5

Diffusion and randomness in transcription

5.1 Random walks and Diffusion

5.1.1 *What is the timescale for homogenizing protein concentrations in a eukaryotic cell with diameter* $20\,\mu m$*. Assume that* $D = 5\,\mu m^2\ s^{-1}$*, or* $D = 0.2\,\mu m^2\ s^{-1}$*.*

Answer A protein initially at point (0,0,0) will at time t be distributed accordingly to a Gaussian distribution with spread $\sigma = \sqrt{2Dt}$. When this becomes comparable to the cell radius, the distribution will be homogenized.

Thus $\sqrt{2Dt} = 10\,\mu\text{m}$, giving $t = 100\,\mu\text{m}^2/2D = 10$ s when using a diffusion constant $D = 5\,\mu\text{m}^2\ \text{s}^{-1}$.

Similarly, $\sqrt{2Dt} = 10\,\mu\text{m}$ giving $t = 100\,\mu\text{m}^2/2D = 250$ s when using a diffusion constant $D = 0.2\,\mu\text{m}^2\ \text{s}^{-1}$.

5.1.2 *What time does it take to homogenize protein concentrations in a neuron by use of diffusion, connecting the brain with tissues that are 1 m away? Assume that* $D = 5\,\mu\text{m}^2\ \text{s}^{-1}$*.*

Answer A protein initially at point (0,0,0) will, at time t, be distributed accordingly to a Gaussian with spreads $\sigma = \sqrt{2Dt}$. When starting the protein at one end of a neuron it similarly spreads in a Gaussian fashion in the positive half plane. The system is homogenized when this becomes comparable to the neuron length. Thus $\sqrt{2Dt} = 1\,\text{m} = 10^6\,\mu\text{m}$, giving $t = 10^{12}\,\mu\text{m}^2/(2 \cdot 5\,\mu\text{m}^2\ \text{s}^{-1}) = 10^{11}$ s $\sim$ 3000 years. Therefore transport

in neurons must be directed; in practice this is done by molecular motors moving by speeds of the order of μm s^{-1} along microtubules.

5.2 Timescales for target location in a cell

5.2.1 *Follow and supplement the formal proof below for diffusion-limited target location as illustrated in Fig. 5.2. In a steady state the current J is independent of r. Use Eq. (5.13) to argue:*

$$\rho(r) \;=\; \frac{J}{D4\pi r} + \rho(\infty)$$

Using the concentration at infinite $\rho(\infty) = N/V$ and $\rho(\epsilon) = 0$ (why are these sensible limits?) one may obtain:

$$\frac{-J}{D4\pi\epsilon} \;=\; \frac{N}{V} \Rightarrow |J| \;=\; 4\,\pi\,\epsilon\,D\,\frac{N}{V} \Rightarrow \tau_{\rm on} \;=\; \frac{1}{|J|} \;=\; \frac{V}{4\pi D\epsilon N}$$

Answer In spherical coordinates:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} \;=\; \frac{1}{r^2} D \frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \frac{\mathrm{d}\rho}{\mathrm{d}r}\right)$$

The current across a spherical shell at radius r is:

$$J \;=\; -D\;4\pi\;r^2\;\frac{\mathrm{d}\rho}{\mathrm{d}r}$$

allowing us to express the diffusion equation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} \;=\; -\;\frac{1}{4\pi r^2} D \frac{\mathrm{d}}{\mathrm{d}r}\,(J)$$

In steady state, the current J is independent of r:

$$\frac{1}{4\pi r^2} D \frac{\mathrm{d}}{\mathrm{d}r} J = 0 \Rightarrow J = \text{constant}$$

From the above equations:

$$\rho(r) \;=\; \frac{J}{D4\pi r} + \rho(\infty)$$

Using that the concentration at infinite $\rho(\infty) = N/V$ and that particles that reach $r = \epsilon$ are instantly absorbed, i.e. $\rho(\epsilon) = 0$, one obtains:

$$\frac{-J}{D4\pi\epsilon} = \frac{N}{V} \Rightarrow |J| = 4\,\pi\,\epsilon\,D\,\frac{N}{V} \Rightarrow \tau_{\rm on} = \frac{1}{|J|} = \frac{V}{4\pi D\epsilon N}$$

which expresses the number of particles that are absorbed on the target per time unit.

5.2.2 *Derive the Smoluchowsky equation in two dimensions (use the two-dimensional analog $J = -2\pi r \mathrm{d}\rho/\mathrm{d}r$ with $\rho(\infty) = N/A$ counting particles per area).*

Answer Again using that the current is independent of distance r from the target:

$$\begin{aligned} J &= -2\pi r D\frac{\mathrm{d}\rho}{\mathrm{d}r} \Rightarrow \\ \frac{\mathrm{d}\rho}{\mathrm{d}r} &= -\frac{J}{2\pi r D} \Rightarrow \\ \rho(R) &= -\int_{\epsilon}^{R} \frac{J\mathrm{d}r}{2\pi D r} = -\frac{J}{2\pi}\cdot\log\left(\frac{R}{D\epsilon}\right) \end{aligned}$$

where J is negative, since it is directed inwards. For a system with area A and radius R, given by $A = 2\pi R^2$, a current from the outskirts of this system with density $\rho \sim N/A$ is given by:

$$\begin{aligned} \frac{N}{\pi R^2 D} &= \frac{-J}{2\pi}\log(R/\epsilon) \Rightarrow \\ (-J) &= \frac{2\pi\cdot D\cdot(N/A)}{\log(R/\epsilon)} \end{aligned}$$

expressing a remarkably slow decay with the size ϵ of the target. Thus the time to find the target only increases logarithmically with decreasing target size:

$$\tau = \frac{\log(R/\epsilon)}{2\pi\cdot D\cdot(N/A)}$$

5.2.3 *Plot $\tau_{\rm off}$ as a function of binding energy ΔG, and find ΔG where $\tau_{\rm off}$ equals the E. coli cell generation time (say, 1 hour).*

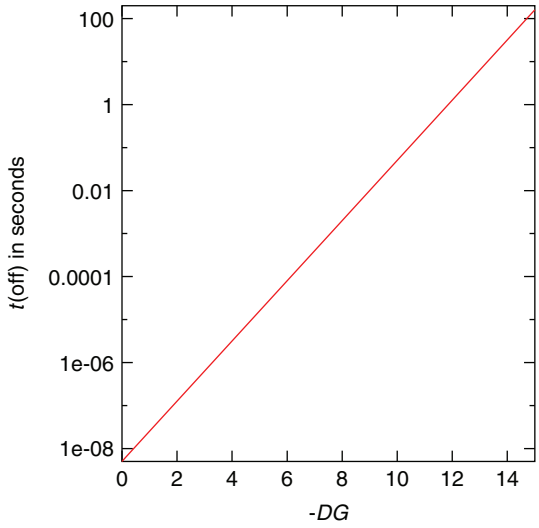

Figure 5.1 Off rates from a binding site as a function of the binding energy in kcal mol^{-1} ($k_{\mathrm{B}}T = 0.62$).

