Chapter 19

Network Organization in Space and Time

"Tout ce qui est vrai pour le Colibacille est vrai pour l'éléphant." (That which is true for *E. coli* is true for the elephant.) - Jacques Monod

Chapter Overview: In Which Statistical Mechanics Is Used to Study Gene Regulation

Specific genes are used only when and where they are needed. For example, we have made much of the classic example of the Lac operon which governs the enzymes responsible for lactose digestion. Similar control is exercised over genes in other bacteria, archaea and eukaryotes. The tools worked out throughout the book leave us poised to consider important quantitative questions about gene regulation such as: how much is a given gene expressed, where in the cell (or the organism) is that gene expressed and at what time during the cell cycle (or life history) of the organism? The key tools we will use to study these questions are statistical mechanics and rate equations. The statistical mechanics approach will use the probability of promoter occupancy as the key quantity of interest, whereas the rate equation approach will examine the concentrations of protein products over time. These same techniques will also be used to examine signaling with special emphasis on the "decisions" cells make about where to go.

19.1 Chemical and Informational Organization in the Cell

Many Chemical Reactions in the Cell Are Linked in Complex Networks

The reality of the chemical reactions that take place in the cell are a far cry from the relatively sterile and simple kinetic processes described in chap. 15. In the discussion given there, we showed how to write the time evolution of the concentrations of a set of reactants and products. That theoretical machinery provides an appealing and useful picture for characterizing many of the beautiful *in vitro* experiments which have powered solution biochemistry. However, biochemistry in living cells has reactants and products linked in a complex set of lineages of biblical proportions where A begets B which begets C which in turn begets D and so on, with the added non-anthropomorphic complication that Z might just beget A again. Indeed, the fact that Z can act back on A reflects the presence of feedback which makes the dynamics even richer. Two of the most important classes of reaction that are central to the functioning of cells are those associated with gene regulation and signaling. Indeed, one of the features that most completely distinguishes the chemistry of a cell from that of solution biochemistry is the way in which the reactants are tuned by up and down regulation. Similarly, the reactions of the cell are also stimulated by external cues in the form of signaling cascades. In this chapter, we consider regulation and signaling by using a variety of tools developed throughout the book.

Genetic Networks Describe the Linkages Between Different Genes and Their Products

One of the most intriguing ways in which the chemistry of the cell cannot be viewed as a bag of reactants and products is the fact that this chemistry is under the strict control of the genetic machinery of the cell. In particular, if left to its own devices, some particular chemical pathway in the cell might just travel a path to eventual equilibrium. On the other hand, because of both external and internal cues, the machinery of the cell can receive orders via signaling pathways which lead, in turn, to the expression of some gene which results in a new reactant in the original chemical pathway which sends it off in some new direction.

The description of the informational pathways that dictate the cellular concentration profiles in both space and time of the various chemical reactants of interest is founded upon a higher-level of abstraction. In particular, there are networks of genes that are linked together in sometimes horrifyingly complex arrays such as that shown in fig. 19.1. This network is an example of a particularly well characterized genetic network which participates in the embryonic development of sea urchins. One important take home message concerning this network is that it is a typical network and should leave the reader with a sense of the implied chemical complexity of these systems. In general, genetic networks like that shown in fig. 19.1 make no reference either to the passage of time, nor to the quantitative distributions of the molecules that mediate these networks. Rather, these networks are an abstraction which shows how genes (and their products) are linked to each other in both space and time. On the other hand, it is important to bear in mind that beneath the surface of these wiring diagrams are actual concentrations of the molecular players of these informational

pathways.

Developmental Decisions Are Made by Regulating Genes

Often, genetic networks serve as the basis of the developmental decisions which send a cell or collections of cells down some path. One of the intriguing features of multicellular organisms is that despite the overwhelming cellular diversity, generally, each cell carries the same genetic baggage. However, in general, cells only express a certain fraction of all the available genes. This differentiation is the basis of the development of embryos and the basis of the different structures found in multicellular organisms. The key point is that not all genes are being expressed all the time.

One of the most famous examples of a "developmental decision" is the λ-switch described in chap. 4 and shown in fig. 4.10 (pg. 196). After infecting an *E. coli* bacterium, lambda phage follows one of two developmental pathways. One pathway (the lytic pathway) results in the assembly of new phage and the lysis of the host cell. The second pathway, the lysogenic pathway, involves incorporation of the lambda genome into that of the host cell. Lysogeny can be reverted by damaging the cell with UV light, which triggers lytic replication.

Another compelling example of the role of developmental decisions is that of embryonic development in fruit flies. One of the most celebrated examples is that of the body plan along the long axis of the fly embryo which is dictated by the distribution of certain proteins along the embryo. Figure 19.2 gives an example of the gradients in four key regulatory proteins which determine the anterior-posterior organization. These proteins determine the pattern of gene expression along the embryo from which the Eve 2 stripe is the most well understood example. These ideas were already introduced in section 2.3.3 (pg. 106).

Part of the hard won wisdom of molecular biology is the recognition that there are many stages in the pathway between DNA and functional protein that can serve as regulatory points. Some of these different regulatory mechanisms are shown in fig. 6.7 (pg. 313). For the purposes of the present discussion, we will focus on one of the most common regulatory mechanisms, namely, transcriptional control where the key decision that is made is whether or not to produce mRNA.

Gene Expression Is Measured Quantitatively in Terms of How Much, When and Where

One of our main arguments is that gene expression is a subject that has become increasingly quantitative. In particular, it is now common to measure how much a given gene is expressed, when it is expressed and where it is expressed. To carry out such measurements, there are a number of useful tools.

- **Experiments Behind the Facts.** Quantitative measurement of gene expression can be made at many stages between the decision to start transcription and the emergence of a functional protein product. As noted earlier, such measurements have provided a quantitative window on how

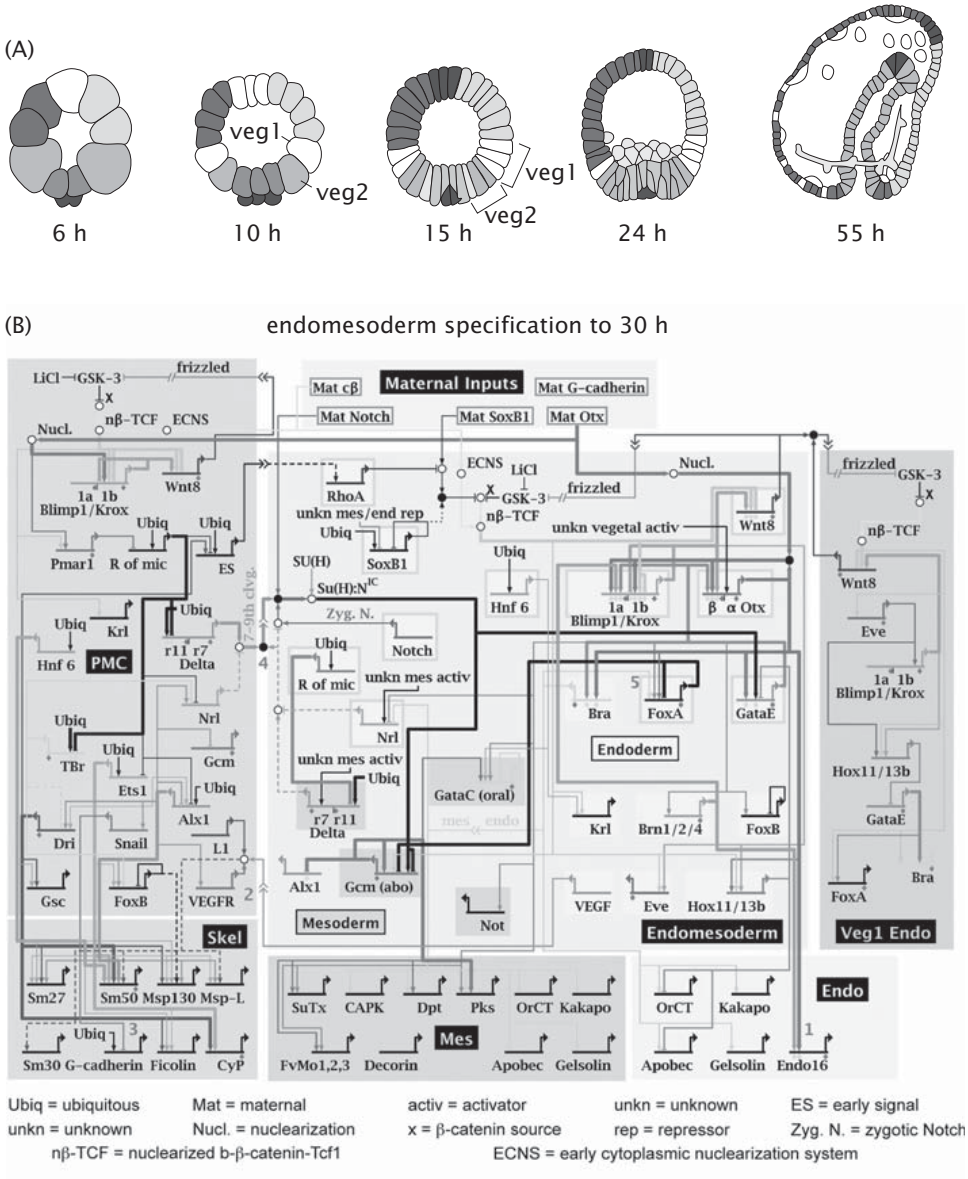

Figure 19.1: Genetic network associated with control of the developmental pathway of the sea urchin embryo. (A) Schematic of stages in the embryonic development of the sea urchin. (B) Genetic network associated with sea urchin development. (Adapted from S. Ben-Tabou de-Leon and E. H. Davidson, Annu. Rev. Biophys. Biomol. Struct., 36:191, 2007.)

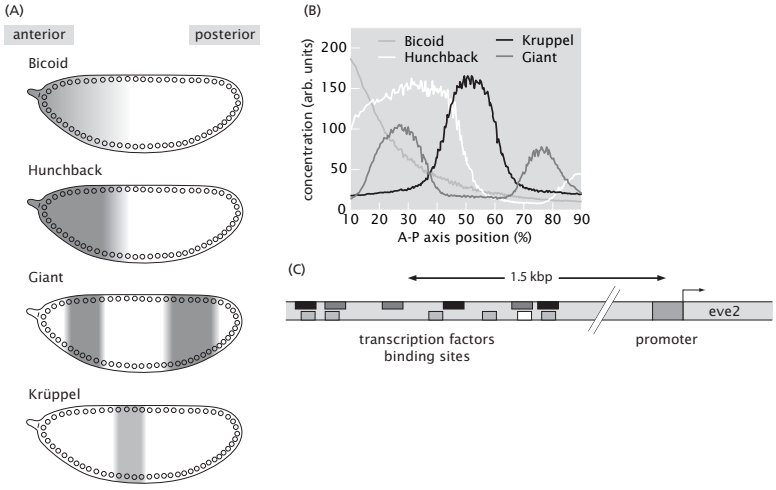

Figure 19.2: Regulatory proteins in the *Drosophila* embryo. The anterior-posterior patterning of the fruit fly is dictated by genes that are controlled by spatially varying concentrations of transcription factors. (A) Schematic of the main transcription factors involved in the regulation of the *even-skipped* stripe 2 gene. (B) Measurement of these morphogen gradients using immunoprecipitation. (C) Regulatory region of the *even-skipped* stripe 2 gene where the binding sites for each transcription factor have been identified. (B, adapted from Myasnikova *et al.*, Bioinformatics, 17:1, 2001; C, Adapted from Small *et al.*, EMBO J., 11:4047, 1992.)

much a given gene is expressed, where it is expressed spatially and when.

One important way to characterize the activity of a gene is by virtue of its protein products. In particular, if the gene product has enzyme activity, that activity can be assayed as a reporter of the extent to which the gene has been expressed as shown in fig. 19.3(A). Recall that beta-galactosidase is the enzymatic product of the *lac* operon, as shown in fig. 4.13 (pg. 200), and that the action of this enzyme is to clip lactose molecules. One of the impressive legacies of years of work on this system is a battery of substrates which respond differently to the enzymatic cleavage. One such substrate (ONPG) turns yellow upon cleavage and measuring the rate at which a solution becomes yellow optically can provide a window on gene expression since it is proportional to the amount of enzyme (over some region of concentrations). By measuring the absorbance at the appropriate wavelengths one has a picture of the amount of active enzyme. To carry out this kind of assay usually requires routine cloning in which sequences encoding the enzyme are inserted into the genome under the control of the transcription factors of interest.

From a molecular biology perspective, this same strategy of inserting a reporter into the gene of interest can be followed, but with the difference that the "reporter" molecule is a fluorescent molecule such as the green fluorescent protein (GFP) rather than an enzyme. This case is shown in fig. 19.3(B). Relative fluorescence levels of reporters such as GFP are easy to characterize.

A second scheme for characterizing the extent to which a given gene is expressed is by measuring how much messenger RNA from the gene of interest is present in the cell. One of the tools of choice for such measurements is the DNA microarray. DNA microarrays are built by labeling a surface with an array of different DNA molecules, each patch of which has small DNAs with the same sequence as shown in fig. 19.4. These sequences are chosen to be complementary to an entire battery of sequences corresponding to the genes of interest in the experiment. Cells are then broken up and their RNA (or DNA copies made from the RNA) are then allowed to flow across the array and hybridize with the molecules on the surface. The various molecules have been fluorescently labeled, so by looking at the fluorescence intensity at each point on the array, it is possible to read off how much RNA was present.

Another scheme for characterizing the amount of RNA is to use quantitative PCR. Once again, the cell is lysed and the mRNA molecules are turned into DNA using a reverse transcription reaction. Then these molecules are used as templates in a PCR reaction, and it is seen how many cycles of PCR are needed before the quantity of DNA in the reaction exceeds some threshold. This threshold value is a direct reflection of the amount of starting molecules since starting with lots of template DNA will result in many more molecules at low cycle numbers than will very little starting

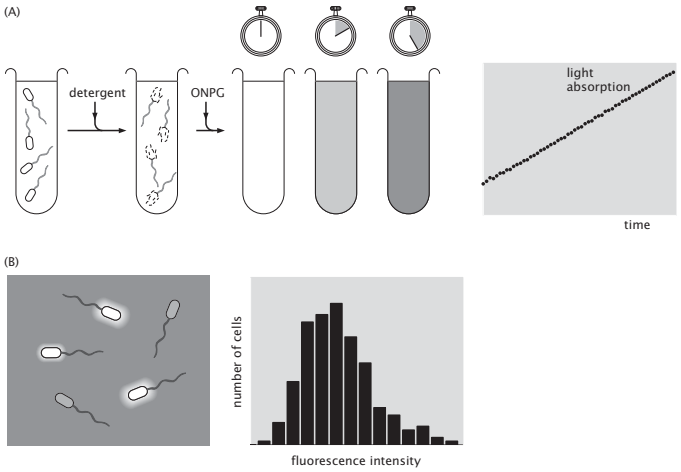

Figure 19.3: Measurement of gene expression. (A) Measurement of gene expression as a result of enzymatic activity. The promoter of interest drives the expression of an enzyme that can cleave a molecule which in the cleaved state is colored. (B) The promoter of interest drives the expression of a fluorescent protein such as the green fluorescent protein (GFP). The amount of fluorescence reports the extent of expression of the gene of interest.

material. With quantitative PCR one can detect mRNA copy numbers as low as 10.

As will be described in the remainder of this chapter, a useful surrogate for the actual question of the extent to which a given gene is expressed is to ask whether or not the promoter for the gene of interest is occupied. There are many *in vitro* and *in vivo* methods for finding out whether or not the promoter is hidden by polymerase binding or not. *Chromatin inmunoprecipitation* and *DNA footprinting* are two methods which are sensitive to promoter occupancy. The idea in these assays is that the part of DNA where the transcriptional apparatus is bound will react differently when the system is exposed to agents such as restriction enzymes. The most common procedure is to try to digest the DNA using a restriction enzyme. They will not be able to access the DNA over which RNAP is situated, leaving a "footprint" of a longer piece of DNA which can be easily detected.

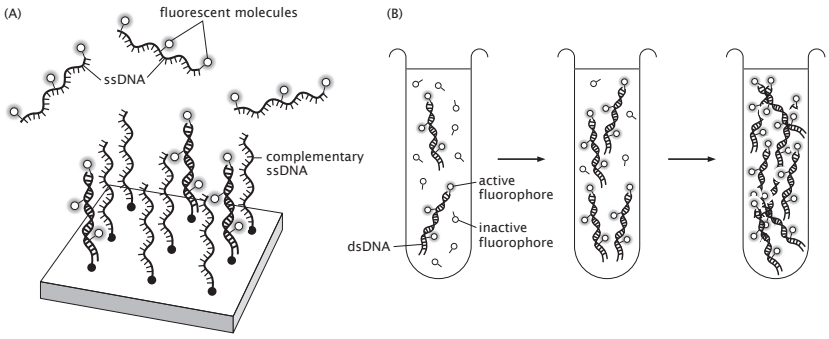

Figure 19.4: Measurement of the mRNA concentration. (A) A DNA microarray uses a collection of different molecules on the surface of a slide, each of which has a sequence complementary to the mRNA (or reverse trasnscribed ssDNA) associated with the gene of interest. By measuring how much hybridization there is between the sample and the molecules on the surface, one can count up the mRNA. (B) Quantitative PCR reactions use a template molecule which is produced from the mRNA using reverse transcription. The amount of template determines how many cycles of PCR it will take to reach a critical threshold.

19.2 Genetic Networks: Doing the Right Thing at the Right Time

In "thermodynamic" models of gene expression, attention is focused on the probability that the promoter is occupied by RNA polymerase. In section 6.1.2 (pg. 312), we showed how the "bare" problem of polymerase molecules interacting with DNA could be solved using these simple ideas from statistical mechanics. However, the shortcoming of that approach is that it ignores the existence of molecular gatekeepers which exercise strict control over the occupancy of promoters. We begin our dissection of gene expression with a consideration of these gatekeepers, which are known as transcription factors.

Promoter Occupancy Is Dictated by the Presence of Regulatory Proteins Called Transcription Factors

In figure 6.8 (pg. 313) we showed a cartoon of some gene of interest and the promoter and DNA upstream from it. As a first cut at the problem of promoter occupancy, we examined the probability of RNA polymerase binding as a competition between this promoter and nonspecific sites, both of which can be occupied by polymerase molecules. We now expand that discussion to account for the presence of a host of important accessory proteins whose presence on the scene can either enhance (activate) or reduce (repress) the probability of promoter occupancy.

As before, we focus primarily on bacteria. What this means concretely is that we will treat RNA polymerase as a single molecule and ask the precise mathematical (but biologically oversimplified) question of whether or not the promoter is occupied by such an RNA polymerase molecule. In the eukaryotic case, this question is less easily posed since the basal transcription apparatus consists of many parts, all of which need to be present simultaneously in order to start transcription.

19.2.1 The Molecular Implementation of Regulation: Promoters, Activators and Repressors

Repressor Molecules Are the Proteins That Implement Negative Control

One of the key control mechanisms of genetic networks is negative regulation of transcription. What this means is that the decision to express the gene of interest is made very early on in the set of processes leading from DNA to protein, namely, at the point where RNA is synthesized. If there is little or no mRNA which codes for a given protein, then clearly the ribosomes are in no position to produce the corresponding protein. The molecular implementation of negative control is through protein molecules known as repressors such as the Lac repressor introduced in figs. 4.13 (on pg. 200) and 8.19 (on pg. 423). In the case of bacteria, repressors can often be viewed as carrying out a blocking action in the sense that through DNA-protein interactions, they occupy the DNA in a region (called the operator) which overlaps the region where RNA polymerase binds (the promoter). The action of such repressor molecules is illustrated schematically in fig. 19.5. Note that the activity of repressors can, in turn, be regulated by small molecules, or inducers, that can bind and generate a conformational (or allosteric) change which alters the binding probability of the transcription factor for the DNA. Later in this chapter, we give a statistical mechanical interpretation of such cartoons.

It is important to recall that the point of cartoons like that in fig. 19.5 is to convey a conceptual picture and not a detailed molecular rendering of the explicit action of the various molecular participants. On the other hand, the fact that such cartoons can be constructed in the first place is often the result of having digested the significance of hard won structural determinations from x-ray crystallography. Indeed, sometimes, not only the structures of the bare repressors are known, but even the structures of these repressors when complexed with DNA. In fact, there are a variety of structural implementations of repression, some famed examples of which are shown in fig. 19.6.

Activators Are the Proteins That Implement Positive Control

A second key mechanism for altering the extent to which a given gene is expressed is known as positive regulation of transcription, or more provocatively, regulated recruitment. Here too, the idea is that the overall process of protein synthesis of a given gene product is regulated very early on where an

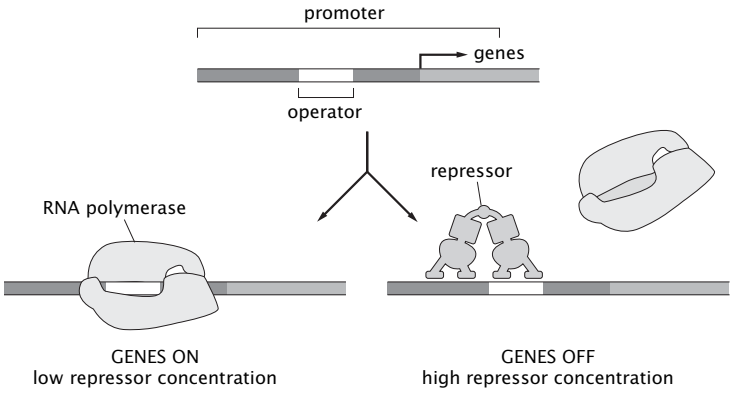

Figure 19.5: The process of repression. Cartoon representation showing the action of repressor molecules in forbidding RNA polymerase from binding to its promoter, or alternatively, if bound, from initiating transcription.

accessory molecule enhances the probability of promoter occupancy by RNA polymerase. This mechanism is built around the idea of proteins other than RNA polymerase that bind to DNA and increase the probability that the RNA polymerase itself will bind the promoter. Just as repressors interfere with the ability of RNA polymerase to bind to its promoter, activators bind in the vicinity of the promoter and have adhesive interactions with RNA polymerase itself which enhance the likelihood of RNA polymerase binding. The key point is that the RNA polymerase molecule interacts not only with the DNA to which it is bound, but also through "glue-like" interactions with the activator molecule. A cartoon representation of the process of regulated recruitment (i.e. activation) is shown in fig. 19.7.

