# 4

# Statistical mechanics of phage $\lambda$

## 4.1 Lifecycle of phage $\lambda$

**4.1.1** *Why is it a good strategy for the $\lambda$ phage to favor lysogeny when a bacterium is infected simultaneously by many $\lambda$s?*

**Answer** By being infected by more than one $\lambda$, the phages have the information that it is likely that there are more phages than bacteria in the surroundings. As lysis reduces the available hosts and just produces even more phages, lysis becomes a less optimal strategy. Notice that the phage particle only lives a relatively short time outside the host, typically less than a week.

**4.1.2** *One theorem in game theory states that for a single player there always exists a deterministic strategy that is no worse than any random strategy. This is apparently not true for $\lambda$. Argue how a random decision could be good, as an alternative to deterministically going lytic or lysogenic as a function of external conditions.*

**Answer** Phages are produced in bursts of several hundreds. By allocating some off-springs to one path, and others to the opposing path, there is a bigger probability that one of the descendants will survive. A more thorough discussion can be found in M. Avlund I.B. Dodd, S. Semsey, K. Sneppen and S. Krishna. Why do phage play dice, J. Virol., 83: 11416, 2009. See also 12.3.2 and 12.3.3.

**4.1.3** *A $\lambda$ lysogen is superinfected with another $\lambda$ phage. Consider the situation where there is already a phage in a lysogenic state inside the bacterium*

*and a second new λ phage infects. Which of the genes in the new λ is transcribed, and which are not? Why does the new phage not induce lysis?*

**Answer** The established lysogens have high concentrations of CI (as well as RexA and RexB). This freely diffusing CI will bind to the operators of the infecting λ phage and establish the same repression as in the integrated phage. Thus the new phage will express CI, RexA and RexB, and nothing else; in particular, it cannot express the proteins needed for replication of itself and therefore the new phage cannot initiate the lytic pathway.

## 4.2 Biological growth and counting

**4.2.1** *There exist virulent phages that always lyse bacteria after infections. An example is the phage T4 which infects* E. coli. *After infection, the phage copies its genome through a rolling circle [169, 170],*[1] *where one copy is generated about every 10 s, giving a phage production rate* $\gamma \approx 6\,\mathrm{min}^{-1}$. *If the bacteria lyse at a latent time* $\tau$, *the number of phage progeny would be* $\beta = \gamma(\tau - \tau_0)$, *where* $\tau_0 \approx 5\,\mathrm{min}$ *is the time needed before phage production starts. If the time needed to find and infect a new* E. coli *is* $1/(\alpha\rho)$, *where* $\alpha \approx 10^{-9} min^{-1} ml$ *is the infection rate per phage per bacterium and* $\rho$ *is the bacteria density. Write an expression for the exponential growth rate,* $g$. *Show that the latent time* ($\tau_{\mathrm{opt}}$) *that maximizes the phage growth rate* ($g$) *satisfies [172]:*

$$\frac{1}{\alpha \cdot \rho} + \tau_{\mathrm{opt}} = (\tau_{\mathrm{opt}} - \tau_0) \cdot \log[\gamma \cdot (\tau_{\mathrm{opt}} - \tau_0)]$$

*Investigate the solution graphically.*

**Answer** One has to count the total length of one phage infection cycle, including the phage latency time and the time it takes the phage to locate a now host. During time $\tau$, one phage becomes $\beta = \gamma(\tau - \tau_0)$ phages. Each of these phages then establishes a new infection within a time $t \sim \tau_{\mathrm{search}} = 1/(\alpha\rho)$. Thus, the total time, amplifying by a factor $\beta$ is:

$$t_\beta = \frac{1}{\alpha\rho} + \tau$$

[1] Here simplified, as the early stage of phage growth often occurs through exponential $\theta$ replication. The final burst time depends on produced holin proteins [171], which puncture the *E. coli* cell when their number reaches a critical threshold.

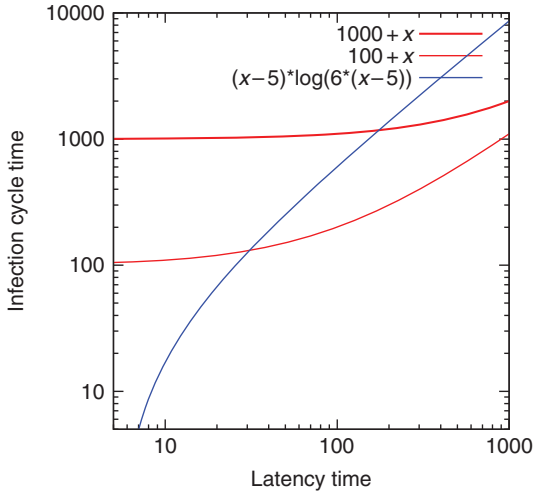


Figure 4.1 Graphical solutions for an average time for locating a bacteria of 100 min, and 1000 min, corresponding to bacterial densities of $10^7$ ml and $10^6$ ml, respectively. The optimal latency time accordingly changes with bacterial density.

