistable. Further, as one varies R across this value, the system is ultra-sensitive. The ultra-sensitivity is less pronounced than in cases where there is recruitment in both directions, as such recruitment represses leakiness.

7.3.4 *Consider a two-state nucleosome model with co-operative recruitment from state S to state A and with passive conversion of nucleosome in the opposite direction. Assume direct transitions β per active recruitment attempt, and that directed drift (toward "y") has strength drift per active recruitment attempt. Derive equations for the model and plot regions in x, y where* $\mathrm{d}x/\mathrm{d}t > 0$ *and* $\mathrm{d}y/\mathrm{d}t > 0$*. Verify that there are parameters where the model is bistable, i.e. where there are two stable fixed points separated by an unstable one.*

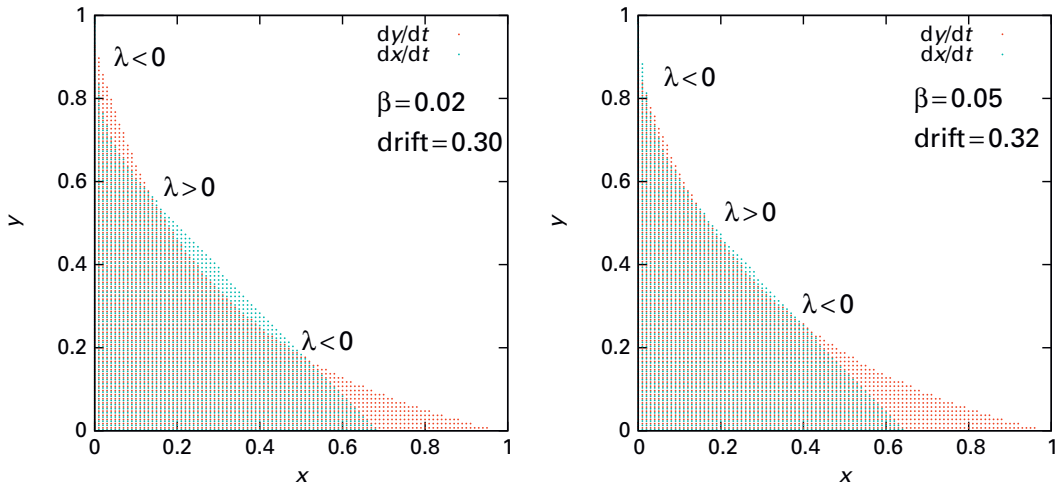

Figure 7.8 Model with recruitment in one direction, but only a drift in the other, corresponding to enzymes acting from the bulk. The region with $dx/dt > 0$ is shown in cyan shading, the region with $dy/dt > 0$ is shown with red shading. Sign of eigenvalues for fixed points are indicated. See question for the equations.

Answer

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\beta}{2}(1-y) - \frac{3\beta}{2}x + (1-y-\mathrm{drift})\cdot x - x^2$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\beta}{2}(1-x) - \frac{3\beta}{2}y + (1-x-y)\cdot \mathrm{drift} - x\cdot y$$

Choosing $drift = 0.30$ and $\beta = 0.02$ indeed gives three fixed points (see Fig. 7.8).

7.4 Coupled epigenetics in olfactoric differentiation

7.4.1 *Consider the model for olfactoric differentiation from Fig. 7.19, with nucleosomes in the silenced state (red in panel C), being partially protected from conversion to the a state (blue) by a protein P, produced by all active genes. Argue for negative feedback through P that reduces the rate of recruitment from S to A:*

$$\mathrm{Rate(S \to A)} \to \mathrm{Rate(S \to A)} \cdot \frac{1}{1 + r \cdot P} \tag{7.3}$$

where P is the total activity of all olfactory genes.

Answer The negative feedback between olfactoric genes is assumed to act through any protein P produced by these genes. This interaction is here assumed to act through conversion of the silenced nucleosomes to a form that cannot be changed by enzymes that makes nucleosomes "active." One way to do this is to let P act catalytically on s:

$$\begin{aligned} s + P \to s^* + P \ \text{ whereas } \ s^* \to s \ \text{ with rate } \ r' \\ \Rightarrow (s - s^*)\cdot P = r' \cdot s^* \Rightarrow \\ s^* = \frac{s\cdot P}{r' + P} \Rightarrow \\ s - s^* = \frac{s\cdot r'}{r' + P} = \frac{s}{1 + r''\cdot P} \end{aligned}$$

where $s - s^*$ is the fraction s_{free} that can be converted to a.