As with the study of repressors, structural biology has permitted a range of atomic-level insights into the mechanisms of transcriptional activation. Fig. 19.8 provides a gallery of some key activators and reveals their sizes relative to the DNA molecule and illustrates the way in which they distort and occlude the DNA when bound.

Genes Can Be Regulated During Processes Other Than Transcription

Our discussion will focus primarily on transcriptional regulation. On the other hand, as shown in fig. 6.7 (pg. 313), there are many points along the route connecting DNA to its protein products where gene expression can be controlled. Two of the most obvious and important ways in which the concentration of active protein are controlled is through post-translational modifications such as phosphorylation and protein degradation. For the moment, we focus on the way in which p_{bound} (the probability that the promoter is occupied by RNA polymerase) can be altered through the action of transcription factors such as repressors and activators.

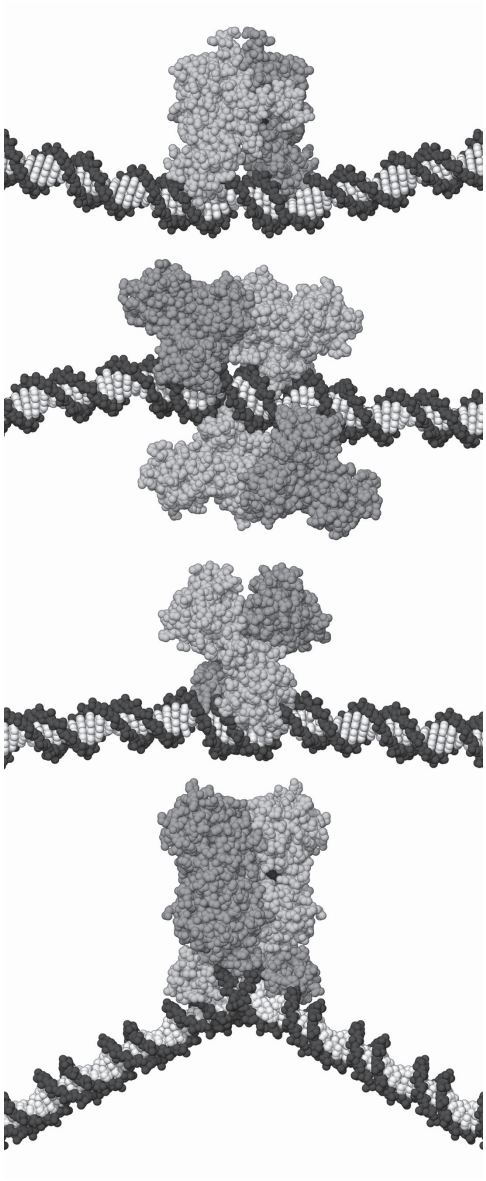

Figure 19.6: Examples of repressor molecules interacting with DNA. From top to bottom the relevant repressors are FadR (pdb 1HW2) , IdeR (pdb 1U8R), TetR (pdb 1QPI) and PurR (pdb 1PNR). The point of the figure is to give an impression of the relative sizes of repressors and their target regions on DNA and to illustrate how these transcription factors deform the DNA double helix in the vicinity of their binding site. These drawings are renditions of actual structures from x-ray crystallography. (Courtesy of David Goodsell.)

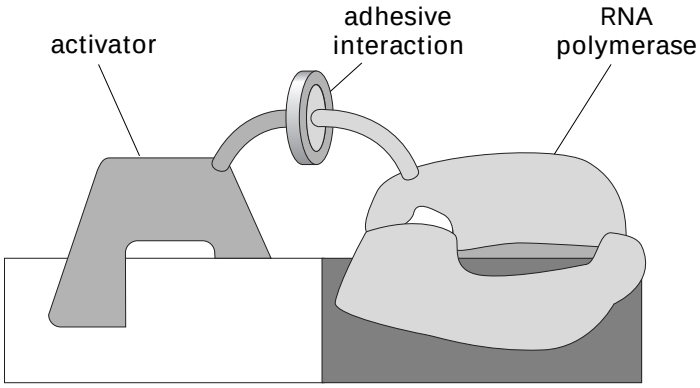

Figure 19.7: The process of activation. Schematic of the way in which activator molecules can recruit the transcription apparatus. Though both the activator and RNA polymerase have their own private interaction energies with the DNA, the enhancement in their occupancies is mediated by the adhesive interaction between them.

19.2.2 The Mathematics of Recruitment and Rejection

Recruitment of Proteins Reflects Cooperativity Between Different DNA Binding Proteins

One of the key general ideas that pervades the description of transcriptional control (and beyond) is the idea of molecular recruitment. In the anthropomorphic terms suggested by the word "recruitment", the basic idea is that a given molecule which is bound on DNA summons some second molecule to the DNA where it can then perform its task. For example, we think of RNA polymerase being summoned (and vice versa) by some activator molecule such as a transcription factor and exemplified by the CAP protein in the case of the *lac* operon. Though this colorful language is suggestive and conjures up a useful physical picture, from the perspective of the rules of statistical mechanics, this is nothing more than the well-worn idea of cooperativity cloaked in different verbal clothing.

Activators are proteins which regulate transcription by binding to a specific site on the DNA so as to recruit an RNA polymerase onto a nearby promoter site. It has been suggested that weak, non-specific binding of the activator protein and the RNA polymerase can greatly enhance the probability of the polymerase binding to DNA, even for very low concentrations of activator proteins typical of the cellular environment. To assess the feasibility of this strategy we compute the probability of the polymerase being bound in the presence of an activator protein using a simple model which is depicted in cartoon form in fig. 19.9. The basic point of this cartoon is to show the different allowed states of polymerase and activator molecules and to use this enumeration of states to compute the

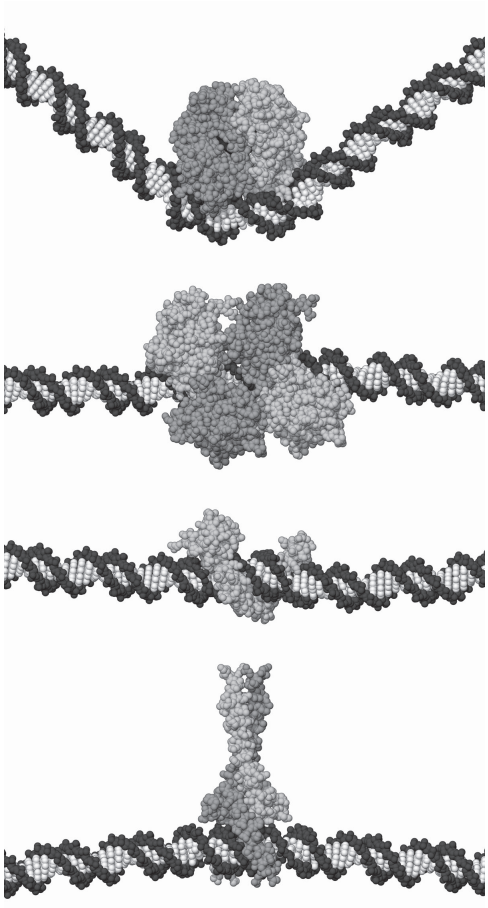

Figure 19.8: Structures of activator molecules. From top to bottom the activators are: CAP (pdb 1CGP), p53 (pdb 1TUP) tumor suppressor, zinc finger DNA binding domain (pdb 2GLI) and leucine zipper DNA binding domain (pdb 1AN2). (Courtesy of David Goodsell.)

probability that the promoter will be occupied. Indeed, this is the same "states and weights" mentality used throughout the book.

The first step in our analysis of this problem is to write the total partition function. Note that the partition function is obtained by summing over all of the eventualities associated with the activators and polymerase molecules being distributed on the DNA (both nonspecific sites and the promoter). As shown in fig. 19.9, there are four classes of outcomes, namely, both the activator site and promoter unoccupied, just the promoter occupied by polymerase, just the activator binding site occupied by activator and finally, both of the specific sites occupied. This is represented mathematically as

$$\begin{aligned} Z_{tot}(P,A;N_{NS}) &= \underbrace{Z(P,A;N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P-1,A;N_{NS})e^{-\beta\varepsilon_{pd}^{S}}}_{\text{RNAP}} \\ &+ \underbrace{Z(P,A-1;N_{NS})e^{-\beta\varepsilon_{ad}^{S}}}_{\text{activator}} \\ &+ \underbrace{Z(P-1,A-1;N_{NS})e^{-\beta(\varepsilon_{ad}^{S}+\varepsilon_{pd}^{S}+\varepsilon_{pa})}}_{\text{RNAP + activator}}. \end{aligned} \tag{19.1}$$

Note that notationally the meaning of $Z(P,A;N_{NS})$ is that it is the partition function for P polymerase molecules and A activator molecules to be bound on the N_{NS} nonspecific sites and is given by

$$Z(P,A;N_{NS}) = \underbrace{\frac{N_{NS}!}{P!A!(N_{NS}-P-A)!}}_{\text{number of arrangements}} \underbrace{e^{-\beta P\varepsilon_{pd}^{NS}}e^{-\beta A\varepsilon_{ad}^{NS}}}_{\text{weight of each state}}. \tag{19.2}$$

We have also introduced the notation ε_{pa} to account for the "glue" interaction between the polymerase and activator. Like in sec. 6.1.2 (pg. 312) for the case of RNA polymerase, we introduce ε_{ad}^{S} and ε_{ad}^{NS} to characterize the binding energy of activator with its specific and nonspecific DNA targets, respectively. Our expression involves a number of terms of the general form

$$\frac{N_{NS}!}{P!A!(N_{NS}-A-P)!}e^{-\beta P\varepsilon_{pd}^{NS}}e^{-\beta A\varepsilon_{ad}^{NS}}. \tag{19.3}$$

As we did earlier, we invoke a simplifying strategy which depends upon the fact that $N_{NS} \gg A+P$ and hence, there will be almost zero chance of RNA polymerase and the activator finding each other on the same non-specific site on the DNA. This permits the approximation $\frac{N_{NS}!}{(N_{NS}-A-P)!} = (N_{NS})^{A+P}$ introduced in section 6.1.2 (see pg. 312).

To compute the probability of promoter occupancy, we construct the ratio of all of those outcomes that are favorable (i.e. polymerase bound to the promoter) to the total set of outcomes ($Z_{tot}(P,A;N_{NS})$), namely,

$$p_{bound}(P,A;N_{NS}) = \frac{Z(P-1,A;N_{NS})e^{-\beta\varepsilon_{pd}^{S}} + Z(P-1,A-1;N_{NS})e^{-\beta(\varepsilon_{ad}^{S}+\varepsilon_{pd}^{S}+\varepsilon_{pa})}}{Z_{tot}(P,A;N_{NS})}. \tag{19.4}$$

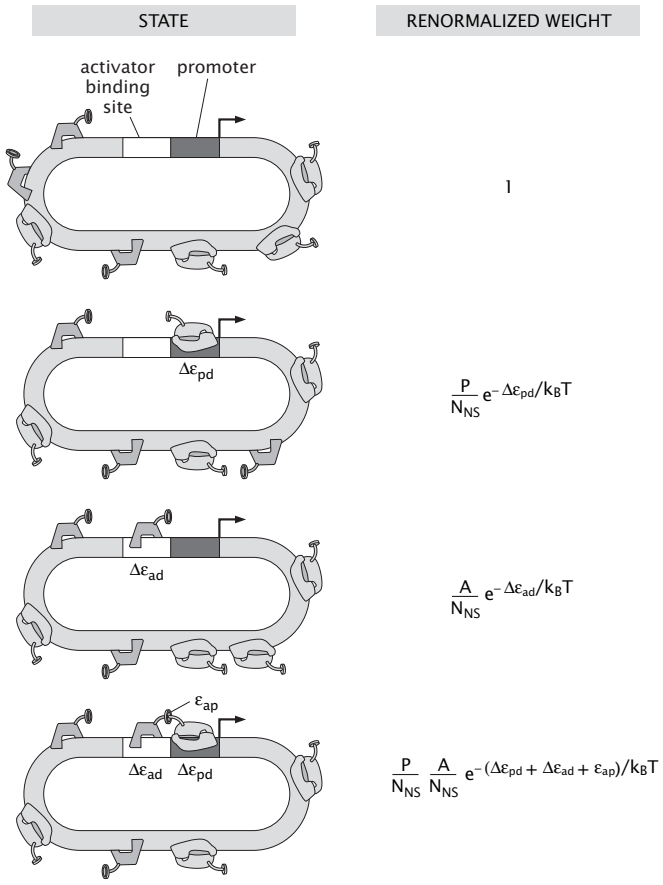

Figure 19.9: Schematic representation of the simple statistical mechanical model of recruitment. The states and weights diagram shows the different binding scenarios in the vicinity of the promoter of interest and the corresponding renormalized statistical weights obtained using statistical mechanics. We make the simplifying assumption that the nonspecific binding energy is *constant*. The large circular DNA is a cartoon representation of the bacterial genome.

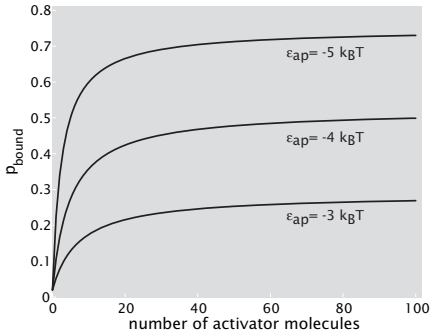

Figure 19.10: Illustration of the recruitment concept. This plot shows the probability of binding when the number of polymerase molecules is $P = 500$ and the binding parameters are $\Delta\varepsilon_{pd} = -5.3\ k_BT$ and $\Delta\varepsilon_{ad} = -13.12\ k_BT$. The three curves correspond to different choices of the adhesive interaction energy between polymerase and the activator.

We now propose to simplify this result by dividing top and bottom by the numerator resulting in

$$p_{bound}(P, A; N_{NS}) = \frac{1}{1 + \frac{N_{NS}}{PF_{reg}(A)} e^{\beta\Delta\varepsilon_{pd}}}, \tag{19.5}$$

where we introduce the regulation factor, $F_{reg}(A)$ which is given by

$$F_{reg}(A) = \frac{1 + \frac{A}{N_{NS}} e^{-\beta\Delta\varepsilon_{ad}} e^{-\beta\varepsilon_{ap}}}{1 + \frac{A}{N_{NS}} e^{-\beta\Delta\varepsilon_{ad}}} \tag{19.6}$$

and where we have defined $\Delta\varepsilon_{pd} = \varepsilon^S_{pd} - \varepsilon^{NS}_{pd}$ and $\Delta\varepsilon_{ad} = \varepsilon^S_{ad} - \varepsilon^{NS}_{ad}$. The details of the derivation are left to the problems at the end of the chapter. Note that in the limit that the adhesive interaction between polymerase and activator goes to zero, the regulation factor itself goes to unity. Further, note that for negative values of this adhesive interaction (i.e. activator and polymerase like to be near each other) the regulation factor is greater than one which is translated into an effective increase in the number of polymerase molecules. The probability of RNA polymerase binding as a function of the number of activators is plotted in fig. 19.10.

The Regulation Factor Dictates How the Bare RNA polymerase Binding Probability Is Altered by Transcription Factors

One of the intriguing claims that we will make is that a simple change in the effective number of RNA polymerase molecules ($P \rightarrow P_{eff}$) will suffice to capture the action of regulatory chaperones such as activators and repressors.

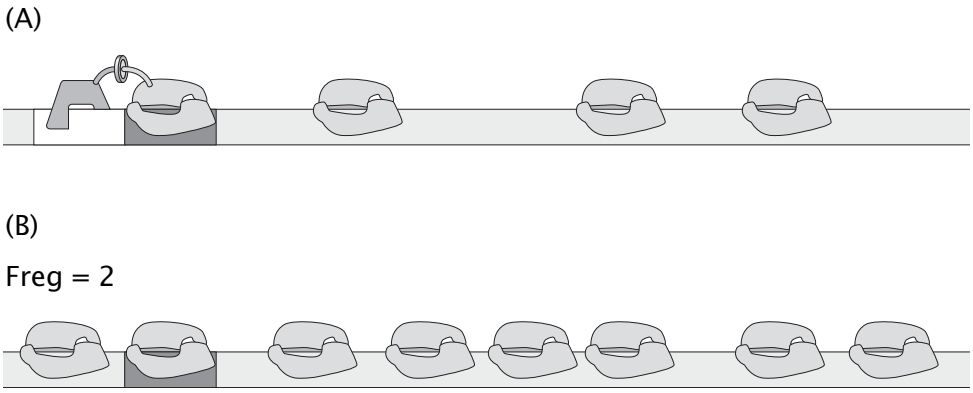

Figure 19.11: Regulation factor and the effective number of polymerase molecules. The presence of activators is equivalent to a problem with just polymerase but a larger number of them. (A) The "bare" problem with activators and polymerase present. (B) The "effective" problem in which the presence of activators is treated as a change in the number of polymerase molecules.

This interpretation of the meaning of the regulation factor is shown in fig. 19.11. As a result of the presence of activators, it is as though the number of RNA polymerase molecules has been changed from P to $F_{reg}P$. For the case of activators, the regulation factor is greater than one and leads to an effective increase in the number of polymerase molecules. By way of contrast, below we will show that when repressors are present, they result in a regulation factor that is less than one and a concomitant decrease in the effective number of polymerase molecules.

In order for our calculations to really carry weight, we need to examine what they have to say about experiments. One of the primary measurables in *in vivo* experiments on regulation is the relative expression for cases in which the transcription factor of interest is present or not. This qualitative notion is made quantitative by introducing the idea of the fold change in activity defined in the activation setting as

$$\text{fold-change} = \frac{p_{bound}(A \neq 0)}{p_{bound}(A = 0)} = \frac{1 + \frac{N_{NS}}{P F_{reg}(A)} e^{\beta \Delta \varepsilon_{pd}}}{1 + \frac{N_{NS}}{P} e^{\beta \Delta \varepsilon_{pd}}}. \quad (19.7)$$

What this expression reveals is how much more expression there is in the presence of activators relative to the "basal" state in which there is no activation.

As before, an inherent assumption in this analysis is the idea that the relative change in what is measured (e.g. protein product, mRNA concentration or promoter occupancy) is equal to the relative change in p_{bound}. Figure 19.12 illustrates the fold-change in gene expression for the problem of simple activation with a choice of parameters dictated by *in vitro* experiments for a value of $\Delta\varepsilon_{ad}$

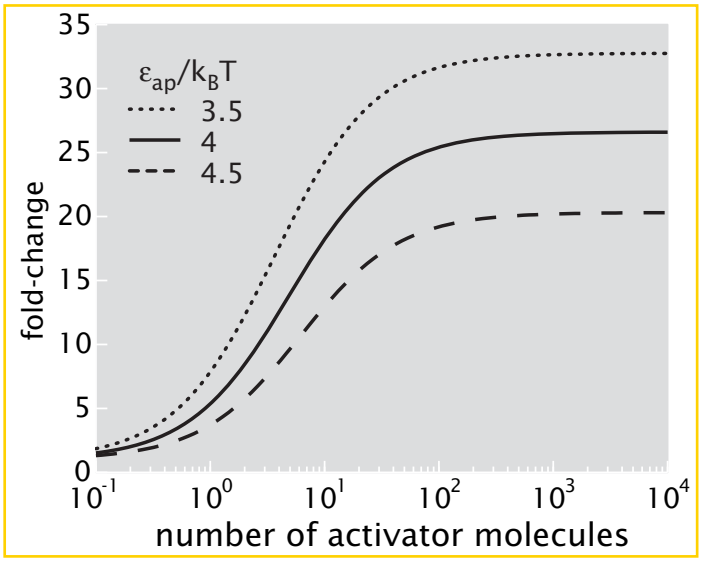

Figure 19.12: Fold-change due to activators. Fold-change in gene expression as a function of the number of activators for different activator-RNA polymerase interaction energies using $P = 500$, $\Delta\varepsilon_{pd} = -5.3\ k_BT$ and $\Delta\varepsilon_{ad} = -13.12\ k_BT$ based on *in vitro* measurements.

in conjunction with an educated guess for ε_{ap} which results in typical fold-changes in activity reported *in vivo* of about 50. Note that a weak promoter satisfies the condition $\frac{N_{NS}}{P} e^{\beta\Delta\varepsilon_{pd}} \gg 1$, which implies that the fold-change in activity can be rewritten as

$$\text{fold-change} \simeq \left(F_{reg}(A)\right)^{-1}. \tag{19.8}$$

Here we are also assumed that $\frac{N_{NS}}{P F_{reg}} e^{\beta\Delta\varepsilon_{pd}} \gg 1$, which means that the promoter isn't too strong even in the regulated case. The conclusion is that in the case of a weak promoter the actual details of the promoter such as its binding energy, factor out of the problem.