The growth rate $g$:

$$\exp(t_\beta \cdot g) = \beta = \gamma(\tau - \tau_0) \Rightarrow g = \frac{\ln(\gamma(\tau - \tau_0))}{\tau + 1/(\alpha\rho)}$$

and maximal growth rate is obtained when:

$$\frac{\mathrm{d}g}{\mathrm{d}\tau} = 0 \Rightarrow \frac{1}{(\tau - \tau_0)(\tau + 1/\alpha\rho)} - \frac{\ln(\gamma(\tau - \tau_0))}{(\tau + 1/\alpha\rho)^2} = 0$$

$$\frac{1}{\alpha \cdot \rho} + \tau_{\mathrm{opt}} = (\tau_{\mathrm{opt}} - \tau_0) \cdot \log[\gamma \cdot (\tau_{\mathrm{opt}} - \tau_0)]$$

which is indeed the desired equation for $\tau = \tau_{\mathrm{opt}}$. For graphic illustration we use a bacterial density $\rho = 10^6\,\mathrm{ml}^{-1}$ giving $1/\alpha\rho = 1000$ min which, with $\gamma = 6\ \ \mathrm{min}^{-1}$ and $\tau_0 = 5$ min gives the equation:

$$1000 + \tau_{\mathrm{opt}} = (\tau_{\mathrm{opt}} - 5) \cdot \log[6 \cdot (\tau_{\mathrm{opt}} - 5)]$$

which can be inspected in Fig. 4.1. Notice that the derivation uses that all phages reach their targets at the same average time. A more correct treatment would take into account that the phages that reach their target early, contribute disproportionally more to the growth.

## 4.3 Chemical binding and counting

**4.3.1** *Consider a titration experiment where one varies CI concentration for a fixed operator concentration in order to probe the first-order chemical reaction* $\mathrm{CI} + \mathrm{O} \leftrightarrow \mathrm{CIO}$. *Compare the fraction of bound [CIO] to the operator as a function of* $[\mathrm{CI_T}]$ *for the case where* $[\mathrm{O_T}] = 10 \cdot K$, *and a case where* $[\mathrm{O_T}] = K/10$.

**Answer** As explained in the main text:

$$\frac{[\mathrm{CIO}]}{[\mathrm{O_t}]} = \frac{[\mathrm{CI}]}{K + [\mathrm{CI}]} \tag{4.1}$$

where $K$ is the CI concentration at which [O] is half occupied. Here the free concentration [CI] is less than the total concentration, as.

$$[\mathrm{CI_{total}}] = [\mathrm{CI}] + [\mathrm{CIO}] \tag{4.2}$$

The equation where we ignore the effect of operators on the free concentration of CI is:

$$\frac{[\mathrm{CIO}]}{[\mathrm{O_t}]} = \frac{[\mathrm{CI_{total}}]}{K + [\mathrm{CI_{total}}]}$$

In contrast, the complete equation is:

$$K = \frac{([\mathrm{CI_{total}}] - [\mathrm{CIO}])([\mathrm{O_t}] - [\mathrm{CIO}])}{[\mathrm{CIO}]} \Rightarrow$$
$$[\mathrm{CIO}] = \frac{[\mathrm{CI_{total}}] + [\mathrm{O_t}] + K}{2} - \sqrt{\frac{([\mathrm{CI_{total}}] + [\mathrm{O_t}] + K)^2}{4} - [\mathrm{CI_{total}}][\mathrm{O_t}]}$$

which we investigate in Fig 4.2 for the case where $K = 1$ and $[\mathrm{O_t}] = 0.1$, and $[\mathrm{O_t}] = 10$ $[\mathrm{CIO}]/[\mathrm{O_t}]$ in both cases plotting.

**4.3.2** *Consider one DNA-binding protein that may bind to* $L = 5\,000\,000$ *possible positions on the E. coli DNA with a binding free energy* $\Delta G = -3$ kcal mol $\mathrm{bp}^{-1}$ *(counted as the energy for each of the* $L = 5\,000\,000$ *positions). Alternately the protein may be anywhere in the volume* $V = 1\,\mu\mathrm{m}^3$. *Express the statistical weight for being in each of these two states, and calculate which is the most likely.*