In the above, the critical number is the protein P, which in turn could be produced from any olfactoric gene. If we assume that the activity is thresholded by its fraction of nucleosome in active state a, then:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \text{const} \cdot \sum_j a_j^h - \frac{P}{\tau}$$

where the Hill coefficient characterizes the threshold function. If, for example, $h = 1$, we simply set the activity of the gene proportional to the likelihood that nucleosomes on promoter sites are in an active state, the assumption used in our model for gene activity. The dynamics equation for P can be simply integrated to a realistic steady state for τ:

$$P = \text{const} \cdot \tau \cdot \sum_j a_j^h \propto \sum_j a_j^h$$

a realistic approximation when τ is short compared to other timescales in the dynamics of individual genes.

7.4.2 *Assume that each gene has an activity proportional to the fraction of active nucleosomes of that gene to an exponent h, and use this to find an expression for the amount of a protein P that is produced due to the activity of all genes. Combine this expression with the results from Question 7.4.1 to argue for the following equation for differentiation between the $i = 1, 2, \ldots N$ genes from Fig. 7.19:*

$$\begin{aligned}\frac{\mathrm{d}a_i}{\mathrm{d}t} = &\frac{1}{1 + r \cdot \sum_j a_j^h} \cdot a_i^2 \cdot (1 - a_i) - \mu \cdot a_i \cdot (1 - a_i)^2 \\ &+ \frac{1}{1 + r \cdot \sum_j a_j^h} \cdot \beta \cdot (1 - a_i) - \beta \cdot a_i \end{aligned} \tag{7.4}$$

Here $\mu < 1$ describes an inherent bias, giving a weaker basic recruitment from the silenced state. In contrast, recruitment from active nucleosomes is reduced due to influence from other genes. β represents passive conversions. Analyze the model for $h = 4$, $r = 3$, $\mu = 1$ and $\beta = 0.03$.

Answer Combining the equations, the fraction of silenced nucleosomes that can be converted to a is:

$$s \to s_{\text{free}} = \frac{s}{1 + r \cdot \sum_j a_j^h}$$

with some constant $r = \text{const} \cdot \tau \cdot r''$.