Activator Bypass Experiments Show That Activators Work by Recruitment

The simple picture of regulated recruitment introduced here is based in part upon a series of classic experiments known as activator bypass experiments. The key idea of such experiments is shown in fig. 19.13. These experiments involve a mix and match approach where the DNA binding domain from one protein is fused with the activator domain of a second protein. A second version of this experiment is based upon direct tethering of the activator and the polymerase. After making the activator bypass constructs, it was found that the gene of interest was still activated. Our ambition here is to consider these experiments

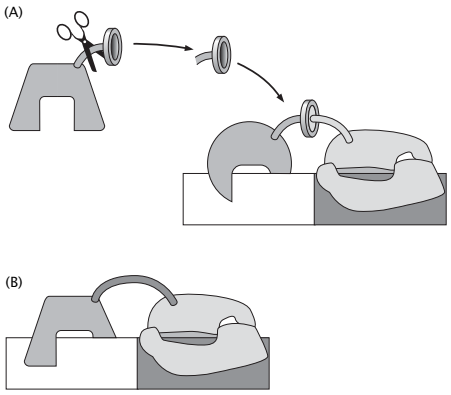

Figure 19.13: Schematic of activator bypass experiments. (A) Activator bypass type 1 in which activation is mediated by proteins with designer DNA binding regions, (B) activator bypass type 2 in which the activator is tethered directly to polymerase.

more quantitatively and to note that if viewed from a mathematical perspective, these two classes of experiments lead to different quantitative outcomes which can be used to further test the full range of validity of the notion of regulated recruitment.

We have already worked out the regulation factor that is associated with activator bypass type 1 experiments. The only change relative to eqn. 19.6 is the fact that by using different proteins, quantities such as $\Delta\varepsilon_{ad}$ and ε_{pa} will have different numerical values which means that the actual level of activation can be different in this experiment relative to its "wild type" value. On the other hand, the entire functional form for the regulation factor is different in the case of activator bypass type 2. In this case, there are only two states we really need consider - polymerase with tethered activator bound at the promoter or not with weights $(P/N_{NS})e^{-\beta(\Delta\varepsilon_{pd}+\Delta\varepsilon_{ad})}$ and 1, respectively. This implies that the probability that polymerase will be bound is

$$p_{bound}(P;N_{NS}) = \frac{1}{1+\frac{N_{NS}}{P}e^{\beta\Delta\varepsilon_{ad}}e^{\beta\Delta\varepsilon_{pd}}}. \tag{19.9}$$

This implies, in turn, that the regulation factor takes the particularly simple form

$$F_{reg} = e^{-\beta\Delta\varepsilon_{ad}}, \tag{19.10}$$

which amounts to the statement that the effective binding energy of polymerase is shifted and nothing more.

Repressor Molecules Reduce the Probability Polymerase Will Bind to the Promoter

The same logic that was introduced above to consider the case of pure activation (i.e. recruitment) can be brought to bear on the problem of repression. Once again, we are faced with considering all of the ways of distributing the repressor and RNA polymerase molecules and it is convenient to introduce the partition function associated with the binding of these molecules to nonspecific sites as

$$Z(P, R : N_{NS}) = \frac{N_{NS}!}{P!R!(N_{NS} - P - R)!} e^{-\beta P \varepsilon_{pd}^{NS}} e^{-\beta R \varepsilon_{rd}^{NS}}, \tag{19.11}$$

which is formally identical to eqn. 19.2, but where we have introduced the notation ε_{rd}^{NS} to describe the nonspecific binding of repressor to DNA (ε_{rd}^{S} will be reserved for the specific binding energy of repressor to its operator). In order to write the *total* partition function for all the allowed states we now need to sum over the states in which the promoter is occupied either by a repressor molecule or by an RNA polymerase molecule. The set of allowed states in this simple model as well as their associated weights is shown in fig. 19.14. Note that in considering this particular model, we do not enter into structural fine points such as whether or not the RNA polymerase can be on its promoter at the same time as the repressor is bound to its operator - the model is intended to be the simplest treatment of the statistical mechanics of the competition between repressors and RNA polymerase.

The total partition function is given by

$$Z_{tot}(P, R; N_{NS}) = \underbrace{Z(P, R; N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P-1, R; N_{NS}) e^{-\beta \varepsilon_{pd}^{S}}}_{\text{RNAP on promoter}} + \underbrace{Z(P, R-1; N_{NS}) e^{-\beta \varepsilon_{rd}^{S}}}_{\text{repressor on promoter}}. \tag{19.12}$$

This result now provides us with the tools in order to evaluate the probability that the promoter will be occupied by RNA polymerase. This probability is given by the ratio of the favorable outcomes to all of the outcomes. In mathematical terms, that is

$$p_{bound}(P, R; N_{NS}) = \frac{Z(P-1, R; N_{NS}) e^{-\beta \varepsilon_{pd}^{S}}}{Z(P, R; N_{NS}) + Z(P-1, R; N_{NS}) e^{-\beta \varepsilon_{pd}^{S}} + Z(P, R-1; N_{NS}) e^{-\beta \varepsilon_{rd}^{S}}}. \tag{19.13}$$

As argued above, this result can be rewritten in compact form using the regulation factor by dividing top and bottom by $Z(P-1, R; N_{NS}) e^{-\beta \varepsilon_{pd}^{S}}$ and by invoking the approximation

$$\frac{N_{NS}!}{P!R!(N_{NS} - P - R)!} \approx \frac{N_{NS}^{P}}{P!} \frac{N_{NS}^{R}}{R!} \tag{19.14}$$

which amounts to the physical statement that there are so few polymerase and repressor molecules in comparison with the number of available sites, N_{NS}, that each of these molecules can more or less fully explore those N_{NS} sites. The resulting probability is

$$p_{bound}(P, R; N_{NS}) = \frac{1}{1 + \frac{\mathrm{N_{NS}}}{\mathrm{P}} \mathrm{e}^{\beta(\varepsilon_{\mathrm{pd}}^{\mathrm{S}} - \varepsilon_{\mathrm{pd}}^{\mathrm{NS}})} (1 + \frac{\mathrm{R}}{\mathrm{N_{NS}}} \mathrm{e}^{-\beta(\varepsilon_{\mathrm{rd}}^{\mathrm{S}} - \varepsilon_{\mathrm{rd}}^{\mathrm{NS}})})}. \tag{19.15}$$

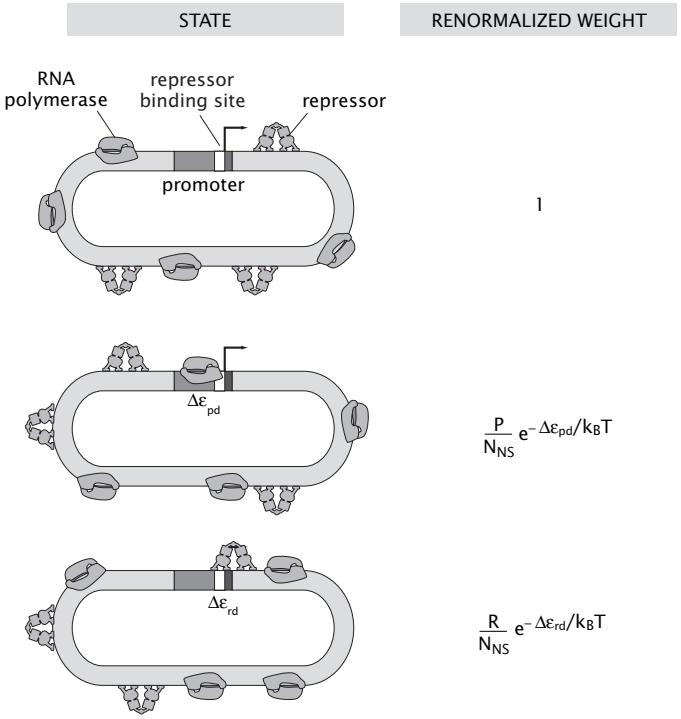

Figure 19.14: States and weights for the case of simple repression. The states of promoter occupancy are empty promoter, RNA polymerase on the promoter and repressor on the promoter.

This result can be couched in regulation factor language with the observation that the regulation factor itself is given by

$$F_{reg}(R) = \left(1 + \frac{R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}}\right)^{-1}, \tag{19.16}$$

with $\Delta\varepsilon_{rd} = \varepsilon_{rd}^{S} - \varepsilon_{rd}^{NS}$. Note that the regulation factor in the case of repression satisfies the inequality $F_{reg} < 1$, which can be interpreted as a reduction in the effective number of RNA polymerase molecules. We explore this in more detail in section 19.2.5 when discussing the particular case of the *lac* operon.

19.2.3 Transcriptional Regulation By the Numbers: Binding Energies and Equilibrium Constants

We have heard it said that "physics isn't worth a damn unless you put in some numbers!" The abstract expressions obtained so far are much more interesting when viewed through the prism of particular measurements. Binding energies quantify the affinity of RNA polymerase or transcription factors for their DNA targets. In particular, RNA polymerase and transcription factors perform molecular recognition as a result of a rank ordering of their preferences for different sequences of nucleotides. Indeed, the sequence associated with a given promoter distinguishes it from some random sequence to which RNA polymerase would bind with a non-specific binding energy ε_{pd}^{NS}. Specific binding energies can also be tuned. For example, even though there might be one very strong consensus promoter, that binding strength can be reduced by introducing mismatches in the sequence. A strong promoter, with a p_{bound} close to one will have a strong level of expression. On the other hand, by weakening a given promoter, cells can broaden their dynamic range by introducing a co-dependency on a battery of transcription factors which effectively tune the range of binding affinities and permit the regulation of promoter occupancy.

Equilibrium Constants Can Be Used To Determine Regulation Factors

In order to compute the regulation factors for the various regulatory scenarios under consideration in this chapter, we need to make estimates for the energy associated with binding protein X to the DNA both specifically and non-specifically; protein X can be a repressor or an activator. Binding energies are determined indirectly in experiments which measure the equilibrium constant for binding X to DNA (D). In particular, we consider the reaction

$$\mathrm{X} + \mathrm{D} \rightleftharpoons \mathrm{XD} \tag{19.17}$$

with an equilibrium binding constant

$$K_{\mathrm{X}}^{(\mathrm{bind})} = \frac{[\mathrm{XD}]}{[\mathrm{X}][\mathrm{D}]} \ . \tag{19.18}$$

Here $[\cdots]$ denotes concentrations of the various species taking part in the reaction.

When a single X binds to DNA there is an overall change of the free energy $\Delta f_{\rm XD}$. The more negative this quantity is, the more likely X will be bound to DNA. Similarly, a larger $K_{\rm X}^{\rm (bind)}$ implies that the bound state is more likely. More precisely, the probability that a particular binding site on the DNA is occupied is equal to the ratio of the number of occupied sites to the total number of sites as was first introduced in section 6.4.1 (pg. 345). In terms of concentrations, this can be written

$$p_{\rm bound} = \frac{[\rm XD]}{[\rm D]+[\rm XD]} = \frac{K_{\rm X}^{\rm (bind)}[\rm X]}{1+K_{\rm X}^{\rm (bind)}[\rm X]} \ , \tag{19.19}$$

where the final expression follows from eqn. 19.18. On the other hand, given that there are $[\rm X]V_{\rm cell}$ copies of protein X in the cell ($V_{\rm cell}$ is the volume of the cell), the probability of a DNA binding site to be occupied is

$$p_{\rm bound} = \frac{[\rm X]V_{\rm cell}e^{-\beta\Delta f_{\rm XD}}}{1+[\rm X]V_{\rm cell}e^{-\beta\Delta f_{\rm XD}}} \ . \tag{19.20}$$

Comparison of the two expressions for $p_{\rm bound}$ allows us to relate the microscopic and macroscopic views of binding through the relation

$$\frac{K_{\rm X}^{\rm (bind)}}{V_{\rm cell}} = e^{-\beta\Delta f_{\rm XD}}. \tag{19.21}$$

Using this relation we can compute the binding free energies for RNA polymerase and the various transcription factors in *E. coli* as the reader is invited to explore in the problems at the end of the chapter. Presently, we use these ideas to tackle the *lac* operon which features both positive and negative regulation.

19.2.4 A Simple Statistical Mechanics Model of Positive and Negative Regulation

Real regulatory architectures in cells often involve both repression and activation simultaneously. In this case, we consider the five distinct outcomes shown in fig. 19.15 and captured through the total partition function

$$\begin{aligned} Z_{tot}(P,A,R;N_{NS}) &= \underbrace{Z(P,A,R;N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P-1,A,R;N_{NS})e^{-\beta\varepsilon^S_{pd}}}_{\text{RNAP}} \\ &+ \underbrace{Z(P,A-1,R;N_{NS})e^{-\beta\varepsilon^S_{ad}}}_{\text{activator}} + \underbrace{Z(P-1,A-1,R;N_{NS})e^{-\beta(\varepsilon^S_{ad}+\varepsilon^S_{pd}+\varepsilon_{pa})}}_{\text{RNAP + activator}} \\ &+ \underbrace{Z(P,A,R-1;N_{NS})e^{-\beta\varepsilon^S_{rd}}}_{\text{repressor}} + \underbrace{Z(P,A-1,R-1;N_{NS})e^{-\beta(\varepsilon^S_{ad}+\varepsilon^S_{rd})}}_{\text{activator + repressor}} \end{aligned} \tag{19.22}$$

Note that the cartoon shows a schematic representation of the different ways that the region in the vicinity of the promoter can be occupied and what the statistical weights are of each such state of occupancy. We can compute the probability of RNA polymerase binding by considering the ratio of favorable outcomes to the total partition function, resulting in

$$p_{bound}(P,A,R;N_{NS}) = \frac{Z(P-1,A,R;N_{NS})e^{-\beta\varepsilon_{pd}^{S}} + Z(P-1,A-1,R;N_{NS})e^{-\beta(\varepsilon_{ad}^{S}+\varepsilon_{pd}^{S}+\varepsilon_{pa})}}{Z_{tot}(P,A,R;N_{NS})}. \tag{19.23}$$

As before, perhaps the simplest way to interpret this result is with reference to the regulation factor, resulting in

$$p_{bound}(P,A,R;N_{NS}) = \frac{1}{1+\frac{N_{NS}}{PF_{reg}(A,R)}e^{\beta(\varepsilon_{pd}^{S}-\varepsilon_{pd}^{NS})}}, \tag{19.24}$$

where the regulation factor itself is now a function of both the number of activators A and the number of repressors R. In particular, the regulation factor is given by

$$F_{reg}(A,R) = \frac{1+\frac{A}{N_{NS}}e^{-\beta(\Delta\varepsilon_{ad}+\varepsilon_{ap})}}{1+\frac{A}{N_{NS}}e^{-\beta\Delta\varepsilon_{ad}}+\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}+\frac{A}{N_{NS}}\frac{R}{N_{NS}}e^{-\beta(\Delta\varepsilon_{ad}+\Delta\varepsilon_{pd})}}. \tag{19.25}$$

The fold-change change in gene expression due to this regulatory architecture in the weak promoter approximation is shown in fig. 19.16. The objective of this figure is to illustrate the combinatorial control that can be reached when different transcription factors act in unison. Perhaps nowhere is this interplay of negative and positive regulation more well known than in our old friend, the *lac* operon.

19.2.5 The *lac* Operon

Both repression and activation are a key part of the equipment of bacteria. Perhaps the most famous example of these effects is provided by the *lac* operon and shown in fig. 4.15 (pg. 204). Indeed, the *lac* operon has served as one of the central workhorses of the entire book and the present section is the *denouement* of that discussion. In this case, the activator is the catabolite activator protein (CAP), also known as cyclic AMP receptor protein (CRP). In order to be able to recruit RNA polymerase, CRP has to be bound to cAMP, a molecule whose concentration goes up when the amount of glucose decreases. The repressor, known as Lac repressor, decreases the level of transcription unless it is bound to allolactose, which is a byproduct of lactose metabolism.

The *lac* Operon Has Features of Both Negative and Positive Regulation

Recall that the *lac* operon oversees the management of the enzymes that are responsible for lactose uptake and digestion. In particular, when *E. coli* cells

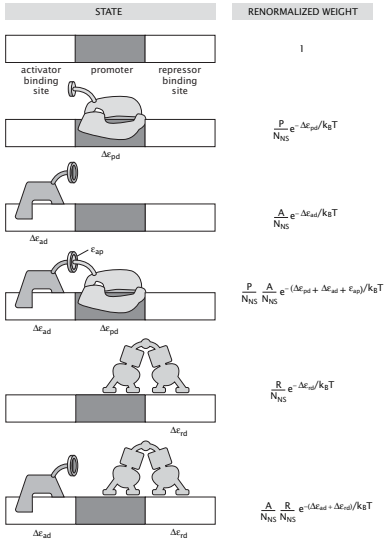

Figure 19.15: Schematic representation of the simple statistical mechanical model of recruitment and repression.

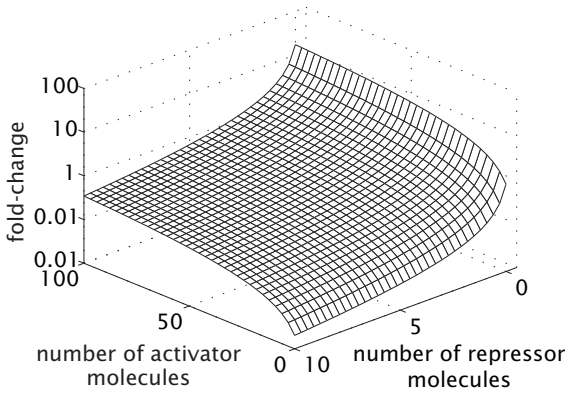

Figure 19.16: Combined regulation by repressor and activator. The fold-change in gene expression as a function of the number of transcription factors shows their combinatorial action. The parameters used are $\Delta\varepsilon_{ad} = -10\ k_BT$, $\varepsilon_{ap} = -3.9\ k_BT$ and $\Delta\varepsilon_{rd} = -16.9\ k_BT$.

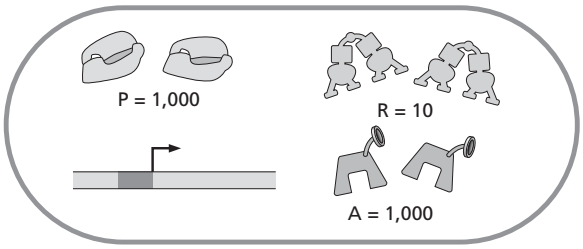

Figure 19.17: Census of the relevant molecular actors in the *lac* operon. The figure shows a rough estimate of the number of polymerase molecules, activators and repressors associated with the *lac* operon.

find themselves simultaneously deprived of glucose and supplied with lactose, the genes of the *lac* operon are turned on so as to take metabolic advantage of the lactose. We have already described the way in which the Lac repressor forbids transcription of the genes associated with lactose digestion by binding on its operator. However, our earlier discussion was a bit too blithe since we said nothing of what happens in the case where glucose and lactose are simultaneously available. If we are to adopt the picture of negative control described above, then our expectation would be that in this case there should be substantive transcription of the genes of the *lac* operon. However, there is a second element of positive control which completes the story. In particular, in the *absence* of glucose, the activator, CAP (catabolite activator protein) binds to a site near the promoter (the RNA polymerase binding site) as shown in fig. 4.15 (pg. 204) and "recruits" RNA polymerase to the promoter. The census shown in fig. 19.17 gives a rough impression of the number of copies of some of the key molecules associated with the *lac* operon and illustrates the striking fact that some of the transcription factors exist with as few as ten copies.

The geometry of the regulatory landscape for the *lac* operon is shown in fig. 19.18. Our discussion of fig. 4.15 was oversimplified in the sense that we ignored the presence of auxiliary binding sites for the Lac repressor which are revealed in fig. 19.18. In particular, there are two other binding sites for the Lac repressor. Specifically, there is a binding site known as O2 located 401 bp downstream from O1 and a second such site known as O3 situated 92 bp upstream. Part of our discussion will center on the subtle ways in which repression takes place in this system. Recall that the repressor itself is a tetramer with two "reading heads" that can each bind to a different operator, looping out the intervening DNA.

One of the most important roles for models like those described here is in providing a conceptual framework for thinking about both *in vivo* and *in vitro* data and in suggesting new experiments. A particularly compelling class of *in vivo* experiments using the *lac* operon measured the repression as a function of the strength and placement of the operator sites which are the targets of Lac repressor. In particular, *E. coli* cells were created which had only one operator

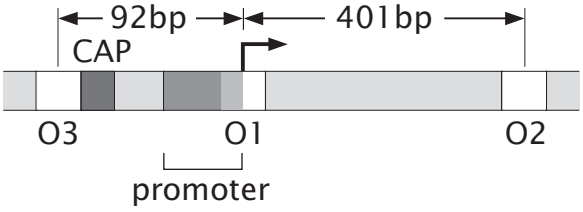

Figure 19.18: Position of the three *lac* operators and CAP binding site relative to the promoter. O1 is the main operator while O2 and O3 are auxiliary binding sites for Lac repressor and are associated with DNA looping.