Answer As stated in the main text, for a target $\epsilon = 5\,\mathrm{nm}$:

$$\tau_{\mathrm{off}} \sim \frac{(10^{-3}\,\mathrm{m}^3)}{4\pi \cdot 5\,\mu\mathrm{m}^2 \cdot 0.005\,\mu\mathrm{m} \cdot\ 6 \cdot 10^{23}} \exp\left(\frac{-\Delta G}{k_{\mathrm{B}}T}\right)$$
$$\sim 5 \cdot 10^{-9}\mathrm{s} \exp\left(\frac{-\Delta G}{k_{\mathrm{B}}T}\right)$$

where the prefactor accordingly corresponds to one escape attempt every 5 ns. See Fig. 5.1.

5.2.4 *Simulate a random walker on a lattice, confined to a box represented by integer positions between* -5 *and* 5 *(included), in all six directions: i.e. the particle at (x,y,z) can move one step in any one of six directions, except when this move brings it outside the box. Compare the time to find point (0,0,0), with the time to find any point in the cube* $[-1:1]^3$ *and also to find the (0,0,0) point in* $[-2:2]^2$*. First use an initial position at the border and investigate all three system sizes. Then simulate the largest system when starting at (0,0,0).*

Answer Define a walk in terms of three co-ordinates, $x(i), y(i)$ and $z(i)$, where i counts the steps of the walk. Each step changes one of the co-ordinates by either $+1$ or -1. Changes that bring walks outside the prescribed box are

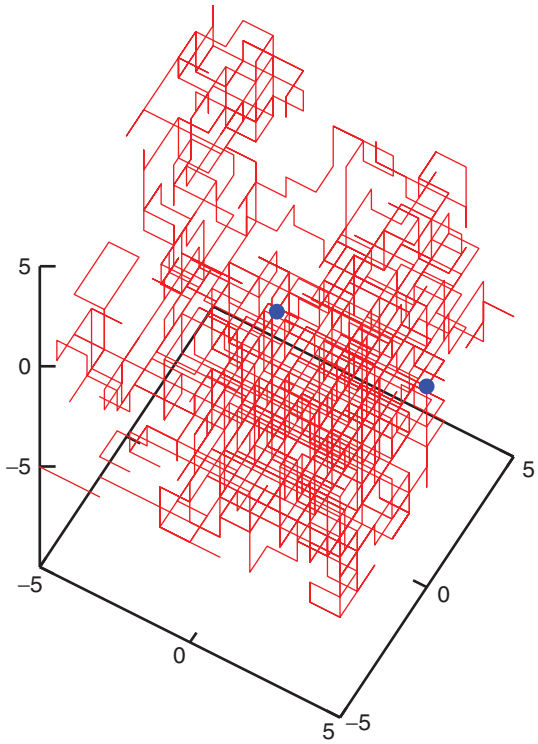

Figure 5.2 Random walker released to the right of a box, and stepping until it locate a target at position (0,0,0). In this particular $[-5:5]^3$ case the walk took 5684 steps to locate the target. Average over 100 walks gave a search time of 6629 steps.

not acceptable. Check whether walk is at position (0,0,0), and if so then abort the walk and record the number of steps i it took to reach this position. An illustration is given in Fig. 5.2.

Sampling 1000 walks from a starting position at $(L, 0, 0)$ until they reach (0,0,0) gives:
Average search time for $L = 1$ is $\langle t\rangle = 132$ (volume $V = 27$) $\langle t\rangle/V = 4.88$).
Average search time for $L = 2$ is $\langle t\rangle = 619$ (volume $V = 125$, $\langle t\rangle/V = 4.95$).
Average search time for $L = 5$ is $\langle t\rangle = 6629$ (volume $V = 1331$, $\langle t\rangle/V = 5$).
Average search time for $L = 10$ is $\langle t\rangle = 47086$ (volume $V = 9261$, $\langle t\rangle/V = 5.08$).

We then investigate the search time starting from (0,0,0) and searching for (0,0,0). For $L = 5$, this gives $\langle t\rangle = 6343$, a number that is remarkably close

to that one starting far away. The chance of geting lost when starting from the target is large anyway.

5.2.5 *Argue that the escape rate to a distance R away from a one-dimensional DNA strand before recapturing scales as: $r \propto \frac{2\pi Dl}{ln(R/b)}$, where b is the diameter of the DNA, l is the length of the DNA and where one assumes that the touching distance $= b$ always leads to absorption (this is the diffusion-limited case, reaction is instant when possible).*

Answer This problem amounts to an escape problem in two-dimensions (perpendicular to the DNA), (see Question 5.2.2). In steady state the rate of gain and loss should be equal, and thus the rate of loss of particles at distance R is:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{2\pi \cdot D \cdot \rho \cdot \ell}{\log(R/b)}$$

where ℓ is the length of the DNA. From this the escape rate per particle is calculated from $((1/\rho) \cdot \mathrm{d}\rho/\mathrm{d}t)$.

5.2.6 *DNA mismatch repair generally repairs about 99% of single nucleotide mismatches after DNA replication. It involves the protein MutS which dimerizes around the mismatch and thereafter, in E. coli, diffuses along the DNA until a hemi-methylated GATC site is located. This "half-methylated" site serves as a detector for new DNA (which ultimately becomes fully methylated). Estimate the minimum diffusion constant that one MutS needs to have in order to reach a GATC site within a characteristic time interval of 2 minutes (the time it takes the protein Dam to methylate the GATC site). Allow for a maximum of* 1% *failure rate and assume that it starts at a position that is 128 base pairs from a GATC site. Hint: with a characteristic time window of 2 min there will be 1% probability that the GATC site is methylated within 1 s.*

Answer To reach a distance $l = 128$ within $t = 120$ s, the diffusion constant should be at least:

$$D = \frac{l^2}{2 \cdot t} = \frac{128^2 \mathrm{bp}^2}{240\mathrm{s}} \sim 70\,\mathrm{bp}^2/\mathrm{s}^{-1} = 6\,\mathrm{nm}^2\,\mathrm{s}^{-1}$$

However, the 99% fidelity before the GATC sites are methylated, implies that it only fails in 1% of the cases. With an average of 2 minutes to methylate,

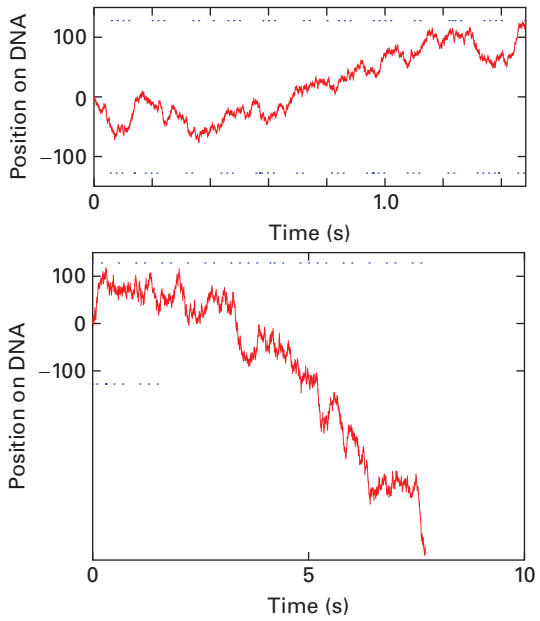

Figure 5.3 Random walker released at position 0 at time $t = 0$ and reaching position $x = 128$ after 1.5 seconds. The diffusion constant $D = 5000$ bp^2 s^{-1} simulated by steps of -1 bp or $+1$ bp every $\tau = 1/(2D) = 0.0001$ s. Simulating 5000 such walks we only fail reach hemi-methylated GATC sites in 1.1% of the simulations. The lower panel shows an example where the GATC sites become methylated before the MutS reaches any of them.