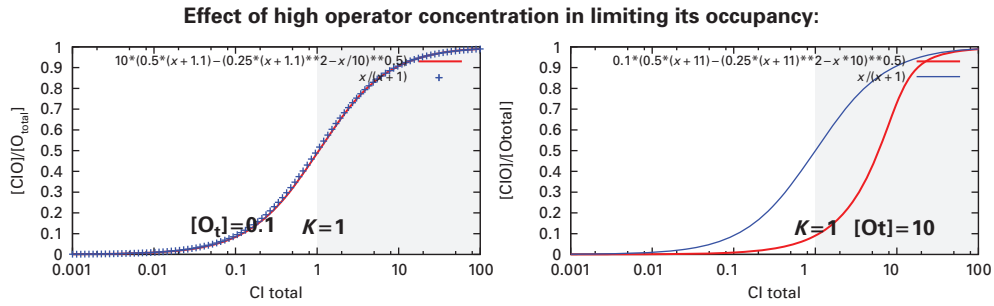


Figure 4.2 Comparison between a case where there is a lot of DNA, and a case with little operator DNA, using $K = 1$.

**Answers** The statistical weight for each binding of a protein C is:

$$Z_{\rm s} = (1/V) \cdot \exp(-\Delta G/k_{\rm B}T) = (1/V) \cdot \exp(-3/0.62)$$

For binding to $L = 5\,000\,000$ sites:

$$Z_{\rm u} = (L/V) \cdot \exp(-\Delta G/k_{\rm B}T) = (L/V) \cdot \exp(3/0.62) = 0.005 \cdot 126 = 0.6$$

using $1/V = 1\,{\rm nM} = 10^{-9}$ M, corresponding to a bacterial volume with 1 000 000 000 states competing with a DNA with 5 000 000 positions. The statistical weight to be free is in this normalization:

$$Z_{\rm unbound} = 1$$

Thus the probability to be bound to the DNA non-specifically is:

$$P({\rm bound–non–specific}) = \frac{0.6}{1+0.6} = 0.38$$

thus it is marginally more likely to be free.

**4.3.3** *If there are $N$ of the above DNA-binding proteins, write the equation for the probability of possible partitions of the proteins in the DNA-bound state.*

**Answers** As there is plenty of space, the binding of each protein to the DNA somewhere is denoted $p$ and the probability to bind $n$ out of $N$ proteins becomes:

$$P(n) = \frac{N!}{(N-n)!n!} \cdot p^n (1-p)^{N-n} \tag{4.3}$$

i.e. a standard binomial expression with an average number of non-specific binders of $N \cdot p$.

**4.3.4** *If a living system demanded* $10^{12}$ *different proteins to work, instead of a diversity of* $\sim 10^3$*, a cell would have a volume* $1\,\mathrm{mm}^3$*. With about one regulator of each type per cell, what would then be the expected binding energy between the proteins and an operator? How much would the time for finding a particular operator be changed?*

**Answers** The binding energy should be on a scale where any binding is accessible by changing the concentration. Thus the binding should not be stronger than $\exp(-\Delta G/k_{\mathrm{B}}T) = 1/V$ and, on the other hand, not so weak that one needs millions of proteins to obtain substantial binding. Overalls we therefore expect a binding energy given by (as $1/V$ is counted in units of M = molar):

$$(\mathrm{e}^{-\Delta G/k_{\mathrm{B}}T} \sim \frac{100}{V} \sim 10^{-7} \Rightarrow \Delta G = -10\,\mathrm{kcal\ mol}^{-1} \tag{4.4}$$

for a bacterium with volume $1\,\mu\mathrm{m}^3$. On the other hand, an increase in protein diversity from 1000 to $10^{12}$ should be associated with a proportionate increase in volume, from $1\,\mu\mathrm{m}^3$ to $1\,\mathrm{mm}^3$. Accordingly the expected binding would be:

$$\mathrm{e}^{-\Delta G/k_{\mathrm{B}}T} \sim \frac{100}{V} \sim 10^{-16} \Rightarrow \Delta G = -23\,\mathrm{kcal\ mol}^{-1} \tag{4.5}$$

When volume is increased by a factor of 1 000 000 000, the time for two molecules to find each other will also be increased by a factor 1 000 000 000. As we will explain later, this time is at present of the order of seconds in an *E. coli*, but would then be of the order of 30 years in this "superbug."

## 4.4 Chemistry and co-operativity as statistical mechanics

**4.4.1** *Develop a computer program to calculate the activity of PRM and PR as a function of CI concentration, given the tabulated free energies for CI to OR sites in Fig. 4.13. To simplify the problem disregard the RNAP in the partition sum, and include only the eight states that bind CI dimers in various ways. Activity is then given by the states that allow for RNAP binding*