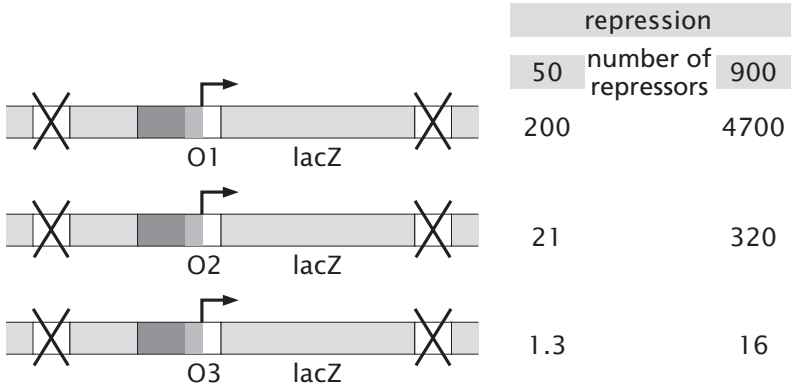

Figure 19.19: Repression in the *lac* operon. The DNA constructs used in these experiments deleted the auxiliary binding sites for repressor and tuned the strength of the main repressor binding site. Repression was measured in each construct for two different concentrations of Lac repressor. (Adapted from S. Oehler *et al.*, EMBO J., 13:3348, 1994.)

for Lac repressor as well as mutants with different spacings between operators (a topic we return to below). The first set of experiments we consider are those in which only one operator was present for Lac repressor binding as shown in fig. 19.19. In these experiments, the repression was measured for cases in which the promoter was repressed by each of the operators O_1, O_2 and O_3 individually. From the standpoint of the models considered here, all that is different from one experiment to the next is the binding energy of repressor for the DNA.

Recall that for a single repressor, the regulation factor is given by eqn. 19.16. What is measured in the experiment is the ratio of the level of gene expression in the absence of repressor to that in the presence of repressor. For the purposes of our model, we replace this definition which is based upon a measure of protein content (i.e. the product of the gene) with a definition based upon examining the probability that the promoter is occupied by RNA polymerase.

The implicit assumption here is that the protein content is linearly related to the probability of promoter occupancy. More precisely, we define repression as the ratio between the probability of binding of RNA polymerase to the relevant promoter in the absence of repressor to the probability of such binding in the presence of repressor, namely

$$\text{Repression} = \frac{p_{bound}(R=0)}{p_{bound}(R\neq 0)}. \tag{19.26}$$

Concretely, this result depends on the number of repressors (R) and their energy of binding to DNA. If we substitute for p_{bound} using eqn. 19.15 we find the repression can be written as

$$\text{Repression}(R) = \frac{1+\frac{P}{N_{NS}}e^{-\beta\Delta\varepsilon_{pd}}+\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}{1+\frac{P}{N_{NS}}e^{-\beta\Delta\varepsilon_{pd}}}. \tag{19.27}$$

For the case of a weak promoter this implies, in turn, that the repression level can be written as

$$\text{Repression}(R) = [\text{fold-change}(R)]^{-1} \simeq [F_{reg}(R)]^{-1} = 1+\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}. \tag{19.28}$$

One of the interesting opportunities afforded by this expression is the possibility of a direct confrontation with experimental data such as is shown in fig. 19.19.

In particular, the data of fig. 19.19 permits us to determine the only unknown in our expression for the repression, namely, the energy parameter $\Delta\varepsilon_{rd}$. Since the data reflects three different choices of binding strength, we find three different binding energies ($\Delta\varepsilon_{rd}$ =-16.9, -14.4 and -11.2 k_BT for O_1, O_2 and O_3, respectively). With these energies in hand, we can predict the outcome of repression measurements in which the number of repressors is tuned to other values as shown in fig. 19.20. Note that once the binding energy difference has been estimated using one data point, it leads to a prediction for the behavior of the system for different numbers of repressor in the cell and will serve as the basis for our analysis of the two-operator case as well.

The Free Energy of DNA Looping Affects the Repression of the *lac* Operon

Our discussion of the *lac* operon from the statistical mechanics perspective has thus far ignored one of the more intriguing features of this system, namely, the presence of DNA looping. The behavior of the *lac* operon has been examined in great detail both *in vitro* and *in vivo*. One beautiful set of experiments that are particularly enlightening with reference to the class of models we have described thus far in the chapter examine the repression of the *Lac* operon as a function of the spacing between the DNA binding sites (the operators) for Lac repressor.

The data on repression as a function of interoperator spacing was introduced in fig. 1.11 (pg. 42) as an example of the sophisticated quantitative data that

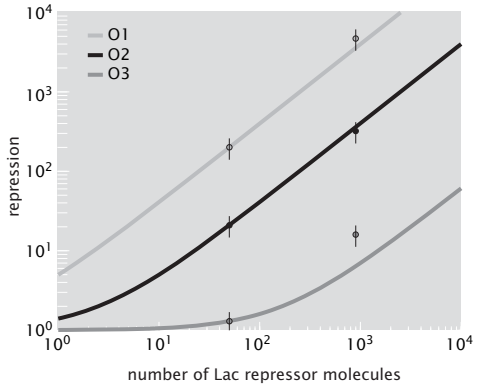

Figure 19.20: Repression model for the *lac* operon. Each curve shows how repression varies as a function of the number of repressor molecules in the cell. Different curves correspond to different main binding sites (operators) for the Lac repressor. (Data from S. Oehler *et al.*, EMBO J., 13:3348, 1994.)

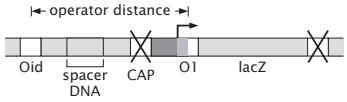

Figure 19.21: Construct used to measure repression in the presence of looping. The binding site for the activator CRP (shown as CAP in the diagram) was deleted as was the theird repressor binding site. (Adapted from J. Müller *et al.*, J. Mol. Biol., 257:21, 1996.)

exists on biological systems in general, and gene expression in particular. These beautiful experiments and others like them show a systematic trend in the promoter activity of the genes in question as a function of the distance between the binding sites for the repressor under consideration. One particular telling feature of such data is the periodicity which results from the twist degrees of freedom and which reflect the need for particular faces of the DNA to be aligned in order to form a loop.

Figure 19.21 shows the DNA construct that was used to examine the *in vivo* consequences of DNA looping. In this construct, both the binding site for CRP and the operator O2 were deleted, while the promoter was replaced with a stronger promoter. The deletion of the CRP site is intended to remove the question of activation from the problem. Note also that this construct permits the insertion of DNA sequences of arbitrary length between O1 and Oid, where Oid has replaced O3. Oid is a much stronger operator than O3, of approximately the same strength as O1. Finally, the deletion of O2 insures that looping will only occur between the two remaining operators.

In order to confront data like that shown in fig. 1.11 (pg. 42), we need to expand our discussion of activators and repressors to include the effect of looping itself. In fig. 19.22, we show a minimal model of the states available to the system when RNA polymerase and Lac repressor are competing for the same region in the vicinity of the promoter. Note that this model permits different repressor molecules to occupy the two operators simultaneously, or a single molecule to occupy both sites and to loop the intervening DNA. We ignore the possibility of activator binding since the activator binding site was eliminated as shown in fig. 19.21. Note that this doesn't unequivocally rule out the possibility of nonspecific CAP binding which might tamper with the results as well.

In order to proceed in quantitative terms, as usual, we need to write down the partition function that corresponds to assigning statistical weights to all of the allowed states depicted in fig. 19.22. Using exactly the same logic as in previous sections, the partition function can be written as

$$\begin{aligned} Z_{tot}(P,R;N_{NS}) &= \underbrace{Z(P,R;N_{NS})}_{\mathrm{P}^{(0)},\ \mathrm{O}^{(0)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(0)}_{\mathrm{aux}}} + \underbrace{Z(P-1,R;N_{NS})e^{-\beta\varepsilon^S_{pd}}}_{\mathrm{P}^{(1)},\ \mathrm{O}^{(0)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(0)}_{\mathrm{aux}}} \\ &+ \underbrace{Z(P-1,R-1;N_{NS})e^{-\beta\varepsilon^S_{pd}}e^{-\beta\varepsilon^S_{rda}}}_{\mathrm{P}^{(1)},\ \mathrm{O}^{(0)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(1)}_{\mathrm{aux}}} \\ &+ \underbrace{Z(P,R-1;N_{NS})e^{-\beta\varepsilon^S_{rdm}}}_{\mathrm{P}^{(0)},\ \mathrm{O}^{(1)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(0)}_{\mathrm{aux}}} + \underbrace{Z(P,R-1;N_{NS})e^{-\beta\varepsilon^S_{rda}}}_{\mathrm{P}^{(0)},\ \mathrm{O}^{(0)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(1)}_{\mathrm{aux}}} \\ &+ \underbrace{Z(P,R-2;N_{NS})e^{-\beta\varepsilon^S_{rdm}}e^{-\beta\varepsilon^S_{rda}}}_{\mathrm{P}^{(0)},\ \mathrm{O}^{(1)}_{\mathrm{main}}\ \text{and}\ \mathrm{O}^{(1)}_{\mathrm{aux}}} \\ &+ \underbrace{Z(P,R-1;N_{NS})e^{-\beta\varepsilon^S_{rdm}}e^{-\beta\varepsilon^S_{rda}}e^{-\beta F_{loop}}}_{\text{repressor/loop}} \end{aligned} \tag{19.29}$$

Our notation has clearly become more cumbersome and deserves explanation. First, we introduce $\mathrm{P}^{(0)}$, $\mathrm{O}^{(0)}_{\mathrm{main}}$ and $\mathrm{O}^{(0)}_{\mathrm{aux}}$ to indicate that the occupancies of the promoter and main and auxiliary operators are zero, respectively. Next, the notation $\mathrm{O}^{(1)}_{\mathrm{main}}$ indicates that the main operator is occupied. The term with $\mathrm{P}^{(0)}$, $\mathrm{O}^{(1)}_{\mathrm{main}}$ and $\mathrm{O}^{(1)}_{\mathrm{aux}}$ indicates the states for which there are distinct repressor molecules bound to the two operators and the final term accounts for the looped state.

One of the terms in the expression includes the looping free energy in the form

$$Z(P,R-1;N_{NS})e^{-\beta\varepsilon^S_{rdm}}e^{-\beta\varepsilon^S_{rda}}e^{-\beta F_{loop}}, \tag{19.30}$$

and the factor $e^{-\beta F_{loop}}$ deserves further comment. Recall that $Z(P,R-1;N_{NS})$ is itself already a sum over all of the possible ways of distributing the P RNA polymerase molecules and the $R-1$ repressor molecules over the N_{NS} nonspecific

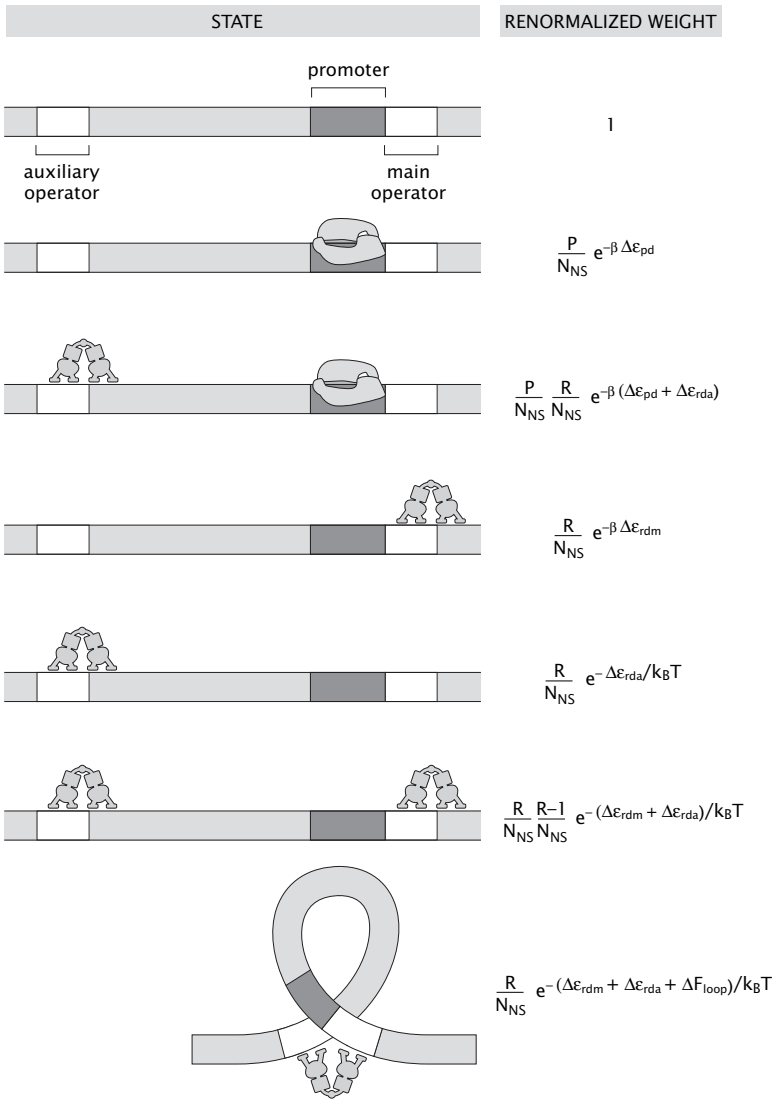

Figure 19.22: Looping states and weights in the *lac* operon. Each state corresponds to a different state of occupancy of the promoter and operators in the operon.

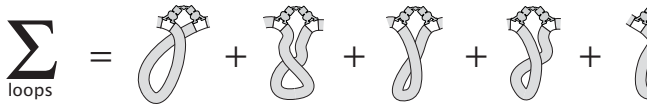

Figure 19.23: Summing over DNA loops. The sum $\sum_{\text{loops}}$ instructs us to sum over all conformations of the DNA loop as indicated schematically here.

binding sites on the DNA, with one of the repressors bound to both operators and looping the intervening DNA. However, for each and every one of these configurations, we have to sum over *all* of the possible geometries of the loop itself. That is, this contribution to the partition function is really of the form

$$Z_{looped}(P, R-1; N_{NS}) = \sum_{loops} Z(P, R-1; N_{NS}) e^{-\beta\varepsilon^S_{rdm}} e^{-\beta\varepsilon^S_{rda}} e^{-\beta\varepsilon_{loop}}, \quad (19.31)$$

where ε_{loop} is the *energy* of a given loop configuration and $\sum_{loops}$ instructs us to sum over all of the possible loop configurations as schematized in fig. 19.23. Since most of the factors are independent of the looping geometry, we can rewrite this as

$$Z_{looped}(P, R-1; N_{NS}) = Z(P, R-1; N_{NS}) e^{-\beta\varepsilon^S_{rdm}} e^{-\beta\varepsilon^S_{rda}} \sum_{loops} e^{-\beta\varepsilon_{loop}}, \quad (19.32)$$

where we have pulled all terms out of the sum that do not depend upon the particular choice of looped state. One way to proceed at this point is to appeal to ideas about elasticity to determine ε_{loop} and use the random walk as the basis for effecting the sum. On the other hand, the simpler scheme is to replace the sum by $e^{-\beta F_{loop}}$ and to treat F_{loop} as a phenomenological parameter as we have already done with the various binding energies.

With the partition function in hand, we can compute the probability of RNA polymerase binding by considering the ratio of favorable outcomes to the total partition function, resulting in

$$\begin{aligned} p_{bound}(P, R; N_{NS}) \quad = \quad & \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} \left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rda}}\right) \\ & \left[1 + \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} \left(1 + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rda}}\right) + \right. \\ & \frac{R}{N_{NS}} \left(e^{-\beta\Delta\varepsilon_{rdm}} + e^{-\beta\Delta\varepsilon_{rda}}\right) + \\ & \frac{R(R-1)}{(N_{NS})^2} e^{-\beta(\Delta\varepsilon_{rdm} + \Delta\varepsilon_{rda})} + \\ & \left. \frac{R}{N_{NS}} e^{-\beta(\Delta\varepsilon_{rdm} + \Delta\varepsilon_{rda} + \Delta F_{loop})} \right]^{-1} \end{aligned} \quad (19.33)$$

where we have defined $\Delta F_{loop} = F_{loop} + \varepsilon_{rd}^{NS}$. From this expression we can obtain the regulation factor

$$\begin{aligned} F_{reg}(R) &= \left(1 + \frac{R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rda}}\right) \\ &\left[\frac{R}{N_{NS}} \left(e^{-\beta \Delta \varepsilon_{rdm}} + e^{-\beta \Delta \varepsilon_{rda}}\right) + \right. \\ &\frac{R(R-1)}{(N_{NS})^2} e^{-\beta(\Delta \varepsilon_{rdm} + \Delta \varepsilon_{rda})} + \\ &\left. \frac{R}{N_{NS}} e^{-\beta(\Delta \varepsilon_{rdm} + \Delta \varepsilon_{rda} + \Delta F_{loop})}\right]^{-1}. \end{aligned} \tag{19.34}$$

To make contact with the results of Müller *et al.*, we now need to write an expression for the repression as a function of the interoperator spacing. Recall that the repression is given by eqn. 19.26 and takes the form

$$\begin{aligned} \text{repression}(N_{bp}) &= (F_{reg})^{-1} \\ &= \left[\frac{R}{N_{NS}} \left(e^{-\beta \Delta \varepsilon_{rdm}} + e^{-\beta \Delta \varepsilon_{rda}}\right) + \right. \\ &\frac{R(R-1)}{(N_{NS})^2} e^{-\beta(\Delta \varepsilon_{rdm} + \Delta \varepsilon_{rda})} + \\ &\left. \frac{R}{N_{NS}} e^{-\beta(\Delta \varepsilon_{rdm} + \Delta \varepsilon_{rda} + \Delta F_{loop})}\right] \\ &\left(1 + \frac{R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rda}}\right)^{-1}. \end{aligned} \tag{19.35}$$

where we have written repression(N_{bp}) as a function of the number of base pairs in the loop (N_{bp}) to signal that the looping free energy (and hence the repression) will depend upon the distance between the two operators. We have invoked the approximation that the promoter is weak (i.e. $\frac{N_{NS}}{P F_{reg}} e^{\beta \Delta \varepsilon_{pd}} \gg 1$). In order to examine the significance of our results on looping, we consider the extent to which the model can be used to interpret existing data and to suggest new experiments. Notice that we already know all the parameters in the weights from the previous experiment with the exception of ΔF_{loop}. We argue that for a given loop size ΔF_{loop} is a parameter that should be indifferent to which combination of operators is used in these two-operator experiments, and as a result, once ΔF_{loop} is determined the model is predictive. The results of a simple fit to the looping free energy are shown in fig. 19.24. To obtain these curves, any single data point is used to obtain the looping free energy itself and then the resulting curves are entirely predictive.

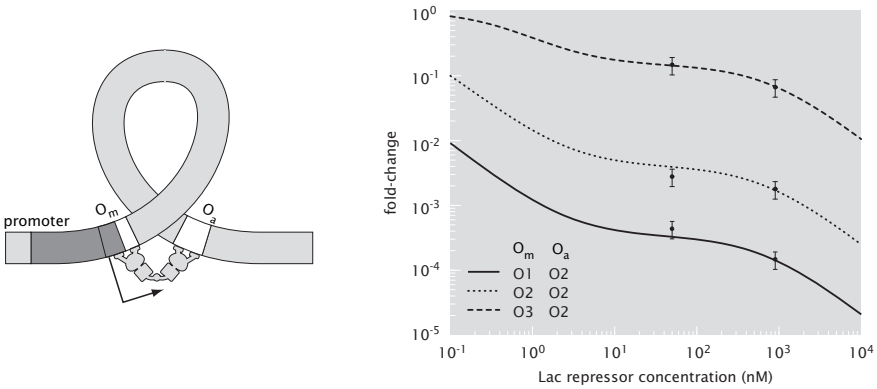

Figure 19.24: Repression and looping. A single fit to ΔF_{loop} giving 8.5 k_BT permits the investigation of multiple configurations of the different operators. (Data from S. Oehler *et al.*, EMBO J., 13:3348, 1994.)

19.3 Regulatory Dynamics

19.3.1 The Dynamics of RNA Polymerase and the Promoter

So far, our treatment of gene regulation has centered on the time-independent output of different regulatory motifs. On the other hand, as is clear from watching the development of any embryo, many of the most beautiful and important questions in regulation center on the orchestration of regulatory decisions over time. Another example that puts questions of the time dependence of gene expression front and center is the study of cells during the cell cycle. As was already shown in chap. 3, entire batteries of genes are expressed at different times during the cell cycle (see fig. 3.17 on pg. 152 for a concrete example in the cell cycle of *Caulobacter crescentus*). Two of the key dynamical motifs that recur in organisms ranging from bacteria to humans are switches and oscillators. In the case of switches, depending upon some environmental cue, for example, a cell can change the regulatory state associated with particular genes from "off" to "on". Even richer behavior is exhibited by regulatory circuits that give rise to oscillations. So that we can see how switches and oscillators are constructed, we now take up the question of time-dependent gene expression.

The Concentrations of Both RNA and Protein Can Be Described Using Rate Equations

Our conceptual starting point for examining the dynamics of gene expression is the rate equation paradigm introduced in chap. 15. In particular, we will proceed by writing rate equations for the time evolution of the concentrations of various molecular participants in the regulatory problem of interest. The

simplest scenario is to consider a dynamical description which refers only to the time development of the concentrations of the relevant proteins. On the other hand, sometimes it is convenient to characterize the time evolution of the mRNA transcripts as well. In either case, our strategy will be to consider some particular regulatory architecture in which different elements are linked and to write down a dynamical description of their concentrations.

Before embarking on an analysis of the dynamics of particular regulatory architectures, we return to one of the most elementary (but important) questions that can be asked about regulatory dynamics. In particular, our use of statistical mechanics in the previous sections was predicated on the idea that the binding of RNA polymerase to the promoter of interest can be thought of as an equilibrium process. In section 15.2.6 (pg. 776) we showed in general terms the conditions under which a dynamical process can be treated as an equilibrium problem. Our interest here is to use that general formalism as a basis for deciding when the equilibrium picture of polymerase binding is appropriate and when it is not.