the methylation rate is $r = 1/(120\,\mathrm{s})$. Thus, within 1.2 s, 1% of the GATC sites are fully methylated. If position 128 is the only GATC site available, the MutS has to reach this site within this 1.2 s. Accordingly the minimal D becomes:

$$D = \frac{l^2}{2 \cdot t} = \frac{128^2 \mathrm{bp}^2}{2\,\mathrm{s}} \sim 8000\ \mathrm{bp}^2\,\mathrm{s}^{-1} = 700\ \mathrm{nm}^2\,\mathrm{s}^{-1} = 0.0007\ \mu\mathrm{m}^2\,\mathrm{s}^{-1}$$

5.2.7 *Simulate the problem in question 5.2.3, starting a random walk at position $x = 0$ and let it walk until it first passes position $+128$ or position -128. Use the simulation to deduce the diffusion constant when one assumes a 99% chance of reaching ± 128 before they are invisible, when methylation of each of these positions occurs at a rate of* $0.01\mathrm{s}^{-1}$.

Answer This requires a basic simulation as described in Fig. 5.3, where a walker is released at position $x = 0$, at time $t = 0$, and subsequently walks one

step to left or right any timestep $\tau = 1/(2D)$, thereby simulating a random walk corresponding to a diffusion constant of D. The two hemi-methylated sites each have a chance of 0.01τ to become methylated at each timestep. When both sites are methylated, the error at position 0 cannot be repaired and the simulation is aborted. Simulations with $D < 5000\,\mathrm{bp}^2\,\mathrm{s}^{-1}$ give more than 1% errors, whereas simulations with higher D give less than 1% errors.

5.3 Traffic on DNA

5.3.1 *Reconsider Eq. (5.26), taking into account that RNAP needs time $\Omega l/v$ to leave the promoter before a new RNAP can bind to it. Let $l = 35\,\mathrm{bp}$ be the length of an elongating RNAP and $v \sim 40\,\mathrm{bp\,s}^{-1}$ the velocity of the RNAP. At what promoter firing strength does this correction become more than a factor of two?*

Answer The steady-state occupancy of the promoter when including that a RNAP that has started to transcribe still occlude the promoter for a time interval $\Delta t = l/v$ is:

$$k_\mathrm{e}\theta = k_\mathrm{on}\cdot\left(1-\theta-k_\mathrm{e}\frac{l}{v}\cdot\theta\right)\Rightarrow$$
$$\theta = \frac{k_\mathrm{on}}{k_\mathrm{e}+k_\mathrm{on}+k_\mathrm{e}\cdot k_\mathrm{on}l/v}\rightarrow$$
$$\Omega = k_\mathrm{e}\theta = \frac{k_\mathrm{on}k_\mathrm{e}}{k_\mathrm{e}+k_\mathrm{on}+k_\mathrm{e}\cdot k_\mathrm{on}l/v}$$

where θ is occlusion associated with bound RNAP and $k_\mathrm{e}\frac{l}{v}\theta$ is occlusion associated, with RNAP that are initiated during the time $\ell l/v$. A factor two correction occurs when:

$$k_\mathrm{e}\cdot k_\mathrm{on}l/v = k_\mathrm{e}+k_\mathrm{on}\Rightarrow l/v = \frac{1}{k_\mathrm{e}}+\frac{1}{k_\mathrm{on}}$$

thus to have a factor of two correction, the basic time for binding to the promoter, plus the time to leave an open complex should be about $l/v = 35/40 = 0.85\,\mathrm{s}$. In this case the full time between elongation initiations $1/\Omega$ would then be 1.7 s.

5.3.2 *Show that $\mathcal{I} = 1+(\Omega^\mathrm{A}/\Omega)\cdot\alpha/(1+\alpha)^2$, where $\alpha = k_\mathrm{on}/k_\mathrm{e}$ is the aspect ratio of the promoter. Which value of α gives maximal interference? Discuss why interference decreases for both very small and very large α.*

Answer The promoter activity without and with opposing promoter:

$$\Omega(\text{without PA}) = \Omega_{\rm S} = \frac{k_{\rm on}k_{\rm e}}{k_{\rm e}+k_{\rm on}} \text{ and } \Omega(\text{with PA}) = \frac{k_{\rm on}k_{\rm e}}{k_{\rm e}+k_{\rm on}+\Omega_{\rm A}} \Rightarrow$$
$$\frac{\Omega(\text{without PA})}{\Omega(\text{with PA})} = \frac{k_{\rm e}+k_{\rm on}+\Omega_{\rm A}}{k_{\rm e}+k_{\rm on}} = 1+\frac{\Omega_{\rm A}}{k_{\rm e}+k_{\rm on}}$$
$$= 1+\frac{\Omega_{\rm A}}{\Omega_{\rm S}}\frac{k_{\rm on}k_{\rm e}}{(k_{\rm e}+k_{\rm on})^2} = 1+\frac{\Omega_{\rm A}}{\Omega_{\rm S}}\frac{k_{\rm on}k_{\rm e}}{(k_{\rm e}+k_{\rm on})^2} = 1+\frac{\Omega_{\rm A}}{\Omega_{\rm S}}\frac{\alpha}{(1+\alpha)^2}$$

where in last equality we divide nominator and denominator by $k_{\rm on}^2k_{\rm e}^2$ and use the dimensionless aspect ratio $\alpha = k_{\rm on}/k_{\rm e}$. Thus, the interference increases with increasing strength between the aggressive promoter, as well as with the symmetry of the sensitive promoter.

5.3.3 *Consider occlusion, the fact that an entering RNAP from pA prevents an RNAP bind to pS for a time given by* $l + r = 35 + 75$ *bp (see Fig. 5.6). (*$r = 75\,\text{bp}$ *is the length an RNAP needs to bind to a promoter.) What is the interference factor* $\mathcal{I}$ *if one includes this occlusion effect? Calculate* $\mathcal{I}$ *for* $\Omega = k_{\rm on}/2 = k_{\rm e}/2 = 0.01\ \text{s}^{-1}$, $\Omega^{\rm A} = 0.1\ \text{s}^{-1}$ *and* $v = 40\text{bp}\,\text{s}^{-1}$. *What is the interference if both promoters are four times stronger?*