*to respective promoters, i.e. PRM with OR3 free and PR with both OR1 and OR2 free. Remember that the PRM activity of states with CI on OR2 is ten times larger than that of states without CI on OR2. Also remember that co-operativity between OR1–OR2 and OR2–OR3 is mutually exclusive. First, calculate PR and PRM activity as a function of CI, assuming that CI dimerization energy is $-\infty$ (all CI in dimers). Second, calculate the same activity using a dimerization constant $K_{\mathrm{Dim}} = [1\,\mathrm{M}] \cdot \mathrm{e}^{-11.1/0.62}$. Third, in the last calculation investigate the effect of removing the co-operative binding between CI bound to OR1 and CI bound to OR2.*

**Answer** In terms of the dimer concentration [C], the statistical weight for the eight states is:

$$
\begin{aligned}
Z(0,0,0) &= 1 \\
Z(1,0,0) &= [\mathrm{C}] \cdot \mathrm{e}^{-\Delta G_3/k_{\mathrm{B}}T} = [\mathrm{C}] \cdot \mathrm{e}^{9.7/0.62} \\
Z(0,1,0) &= [\mathrm{C}] \cdot \mathrm{e}^{-\Delta G_2/k_{\mathrm{B}}T} = [\mathrm{C}] \cdot \mathrm{e}^{11.0/0.62} \\
Z(0,0,1) &= [\mathrm{C}] \cdot \mathrm{e}^{-\Delta G_3/k_{\mathrm{B}}T} = [\mathrm{C}] \cdot \mathrm{e}^{12.8/0.62} \\
Z(1,1,0) &= [\mathrm{C}]^2 \cdot \mathrm{e}^{-(\Delta G_3+\Delta G_2+\Delta G_{2-3})/k_{\mathrm{B}}T} = [\mathrm{C}]^2 \cdot \mathrm{e}^{24/0.62} \\
Z(1,0,1) &= [\mathrm{C}]^2 \cdot \mathrm{e}^{-(\Delta G_3+\Delta G_1)/k_{\mathrm{B}}T} = [\mathrm{C}]^2 \cdot \mathrm{e}^{22.5/0.62} \\
Z(0,1,1) &= [\mathrm{C}]^2 \cdot \mathrm{e}^{-(\Delta G_2+\Delta G_1+\Delta G_{1-2})/k_{\mathrm{B}}T} = [\mathrm{C}]^2 \cdot \mathrm{e}^{26.4/0.62} \\
Z(1,1,1) &= [\mathrm{C}]^3 \cdot \left(\mathrm{e}^{-(\Delta G_3+\Delta G_2+\Delta G_1+\Delta G_{1-2})/k_{\mathrm{B}}T} + \mathrm{e}^{-(\Delta G_3+\Delta G_2+\Delta G_1+\Delta G_{2-3})/k_{\mathrm{B}}T}\right) \\
&= [\mathrm{C}]^3 \cdot (\mathrm{e}^{36.8/0.62} + \mathrm{e}^{36.1/0.62})
\end{aligned}
$$

where the double sum in $Z(1,1,1)$ expresses the alternating cooperativity: that CI at OR2 can only bind to OR1 or to OR3, but not to both simultaneously. The promoter activities are:

$$
\mathrm{PRM} = \mathrm{Basal}_{\mathrm{PRM}} \cdot \frac{Z(0,0,0) + Z(0,0,1) + 9 \cdot Z(0,1,0) + 9 \cdot Z(0,1,1)}{\sum_{i,j,k} Z(i,j,k)}
$$

$$
\mathrm{PR} = \mathrm{Basal}_{\mathrm{PR}} \cdot \frac{Z(0,0,0) + Z(1,0,0)}{\sum_{i,j,k} Z(i,j,k)}
$$

where we use that an unstimulated PRM is nine times weaker than a PRM where CI on OR2 can recruit the RNAP.[2] As given in the main text, the dimer concentration:

$$\begin{aligned}[\mathrm{C}] &= \frac{1}{2}([\mathrm{CI_T}] - [\mathrm{CI_M}]) \\ &= \frac{1}{2}\left([\mathrm{CI_T}] + \frac{K_\mathrm{D}}{4} - \frac{K_\mathrm{D}}{4}\sqrt{1 + \frac{8}{K_\mathrm{D}}[\mathrm{CI_T}]}\right)\end{aligned}$$

with $K_\mathrm{D} = \mathrm{e}^{-11.1/0.62} = 1.7 \cdot 10^{-8}$ i.e. 17 nM. The results can be inspected in Fig. 4.3. Notice that the wild-type $\lambda$ provides the sharpest onset of PRM activity (positive feedback), as well as negative feedback around lysogenic levels of CI that will stabilize this state. The cyan and blue curves lack co-operativity between CI bound to OR2 and CI bound to neighbor operators. The magenta curve has infinitely strong dimerization, and therefore stronger, but less co-operative, CI binding.

The lower panels examine the full model with proper inclusion of RNAP binding as an active state. Qualitatively, the results are similar, although the repression of PR is weaker.