- **Estimate: Time Scales for Promoter Occupancy and Escape.** So far we have used equilibrium statistical mechanics to compute the probability of finding RNA polymerase bound to the promoter and how this probability is modulated by a variety of transcription factors. However, the validity of this calculation rests on the assumption that the binding of RNA polymerase to DNA is in equilibrium.

  The simplest model of promoter binding by RNA polymerase leading to transcript elongation is as follows

$$\mathrm{P} + \mathrm{D} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{PD_C} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{PD_O} \overset{k_{\mathrm{clear}}}{\rightarrow} \mathrm{Elongation}. \tag{19.36}$$

  Here, P is free RNA polymerase and D is unbound promoter. $\mathrm{PD_C}$ is the closed complex, which corresponds to the species generated upon RNA polymerase binding to the promoter. This leads to the formation of the open complex $\mathrm{PD_O}$, where the DNA strands are opened. Finally, RNA polymerase leaves the promoter and begins transcription, which can lead to a full transcript elongation or to abortive initiation.

  If we want to describe the binding and unbinding of polymerase to form a closed complex using equilibrium statistical mechanics the most stringent condition $k_{\pm 1} \gg k_{\pm 2}$ should be satisfied. If RNA polymerase has time to bind and unbind from the promoter multiple times before open complex formation then we can think of the first step as effectively an equilibrium step and define an equilibrium constant $K_B = \frac{k_{-1}}{k_{+1}}$.

19.3.2 Genetic Switches: Natural and Synthetic

Switches are an important part of the genetic repertoire of all organisms. To explore the behavior of these switches more carefully, recently a synthetic version of such a switch was constructed in *E. coli* which had the convenient property

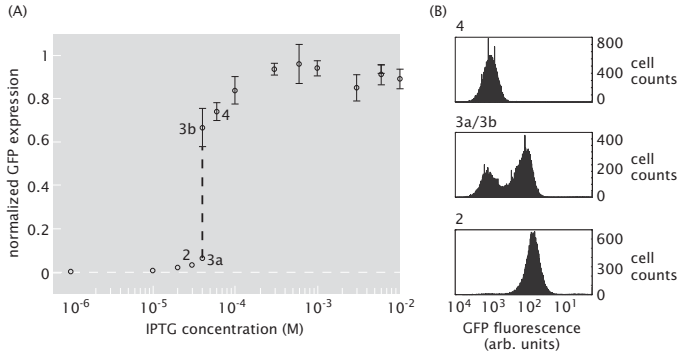

Figure 19.25: Data illustrating the flipping of the genetic switch in *E. coli* cells. (A) Average fluorescence of a population of *E. coli* cells harboring the genetic switch as a function of the concentration of an inducer molecule which flips the switch. (B) Flow cytometry data showing the single cell fluorescence distribution for different inducer concentrations. The labels correspond to points in the curve shown in part (A). Bistability is revealed through the fact that there are two populations of cells at the same inducer concentration. (Adapted from T. S. Gardner *et al.*, Nature, 403:339, 2000.)

that the gene product of the switch is a fluorescent reporter protein such that flipping of the switch can be read out by observing the fluorescent state of the cells. Data from this synthetic switch is shown in fig. 19.25.

The switch described above was constructed by using two repressor proteins whose transcription is mutually regulated as shown in fig. 19.26. The protein that is an output from the first gene serves as a repressor for the second gene. Conversely, the protein that is the output from the second gene serves as a repressor of the first gene. We denote the concentrations of the two protein species by c_1 and c_2. We are interested in writing equations for dc_1/dt and dc_2/dt. We consider two classes of processes that can alter the concentrations of these proteins. First, the proteins can be degraded over time. The change in concentration resulting from degradation can be written as $dc_1/dt = -Kc_1$. Second, protein 2 can bind onto the promoter for protein 1 and repress its production and vice versa. To capture this effect, we introduce a term of the form $\gamma(1 - p_{bound})$ where γ is the basal rate of production and p_{bound} is the probability that the repressor of interest will be bound - when $p_{bound} = 1$, there is no protein production and when $p_{bound} = 0$, the rate of protein production takes its basal rate. Recall from chap. 6 that for binding described by a Hill function, we have

$$p_{bound}(c_1) = \frac{K_b c_1^n}{1 + K_b c_1^n}, \tag{19.37}$$

where K_b is the binding constant for the repressor. This implies in turn that

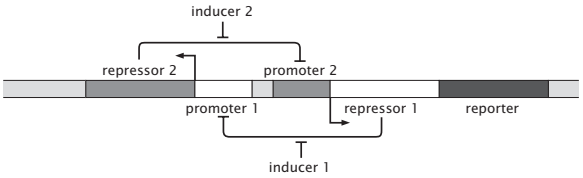

Figure 19.26: Regulatory architecture for a genetic switch. There are two promoters which are under the transcriptional control of the gene product of the partner promoter.

the protein production rate for protein 2 is

$$\gamma(1 - p_{bound}) = \frac{\gamma}{1 + K_b c_1^n}. \tag{19.38}$$

The choice to model p_{bound} using a Hill function rather than our statistical mechanical treatment is so that our treatment is consonant with the original literature. The reader will have the chance to explore the behavior of this circuit using p_{bound} as it has been considered throughout the book in the problems at the end of the chapter. Notice that our treatment of the binding constant here is slightly different than that favored in sec. 6.4.3, also for the purposes of consistency with the original literature.

Using the conceptual framework introduced above, the chemical rate equations for the genetic switch are

$$\begin{aligned} \frac{dc_1}{dt} &= -Kc_1 + \frac{\gamma}{1 + K_b c_2^n} \\ \frac{dc_2}{dt} &= -Kc_2 + \frac{\gamma}{1 + K_b c_1^n} \ . \end{aligned} \tag{19.39}$$

The first term on the right side of both equations corresponds to protein degradation, and for simplicity we assume that the degradation rate (characterized by the parameter K) of both proteins is the same. For proteins that are stable over time scales longer than the cell cycle (as is the case in the repressors used in this circuit) the dilution rate is determined by the cell doubling time and the subsequent dilution of the protein between the two daughter cells. Therefore, under these conditions the effective protein degradation rate is the same and is set by the cell division time. The second terms on both sides characterize the rate of protein production. As introduced above, the *basal* rate of production is captured in the parameter γ. However, this rate is reduced when the repressor is bound to the promoter of interest as shown above. For simplicity, we assume that the basal production rates and the binding constants that characterize the affinity of the repressors for their binding site are the same for both genes. For a realistic circuit these assumptions are not necessarily true, but will suffice here to describe the basic operation of the circuit. Another conceptual simplification

implicit in these rate equations is the idea that the binding of the repressors is characterized by a Hill function with Hill coefficient n.

From a mathematical perspective, we wonder whether equations like those given in eqn. 19.39 yield switch-like solutions. Our assertion is that there are two regions in the space of parameters, one with a single stable solution corresponding to equal concentrations of the two species (decidedly not a switch), and another, more interesting regime, where we find two stable solutions distinguished by having one of the protein concentrations much larger than the other. For values of the parameters where the stable solutions are of this variety, the genetic network exhibits switch-like behavior.

In order to simplify the mathematical analysis of the circuit we resort to a dimensionless form for eqn. 19.39. This is achieved by measuring c_1 and c_2 in units of $K_b^{-1/n}$ and time in units K^{-1}. This reduces the circuit equations to

$$\begin{aligned} \frac{du}{dt} &= -u + \frac{\alpha}{1+v^n} \\ \frac{dv}{dt} &= -v + \frac{\alpha}{1+u^n} \ , \end{aligned} \tag{19.40}$$

where the parameters $\alpha = \gamma K_b^{1/n}/K$ and the Hill coefficient n are the only remaining dimensionless parameters. We have introduced the notation u for the dimensionless concentration of c_1 and v for the dimensionless concentration of c_2. At this point, our goal is to find the steady state solutions of eqn. 19.40 and analyze their stability for different values of α and the Hill coefficient.

To find the steady state solutions to the rate equations we set the time derivatives to zero. Since the equations are symmetric with respect to u and v we immediately conclude that

$$u^* = v^* = \frac{\alpha}{1+v^{*n}} \tag{19.41}$$

is always a solution. Clearly, this result does *not* exhibit the properties of a switch since the concentrations of both proteins in this case are the same. Are there other solutions that exhibit switching behavior? The equations that determine the steady state u^* and v^* are of the form $x = f(f(x))$, where $f(x) = \alpha/(1+x^n)$. To see this, solve the first equation for u and substitute that result into the second equation. Since the function f is monotonically decreasing (that is, larger values of x imply $f(x)$ is smaller) the composition $f \circ f$ will be a monotonically *increasing* function, like the function x itself. Therefore there is the possibility that the two curves x and $f(f(x))$ intersect at more than one point, leading to multiple steady states.

To make these considerations explicit we consider the case when the Hill coefficient is $n = 2$, which lends itself to analytic treatment. The steady state equation for the repressor concentration u^* is

$$u^* = \frac{\alpha}{1+\left(\frac{\alpha}{1+u^{*2}}\right)^2} \ , \tag{19.42}$$

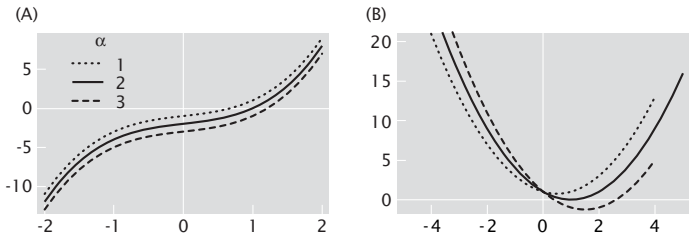

Figure 19.27: Steady-state solutions for protein concentrations in the genetic switch. (A) The function $y = u^3 + u - \alpha$ plotted for various values of α. The solution u^* corresponds to the point at which the curve crosses the u-axis. (B) The function $y = u^2 - \alpha u + 1$ plotted for various values of α. Depending upon the choice of α, there can be 0, 1 or 2 crossings of the u-axis.

and the same equation holds for v^*. A little bit of algebra transforms the above equation to a much simpler form given by a product of two polynomials

$$(u^{*2} - \alpha u^* + 1)(u^{*3} + u^* - \alpha) = 0 \ . \tag{19.43}$$

The steady state solutions to the rate equations for the genetic switch, eqn. 19.40 are therefore zeroes of the two polynomials appearing in the above equations.

The cubic polynomial has one real zero, which can be seen from fig. 19.27(A) where we plot the polynomial for different values of α. A mathematically rigorous way to show this is to note that the first derivative of this polynomial, $3u^{*2}+1$, is always positive, which implies that the function is strictly increasing and can therefore intercept the u^*-axis at most once. The equilibrium state that corresponds to the zero of the cubic polynomial has equal concentrations of the two repressor species since the equation $u^{*3} + u^* - \alpha = 0$ can be rewritten as $u^* = \alpha/(1 + u^{*2})$, and the right hand side of this equation is v^*.

The quadratic polynomial in eqn. 19.43 can have one, two, or no zeroes depending on the value of α, as observed in fig. 19.27(B). For $\alpha < 2$ there are no zeroes, for $\alpha > 2$ the polynomial has two zeroes, while for $\alpha_c = 2$, the critical value of α, it has one zero at $u^* = 1$. For the two solution case the two steady state values of u^* and v^* correspond to the two different ways of assigning the two roots to each of the dimensionless repressor concentrations. Namely, for a given u^* the corresponding value of v^* can be calculated using $v^* = \alpha/(1+u^{*2})$. For these values of u^* and v^* the equality $u^* + v^* = \alpha$ is satisfied assuming u^* is one of the zeroes of the quadratic polynomial in eqn. 19.43.

In light of the general analysis done above, we see that for $\alpha < 2$ the genetic switch exhibits only one stable equilibrium state with $u^* = v^*$, while for $\alpha > 2$ it has two stable and one unstable state. In the latter case the unstable state is the one in which the concentrations of the two repressors are equal, while stable equilibrium states have either repressor u or repressor v in excess.

The dynamical behavior of a system of rate equations like those given in eqn. 19.40 can be examined in a different way graphically using the idea of a

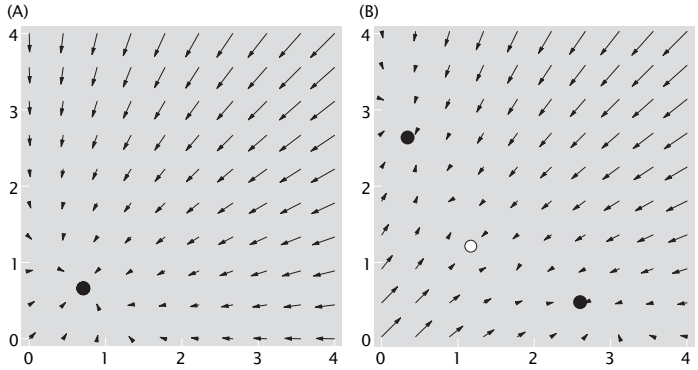

Figure 19.28: Graphical representation of the dynamics of the genetic switch. (A) The phase portrait of the genetic switch for $\alpha = 1$ and (B) $\alpha = 3$. Stable equilibria are represented by filled circles while the unfilled circle corresponds to an unstable state.

phase portrait (the mathematics is explained in the "Tricks Behind the Math" box at the end of the section). The idea is that we can think of du/dt and dv/dt as the two components of a velocity vector and we can plot the velocity field at every point (u, v). The steady state solutions will correspond to those points in the phase portrait where the vectors are zero. The solutions represented by those points are stable if for any small excursion away from that point, all the velocity vectors point towards the solution point. An example of this idea for several choices of α is shown in fig. 19.28. The phase portrait provides a convenient graphical representation of the dynamics of the genetic switch. Namely, for a given initial condition u_0, v_0 in order to see how the concentrations will evolve with time all one has to do is follow the flow depicted by the arrows in the phase portrait. We therefore conclude that the stable steady states of the rate equations are associated with positions in the (u, v) plane where the phase flow converges from all directions, while diminishing in size, while unsteady states have at least one direction along which the flow is diverging.

- **The Tricks Behind the Math: Phase Portraits and Vector Fields.** As we have seen repeatedly in the book, there are many circumstances in which the dynamics of some system of interest involves *coupled* rate equations of the form

$$\frac{dx}{dt} = f(x, y) \tag{19.44}$$

$$\frac{dy}{dt} = g(x, y) \ , \tag{19.45}$$

where, in general, $f(x, y)$ and $g(x, y)$ are nonlinear functions. The idea of the phase portrait is to graphically depict the "flows" implied by the

rate equations. In particular, we imagine a velocity vector field $\mathbf{v}(x, y) = (dx/dt, dy/dt)$, which depicts which way the system will "move" in the next time step. For a given initial condition (x_0, y_0), we can find the subsequent dynamics of the system by following the arrows. From the standpoint of stability analysis, the most interesting points in a phase portrait are the fixed points. These are the points at which the vector field satisfies the condition $\mathbf{v}(x^*, y^*) = 0$. In other words, if we choose (x^*, y^*) as an initial condition, the system will stay put. Stability is determined by the directions of the arrows in the neighborhood of the fixed point. If the arrows *all* point back towards that fixed point, the point is said to be a stable fixed point. Otherwise, it is unstable. This type of graphical analysis is a powerful qualitative tool for examining the dynamics of nonlinear, coupled equations.

19.3.3 Genetic Networks That Oscillate: The Repressilator

A second key class of regulatory motif leads to oscillations in the concentration of protein products. Perhaps the most notable biological example of such oscillators are those driving the cell cycle. However, as with the analysis of genetic switches, the analysis of oscillatory genetic networks was serviced particularly by the creation of a synthetic oscillator known as the repressilator. The idea is to use various genetic regulatory elements, again in conjunction with a reporter gene that expresses green fluorescent protein, to construct a set of coupled genes whose outputs interact with one another in such a way as to lead to a time varying concentration of GFP. From the perspective of coupled rate equations, the existence of oscillatory solutions is nothing new. However, the intriguing feature of the experiment to be described here is the construction of a concrete implementation of such coupled rate equations in the form of a genetic network in such a way that these oscillations are realized *in vivo* in bacterial cells.

As shown in fig. 19.29, the synthetic network constructed to exhibit oscillatory behavior used three interacting promoters in which the gene product of the first promoter resulted in a repressor on the second. Similarly, the gene product of the second promoter repressed the third. Finally, the gene product of the third promoter acted back on the first promoter by repressing it. This circuit was constructed on a plasmid. A second plasmid contained a reporter gene that expressed GFP and was under the control of one of the repressors from the oscillatory network.

The experimental consequences of the construction described above are shown in fig. 19.30. In particular, fig. 19.30(A) shows a series of snapshots of *E. coli* cells at various stages in the experiment. Note that the time scale of the experiment is set by the oscillation period associated with expression of the green fluorescent protein and is of order an hour. Fig. 19.30(B) is a more quantitative rendering of the experimental outcome in the form of a plot of the time evolution of the average fluorescence intensity of a single cell. The oscillatory character of the experiment is clearly revealed as is a background increase in the overall

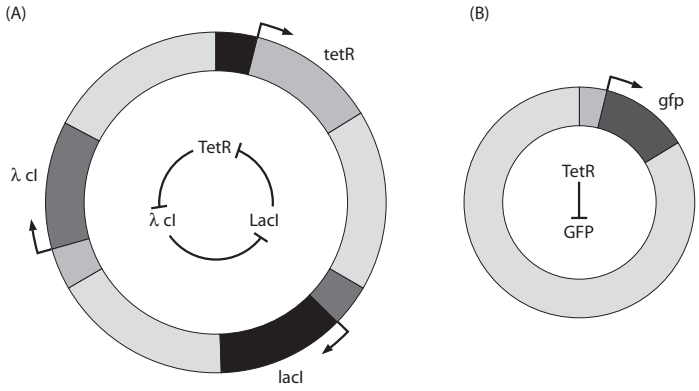

Figure 19.29: Schematic illustration of the genes and gene products used to construct the repressilator. (A) the plasmid containing the relevant regulatory elements and (B) the plasmid containing the reporter gene. (Adapted from M. B. Elowitz and S. Leibler, Nature, 403:335, 2000.)

intensity which is tied to the protein accumulation in the cells.

As constructed, this oscillatory network consists of three repressor proteins which inhibit the expression of each other: protein 1 represses the expression of gene 2, whose product, protein 2, represses gene 3, while protein 3 represses gene 1. The rate equations that characterize the time evolution of mRNA and protein are

$$\begin{aligned} \frac{dm_i}{dt} &= -K_m m_i + \frac{\gamma}{1 + K_b p_{i-1}^n} + \gamma_0 \\ \frac{dp_i}{dt} &= -K_p p_i + T m_i \end{aligned} \tag{19.46}$$

where, m_i and p_i ($i = 1, 2, 3$) are concentrations of mRNA and protein, respectively (note that in the equation for m_1, $p_0 \equiv p_3$). For simplicity, we have chosen the rate constants for protein and mRNA degradation (K_p and K_m), as well as the rate constants associated with mRNA transcription (T), repressor binding (γ and K_b), and the leakiness of the repressor (γ_0), to be the same for all three species. The final parameter in the model is n which is the Hill coefficient expressing the degree of cooperativity of repressor binding.

As we did with the genetic switch, we simplify the math by writing p_i in units of $K_b^{-1/n}$, m_i in units of $K_p/(TK_b^{1/n})$, and time in units of K_m^{-1}. In these units eqn. 19.46 becomes

$$\begin{aligned} \frac{dm_i}{dt} &= -m_i + \frac{\alpha}{1 + p_{i-1}^n} + \alpha_0 \\ \frac{dp_i}{dt} &= -\beta p_i + \beta m_i \ , \end{aligned} \tag{19.47}$$

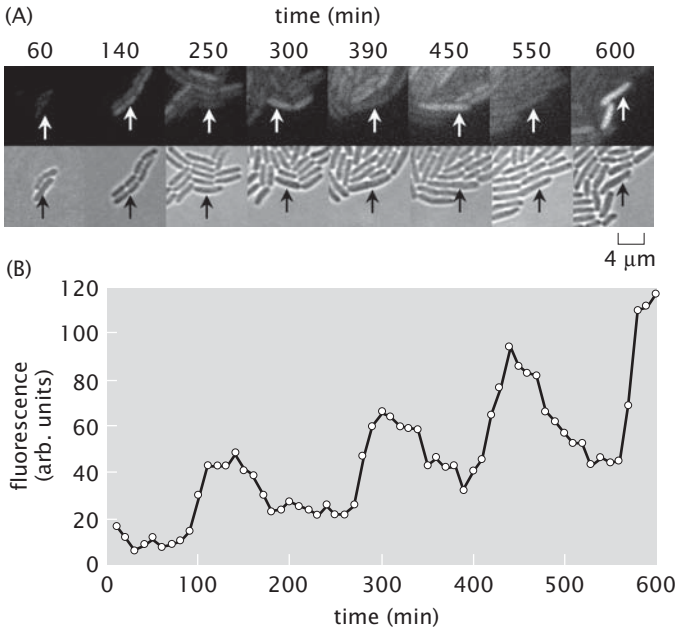

Figure 19.30: Cells at various phases in the represillation cycle. (A) Snapshots of bacteria dividing as a function of time. The top row shows fluorescence images and the bottom row shows bright field images of the cells. (B) Measured GFP fluorescence intensity of the single cell marked by an arrow as a function of time. (Adapted from M. B. Elowitz and S. Leibler, Nature, 403:335, 2000.)

with new parameters

$$\begin{aligned} \alpha &= \frac{\gamma K_b^{1/n} T}{K_p K_m} \\ \alpha_0 &= \frac{\gamma_0 K_b^{1/n} T}{K_p K_m} \\ \beta &= \frac{K_p}{K_m} \ . \end{aligned} \tag{19.48}$$

In order to map out the phase diagram described by the repressilator equations we employ the tactic of first finding the steady state solutions. To find the steady state solutions we set all the time derivatives appearing on the left hand sides of the rate equations to zero. This immediately implies $p_i = m_i$ for each protein and mRNA species. We can then rewrite the equations for the mRNA concentration as:

$$m_i = f(f(f(m_i))) \tag{19.49}$$

where the function

$$f(x) = \frac{\alpha}{1 + x^n} + \alpha_0 \ . \tag{19.50}$$

Since the function f is monotonically decreasing the function $f \circ f \circ f$ is also a monotonically decreasing function implying that there will be a single solution to eqns. 19.49 given by

$$m = \frac{\alpha}{1 + m^n} + \alpha_0 \ . \tag{19.51}$$

Next we investigate the stability of the steady state solution.