Answer The steady-state balance between the RNAP leaving the promoter and the RNAP entering the promoter reads:

$$(k_{\rm e}+\Omega_{\rm A})\theta = k_{\rm on}\cdot(1-\theta)\cdot\exp\left(-\frac{L}{v}\cdot\Omega_{\rm A}\right)$$

where the last two factors count respectively the probability that there is no RNAP bound to the PS, and the probability that there is no RNAP from PA that moves across the promoter. The exponential is derived by the assumption that each time interval $\text{d}t$ has a probability $\exp(-\Omega_{\rm A}\text{d}t)$ not initiating an RNAP from PA. By multiplying all probabilities, we implicitly assume that all RNAP initiations are independent:

$$\left(k_{\rm e}+\Omega_{\rm A}+k_{\rm on}\exp\left(-\frac{L}{v}\cdot\Omega_{\rm A}\right)\right)\theta = k_{\rm on}\cdot\exp(-\frac{L}{v}\cdot\Omega_{\rm A}) \Rightarrow$$
$$\theta = \frac{k_{\rm on}\cdot\exp(-\frac{L}{v}\cdot\Omega_{\rm A})}{k_{\rm e}++\Omega_{\rm A}+k_{\rm on}\exp(-\frac{L}{v}\cdot\Omega_{\rm A})} \Rightarrow$$
$$\Omega(\text{with PA}) = \frac{k_{\rm e}k_{\rm on}\cdot\exp(-\frac{L}{v}\cdot\Omega_{\rm A})}{k_{\rm e}+\Omega_{\rm A}+k_{\rm on}\exp(-\frac{L}{v}\cdot\Omega_{\rm A})}$$

which should be compared to the activity without an interfering promoter

$$\Omega(\text{without PA}) = \Omega_{\rm S} = \frac{k_{\rm on}k_{\rm e}}{k_{\rm e}+k_{\rm on}}$$

The interference factor:

$$\begin{aligned}
\frac{\Omega(\text{without P})}{\Omega(with\ pA)} &= \frac{k_{\rm on}k_{\rm e}}{k_{\rm e}+k_{\rm on}} \cdot \frac{k_{\rm e}+\Omega_{\rm A}+k_{\rm on}\exp(-\frac{L}{v}\cdot\Omega_{\rm A})}{k_{\rm e}k_{\rm on}\cdot\exp(-\frac{L}{v}\cdot\Omega_{\rm A})} \\
&= \exp\left(\frac{L}{v}\cdot\Omega_{\rm A}\right)\cdot\left(1+\frac{\Omega_{\rm A}}{k_{\rm e}+k_{\rm on}}-\frac{k_{\rm on}(1-\exp(-\frac{L}{v}\cdot\Omega_{\rm A}))}{k_{\rm e}+k_{\rm on}}\right) \\
&= \exp\left(\frac{L}{v}\cdot\Omega_{\rm A}\right)\cdot\left(1+\frac{\Omega_{\rm A}}{\Omega_{\rm S}}\frac{\alpha}{(1+\alpha)^2}-\frac{k_{\rm on}(1-\exp(-\frac{L}{v}\cdot\Omega_{\rm A}))}{k_{\rm e}+k_{\rm on}}\right) \\
&\sim \exp\left(\frac{L}{v}\cdot\Omega_{\rm A}\right)\cdot\left(1+\frac{\Omega_{\rm A}}{\Omega_{\rm S}}\frac{\alpha}{(1+\alpha)^2}-\frac{\alpha\frac{L}{v}\cdot\Omega_{\rm A}}{1+\alpha}\right)
\end{aligned}$$

where the last two expressions use $\alpha = k_{\rm on}/k_{\rm e}$ (the detailed derivation for the $\Omega_{\rm A}/\Omega_{\rm S}$ term is given in an earlier question). One notices that the interference due to occlusion becomes especially large when $\Omega_{\rm A} \sim v/L > (40\,\text{bp}\,\text{s}^{-1})/105\,\text{bp} \sim 0.4\,\text{s}^{-1}$. Thus a promoter that initiates RNAP every 2.5 s occludes PS by about a factor of 1/e in addition to the interference associated with the "sitting duck" collisions.

For a concrete example:

$$\begin{aligned}
\frac{\Omega(\text{without PA})}{\Omega(\text{with PA})} &\sim \exp\left(\frac{105\,\text{bp}}{40\,\text{bp}\,\text{s}^{-1}}\cdot 0.1 s^{-1}\right)\cdot\left(1+\frac{0.1}{0.01}\frac{1}{(1+1)^2}-\frac{\frac{105}{40}\cdot 0.1}{1+1}\right) \\
&= \text{e}^{0.26}\cdot\left(1+10\cdot\frac{1}{4}-\frac{0.26}{2}\right) = \text{e}^{0.26}\cdot(3.5-0.13) = 4.4
\end{aligned}$$

In case both promoters are four times faster,

$$\begin{aligned}
\frac{\Omega(\text{without PA})}{\Omega(\text{with PA})} &\sim \exp\left(\frac{105bp}{40\,\text{bp}\,\text{s}^{-1}}\cdot 0.4\,\text{s}^{-1}\right)\cdot\left(1+\frac{0.1}{0.01}\frac{1}{(1+1)^2}-\frac{\frac{105}{40}\cdot 0.4}{1+1}\right) \\
&= \text{e}^{1.04}\cdot\left(1+10\cdot\frac{1}{4}-\frac{1.04}{2}\right) = \text{e}^{1.04}\cdot(3.5-0.52) = 8.4
\end{aligned}$$

5.3.4 *Assume that RNAP from pS (see Fig. 5.6) has to travel a distance* $L-40$ *before it has escaped possible collision with RNAP from pA. How does* $\mathcal{I}$ *change with increasing* N*?*

Answer The moving RNAP can collide with RNAP from PA that is initiated at any time from $(N-40)/v$ before it starts elongating itself, to a time $(N-40)/v$ after it started (the time where it passed pA). Thus it is vulnerable to opposing RNAP during a time window $\Delta t = 2(N-40)/v$, where $v = 40\,\mathrm{bp\,s^{-1}}$. The probability that pA does not fire during this time is:

$$P(\text{no collision}) = \exp(-\Omega_{\mathrm{A}} \cdot \Delta t)$$

if one assume that all initiates is independent of each other. Thus the interference factor:

$$\begin{aligned}\frac{\Omega(\text{without pA})}{\Omega(\text{with pA})} &= \frac{1}{P(\text{no}-\text{collision})} \cdot (\text{occlusion}) \cdot (\text{sitting duck}) \\ &= \exp(2\Omega_{\mathrm{A}} \cdot (N-40)/v) \cdot (\text{occlusion}) \cdot (\text{sitting duck})\end{aligned}$$

5.3.5 *Implement a stochastic model for promoter interference on a computer, including only the collision effect. Thus, each promoter starts one RNAP at each of their respective initiation rates, and each RNAP moves one step at a time towards the right, or left respectively. RNAP are assumed to fall off when colliding. Use a promoter strength of* $\mathrm{pA} = 0.1\,\mathrm{s^{-1}}$ *and* $\mathrm{pS} = 0.05\,\mathrm{s^{-1}}$ *and an RNAP speed of* $40\,\mathrm{bp\,s^{-1}}$*. Calculate the promoter activity for distances* $L = 100\,\mathrm{bp}$*, 1000 bp and 2000 bp. Compare with the analytical predictions from the previous question. Hint: Use* $\mathrm{d}t = 0.01\,\mathrm{s}$ *as the time unit, and let each promoter initiate a new RNAP with probability* $\mathrm{d}t \cdot pA$*, or* $\mathrm{d}t \cdot pS$*. At each time step move the RNAPs in their specified direction with probability* $v \cdot \mathrm{d}t = 0.4$*. One may speed up the simulation by implementing steps of 10 bp, thus using time steps* $\mathrm{d}t = 0.1$ *and accordingly adjusted promoter initiation probabilities per time step.*