**4.4.2** *If CI is cleaved by RecA, its dimerization is prevented. What concentration of CI monomers is needed to maintain a similar probability of having OR1 occupied as 100 nM of CI dimers does. Assume that monomer–OR1 binding is half of the dimer CI–OR1 binding (= $-12.8\,\mathrm{kcal\,mol^{-1}}$), and also simplify the problem by assuming that only one CI monomer can bind to the operator.*

**Answer** The dimer concentration is (with $[\mathrm{CI_T}] = 100$ nM):

$$\begin{aligned}[\mathrm{C}] &= \frac{1}{2}([\mathrm{CI_T}] - [\mathrm{CI_M}]) \\ &= \frac{1}{2}\left([\mathrm{CI_T}] + \frac{K_\mathrm{D}}{4} - \frac{K_\mathrm{D}}{4}\sqrt{1 + \frac{8}{K_\mathrm{D}}[\mathrm{CI_T}]}\right) = 37\,\mathrm{nM}\end{aligned}$$

[2] The basal activities of the two promoters are quite different, with unstimulated PRM activity every 150 s, whereas non-repressed PR should have one initiation every 15 s.

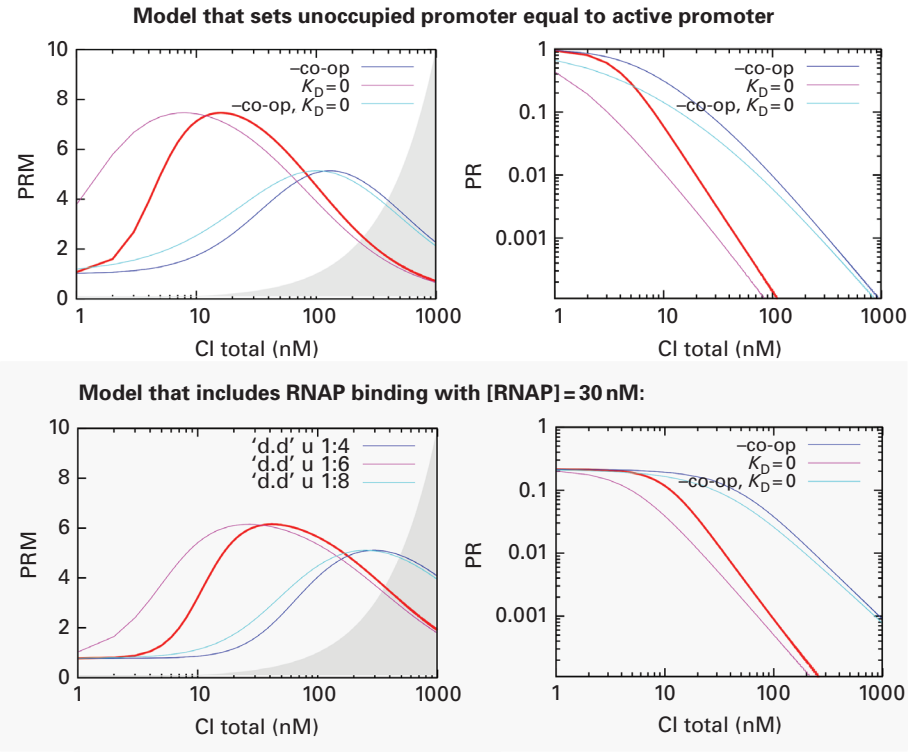


Figure 4.3 Promoter activity in phage $\lambda$. The upper two panels are for the simplified model specified in Question 4.3.1. The lower two panels are for the full model with RNAP binding as an explicit state, assuming that RNAP binding to OR3 and CI binding to OR2 provides nine times the activity of non-stimulated RNAP binding to OR3. The Gray area refers to a dilution rate associated with cell division, which should correspond to about 300 nM CI in a wild-type lysogen.

using $K_D = 17$ nM. Each monomer binds with half the energy of a dimer, and therefore we require a monomer concentration $X$, determined from:

$$\frac{37 \cdot 10^{-9} \exp(12.8/0.62)}{1 + 37 \cdot 10^{-9} \exp(12.8/0.62)} = \frac{X \cdot \exp(6.4/0.62)}{1 + X \cdot \exp(6.4/0.62)} \Rightarrow$$

$$X = 37\ nM \cdot \exp((12.8 - 6.4)/0.62) = 1\,130\,000\,\text{nM}$$

thus similar repression demands an increase from 100 CI proteins to about one million CI monomers per *E. coli* cell.