One of the standard tools for analyzing the qualitative features of a set of dynamical equations like those given above is to perform a linear stability analysis about the fixed points. That is, we seek the steady state (for which the left side of the dynamical equations is zero) and then ask how the system responds to *small* perturbations about these steady states. To check for the stability of the steady state solution we consider the effect of a small deviation in the protein and mRNA concentrations from the steady state value given by eqn. 19.51. If we label the small deviations by δm_i and δp_i, the chemical kinetics equations, eqn. 19.47, to lowest order in the deviations take on a linear form,

$$\begin{aligned} \frac{d\delta m_i}{dt} &= -\delta m_i + X \delta p_{i-1} \\ \frac{d\delta p_i}{dt} &= -\beta(\delta p_i - \delta m_i) \ , \end{aligned} \tag{19.52}$$

where the parameter X is the derivative of $f(x)$ evaluated for the steady state value $x = m$. This set of six first order differential equations can be written in matrix form as

$$\frac{d\mathbf{v}}{dt} = \hat{\mathbf{A}}\mathbf{v} \tag{19.53}$$

where

$$\hat{\mathbf{A}} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & X \\ 0 & -1 & 0 & X & 0 & 0 \\ 0 & 0 & -1 & 0 & X & 0 \\ \beta & 0 & 0 & -\beta & 0 & 0 \\ 0 & \beta & 0 & 0 & -\beta & 0 \\ 0 & 0 & \beta & 0 & 0 & -\beta \end{pmatrix} \tag{19.54}$$

and

$$\mathbf{v} = \begin{pmatrix} \delta m_1 \\ \delta m_2 \\ \delta m_3 \\ \delta p_1 \\ \delta p_2 \\ \delta p_3 \end{pmatrix} . \tag{19.55}$$

The solution to the linearized rate equations is

$$\mathbf{v}(t) = e^{\hat{\mathbf{A}}t}\, \mathbf{v}(0) \tag{19.56}$$

and for the steady state solution to be stable $\mathbf{v}(t)$ must go to zero in the long-time limit for arbitrary $\mathbf{v}(0)$. This will be the case if all the eigenvalues of the matrix A are negative; for complex eigenvalues we demand that the real part be negative. If this condition is not fulfilled the deviations from the steady state solution will grow and new behavior, such as oscillations of the protein and mRNA concentrations will be obtained.

The eigenvalues λ of the matrix $\hat{\mathbf{A}}$ satisfy the equations

$$\hat{\mathbf{A}} \begin{pmatrix} \delta\mathbf{m} \\ \delta\mathbf{p} \end{pmatrix} = \lambda \begin{pmatrix} \delta\mathbf{m} \\ \delta\mathbf{p} \end{pmatrix} \tag{19.57}$$

where $\delta\mathbf{m} = (\delta m_1, \delta m_2, \delta m_3)^{\mathrm{T}}$ and similarly for $\delta\mathbf{p}$; the symbol $^{\mathrm{T}}$ is for "transpose", reminding us that we are in fact dealing with a column-vector. The eigenvalues λ are readily computed. In particular, from eqn. 19.54 and eqn. 19.57 it follows that

$$\begin{pmatrix} -\delta\mathbf{m} + \hat{\mathbf{X}}\delta\mathbf{p} \\ \beta\delta\mathbf{m} - \beta\delta\mathbf{p} \end{pmatrix} = \lambda \begin{pmatrix} \delta\mathbf{m} \\ \delta\mathbf{p} \end{pmatrix} , \tag{19.58}$$

where

$$\hat{\mathbf{X}} = \begin{pmatrix} 0 & 0 & X \\ X & 0 & 0 \\ 0 & X & 0 \end{pmatrix} . \tag{19.59}$$

These equations can be solved in terms of $\delta\mathbf{p}$ as

$$\begin{aligned} \delta\mathbf{m} &= \frac{\lambda+\beta}{\beta}\, \delta\mathbf{p} \\ \hat{\mathbf{X}}\, \delta\mathbf{p} &= \frac{(\lambda+1)(\lambda+\beta)}{\beta}\, \delta\mathbf{p} \, . \end{aligned} \tag{19.60}$$

We conclude that the eigenvalues λ of the matrix $\hat{\mathbf{A}}$ can be computed from the eigenvalues τ of the matrix $\hat{\mathbf{X}}$ by solving the quadratic equation

$$\frac{(\lambda+1)(\lambda+\beta)}{\beta} = \tau \ , \tag{19.61}$$

which relates the two. Since the matrix $\hat{\mathbf{X}}$ is proportional to a cyclic permutation of three indices, its eigenvalues are multiples of the cubic roots of unity, $\tau_{1,2,3} = X,\ X\exp(i2\pi/3),\ X\exp(i4\pi3)$. For each τ there are two eigenvalues of the matrix $\mathbf{A}$, namely,

$$\lambda = \frac{1}{2}\left(-(1+\beta) \pm \sqrt{(1-\beta)^2+4\beta\tau}\right) \ , \tag{19.62}$$

corresponding to the solutions of eqn. 19.61. In order for the steady state solution of the rate equations to be stable the real parts of all six eigenvalues λ have to be negative.

First we consider the pair of eigenvalues

$$\lambda_{1,2} = \frac{1}{2}\left(-(1+\beta) \pm \sqrt{(1-\beta)^2+4\beta X}\right) \ , \tag{19.63}$$

obtained from τ_1. These eigenvalues are both real and for them to be both negative we must insist on $X < 1$. Since, as defined, X is negative, this condition is automatically fulfilled.

Next we turn to the two complex eigenvalues $\tau_{2,3}$. As we will see shortly they both lead to the same condition on X and β. Substituting, say, $\tau_2 = X\exp(i2\pi/3)$ into eqn. 19.62 we end up with the condition

$$Re\left(-(1+\beta) \pm \sqrt{(1-\beta)^2 - 2\beta X + 2\sqrt{3}\beta X i}\right) < 0 \ . \tag{19.64}$$

Writing the complex number under the square-root in polar form, $(1-\beta)^2 - 2\beta X + 2\sqrt{3}\beta X i = A\exp(i\phi)$, the above inequality reduces to

$$A\cos\frac{\phi}{2} < 1+\beta \tag{19.65}$$

where

$$\begin{aligned} A\cos\phi &= (1-\beta)^2 - 2\beta X \\ A\sin\phi &= 2\sqrt{3}\beta X \ , \end{aligned} \tag{19.66}$$

and in eqn. 19.64 we have chosen the positive sign; the inequality with the negative sign will be automatically fulfilled. Using the trigonometric identity, $\cos^2(\phi/2) = (1+\cos\phi)/2$, and substituting the values for A and $\cos\phi$ from eqn. 19.66, we obtain the condition

$$(4+2X)(1+\beta)^2 - 3\beta X > 0 \ . \tag{19.67}$$

Note that the result does not depend on the sign of $A \sin \phi$ which is the only difference between the eigenvalues obtained from τ_2 and τ_3. Therefore, eqn. 19.67 is the sole condition for solutions to the linearized repressilator equations, eqn. 19.56, to be *decaying* for arbitrary choice of initial conditions. Here initial conditions correspond to the initial deviations of the protein and mRNA concentrations from their steady state value. Further note that the first eigenvalues of the linearized equations to go unstable come in complex conjugate pairs indicating that the instability is towards oscillating solutions. This can be confirmed by solving the full non-linear repressilator equations numerically.

The key point of this analysis is the demonstration that genetic circuits like the repressilator can exhibit stable oscillations. Further, this linear stability analysis is a reminder of the unexpected way that "springs" can arise in the mathematical analysis of all sorts of problems. This idea was introduced in fig. 1.12 (pg. 45). The two examples (i.e. genetic switch and repressilator) given here illustrate how to put time into our analysis of genetic networks. This step in the direction of a full treatment of the dynamics of regulatory networks must be supplemented by a treatment of how gene expression varies in space as well.

19.3.4 Putting Space in the Model: Reaction-Diffusion Models

So far, our use of rate equations has assumed that the concentration is the same at all points in space. On the other hand, there are many instances in which the concentrations of transcription factors or gene products vary in space. One of the most familiar and conceptually important examples concerns the establishment of the anterior-posterior patterning of the fly body plan already introduced in section 2.3.3 (pg. 106). An interesting question raised by the developmental processes in the early embryo of the fruit fly, is how are the spatial patterns of morphogens generated. For example, the formation of the second stripe of Eve is regulated by the minimal stripe element which converts the existing concentration profile of the Bicoid, Hunchback, Giant and Krüpel (see fig. 19.2) into the expression of the *eve* gene. The patterns that serve as input to *eve* expression are themselves outputs of previous steps in the developmental program of the fruit fly. We are thus lead to the question, what is the initial condition, and how is it set up?

The first morphogen whose concentration shows patterning is Bicoid. The Bicoid protein is produced by the maternal *bicoid* mRNA which is found at the anterior end of the embryo. Once Bicoid is produced at the anterior end it diffuses throughout the embryo. This protein is also degraded in time and, as we show below, this sets up an exponentially decaying profile of Bicoid concentration from anterior to posterior end.

In order to simultaneously account for the space and time dependence of transcription factors such as Bicoid, one useful approach is so-called "reaction-diffusion equations". These equations merge the thinking on diffusion developed in chap. 13 with the ideas on rate equations introduced in chap. 15. The idea is that the various reactants and products can diffuse around and change their

identity. Conceptually, the starting point of this analysis is that the concentrations of the various species of interest must be promoted to full functions of space and time. For example, in the case of the Bicoid, we introduce the field $[\text{Bcd}](\mathbf{r}, t)$. For simplicity, we imagine that the concentration profile is one-dimensional (i.e. consider a one-dimensional fly embryo) so that the concentration can be written as $[\text{Bcd}](x, t)$. Further, the only "reaction" to which the Bicoid is subjected is that it can decay over time.

The rate equation for the concentration of Bicoid along the anterior-posterior axis is

$$\frac{\partial[\text{Bcd}]}{\partial t} = D\frac{\partial^2[\text{Bcd}]}{\partial x^2} - \frac{1}{\tau}[\text{Bcd}] \ . \tag{19.68}$$

Here D is the diffusion constant of the protein, while τ is the mean lifetime. The second term on the right hand side of the equation is a "source" term and accounts for the fact that even in the absence of diffusion, the concentration of Bicoid can change simply by virtue of decay processes characterized by the mean lifetime τ. Since the anterior region ($x = 0$) of the embryo acts as a source of Bicoid we expect a steady state to develop, characterized by a concentration profile which does not change in time. In this case the Bicoid concentration satisfies the equation

$$D\frac{d^2[\text{Bcd}]}{dx^2} - \frac{1}{\tau}[\text{Bcd}] = 0 \ , \tag{19.69}$$

which has the solution

$$[\text{Bcd}] = [\text{Bcd}]_0 e^{-\frac{x}{\lambda}} \tag{19.70}$$

where $[\text{Bcd}]_0$ is the concentration at the anterior end, while $\lambda = \sqrt{D\tau}$ is the characteristic length over which the concentration decays by a factor of e.

The predicted exponential profile has been measured in the fruit fly embryo with $\lambda = 100$ μm. An interesting observation is that for embryos of different fly species that differ in length by as much as a factor of ten, the characteristic length is roughly proportional to the embryo length. This is necessary for proper scaling of the morphogens stripes and other patterns with embryo size (this in turn leads to the development of a proportional fly). Given the simple diffusion-degradation model we have introduced it is not clear at all how this comes about since it would seem that the diffusion constant and the degradation rate should both be independent of embryo size. This highly simplified example is intended to whet the reader's appetite for the necessity of a full space-time description of the regulatory networks described thus far in the chapter.

19.4 Cellular Fast Response: Signaling

Gene regulatory networks are clearly of central importance to the functioning of organisms of all types. Of course, there are many aspects of biology where dynamics of regulation is critical that do not involve gene transcription as an ultimate outcome. This is particularly obvious for biological behaviors that simply occur too quickly for transcription of new genes to have any useful impact.

Rather, these *signaling* networks involve batteries of proteins and their partner ligands connected together such that their interactions affect the activity of some enzyme. For example, a membrane-spanning receptor might bind a ligand in the extracellular space. As a result of this binding event, there will be a concomitant structural change on the intracellular domain of this same protein, activating a protein kinase enyzme activity, which results in the phosphorylation of some other protein rendering it active. The goal of the remainder of this chapter is to examine some examples of this kind of signaling and to construct simple models of their behavior.

19.4.1 Bacterial Chemotaxis

One fascinating and fairly well understood case that we have mentioned briefly already is the case of bacterial chemotaxis. Bacteria import small nutrients such as sugars and amino acids to use as building blocks as we calculated in chap. 3. A bacterial cell must take up a huge number (in excess of 10^9) of glucose molecules to go through a cycle of cell division. Obviously, this can be done more rapidly in areas of higher ambient glucose concentration. It therefore behooves the bacterium to actively seek out regions of its watery environment that contain the highest accessible concentration of glucose. An elegant and extraordinarily efficient system has evolved for this purpose.

As we mentioned in section 4.4.4 (pg. 205), the motor used for swimming by the class of bacteria including *E. coli* and *Salmonella* is a rotary propellor that spins a long flagellum (each bacterial cell has a several flagella that all work in synchrony). The only control point the bacterium has for the rotor is to alter its direction of spin to be either clockwise or counterclockwise. Counterclockwise rotation of the flagella drives the bacterium forward in a nearly straight "run", while clockwise rotation causes the flagellar bundle to become disorganized and the bacterium "tumbles", randomly changing its direction. The chemotactic signal transduction machinery regulates this directional switching. If desirable nutrients are present at high concentrations, the bacterium tends to keep moving in a straight line. If nutrient concentrations are low, the bacterium tends to tumble. *E. coli* is able to use the patterns of directional switching generated by this signal transduction network to swim up gradients of desirable nutrients.

How can a binary switch be used to detect the direction of a gradient? We can imagine at least two possibilities. First, the bacteria might be able to compare the signal coming from receptors located at the opposite poles of the cell, and switch in such a way as to swim toward the end with the higher signal, i.e., sensing the gradient in space. Alternatively, the bacteria might be able to compare the signal being received at a given moment in time with the strength of the signal it received in the recent past, i.e., sensing the gradient in time. As we will discuss below, the bacteria appear to use the time-based mechanism. The reader will have a chance to explore and compare these two possible schemes in the problems at the end of the chapter.

The cellular decision-making that attends chemotaxis is mediated by a signal transduction network that has been extremely well-characterized. Our com-

ments will center on the particular features of the *E. coli* chemotaxis network. The key elements in this system are: i) membrane-spanning receptors that interact with the molecules in the environment (sugars, amino acids, etc.); ii) CheW and CheA, proteins that bind to the intracellular domain of the receptor and change their activity depending on whether or not the receptor has a ligand bound (CheA is a protein kinase that can catalyze the attachment of phosphate groups to other target proteins, and CheW modulates CheA activity); iii) a messenger molecule known as CheY that, when phosphorylated by CheA, can interact with the flagellar rotary motor to induce it to switch to clockwise (tumbling) rotation, iv) CheZ, a phosphatase that can remove the phosphate from CheY, and v) a pair of enzymes known as CheR and CheB that can respectively methylate and demethylate the receptors themselves, effectively tuning their affinity for their binding partners.

Even for this relatively simple network, it is hard to avoid getting lost in the alphabet soup of names, so we try to examine how the network works conceptually without focusing on the names of the molecules. In addition, we will take a hierarchical view, first explaining the overall functioning of the network and then taking up the fancy bells and whistles that make it work over such a wide range of concentrations, a phenomenon known as adaptation. In simplest terms, the question of whether or not the cell will make a tumble (and hence a change of direction) comes down to the state of phosphorylation of the messenger molecule CheY. In order to be responsive to changes in the environment, the phosphorylation of CheY must be sensitive to whether or not there is a ligand bound to the receptor. In the presence of desirable attractant molecules such as glucose or aspartate, the cell should repress tumbling, so we expect that the ligand-bound receptor will tend to be in the "off" form, where CheY is not phosphorylated, and the unbound receptor will tend to be in the "on" form, where CheY is phosphorylated. (Although *E. coli* is actually able to use the same chemotactic network to swim away from noxious chemicals, here we will only consider the happier problem of swimming toward delicious ones.) An idealization of these elements is shown in fig. 19.31(A), where we have combined the transmembrane receptor, CheW and CheA into a single unit, and for the moment are ignoring the the other components of the pathway.

We can treat this complex process approximately by appealing to our usual statistical mechanics formulation in which we imagine a rapid preequilibrium of the state of activity of the receptor. In particular, the quantity p_{on} measures the ability of the receptor to produce phosphorylated Che-Y, resulting in a change in the motors direction of rotation. As we have done throughout the book, the statistical mechanics of this system can be examined by appealing to a states and weights diagram like that shown in fig. 19.32. The probability that the receptor will be active is gotten by constructing the ratio

$$p_{on} = \frac{\frac{\Omega^L}{L!}e^{-\beta L\varepsilon_{sol}}e^{-\beta\varepsilon_{on}} + \frac{\Omega^{L-1}}{(L-1)!}e^{-\beta(L-1)\varepsilon_{sol}}e^{-\beta\varepsilon_{on}}e^{-\beta\varepsilon_b^{on}}}{\frac{\Omega^L}{L!}e^{-\beta L\varepsilon_{sol}}e^{-\beta\varepsilon_{off}} + \frac{\Omega^{L-1}}{(L-1)!}e^{-\beta(L-1)\varepsilon_{sol}}e^{-\beta\varepsilon_{off}}e^{-\beta\varepsilon_b^{off}} + \frac{\Omega^L}{L!}e^{-\beta L\varepsilon_{sol}}e^{-\beta\varepsilon_{on}} + \frac{\Omega^{L-1}}{(L-1)!}e^{-\beta(L-1)\varepsilon_{sol}}e^{-}} \tag{19.71}$$

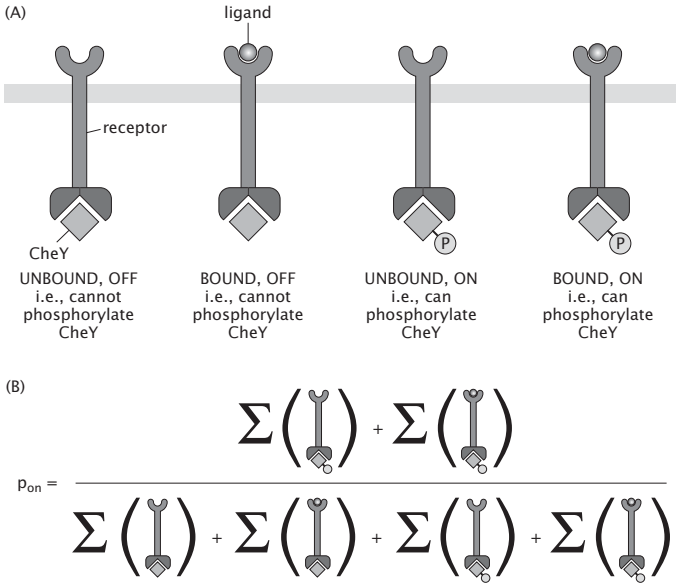

Figure 19.31: Probability that receptor will be "on". (A) The receptor and its states of occupancy and activity. The receptor can either have a bound ligand or not. Similarly, the receptor can either be "on" or "off", where this state of activity determines whether or not it is able to phosphorylate the messenger CheY. (B) The probability that the receptor will be on is constructed as a ratio of the "on" states, appropriately weighted by their Boltzmann factors to the sum over the statistical weights of all states.

This result can be simplified by multiplying through the top and bottom of the equation by $L!/\Omega^L$, resulting in

$$p_{on} = \frac{e^{-\beta\varepsilon_{on}}(1 + \frac{L}{\Omega}e^{-\beta\Delta\varepsilon_{on}})}{e^{-\beta\varepsilon_{on}}(1 + \frac{L}{\Omega}e^{-\beta\Delta\varepsilon_{on}}) + e^{-\beta\varepsilon_{off}}(1 + \frac{L}{\Omega}e^{-\beta\Delta\varepsilon_{off}})}. \tag{19.72}$$

Here we have defined $\Delta\varepsilon_{on}$ as the difference in energy between a single ligand bound to the "on" state of the receptor and the same ligand in solution, and $\Delta\varepsilon_{off}$ equivalently for ligand binding to receptor in the "off" state. Throughout the book, we have repeatedly translated back and forth between the statistical mechanics language used above and the thermodynamics language using equilibrium constants. By exploiting the relationship between energy differences and biochemical dissociation constants derived in section 6.4.1 (pg. 345), our expression for the probability that the receptor will be "on" can be rewritten using the dissociation constants as

$$p_{on} = \frac{1}{1 + e^{-\beta(\varepsilon_{off} - \varepsilon_{on})}\frac{(1+\frac{[L]}{K_d^{off}})}{(1+\frac{[L]}{K_d^{on}})}}. \tag{19.73}$$

This formula suggests that the probability of the "on" state depends on a few biologically important variables; the energy difference between the "on" and "off" states of the receptor in the absence of ligand, the affinity of the ligand for the "on" state and the "off" state of the receptor respectively, and the amount of ligand itself. For attractive substances, binding of the ligand will tend to favor the "off" state (where CheY is not phosphorylated), that is, $K_d^{off} < K_d^{on}$.