Answer The main complication in the model is taking into account that there may be multiple RNAPs from each promoter on the DNA at any given time. This may be accomplished by having an array of RNAPs from PA and another array of RNAPs from PS. The array specifies the position of the RNAP between 1 and the system size N. When a new RNAP is introduced it is inserted in the first empty position in the corresponding array. When

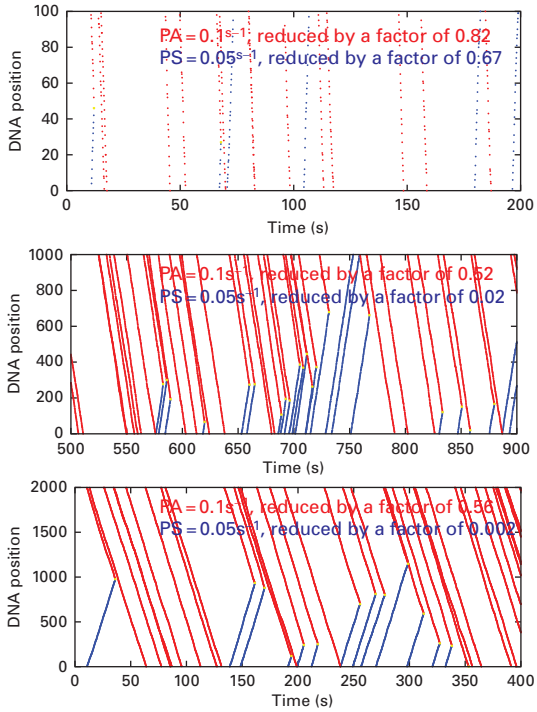

Figure 5.4 Stochastic simulation of traffic of RNAP between two opposing promoters at distances 140, 1040 and $N = 2040$ (the front ends of starting RNAPs are 20 bp ahead of the promoter start site. The promoters are a factor of two different in strength, and are modeled as a random stochastic initiation with probability $p \cdot \mathrm{d}t$ per time step $\mathrm{d}t = 0.01$s. At each timestep each RNAP moves 1 bp forward with probability $v \cdot \mathrm{d}t = 0.4$ as we simulate a velocity $v = 40\, bp/\mathrm{s}$. When two elongating RNAP collide, both are assumed to be removed from the DNA.

a RNAP moves outside the interval $[1, N]$ the corresponding position in the array is emptied. All RNAP are updated at each time step, and each attempts a move with probability $v \cdot \mathrm{d}t$. When an RNAP is moved to a position where there is another RNAP from same promoter, the move is not performed (one may keep a short-term memory of new attempted positions of all RNAP to test collisions and avoid any where RNAPs overtake each other). When RNAPs from opposing promoter occupy the same position or move past each other, both are removed.

The result of the simulation is shown in Fig. 5.4. The predicted interference effect becomes (here listing $1/\mathcal{I}$):

$$
\begin{aligned}
N = 100: \quad & \frac{PA}{PA_0} = 0.82 \text{ whereas } \mathrm{e}^{-2\cdot(100/40)\cdot 0.05} = 0.78 \\
& \frac{PS}{PS_0} = 0.67 \text{ whereas } \mathrm{e}^{-2\cdot(100/40)\cdot 0.1} = 0.61 \\
N = 1000: \quad & \frac{PA}{PA_0} = 0.52 \text{ whereas } \mathrm{e}^{-2\cdot(1000/40)\cdot 0.05} = 0.082 \\
& \frac{PS}{PS_0} = 0.022 \text{ whereas } \mathrm{e}^{-2\cdot(1000/40)\cdot 0.1} = 0.0067 \\
N = 2000: \quad & \frac{PA}{PA_0} = 0.56 \text{ whereas } \mathrm{e}^{-2\cdot(2000/40)\cdot 0.05} = 0.0067 \\
& \frac{PS}{PS_0} = 0.002 \text{ whereas } \mathrm{e}^{-2\cdot(2000/40)\cdot 0.1} = 0.000045
\end{aligned}
$$

In all cases the numerical simulations predict less interference than the analytical equation, reflecting a shielding effect where the first successful RNAP paves the way for subsequent RNAPs from the same promoter.

5.4 Bursty transcription initiation

5.4.1 *Make a computer program to select random numbers from a Poisson disribution with average 3. Test it against the real distribution. Hint: calculate cumulative distribution, and select random numbers between* $[0;1]$ *to select numbers from the cumulative distribution.*

Answers The Poisson distribution with mean 3 reads:

$$p(k) = \frac{3^k}{k!} \cdot \mathrm{e}^{-3}$$

The corresponding cumulative distribution is:

$$P(n) = \sum_{k=0}^{n} \frac{3^k}{k!} \cdot \mathrm{e}^{-3}$$

Select a random number $r \in [0,1]$ and find the lowest n where:

$$P(n) > r$$

The correspondingly selected k is then $k = n$. Results for 1000 random numbers are shown in Fig. 5.5.

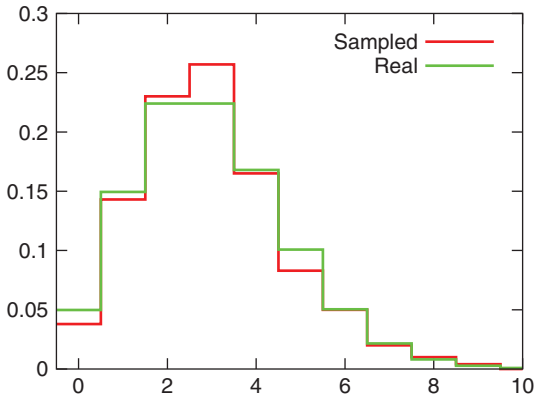

Figure 5.5 Stochastic simulation of Poisson distribution, compared with actual distribution.

5.4.2 *Consider a simplified production of proteins, where each mRNA on average gives one protein, and where one mRNA is produced every 10 minutes, whereas the decay time for proteins is 30 minutes. First assume that there is always exactly one protein per mRNA. Second assume that the mRNA lifetime is exponentially distributed, and that proteins are produced at random intervals as long as the mRNA survives.*

Answer Average number of proteins would be $\langle m \rangle = 3$ (production rate divided by decay rate). To simulate the first process, at each timestep of $\mathrm{d}t = 1$ we produce an mRNA $m \to m + 1$ with probability $\mathrm{prod} \cdot \mathrm{d}t = \mathrm{d}t$. Each mRNA is instantly converted to exactly one protein. At each time step we remove any of the current proteins with probability $\mathrm{decay} \cdot \mathrm{d}t/3$. The result is shown in the left-hand panels of Fig. 5.6, where it is also seen that the resulting distribution is close to a Poisson distribution with average $= 3$.