**4.4.3** *Often there are very few molecules of a given transcription factor inside a cell. Accordingly, it may be useful to consider the partition function for the case where occupancy of the single operator depletes the amount of free CI substantially. Argue that for a state characterized by $n_{\rm M}$ free monomers, $n_{\rm D}$ free dimers and an operator state $s$ where $i = i(s)$ dimers are bound to operators, the exact statistical weight is [175]:*

$$Z(s, n_{\rm M}) = \frac{V^{n_{\rm D}}}{n_{\rm D}!} \cdot \frac{V^{n_{\rm M}}}{n_{\rm M}!} \cdot K_{\rm D}^{n_{\rm D}+i(s)} \cdot {\rm e}^{G(s)/k_{\rm B}T}$$

*where $n_{\rm D} = (N - n_{\rm M} - 2i)/2$, since the number of free dimers in the cell is fixed by the conservation requirement: $N = n_{\rm M} + 2n_{\rm D} + 2i$, with $N$ being the total numbers of repressors in the cell. Hint: each of the $n_{\rm D} + i$ dimers contribute by their dimerization binding free energy through $K_{\rm D} = {\rm e}^{\Delta G_{\rm D}/k_{\rm B}T}$, whereas only the dimers bound to operators contribute to $-\Delta G(s)$. Again the total partition function can be written as a sum over all the states $s, n_{\rm M}$ that the $N$ molecules can be in: $Z = \sum Z(s, n_{\rm M})$.*

**Answer** Consider binding of one molecule out of $N$ molecules in a volume $V$, The statistical weight for unbound versus bound option being normalized in various equivalent forms:

$$Z_{\rm free} = 1 \text{ and } Z_{\rm bound} = \frac{N}{V}{\rm e}^{-\Delta G/k_{\rm B}T}$$

$$Z_{\rm free} = \frac{V^N}{V!} \text{ and } Z_{\rm bound} = \frac{V^{N-1}}{(N-1)!} \cdot {\rm e}^{\Delta G/k_{\rm B}T}$$

Iterating the above to a situation where the bound state involves $i$ bound C molecules with a total binding energy $\Delta G_i$, then:

$$Z_i = \frac{V^{N-i}}{(N-i)!} \cdot {\rm e}^{\Delta G_i/k_{\rm B}T}$$

If there are both monomers and dimers, then each dimer contributes to binding with $\exp(\Delta G_{\rm dimer}/k_{\rm B}T) = K_{\rm D}$ Thus, partitioning $N$ CI proteins into $n_{\rm D}$ dimers and $n_{\rm M} = N - 2n_{\rm D}$ monomers is associated with a statistical weight:

$$Z(n_{\rm D}) = \frac{V^{n_{\rm D}}}{n_{\rm D}!} \cdot \frac{V^{n_{\rm M}}}{n_{\rm M}!} \cdot K^{n_D} \tag{4.6}$$

The statistical weight for a state $s$ where $i$ dimers are bound to operators with total binding energy $\Delta G_i = \Delta G(s)$, $n_{\rm D}$ dimers are free and the remaining $n_{\rm M} = N - 2n_{\rm D} - 2i$ are freely floating monomers:

$$Z(i, n_{\rm D}) = \frac{V^{n_{\rm D}}}{n_{\rm D}!} \cdot \frac{V^{n_{\rm M}}}{n_{\rm M}!} \cdot K^{n_{\rm D}+i} \cdot \exp(\Delta G_i / k_{\rm B}T)$$

The message here is that at any instant there will be cell to cell variation in actual occupancy of the various states.

## 4.5 Distant DNA in gene regulation

**4.5.1** *The linear dimensions of a human cell are about 10 times that of a bacterium. Human DNA consists of* $3 \cdot 10^9$ *base pairs. If one (wrongly!) assumes that the human cell can be viewed as one big "bag" of proteins and DNA, what would be the non-specific binding energy that makes a protein as equally likely to be on the DNA as in the cell volume?*

**Answer** Equal statistical weight for being free is $Z_{\rm free} = 1$, whereas weight for being bound is:

$$Z_{\rm bound} = \frac{L}{V} \cdot {\rm e}^{-\Delta G/k_{\rm B}T} = \frac{3 \cdot 10^9}{10^{12}} \cdot {\rm e}^{\Delta - G_{\rm u}/k_{\rm B}T}$$

Where we set cell volume to be 1000 times that of *E. coli*, and thus $1/{\rm v} = 0.001{\rm nM}$. Setting them equal gives:

$${\rm e}^{\Delta - G_{\rm u}/k_{\rm B}T} = 333 \Rightarrow \Delta G_{\rm u} = -0.62 \cdot \log(333)\ {\rm kcal\ mol}^{-1} = -3.6\,{\rm kcal\,mol}^{-1}$$

a slightly larger binding affinity than the similar estimate *E. coli*.