Let us consider the implications of this result. In the absence of ligand (if $[L] = 0$), the equation simplifies to the familiar result for a two-state system such as an ion channel. If ε_{off} is lower than ε_{on}, then the "off" state predominates in the absence of ligand (i.e p_{on} is low), and addition of ligand will simply drive p_{on} to be even lower. The system will be sensitive to concentrations of ligand that are comparable to or greater than K_d^{off}. However, if ε_{off} is higher than ε_{on}, the "on" state predominates in the absence of ligand, and addition of ligand will drive p_{on} down. In this regime, the system will only be sensitive to concentrations of ligand that are comparable to or greater than K_d^{on}.

In order to modulate its response over a wide range of ligand concentrations and conditions, *E. coli* is actually able to change the relative values of ε_{on} and ε_{off}, and also the relevant K_ds, by performing regulated covalent modifications of the receptor protein itself. This is the job of the methylase CheR and the demethylase CheB, which add and remove methyl groups on a series of glutamate residues present in the intracellular domain of the membrane-spanning receptor protein. The more highly methylated the receptor protein, the more likely it is to be in the "on" state. These modifications permit two impressive consequences. First, as mentioned above, *E. coli* can detect gradients of chemoattractants by comparing the strength of the signal it currently senses to the strength of the signal it detected in the recent past. Second, the bacterium

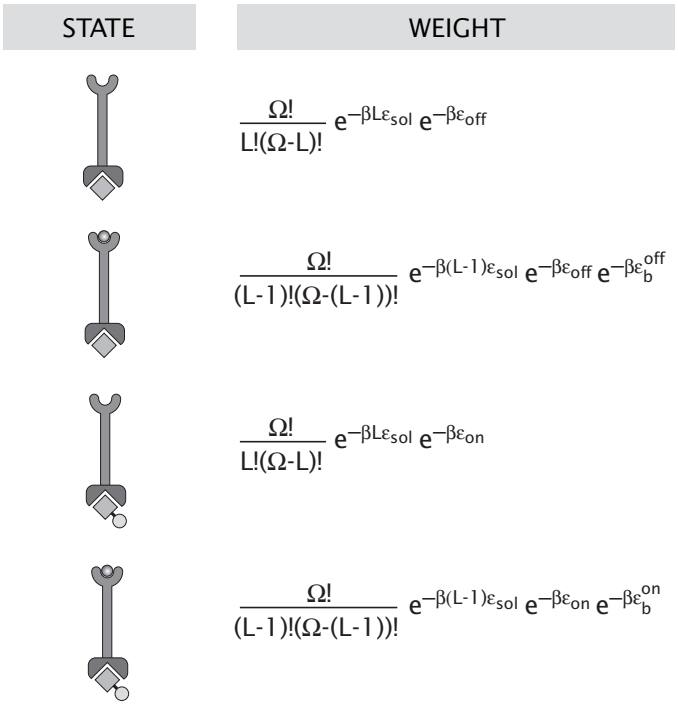

Figure 19.32: States and weights for a simple model of bacterial chemotaxis.

is able to detect gradients in concentration over many orders of magnitude of absolute concentrations, a phenomenon known as adaptation. This corresponds to our own ability to be able to whisper to someone else even in a crowded and noisy room, or our ability to see our surroundings either inside a darkened room or after stepping out into the bright sunshine. For the bacteria, both adaptation and time-sensing depend on the fact that the demethylase, CheB, is itself regulated by phosphorylation by CheA, and therefore depends on ligand binding to the receptor. If CheB is phosphorylated (i.e. if the receptor is "on"), CheB will be more active as a demethylase, and will tend to convert the receptor into an "off" state, damping the response. Conversely, if CheB is dephosphorylated (i.e. the receptor is "off"), more methyl groups will accumulate, tending to switch the receptor "on". This sequence of events takes some time, a few seconds. At the same time, ligand binding influences the activity state of the receptor. Therefore, receptor occupancy by ligand reflects current conditions, and the methylation state of the receptor reflects the past conditions of a few seconds ago. The cell is able to swim up concentration gradients essentially by comparing these two signals.

Our calculations so far illustrate the key ideas, but they will not suffice to capture the full complexity of chemotactic behavior as revealed in fig. 19.33(A). In addition to the precise adaptation already discussed, the system exhibits a high degree of cooperativity. To account for cooperativity, our previous results can be amended to the form

$$p_{on} = \frac{1}{1 + e^{-n\beta(\varepsilon_{off} - \varepsilon_{on})} \frac{(1+\frac{[L]}{K_d^{off}})^n}{(1+\frac{[L]}{K_d^{on}})^n}}. \tag{19.74}$$

The inclusion of cooperativity sharpens the response of the system. Previously, we have considered cases of cooperativity such as oxygen binding to hemoglobin where a single protein has multiple ligand binding sites. In chemotaxis, the *E. coli* cell clusters essentially all its membrane-spanning receptors together in a single patch at one pole in a tight cluster, such that binding of one ligand to one receptor can influence the conformational state of many other receptors, including distinct receptors that are able to detect different substances. A fully detailed mathematical model that incorporates adaptation and cooperativity in mixed receptor clusters along with the basic two-state model derived above is able to reproduce many of the complex features of chemotactic receptor response, as illustrated in fig. 19.33(B).

19.4.2 Biochemistry on a Leash

One of the most fundamental features of living organisms is movement. As noted in our discussion of chemotaxis, cells make "decisions" about where to go and these decisions in eukaryotes are implemented in the form of polymerization of actin filaments. Examples of actin polymerization organized in both space and time were shown in figs. 15.2 (pg. 755) and 15.3 (pg. 756). What chains

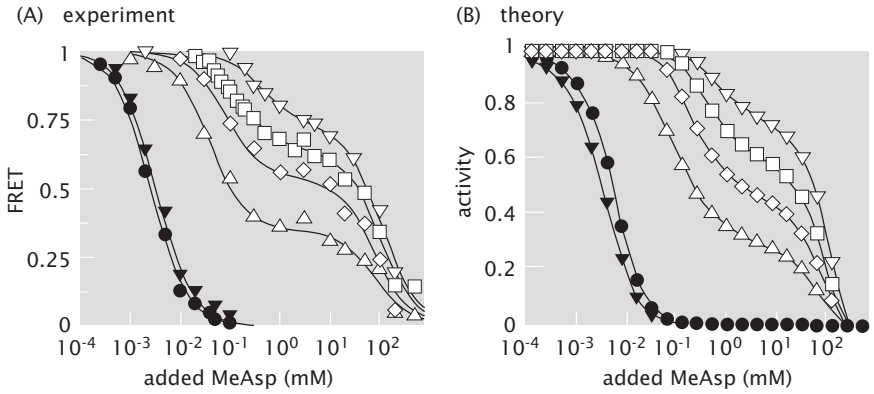

Figure 19.33: Probability of receptor being "on". (A) Graph of concentration-dependence of the on probability based on *in vivo* FRET measurements. The different curves correspond to different bacterial strains. The wild-type response is shown as dark circles. The white symbols are for mutants that correspond to different states of receptor methylation, increasing from left to right. (B) The results of a calculation of the probability of the receptor being active as a function of the concentration of chemoattractant. The model reproduces many aspects of the living cell responses, including the complex behaviors of the methylation mutants. (Adapted from J. .E. Keymer *et al.*, Proc. Nat. Acad. Sci., 103:1786, 2006.)

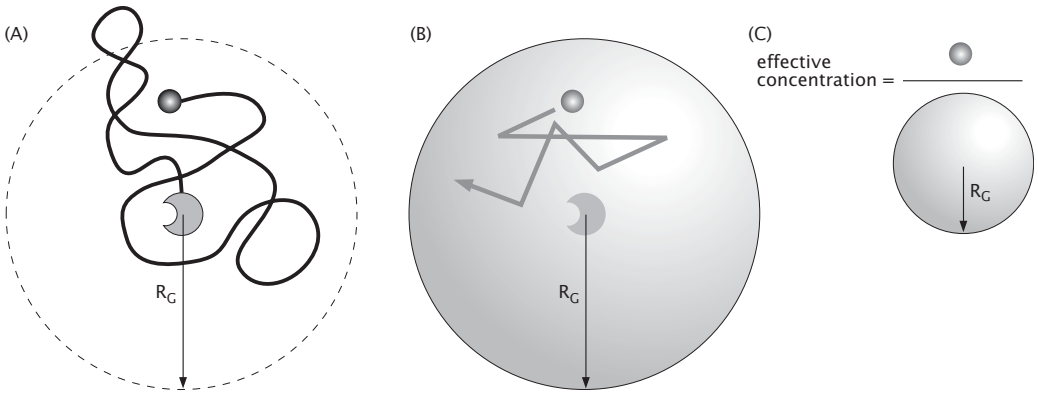

Figure 19.34: Tethering and effective concentration. (A) As a result of tethering, the ligand can only explore a limited region of space. (B) The concentration of the tethered ligand can be estimated by considering a sphere with a radius given by the radius of gyration of the tether. (C) To compute the effective concentration due to tethering, consider one ligand per volume given by a sphere with a radius equal to that of the radius of gyration.

of events link the detection of some external cue and the formation of new actin filaments in a motile cell? The advent of video microscopy in conjunction with a host of different classes of fluorescent markers has made the study of cell motility one of the most exciting areas of current research. As a particular case study that will allow us to flex several sets of muscles we have developed throughout the book, we consider molecules that have the interesting feature that they include a tethered ligand and receptor pair that compete with free ligands. These tethering motifs are a common feature of signaling molecules.

Tethering Increases the Local Concentration of a Ligand

One simple way to see the significance of tethering is illustrated in fig. 19.34. The idea is that the tethered ligand is confined to a volume dictated by the length of the tether. In particular, if the tether has a length L, then the effective concentration of the tethered ligand can be estimated as

$$\text{effective concentration} = \frac{1}{\frac{4}{3}\pi L^3}. \tag{19.75}$$

To develop an intuitive sense of the significance of this tethering, this estimate can be used to roughly determine the concentration at which the free ligands compete with the tethered ligand. In particular, for the case in which a tethered ligand competes with free ligands for the attention of a tethered receptor, clearly at high enough concentrations, the free ligands will dominate the binding.

Signaling Networks Help Cells Decide When and Where to Grow Their Actin Filaments For Motility

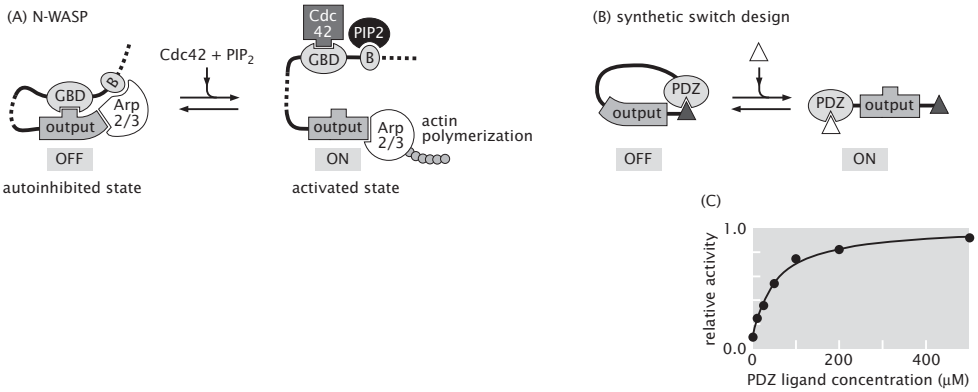

Figure 19.35: Schematic of signaling process leading to actin polymerization. (A) Activation of Arp2/3 by ligands Cdc42 and PIP_2, (B) synthetic switch constructed to activate Arp2/3 as a result of the presence of alternative ligand, (C) activity of the synthetic switch as a function of the signaling ligand. (Adapted from J. E. Dueber *et al.*, Science, 301:1904, 2003.)

The case of bacterial chemotaxis described above is but one of many examples where the motility of cells is dictated by the presence of environmental cues. In many cases, these environmental cues have the effect of inducing actin polymerization which leads to changes in cell shape which are then coupled to motility. From the standpoint of cell signaling, small signaling molecule can relay information to N-WASP, a protein that can interface with a complex of proteins called the Arp2/3 complex to create new actin filaments. The way in which this works is shown in fig. 19.35(A). In particular, the presence of two ligands, Cdc42 and PIP_2 activate N-WASP by binding to this protein in a way that then permits it to activate Arp2/3. The presence of Cdc42 and PIP_2 leads to the unbinding of GDB and B domains from the C-domain and Arp2/3, and N-WASP begins to stimulate actin polymerization by recruiting (and perhaps appropriately orienting) actin monomers to the proximity of the Arp2/3. With the help of activated N-WASP, Arp2/3 promotes actin polymerization by providing heterogeneous nucleation sites. Here, our aim is to study this process quantitatively.

Synthetic Signaling Networks Permit a Dissection of Signaling Pathways

As with the analysis of genetic networks, one exciting way in which signaling pathways have been dissected is by rewiring such pathways to form various synthetic signaling networks. Fig. 19.35(B) shows a synthetic activator of Arp2/3 in which a domain known as a PDZ domain is attached to the output domain that activates Arp2/3. On the other end of the construct is a peptide sequence that binds to PDZ. This synthetic protein mimics N-WASP and can be activated

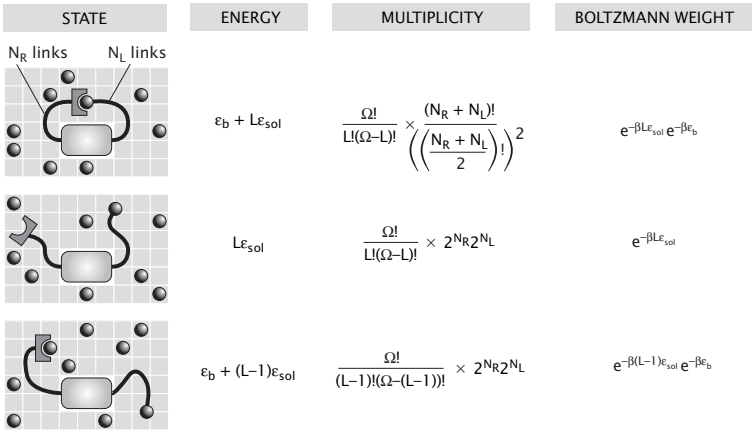

Figure 19.36: States and weights for the synthetic signaling problem.

by soluble ligands that bind to the PDZ domain.

To analyze the function of this signaling process, we invoke statistical mechanics in the same spirit as we have earlier for considering gene regulation. The goal of our statistical mechanical model of the synthetic switch is to work out the probability that the molecule is in the active state. In particular, the active state corresponds to the case in which the tethered receptor is not bound to the tethered ligand. That is, the tethered ligand and receptor are separately flopping around freely. As usual, we resort to a states and weights diagram to work out the probability of the active state. As shown in fig. 19.36, there are three classes of states, each with their own corresponding statistical weights: i) the switch is in the autoinhibitory state and the tethered ligand and receptor are bound to each other, ii) the tethered ligand and receptor are both flopping around freely and the receptor has no bound free ligands, iii) the tethered ligand and receptor are both flopping around freely, and the receptor has bound one of the free ligands. Our aim is to make falsifiable predictions for the signal dependence on e.g. the linker length and ligand concentration.

To develop an intuitive sense of how this situation plays out, the probability of finding the switch in the active state is represented schematically in fig. 19.37. The essence of the situation is that as the concentration of free ligand is increased, the probability that the receptor will be bound by one of the free ligands will increase until this outcome dominates the probability. From the standpoint of testing our understanding of such systems, one of the other design parameters that can be varied is the length of the flexible tethers. As will be shown explicitly when we demonstrate the contributions of the autoinhibitory state to the overall partition function, the length of the tether is a significant part of the overall free energy budget.

To make this calculation concrete, we resort here to simple one-dimensional ideas on the random walk introduced in chap. 8 and show how the calculation

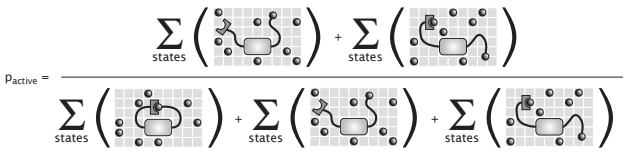

Figure 19.37: Probability of activation of Arp2/3.

generalizes to three-dimensions, but leave the details for the reader as a problem at the end of the chapter. Our strategy will be to break the *total* partition function for this system down into three parts as reflected in fig. 19.36, where the sum can be written as

$$Z_{tot}(L, N_R, N_L) = \underbrace{Z_1(L, N_R, N_L)}_{\text{autoinhibitory state}} + \underbrace{Z_2(L, N_R, N_L)}_{\text{free tethers}} + \underbrace{Z_3(L, N_R, N_L)}_{\text{tether with ligand}}. \tag{19.76}$$

The parameter L is the number of ligands in the system, N_R is the number of Kuhn segments in the polymer tether that has the tethered receptor and N_L is the number of Kuhn segments in the polymer tether that has the tethered ligand. Given these decompositions, we can then write the probability that the switch will be in the active state as

$$p_{active} = \frac{Z_2 + Z_3}{Z_1 + Z_2 + Z_3}. \tag{19.77}$$

The separate contributions to the total partition function can be worked out in much the way we have done similar problems throughout the book. The key point is that each class of state has a number of microscopically equivalent configurations and to find their contribution to the overall partition function, we need to multiply the Boltzmann weight for each class of state by its corresponding microscopic degeneracy (obtained by adding up all of the different ways of arranging the system). For example, the contribution from the states in which the tethers are flopping around freely and there is no free ligand bound is given by

$$Z_2 = \underbrace{\frac{N!}{L!(N-L)!}}_{\text{solution ligands}} \times \underbrace{2^{N_R} 2^{N_L}}_{\text{tether configs.}} \times \underbrace{e^{-\beta L \varepsilon_{sol}} e^{-\beta \varepsilon_{sol}^{lig}}}_{\text{Boltzmann weight}}. \tag{19.78}$$

The treatment of the tether degrees of freedom is based on the simplest one-dimensional random walk in which we imagine that every segment in the tether

can point either to the left or right and we don't worry about self-avoidance. It is straightforward to use a more robust model of the tethers, but we use this one for simplicity. What this means precisely is that each tether can be in one of 2^N different configurations, where N is the number of Kuhn segments in the tether of interest. We have also introduced the energy ε_{sol} for the energy of the ligands when they are free in solution and the parameter ε_{sol}^{lig} for the energy of the tethered ligand when it is in solution. The most interesting class of states are those associated with the autoinhibition of the switch and which involve the tethering ligand and receptor being linked up. In this case, the contribution to the partition function is

$$Z_1 = \underbrace{\frac{N!}{L!(N-L)!}}_{\text{solution ligands}} \times \underbrace{\frac{(N_R+N_L)!}{((\frac{N_R+N_L}{2})!)^2}}_{\text{tether closure}} \times \underbrace{e^{-\beta L \varepsilon_{sol}} e^{-\beta \varepsilon_B}}_{\text{Boltzmann weight}}, \tag{19.79}$$

where we have used the result from section 8.2.4. The contribution from tether closure is the number of ways of making a closed loop out of a polymer of length $N_R + N_L$ Kuhn segments. The last contribution to the total partition function arises from those microstates in which one of the free ligands attaches to the tethered receptor. This means that the solution contribution to the partition function will only involve $L-1$ ligands. This term can be written as

$$Z_3 = \underbrace{\frac{N!}{(L-1)!(N-(L-1))!}}_{\text{solution ligands}} \times \underbrace{2^{N_R} 2^{N_L}}_{\text{tether configs.}} \times \underbrace{e^{-\beta(L-1)\varepsilon_{sol}} e^{-\beta \varepsilon_B}}_{\text{Boltzmann weight}}. \tag{19.80}$$

The actual formula for p_{active} can now be obtained by substituting the values for Z_1, Z_2 and Z_3 obtained above into eqn. 19.77. The resulting expression is considerably simpler if we use an alternative form of this equation, namely,

$$p_{active} = \frac{1 + \frac{Z_3}{Z_2}}{1 + \frac{Z_1}{Z_2} + \frac{Z_3}{Z_2}}. \tag{19.81}$$

This leads to an expression for p_{active} of the form

$$p_{active} = \frac{1 + \frac{c}{c_0} e^{-\beta \Delta \varepsilon_1}}{1 + p_{loop} e^{-\beta \Delta \varepsilon_2} + \frac{c}{c_0} e^{-\beta \Delta \varepsilon_1}} \tag{19.82}$$

where we have introduced $c = L/(Nv)$, $c_0 = 1/v$ and p_{loop} which is the probability of forming a loop. For the one-dimensional model considered above, we have

$$p_{loop} = \frac{\frac{(N_R+N_L)!}{((\frac{N_R+N_L}{2})!)^2}}{2^{N_R+N_L}} \tag{19.83}$$

which amounts to the ratio of the number of closed configurations for the polymer of length $N_R + N_L$ to the *total* number of configurations. However, the

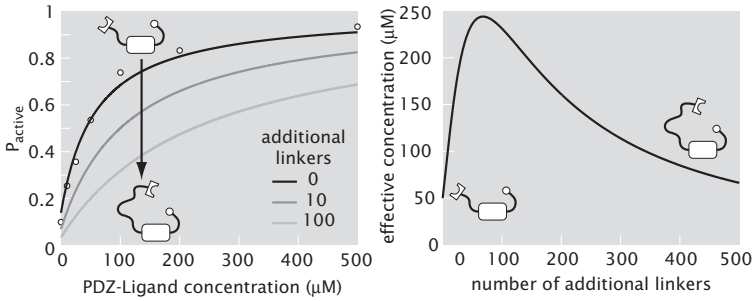

Figure 19.38: Prediction of dependence of activation on effective tail length. (A) p_{active} as a function of ligand concentration for different tether lengths. Experimental data shown as small circles. (B) The effective concentration of tethered ligand as seen by the tethered PDZ domain as a function of tether length. (Data from J. E. Dueber *et al.*, Science, 301:1904, 2003.)

one-dimensional model has outlived its usefulness and we can just as well use the result of a full three-dimensional analysis of p_{loop} using the Gaussian model of a polymer, for example. This calculation is left as an exercise for the reader.