The second simulation takes into account the varying proteins produced from the mRNA. To simulate this we select an mRNA lifetime t_m from an exponential $\exp(-t_\mathrm{m})$. This is done by selecting a random number $r \in [0, 1]$ and assigning a lifetime $t_\mathrm{m} = -\log(r)$. Given a selected t_m, we start production of individual proteins at subsequent times t, separated by exponentially distributed time intervals Δt (also from a distribution $\exp(-\Delta t)$. When accumulated time $t = \sum_i \Delta t_i$ for these production events exceeds t_m, the protein burst is terminated (proteins cannot be produced after the mRNA is degraded). Notice that we assume that mRNA is short-lived compared to the protein lifetime, and accordingly finishes its total protein production

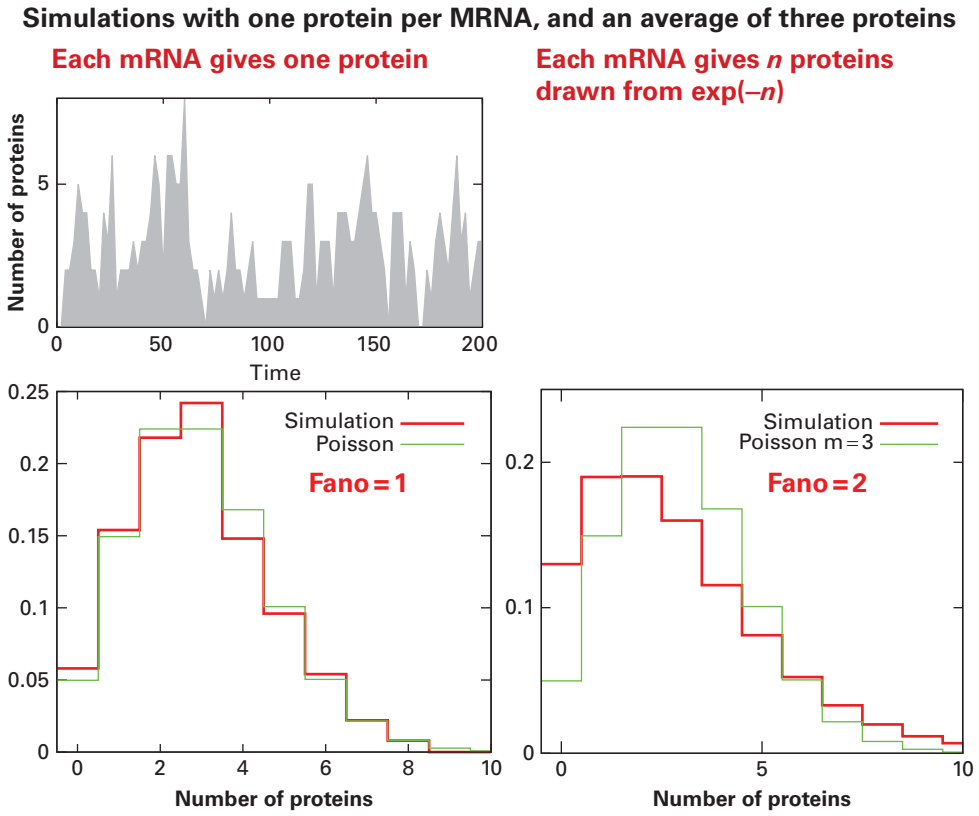

Figure 5.6 Stochastic simulation of production and decay of mRNA, comparing the steady-state distribution with a Poisson distribution with average 3.

without changing anything else in the system. The right-hand panel of Fig. 4.2 shows the resulting distribution. Notice that this distribution is substantially broader than the distribution in the left-hand panel (Fano factor of 2 instead of Fano factor of 1).

5.4.3 *Repeat as in 5.4.2, but now with 10 proteins produced simultaneously at each event (as it would be if we looked at proteins). In addition compare with situations where each mRNA gives an exponentially distributed "'burst" of proteins.*

Answer The simulations is as in the previous questions, except that one produces 10 proteins per production event, i.e. $p \to p+10$. Results are shown in Fig. 5.7. In the right-hand panel we show corresponding behavior when

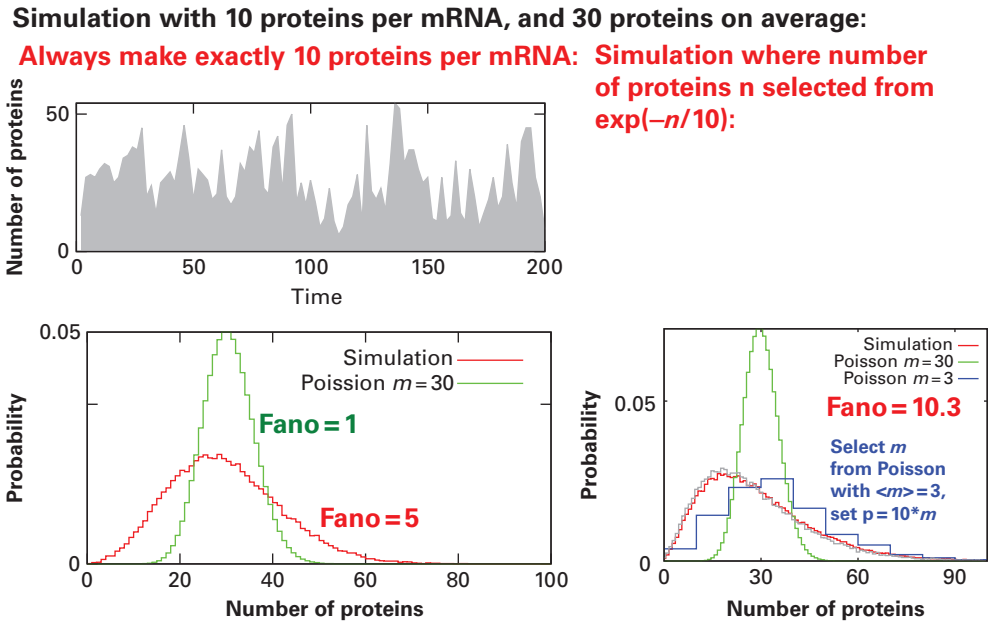

Figure 5.7 Stochastic simulation of production and decay of proteins where they are produced in bursts of size 10. The Fano factor of the final distribution is 5. The final distribution is compared to the steady-state distribution a with Poisson distribution with average 30, and a Fano factor of 1. The Right-hand panel takes into account the exponential distribution of protein for a given mRNA, and has a Fano factor ~ 10.

protein bursts are exponentially distributed with mean 10. The red curve is produced as in Question 5.4.2, just using subsequent time intervals between protein productions to be 1/10 of the mRNA lifetime. The gray curve is simply produced by setting protein bursts $p \to p - 10 \cdot \ln(r)$ with $r \in [0, 1]$ uniformly selected.