**4.5.2** *Consider a DNA-binding protein in an E. coli cell that binds to 90% of the its DNA with* $\Delta G = -3\,{\rm kcal\,mol}^{-1}$ *and to 10% of its DNA with* $\Delta G = -5\,{\rm kcal\,mol}^{-1}$ *What is the probability that such a protein will be free in the cell? Hint: The statistical weight for one repressor to be bound to a sample of binding sites with energies* $\Delta G_i, i = 1, \ldots$ *is* $Z = \sum_i {\rm e}^{-\Delta G_i/k_{\rm B}T}$.

**Answer** The statistical weight for being bound is:

$$Z_{\rm bound} = \frac{0.9 \cdot 5000000}{10^9} \cdot {\rm e}^{3/0.62} + \frac{0.1 \cdot 5000000}{10^9} \cdot {\rm e}^{5/0.62} = 2.2$$

when setting $k_{\rm B}T = 0.62\,{\rm kcal\ mol}^{-1}$.

**4.5.3** *The interaction between a DNA-binding protein and the DNA can be written as a sum of individual interactions between amino acids and the base pairs at the corresponding positions [200, 201]. Assume, for a repressor in E. coli, that each of the 5 000 000 non-specific bindings are drawn from a Gaussian distribution with mean* $-3\,\mathrm{kcal\,mol^{-1}}$ *and standard deviation* $2\,\mathrm{kcal\,mol^{-1}}$*. What must the binding energy to the specific operator site O be in order that a protein should spend at least half its time at O?*

**Answer** The normalized Gaussian distribution is:

$$P(G) = \frac{1}{2\sqrt{2\pi}} \cdot \exp(-(G+3)^2/8)$$

Thus, the statistical weight for being bound to $L = 5\ 000\ 000$ potential sites is:

$$\begin{aligned}
Z_{\mathrm{non-spec}} &= \frac{5\,000\,000}{V \cdot \sqrt{2\pi} \cdot 2} \cdot \int_{-\infty}^{\infty} \exp(-(G+3)^2/8) \cdot \exp(-G/k_{\mathrm{B}}T) \cdot \mathrm{d}G \\
&= \frac{5\ 000\ 000}{V \cdot \sqrt{2\pi} \cdot 2} \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{8}G^2 - \left(\frac{3}{4} + \frac{1}{k_{\mathrm{B}}T}\right) G - \frac{9}{8}\right) \cdot \mathrm{d}G \\
&= \frac{5\ 000\ 000}{V \cdot \sqrt{2\pi} \cdot 2} \cdot \int_{-\infty}^{\infty} \cdot \exp\left(-\frac{1}{8}\left(G + 3 + \frac{4}{k_{\mathrm{B}}T}\right)^2\right. \\
&\quad \left. + \frac{1}{8}\left(3 + \frac{4}{k_{\mathrm{B}}T}\right)^2 - \frac{9}{8}\right) \mathrm{d}G \\
&= \frac{5\ 000\ 000}{V} \exp\left(\frac{1}{8}\left(3 + \frac{4}{k_{\mathrm{B}}T}\right)^2 - \frac{9}{8}\right) \\
&= \frac{5\ 000\ 000}{V} \cdot \mathrm{e}^{10.0} = \frac{5\ 000\ 000}{10^9} \cdot 23 \cdot 10^3 = 115
\end{aligned}$$

where we use $k_{\mathrm{B}}T = 0.62$. This binding to DNA in general should be compared to the non-specific binding of:

$$Z_{\mathrm{off}} = 1 \tag{4.7}$$

and a specific binding:

$$Z = \frac{1}{V} \cdot \exp(-\Delta G_{\mathrm{spec}}/k_{\mathrm{B}}T) = 10^{-9} \exp(-\Delta G/k_{\mathrm{B}}T)$$

To have 1/2 occupation of a specific binding site:

$$\frac{Z_{\text{spec}}}{Z_{\text{spec}} + Z_{\text{non-spec}} + Z_{\text{off}}} = \frac{1}{2} \Rightarrow$$
$$Z_{\text{spec}} = Z_{\text{non-spec}} + Z\text{off} = 116 \cdot 10^9 \Rightarrow \Delta G = -15.8\,\text{kcal}\,\text{mol}^{-1}$$

**4.5.4** *Repeat above question if the typical non-specific binding was +3kcal/mol. (trick question, remember that the protein may also be free)*

**Answer** The non-specific binding is now a factor of $\exp(6/k_{\text{B}}T) = \exp(6/0.62) = 16\,000$ lower than in the previous question, so $Z_{\text{non-spec}} = 115/16000 << 1$. Accordingly non-specific DNA binding does not contribute and 1/2 occupancy occurs when:

$$\frac{1}{V} \cdot \exp(-\Delta G/k_{\text{B}}T) = 1 \Rightarrow \Delta G = -0.62 \cdot \log(10^9)\ \text{kcal mol}^{-1}$$
$$= 12.9\ \text{kcal}\,\text{mol}^{-1}$$

**4.5.5** *Simulate a lattice polymer on a 3-dimensional cubic lattice. Let it start at position (0,0,0) and make 12 steps, each in a random, uncorrelated direction compared to its previous step. What is the likelihood that it returns to the origin after exactly 12 steps and exactly 20 steps. Repeat the above questions for a polymer where we do not allow self-interactions, apart from the last point.*

**Answer** This question demands simulations on a cubic lattice, starting each walk at position 0, and then walking 12, or 20 steps. One finally scores whether the walk ended at the start point or not. Fig. 4.4 shows such walks, comparing in both cases the walk that returns (thick line) with a typical walk, where we allow self-intersections: $L = 12$, $P = 0.015$; $L = 20$, $P = 0.0073$.