The outcome of this kind of analysis is shown in fig. 19.38. There are several subtleties that were not accounted for in the calculation described above. First, as shown in the figure, the tethers do not emanate from the same point. This results in a fundamental difference in the behavior of p_{loop} as a function of tether length as shown in fig. 19.38(B). Second, we used a three-dimensional Gaussian model for the tethers rather than the one-dimensional example worked out above. Both of these features are left as an exercise for the reader in the problems at the end of the chapter.

19.5 Summary and Conclusions

Regulation and signaling are two of the most important manners in which cells orchestrate their behavior in space and time. The goal of this chapter has been to take stock of some of the key architectures of regulatory and signaling networks and to show how simple models using statistical mechanics and rate equations can be put forth to develop intuition and to make predictions about how these networks work. The so-called "thermodynamic models" of gene expression are predicated on the idea of using equilibrium statistical mechanics to examine the probability of promoter occupancy. A dynamical interpretation of these same questions uses rate equations to compute the concentration of both mRNA and their associated proteins.

19.6 Appendix: Stability Analysis for the Genetic Switch

We search for additional steady state solutions of the genetic switch by analyzing the case of α large and α small. First, assume $\alpha \gg 1$. We also assume that the solutions to the steady state equations, namely,

$$\begin{aligned} u^* &= \frac{\alpha}{1+v^{*n}} \\ v^* &= \frac{\alpha}{1+u^{*n}} \ , \end{aligned} \tag{19.84}$$

are such that $u^* \ll 1$. Then $1+u^{*n} \approx 1$ and the steady state values for the two concentrations that follow from eqn. 19.84, to lowest order in $1/\alpha$, are

$$\begin{aligned} u^* &= \alpha^{1-n} \\ v^* &= \alpha \ , \end{aligned} \tag{19.85}$$

consistent with the assumptions we have made. Similarly, by assuming that the solution to eqn. 19.84 has the property $v^* \ll 1$, from which $1+v^{*n} \approx 1$ follows, we find a new solution

$$\begin{aligned} u^* &= \alpha \\ v^* &= \alpha^{1-n} \ , \end{aligned} \tag{19.86}$$

for which the roles of u and v are exchanged. Assuming that both u^* and v^* are large leads to $u^* = v^* = \alpha^{1/1+n}$ while the assumption that both are small is inconsistent with eqn. 19.84. We conclude that in addition to the $u^* = v^*$ case, there are two other steady state protein concentrations. Interestingly, the additional solutions are characterized by very different values for u^* and v^* providing the necessary ingredients for a genetic switch.

Next we analyze the case $\alpha \ll 1$. Following the same analysis as above we do not find any additional solutions. Namely, assuming $u^* \ll 1$ we compute from eqn. 19.84 $v^* = \alpha$ and $u^* = \alpha$, since now $1+\alpha^n \approx 1$. The same conclusions are reached assuming $v \ll 1$ while the assumptions $u^* \gg 1$ or $v^* \gg 1$ are not consistent with eqn. 19.84. We conclude that there is a critical value of the parameter α, which will be of order one and dependent on the value of the Hill coefficient n, such that for values of α below the critical value the steady state solution is unique, while for larger α's there will be three steady states. Now we examine the stability of these solutions, paying particular attention to the case when very different values for u and v are obtained in the steady state.

One of the most important requirements in carrying out an analysis like that given above is to assess the stability of the solutions to a given problem. What this means is that if we perturb the system slightly from the steady state (i.e. $u = u^* + \delta u$ and $v = v^* + \delta v$), we ask do the perturbations grow or shrink in time. If the perturbations grow in time, the system is said to be unstable. If the perturbations shrink in time, the system is said to be stable. A favorite

example for depicting this idea is to consider a particle on some potential energy landscape. If the particle is at the bottom of a well (i.e. the potential energy is locally of the form $\frac{1}{2}kx^2$) a small disturbance of the particle from its equilibrium position will result in jiggling around the equilibrium point. Alternatively, if the particle is balanced at the $x = 0$ on a potential energy landscape of the form $-\frac{1}{2}kx^2$, any slight disturbance to the particle will cause it to wander away from the equilibrium. The idea of our stability analysis in this case is the same - we ask, does a slight disturbance away from the steady state concentration lead to solutions that grow or decay in time.

To asses the stability of the steady state we analyze the linear equations for the small deviations $(\delta u, \delta v)$ of the repressor concentrations away from their steady state values. In particular, in eqn. 19.40, we substitute $u = u^* + \delta u(t)$ and $v = v^* + \delta v(t)$ and then exploit the fact that $\delta u(t)$ and $\delta v(t)$ are small and Taylor expand the nonlinear Hill functions in powers of δu and δv. The result of this analysis is

$$\frac{d}{dt}\begin{pmatrix} \delta u \\ \delta v \end{pmatrix} = \hat{\mathbf{A}}\begin{pmatrix} \delta u \\ \delta v \end{pmatrix}. \tag{19.87}$$

As noted above, the matrix

$$\hat{\mathbf{A}} = \begin{pmatrix} -1 & f'(v^*) \\ f'(u^*) & -1 \end{pmatrix} \tag{19.88}$$

results from linearizing the rate equations, eqn. 19.40, around the steady state solution (u^*, v^*), and

$$f'(x) = -\frac{n\alpha x^{n-1}}{(1+x^n)^2}\ . \tag{19.89}$$

At this point, the stability of this *linear* set of equations is queried by assuming solutions of the form $\delta u(t) = \delta u_0 e^{\lambda t}$ and $\delta v(t) = \delta v_0 e^{\lambda t}$. The essence of the analysis (which is left as a problem at the end of the chapter) is to examine the *sign* of the parameter λ. If $\lambda < 0$, the perturbations decay in time and if $\lambda > 0$ the perturbations grow in time. The behavior of λ is revealed by examining the eigenvalues of the matrix $\mathbf{A}$. The eigenvalues of $\hat{\mathbf{A}}$ are both real and are given by

$$\lambda_{1,2} = -1 \pm \sqrt{f'(u^*)f'(v^*)}. \tag{19.90}$$

For the steady state solution to be stable both λ_1 and λ_2 need to be negative. This will be the case if

$$f'(u^*)f'(v^*) < 1\ . \tag{19.91}$$

RP: but greater than zero, right

Given this condition for the stability of the solutions, we can now revisit the different solutions found above and explicitly examine their stability. First we consider the single steady state, $u^* = v^* = \alpha$, that we found for $\alpha \ll 1$. In this case, using eqn. 19.89, we find $f'(u^*)f'(v^*) = n^2\alpha^{2n} \ll 1$, and the stability condition, eqn. 19.91, is satisfied. Next, we consider the three steady state solutions found for $\alpha \gg 1$. For the solution $u^* = v^* = \alpha^{1/(1+n)}$, we find that

$f'(u^*)f'(v^*) = n^2$. Since the Hill coefficient satisfies the condition $n > 1$, we conclude that this solution is unstable. A small perturbation will drive it to one of the other two solutions, which are stable. Namely, for $u^* = \alpha^{1-n}$ and $v^* = \alpha$, we see that $f'(u^*)f'(v^*) = n^2\alpha^{-n(n-1)} \ll 1$, and we conclude that the solution is stable. Since the third solution is obtained by u^* and v^* switching roles, it too will be stable.

The analysis above leads to the phase diagram shown in fig. 19.28 in terms of the parameter α. For α less than some critical value (which is of order one) the rate equations at long times lead to a unique steady state in which the concentrations of the two repressor proteins are equal. On the other hand, for α larger than the critical value, at long times the system will settle into one of two stable states, with the concentration of one repressor dominating over the other. Which of the two steady states is reached will depend on the initial conditions. In this regime the rate equations, eqn. 19.39, describe a genetic switch.

19.7 Problems

1. Strong and weak promoters.

In the chapter we introduced repression as a quantitative measure of the reduction in the level of gene expression due to the action of a repressor molecule. For the simple model of repression introduced on pg. 1025 make a plot comparing repression in the case of a weak and a strong promoter. Show that, unlike the weak promoter case, in the case of the strong promoter, the repression depends upon the number of polymerase molecules in the cell.

2. Lac Repressor and the *lac* Operon

A beautiful set of quantitative experiments on the *lac* operon were done by the Muller-Hill group in the 90's, where repression of expression of the *lacZ* gene was measured in a population of different mutant *E. coli* cells. Different mutant cells differed in the number, sequence, and position of the operator sites that bind the Lac repressor. In this problem we explore how, using thermodynamic models of gene expression, this data can be used to obtain a number of quantities characterizing the Lac repressor-DNA interaction as well as DNA looping.

(a) Using the data from Oehler *et al.* shown in fig. 19.19 determine the *in vivo* binding energy of Lac repressor to each one of its operators and reproduce fig. 19.20.
(b) Using your results from (a), and the repression measured by Oehler *et al.* in cells with two operators present, which leads to DNA looping, in order to determine the looping energy and to reproduce fig. 19.24.
(c) As mentioned many times throughout the book, Müller *et al.* performed a beautiful experiment where the repression level was measured as a function of the distance between operators. The experiment and its results are shown in fig. 1.11. Based on their repression data and the thermodynamic models from

the chapter make a plot of the looping energy as a function of the interoperator distance. Show analytically that a maximum in repression corresponds to a minimum in looping energy. At what interoperator distance is the inferred looping free energy at a minimum? Is this consistent with the measured persistence length of DNA *in vitro*, which is 50 nm.
(d) Fit the looping energy obtained in (c) to the functional form $\Delta F_{loop} = a/N_{bp} + b\ln(N_{bp}) + cN_{bp} + e$. Use this looping energy to make predictions about the outcome of a hypothetical experiment similar to the one performed by Müller *et al.*, but now using cells bearing 10, 200 and 900 Lac repressor molecules per cell.
Relevant data for this problem is provided on the book website.

3. Activation and Squelching

In the chapter we computed the regulation factor for a simple activator and found that it describes an effective increase of the number of RNA polymerases in the cell. Furthermore we found that beyond a certain point increasing the number of activators no longer has an effect on gene expression, since the regulation factor asymptotes to a constant; see fig. 19.12. This last conclusion is not quite true and in fact an exceedingly large number of activators can begin to repress gene expression. This effect is known as squelching. Here we study a simple model of squelching.
As discussed in the chapter, assume that all the RNA polymerases and all the activators are bound to DNA nonspecifically. RNA polymerase and the activator can be on their own, or in a AP complex with binding energy ϵ_{ap}. Compute the average number of AP complexes, and the average number of free RNA polymerases. Show that in the limit of a very large number of activators the number of free RNA polymerases goes to zero. Explain why this leads to squelching.

4. Sensitivity of the regulation factor.

An important concept in gene regulation is the sensitivity, that is, how steep is the change in gene expression in response to a change in the number of transcription factors. It is defined as the slope of the gene expression with respect to the number of transcription factor on a log-log plot. Using thermodynamics models of gene regulation determine how the sensitivity depends on the relevant parameters for the following regulation motifs in the case of a weak promoter:
(a) Simple repression.
(b) Repression in the presence of DNA looping.
(c) Two repressors that can recruit each other to their respective operator sites and repress RNAP independently. What happens when the interaction is turned off?
(d) Two repressors competing for the same operator site which overlaps the promoter.
(e) Simple activation.

5. The transcriptional machinery in eukaryotes.

In the thermodynamic models of gene regulation discussed in the chapter the RNA polymerase is treated as a single molecular species. This is a reasonable assumption for transcription in prokaryotes, while in eukaryotes some 60 different molecules need to come together in order to form the transcriptional machinery. The objective of this problem is to develop intuition about the requirements for our simple model to apply in such a complex case by assuming that the transcriptional machinery is made out of two different subunits, A and B, that come together at the promoter.
(a) Calculate the probability of finding the complex A+B bound to the promoter in the case where unit A binds to DNA and unit B binds to A. Can you reduce this to an effective one-molecule problem such as in the bacterial case?
(b) Investigate the same question for the case when B binds to a site on the DNA which is near the A binding site, and there is an interaction energy between A and B.

6. Induction of transcription factors.

Even though experiments where the concentration of a transcription factor is varied are easier to interpret in terms of models, the experiments that are the easiest to perform are those where the affinity of the transcription factor to its specific binding sites on the DNA is regulated by an inducer molecule. In the case of Lac repressor, for example, allolactose or any of its analogues (IPTG, for example) can be used to reduce its specific binding energy down to values similar to its non-specific binding to DNA.

Assume a simple model of induction where one inducer molecule binds to the repressor which then looses its ability to bind specifically to its operator site. Calculate repression in this case and plot it as a function of the number of inducer molecules in the cell.

7. Separation of time scales and transcriptional regulation

For transcription to start the RNA polymerase bound to the promoter needs to undergo a conformational change to the so called open complex. The rate of open complex formation is typically much smaller that the rates for the polymerase binding and falling off the promoter. Here we investigate within a simple model how this state of affairs might justify the equilibrium assumption underlying thermodynamic models of gene regulation, namely that the equilibrium probability that the promoter is occupied by the RNA polymerase determines the level of gene expression.
a) Write down the chemical kinetics equation for this situation. Consider three states: RNA polymerase bound non-specifically on the DNA (N), RNA polymerase bound to the promoter in the closed complex (C), and RNA polymerase bound to the promoter in the open complex (O). To simplify matters take both the rate for N→C and the rate for C→N to be k. Assume that the transition C→O is irreversible, with rate Γ.
b) For $\Gamma = 0$ show that in the steady state there is an equal number of RNA polymerases in the N and C state. What is the steady state in the case $\Gamma \neq 0$.

c) For the case $\Gamma \neq 0$ show that for times $1/k \ll t \ll 1/\Gamma$ the number of RNA polymerases in the N and C state is equal, as would be expected in equilibrium.

8. Genetic switch.

In the chapter we introduced the synthetic genetic switch, consisting of two promoters which repress each other via the transcription factors whose production they control. In this problem we investigate further investigate the rate equations introduced in the chapter to describe this synthetic gene circuit.
a) Show that the critical value of the dimensionless control parameter α is given by

$$\alpha_{crit} = \frac{n}{(n-1)^{\left(\frac{n+1}{n}\right)}}. \tag{19.92}$$

with n being the Hill coefficient. Make a phase portrait of the genetic switch for the case $n = 4$.
b) Consider the simple model of repression where a single repressor protein binds to an operator site that overlaps with the promoter site. For this case write down the rate equations for the genetic switch introduced in the chapter.
c) Show that the rate equations have a single steady state solution that corresponds to an equal number of repressor$_1$ and repressor$_2$ proteins, and therefore will not exhibit bistability necessary for switching.

9. Statistical mechanics and the repressilator.

In our treatment of the repressilator, the Hill function was used to characterize binding of transcription factors to DNA. Replace that treatment by one in which p_{bound} is calculated as we did throughout the book and numerically integrate the equations.

10. Chemotaxis of *E. coli*.

In chemotaxis experiments a source of nutrient molecules is typically introduced into the medium containing bacteria via a micropipette. The outward diffusion of the nutrient molecules creates a position dependent concentration gradient and the chemotactic response of the bacteria can be observed under a microscope.
a) Estimate the nutrient gradient in steady state as a function of the distance from the micropipette r by assuming that it keeps the concentration fixed at c_0 for distances $r < r_0$. Make a plot of the concentration gradient as a function of r for typical values $c_0 = 100$ μM and $r_0 = 1$ μm.
b) Assuming that the bacterium makes two measurements of the concentration using one receptor protein at one of its ends and another receptor at the other, estimate the maximum distance from the nutrient source for which the bacterium is still able to detect a gradient. Assume that the receptor counts the number of molecules present in a spherical volume of radius $a = 2$ nm. To solve this problem you should recall that the counting error for counting N molecules is roughly $\sqrt{N}$, and in order to detect the difference in concentration between

the two ends of the bacterium, the measurement error should be less than the difference itself.
c) Now assume a different strategy, where one receptor is employed but the bacterium compares the concentration at two different positions along a run, separated by a distance of 10 μm. Compute the maximum distance from the nutrient source at which the bacterium will be able to detect the gradient in this case.

11. N-wasp and biochemistry on a leash.

In the last section of the chapter we considered the action of N-wasp using a simple one-dimensional random walk model to treat the statistical mechanics of looping. Redo that analysis by using the Gaussian model of a polymer chain. First, assume that the loop has to close on itself and then account for the finite size of the protein domain. Compare your results to those obtained in the chapter.

19.8 Further Reading

L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman and R. Phillips, "Transcriptional regulation by the numbers: applications", Curr. Opin. Genet. Dev. **15**, 125 (2005). Application of thermodynamic models to several different regulatory architectures.

N. E. Buchler, U. Gerland and T. Hwa, "On schemes of combinatorial transcription logic", Proc. Natl. Acad. Sci. **100**, 5136 (2003). Excellent general discussion of thermodynamic models of gene regulation.

C. T. Walsh, **Posttranslational Modification of Proteins - Expanding Nature's Inventory**, Roberts and Company Publishers, Englewood: Colorado, 2006. Besides the cool subtitle, this book is full of interesting insights into the phenomenology of posttranslational modification. This is a reminder that there is more to regulation than transcriptional control.

J. L. Cherry and F. R. Adler, "How to make a Biological Switch", J. theor. Biol. **203**, 117 (2000). This article presents a nice discussion of the issues that arise in designing biological switches.

M. Ptashne, **A Genetic Switch**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, 2004. A beautiful book that focuses on ideas as opposed to facts and paints a picture of how gene regulation works.

M. Ptashne and A. Gann, **Genes and Signals**, Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, 2002. Like in his previous book, Ptashne and Gann have created an outstanding description of their ideas also,

call attention to the book review of their book.

B. Müller-Hill, **The Lac Operon**, Walter de Gruyter and Co., Berlin: Germany, 1996. Muller-Hill's book is a fascinating and idiosyncratic account of the development of thinking on the gene regulation in general and the *lac* operon in particular. The book is full of interesting touches such as fig. 3, for example, which illustrates the ways in which synthetic analogs of lactose have played a role in the development of molecular biology.

E. H. Davidson, **Genomic Regulatory Systems**, Academic Press, San Diego: California, 2001. Davidson's book is full of both interesting facts and provocative ideas. We particularly recommend it as a way to explore the complexity associated with eukaryotic gene regulation.

J. Gerhart and M. Kirschner, **Cells, Embryos and Evolution**, Blackwell Science, Malden: Massachusetts, 1997. This is a thought provoking book, much of which remains over our heads. It seems clear that trying to work out the physics of some of the problems they discuss is a ripe area for effort.

U. Alon, **An Introduction to Systems Biology**, Chapman & Hall/CRC, Boca Raton: Florida, 2007. Alon's book gives a comprehensive and thoughtful discussion of regulation.

S. P. Ellner and J. Guckenheimer, **Dynamic models in biology**, Princeton University Press, Princeton: New Jersey, 2006. This book examines dynamical models and their relevance to biology and has a treatment of both the genetic switch and the repressilator.

T. Gregor, W. Bialek, R. R. de Ruyter van Steveninck, D. W. Tank and E. F. Wieschaus, "Diffusion and scaling during early embryonic pattern formation", Proc. Nat. Acad. Sci., **102**, 18403 (2005). This intriguing paper reports observations on the stripes along the anterior-posterior axis of different flies.

H. C. Berg and D. A. Brown, "Chemotaxis in *Escherichia coli* analysed by Three-dimensional tracking", Nature, **239**, 500 (1972). This paper used a novel three-dimensional tracking technique to follow individual bacterial cells during chemotaxis and demonstrated how bacteria find their way by altering the timing of runs and tumbles.

19.9 References

S. Ben-Tabou de-Leon and E. H. Davidson, "Gene regulation: gene control network in development", Annu. Rev. Biophys. Biomol. Struct. **36**, 191 (2007).

J. E. Dueber, B. J. Yeh, K. Chak, and W. .A. Lim, "Reprogramming control of an allosteric signaling switch through modular recombination", Science **301**, 1904 (2003).

M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators", Nature **403**, 6767 (2000).

T. S. Gardner, C. R. Cantor and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*", Nature **403**, 339 (2000).

J. E. Keymer, R. G. Endres, M. Skoge, Y. Meir and N. S. Wingreen, "Chemosensing in *Escherichia coli*: two regimes of two-state receptors", Proc. Natl. Acad. Sci. **103**, 1786 (2006).

J. Müeller, S. Oehler and B. Müller-Hill, "Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator", J. Mol. Biol., **257**, 21 (1996).

E. Myasnikova, A. Samsonova, K. Kozlov, M. Samsonova and J. Reinitz, "Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods", Bioinformatics **17**, 3 (2001).

S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, B. Müller-Hill, "Quality and position of the three *lac* operators of *E. coli* define efficiency of repression", EMBO J. **14**, 3348 (1994).

S. Small, A. Blair and M. Levine, "Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo", EMBO J. **11**, 4047 (1992).