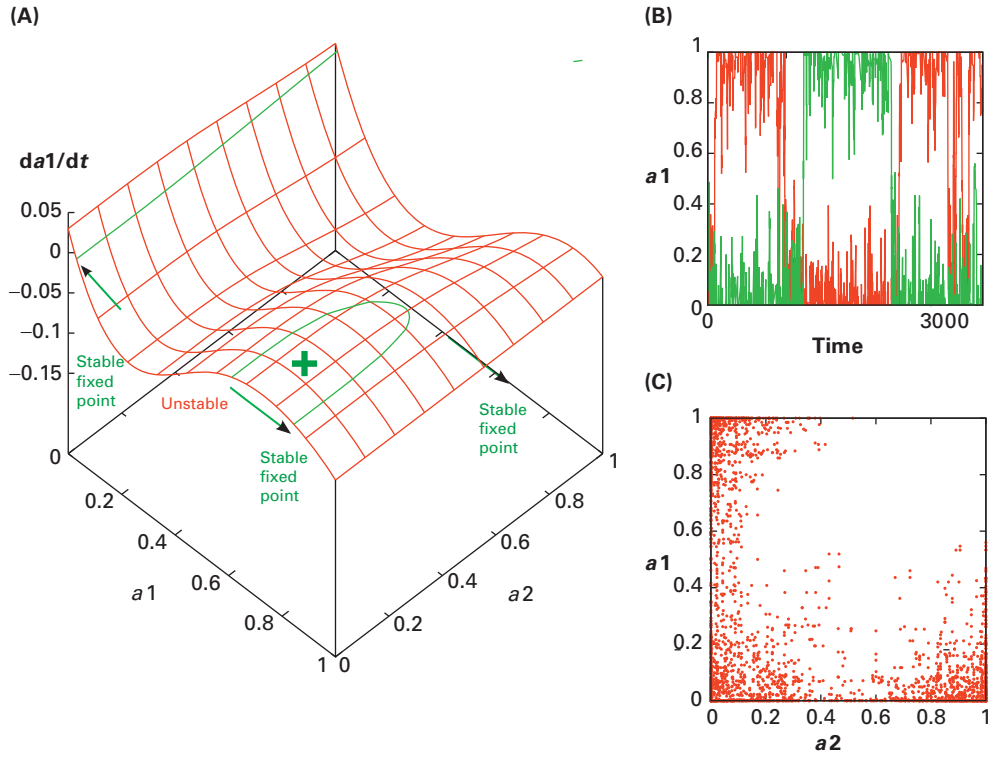

Figure 7.9 Two-gene model for olfactoric differentiation.

The recruitment acting from a on s is then:

$$\text{rate}(s \to a) = a^2 \cdot (1-a) \cdot s_{\text{free}} = a^2 \cdot \frac{1-a}{1 + r \cdot \sum_j a_j^h}$$

whereas the recruitment in opposing directions is:

$$\text{rate}(a \to s) = \mu \cdot (1-a)^2 \cdot a$$

Adding the small passive drift terms to this recruitment rate leads to the deterministic equations:

$$\begin{aligned}\frac{\mathrm{d}a_i}{\mathrm{d}t} &= \frac{1}{1 + r \cdot \sum_j a_j^h} \cdot a_i^2 \cdot (1-a_i) - \mu \cdot a_i \cdot (1-a_i)^2 \\ &+ \frac{1}{1 + r \cdot \sum_j a_j^h} \cdot \beta \cdot (1-a_i) - \beta \cdot a_i\end{aligned}$$

where a_i is the fraction of active nucleosomes in gene i, and $\mu < 1$ is the bias that favors the active state of all genes.

Figure 7.9 analyzes the $N = 2$ case for the model:

$$\frac{\mathrm{d}a_i}{\mathrm{d}t} = \frac{1}{1 + 3 \cdot \sum_j a_j^4} \cdot a_i^2 \cdot (1 - a_i) - \cdot a_i \cdot (1 - a_i)^2 + \frac{1}{1 + 3 \cdot \sum_j a_j^4} \cdot 0.03 \cdot (1 - a_i) - 0.03 \cdot a_i$$

This model is indeed bistable, selecting one and only one of the genes as the active one. In Fig. 7.9 panel (A) it can thus be seen that $\mathrm{d}a_1/\mathrm{d}t$ is positive for high a_1 values, allowing for a stable fixed point, provided that a_2 is small. However, if a_2 is big then there is a stable fixed point for a_1 that is larger than 0.1.

7.4.3 *Simulate the two-gene version of the equations from Question 7.4.2, using a Gillespie algorithm on the four different terms in the equations (making a total of eight rates for the two genes). Use an update where each of the state fractions is changed in steps of size 0.1.*

Answer Assign eight discrete processes, with rates:

$$\begin{aligned}
\mathrm{rate}(i,1) &= \frac{1}{1 + 3 \cdot \sum_j a_j^4} \cdot a_i^2 \cdot (1 - a_i) \\
\mathrm{rate}(i,2) &= a_i \cdot (1 - a_i)^2 \\
\mathrm{rate}(i,3) &= \frac{1}{1 + 3 \cdot \sum_j a_j^4} \cdot 0.03 \cdot (1 - a_i) \\
\mathrm{rate}(i,4) &= 0.03 \cdot a_i
\end{aligned}$$

and assign interaction times:

$$\mathrm{time}(i,k) = -\delta \cdot \ln(\mathrm{random})/\mathrm{rate}(i,k)$$

where random is a random number drawn uniformly between 0 and 1. Select minimal time, and change $a_i \to a_i + \delta$ if $k = 1, 3$ and assign $a_i \to a_i - \delta$ if $k = 2, 4$. Use $\delta = 0.1$. For panel C, we use variable values of step size $-\delta \cdot \ln(\mathrm{random})$ in panel C. The simulation is shown in Fig. 7.9 B,C).