5.4.4 *Show that the large N, small p limit of a binomial distribution is a Poisson distribution. Hint:*

$$p(n) = \frac{N!}{(N-n)!n!} \cdot p^n \cdot (1-p)^{N-n}$$

and use $(1 - \frac{\lambda}{N})^{-n} \sim \mathrm{e}^{-n\lambda/N} \sim 1$, $(1 - \frac{\lambda}{N})^{N} \approx \mathrm{e}^{-\lambda}$ *and* $\frac{N!}{(N-n)!N^n} \sim 1$.[1]

[1]To derive the last of these approximations one may employ Stirling's equation $(N! = (N/e)^N)$ for $(N-n)! = ((N-n)/e)^{N-n}$

Answer Using Stirling's equation:

$$\begin{aligned}\frac{N!}{(N-n)!N^n} &= \frac{(N/\mathrm{e})^N}{((N-n)/\mathrm{e})^{N-n}\cdot N^n}\\ &= \mathrm{e}^{-n}\cdot\frac{N^{N-n}}{(N-n)^{N-n}}\\ &= \mathrm{e}^{-n}\cdot\frac{1}{(1-n/N)^{N-n}} = \mathrm{e}^{-n}\cdot\mathrm{e}^{n\cdot(N-n)/N} = \mathrm{e}^{-n^2/N}\sim 1\end{aligned}$$

when $n << \sqrt{N}$. Then use $\lambda = pN$ in the accordingly modified bimodial distribution to obtain

$$p(n) = \mathrm{e}^{-n^2/N}\cdot\frac{\lambda^n}{n!}\cdot\mathrm{e}^{-\lambda}\cdot\mathrm{e}^{+\lambda n/N}\sim\frac{\lambda^n}{n!}\cdot\mathrm{e}^{-\lambda}$$

5.4.5 *Simulate a single promoter with* $\alpha = 0.1$ *and overall firing activity of one elongation initiation every 20 s. Assume an mRNA decay rate of* $0.01\,\mathrm{s}^{-1}$ *(100 s lifetime). Plot mRNA as a function of time, and calculate the Fano factor for the number of mRNA. Repeat the simulation for* $\alpha = 1$ *and calculate the Fano factor. Why is the Fano factor smaller than 1? Finally, repeat the simulation for an mRNA decay rate of* $0.001\,\mathrm{s}^{-1}$ *and* $\alpha = 0.1$*, and convince yourself that the Fano factor remains around 1.*

Answer The promoter is simulated in terms of two states, either free or occupied. The on-rate for binding to the promoter is:

$$k_{\mathrm{on}} = \Omega(1+\alpha)\ \text{ whereas }\ k_{\mathrm{e}} = k_{\mathrm{e}}/\alpha \tag{5.1}$$

where $\Omega = 0.05\,\mathrm{s}^{-1}$ is total initiation rate of promoter. For $\alpha = 0.1$, then $k_{\mathrm{on}} = 0.05\cdot 1.1 = 0.055$ and $k_{\mathrm{e}} = 0.055/0.1 = 0.55\,\mathrm{s}^{-1}$. For $\alpha = 1$, then $k_{\mathrm{on}} = 0.05\cdot 2 = 0.1$ and $k_{\mathrm{e}} = 0.1/1 = 0.1\,\mathrm{s}^{-1}$. The simulation can be performed in discrete timesteps, $\mathrm{d}t = 1\,\mathrm{s}$. At each timestep one tests whether the promoter is occupied or not. For $\alpha = 0.1$, the code reads: if occupied, an mRNA is produced with probability $k_{\mathrm{e}}\mathrm{d}t = 0.55$ and if so the promoter is left unoccupied and the mRNA count $m \to m+1$; if un-occupied, the promoter becomes occupied with probability $k_{\mathrm{on}}\mathrm{d}t = 0.055$. At each timestep each mRNA is removed by probability decay $\cdot\,\mathrm{d}t$: $m \to m-1$ for each m (where decay = 0.01). Results are shown in Fig. 5.8.

5.4.6 *Repeat the simulation above, with decay time 100 s,* $k_{\mathrm{on}} = 0.00505\,\mathrm{s}^{-1}$ *and* $k_{\mathrm{e}} = 0.505\,\mathrm{s}^{-1}$*, but now allow the elongating RNAP to recruit a new*

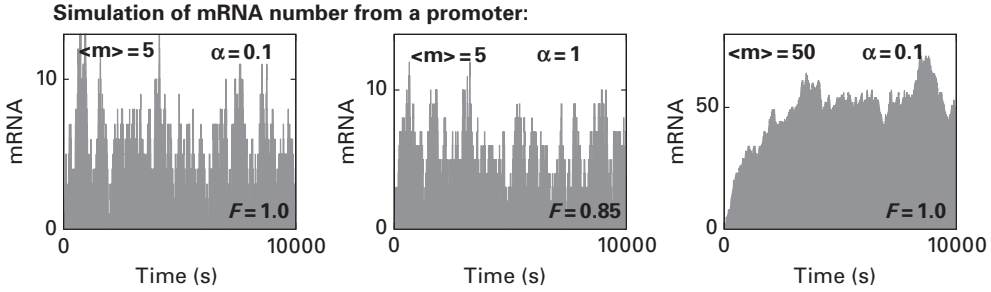

Figure 5.8 Stochastic simulation of production and decay of mRNA from promoters with a total elongation initiation rate of 0.05 s (one mRNA every 20 seconds). For the promoter $\alpha = k_{\rm on}/k_{\rm e} = 0.1$ On two the left panels the mRNA lifetime is 100 s, on the right-most the mRNA lifetime is 1000 s.

RNAP to the open complex with 90% probability. Thus the average promoter activity remains close to that in the previous question, but the fluctuations become larger. Calculate the Fano factor for the number of mRNA produced.

Answer Each time an RNAP binds to the promoter there are on average 10 mRNA produced. Therefore an unbound promoter must be about $\Omega_0 = 0.005$; To model this overall reduction one may assume that $k_{\rm on} = 0.00505\,{\rm s}^{-1}$ and $k_{\rm e} = 0.505\,{\rm s}^{-1}$, implying that it takes about 200 s for an unbound promoter to recruit a new RNAP. The simulation proceeds in time steps dt where at each step one checks the promoter: if occupied, an mRNA is produced with probability $k_e{\rm d}t = 0.505$ and if so the promoter is left unoccupied with probability 0.1 (=1 − 0.9) and the mRNA count $m \to m + 1$. If unoccupied, the promoter becomes occupied with probability $k_{\rm on}{\rm d}t = 0.00505$. At each timestep each mRNA is removed with probability decay · dt: $m \to m - 1$ for each m (where decay = 0.01). Results are shown in Fig. 5.9.

5.4.7 *Consider the promoter from 5.4.5, but now also consider that RNAP binds competitively with a repressor that has an on rate of* $1\,{\rm s}^{-1}$ *and an off rate of* $0.05\,{\rm s}^{-1}$. *Vary* α *and discuss why repression varies. Calculate the Fano factor for the number of mRNA in each case.*

Answer The simulation can be performed in discrete time steps, dt = 1 s. At each time step one tests whether the promoter is occupied or not with state = −1 for occupied by transcription factor, state 0 for free, and state +1 for promoter occupied by RNAP. The code reads: if occupied by RNAP, an mRNA is produced with probability $k_{\rm e}{\rm d}t$ and if so the promoter is left

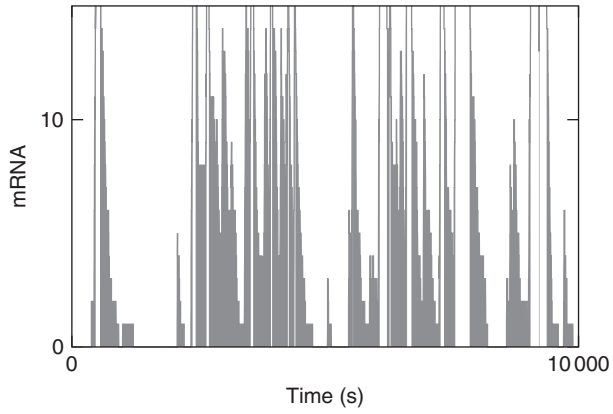

Figure 5.9 Stochastic simulation of production and decay of mRNA from intermittent promoter, with $k_{\rm on} = 0.00505\,{\rm s}^{-1}$, $k_{\rm e} = 0.505\,{\rm s}^{-1}$ and with the rule that a elongating RNAP recruits a new RNAP ready for elongation with probability 0.9. The mRNA lifetime is 100 s, giving an average of 4.5 mRNA in the system with a Fano factor of $F = 7$.