Self-avoiding walk: $L = 12$, $P = 0.0037$, $L = 20$, $P = 0.0017$.

One can observe that longer walks are less likely to return to the origin.

**4.5.6** *Other temperate phages are also governed by two antagonistic promoters, with one promoter directing the production of a lysogen-maintenance protein that one may call CI. In [202] the CI promoter $PL$ in phage 186 shows an activation curve with respect to its product CI, which resembles*

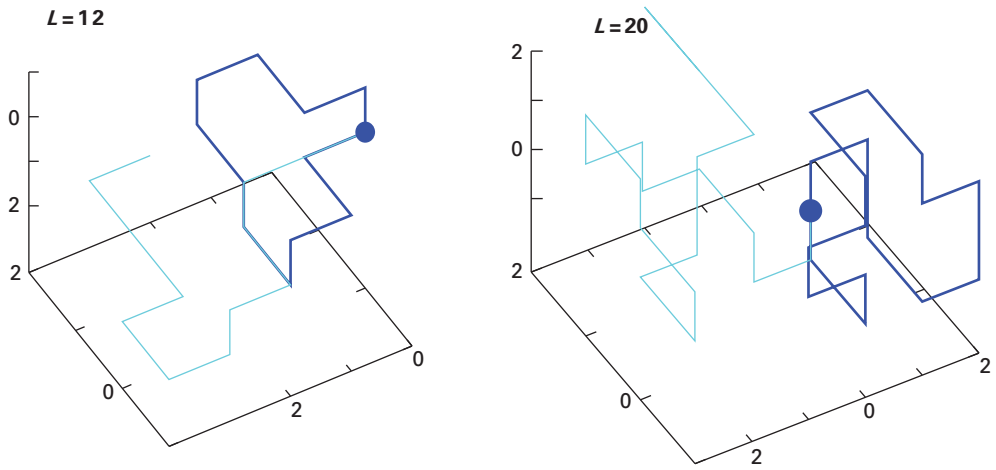


Figure 4.4 Looping of a polymer of 12 kuhn lengths, and 20 kuhn lengths, respectively (a kuhn length is the double persistance length, and is defined such that a random walk on a corresponding cubic lattice represents the random configurations of the polymer.)

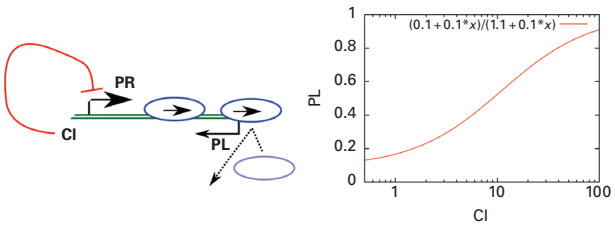


Figure 4.5 PL activity in a simple model for indirect interference from a PR promoter. The left part of the figure illustrates the model as it would work in the wild-type phage, whereas the right part illustrates the predicted PL activity as a function iof CI, when CI is given from outside.

*that of PRM in phage λ in Fig. 4.17. In 186 the self-activation comes from an indirect effect of CI that represses a PR promoter that would otherwise initiate RNAP, which removes RNAP that is bound to the promoter for CI. Assume that this interference is of the form* $PL = PL_0/(PR + PL_0)$ *with* $PR = PR_0/(CI + 1)$ *and deduce the activity of* $PL$ *as a function of CI. Set the base promoter activities to* $PR_0 = 10 \cdot PL_0$.

**Answer** Express:

$$PL = \frac{PL_0}{PR + PL_0} = \frac{PL_0}{PR_0/(\mathrm{CI}+1) + PL_0} = \frac{PL_0 \cdot (\mathrm{CI}+1)}{PR_0 + PL_0 \cdot (\mathrm{CI}+1)} \tag{4.8}$$

$$= \frac{PL_0 + PL_0 \cdot CI}{PR_0 + PL_0 + PL_0 \cdot CI} \tag{4.9}$$

$$= \frac{0.1 + 0.1 \cdot CI}{1.1 + 0.1 \cdot CI} \tag{4.10}$$

This function is shown in Fig. 4.5.