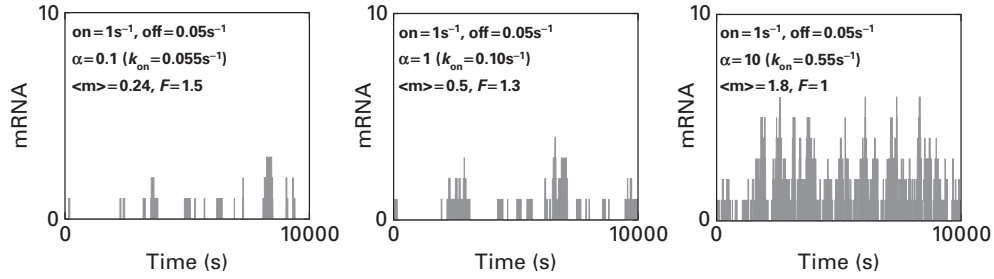

Figure 5.10 Stochastic simulation of production and decay of mRNA from a promoter that can also be occluded by a transcription factor. The on rate for the transcription factor is $r_{\rm on} = 1\,{\rm s}^{-1}$, whereas the of rate is $r_{off} = 0.05\,{\rm s}^{-1}$, giving a maximal repression capacity of 20. However this repression is only obtained for a very low $k_{\rm on}$ for the RNAP ($\alpha = 0.1$ case).

unoccupied and the mRNA count $m \to m + 1$; if unoccupied, the promoter becomes occupied by either RNAP or transcription factor with probability $(k_{\rm on} + r_{\rm on}){\rm d}t$. Subsequently one of the actual states of the promoter is chosen with a probability given by the relative weights of the two on rates. If the promoter is occupied by transcription factor, this factor is removed with probability $r_{\rm off}{\rm d}t$. At each timestep, each mRNA is removed with probability decay $\cdot\, {\rm d}t$: $m \to m - 1$ for each m. Results are shown in Fig. 5.10.

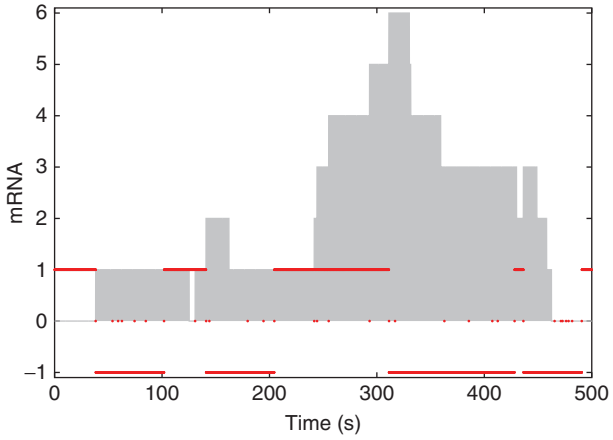

Figure 5.11 Stochastic simulation of production and decay of mRNA from a promoter that can also be occluded by a transcription factor. The on rate for the transcription factor is $r_{\text{on}} = 1/\,\text{s}^{-1}$, whereas the of rate is $r_{off} = 0.01\,\text{s}^{-1}$, giving a maximal repression capacity of 100. The red dots show promoter status, with -1 being occluded by transcription factor, whereas 0 is unoccupied and 1 is occupied by RNAP.

5.4.8 *Consider the promoter from 5.4.5 and 5.4.7, but now also consider that RNAP binds competitively with a repressor that has an on rate of* $1\,\text{s}^{-1}$ *and an off rate of* $0.01\,\text{s}^{-1}$. *Set* $\alpha = 10$ *and plot the dynamics on mRNA and promoter status. The lifetime of mRNA can again be assumed to be 100 s.*

Answer See Fig. 5.11.

5.5.1 *For a single promoter producing one mRNA (m), which encodes for one protein (P) the final protein concentration will have a distribution with width* σ_P *described by [223, 228]:*

$$\frac{\sigma_P^2}{\langle P\rangle^2} = \frac{\gamma_P \cdot \gamma_m}{\Omega \cdot \omega} + \frac{\gamma_P \cdot \gamma_m}{\omega \cdot (\gamma_P + \gamma_m)}$$

where $\langle P\rangle$ *is the average protein level,* γ_P *and* γ_m *are degradation rates of mRNA and its encoded protein, and* Ω *and* ω *are transcription and translation initiation rates, respectively. Reformulate this equation in terms of average proteins per message etc.*

Answer $\frac{\Omega}{\gamma_{\text{m}}}$ is the rate of mRNA production multiplied by the lifetime of mRNA ($1/\gamma_{\text{m}}$). Thus the ratio is the average number of mRNAs present in the cell.

$\frac{\omega}{\gamma_\mathrm{P}}$ is the protein production rate per mRNA multiplied by the protein life time, giving the number of proteins in the cell for each existing mRNA in the cell.
$\frac{\Omega}{\gamma_\mathrm{m}} \cdot \frac{\omega}{\gamma_\mathrm{P}}$ is thus the number of proteins in the cell $= \langle \mathrm{P} \rangle$.
$\frac{\gamma_\mathrm{m}}{\Omega}$ is 1 divided by number of mRNA in the cell.

$$\frac{\gamma_\mathrm{m}}{\gamma_\mathrm{p} + \gamma_\mathrm{m}} = t_\mathrm{protein}/(t_\mathrm{protein} + t_\mathrm{mRNA})$$

is the relative lifetime of proteins to the sum of protein and mRNA life times. Thus for very short mRNA lifetimes, this factor is ~ 1, whereas for long mRNA lifetimes and short protein lifetimes the last factor is ~ 0. As an overall result, the protein level is distributed by:

$$\frac{\mathrm{Variance}}{\langle \mathrm{P} \rangle^2} = \frac{1}{\langle \mathrm{P} \rangle} + \frac{1}{\langle \mathrm{mRNA} \rangle} \cdot \frac{t_\mathrm{protein}}{t_\mathrm{protein} + t_\mathrm{mRNA}}$$

an equation that expresses the variance of P in terms of a sum that represents the two contributions to the variance. Here, the latter, $1/\langle \mathrm{mRNA} \rangle$, will typically be the bigges contribution, because each mRNA makes many proteins and thus $\langle \mathrm{mRNA} \rangle << \langle \mathrm{P} \rangle$.

In the special case where proteins have shorter or comparable lifetimes to mRNA, it is not all proteins from a given mRNA that contribute to the proteins present in the cell. In fact only a fraction $\frac{\gamma_\mathrm{m}}{\gamma_\mathrm{p}+\gamma_\mathrm{m}}$ of the proteins produced from a given mRNA will be present at a given time. This is the reason for the reduction factor $\frac{t_\mathrm{protein}}{t_\mathrm{protein}+t_\mathrm{mRNA}}$.