ELSEVIER

DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise

Jose MG Vilar[1] and Leonor Saiz[1,2]

The formation of DNA loops by the binding of proteins and protein complexes at distal DNA sites plays a central role in many cellular processes, such as transcription, recombination and replication. Important thermodynamic concepts underlie the assembly of macromolecular complexes on looped DNA. The effects that this process has on the properties of gene regulation extend beyond the traditional view of DNA looping as a mechanism to increase the affinity of regulatory molecules for their cognate sites. Recent developments indicate that DNA looping can also lead to the suppression of cell-to-cell variability, the control of transcriptional noise, and the activation of cooperative interactions on demand.

Addresses
[1] Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street, New York, NY 10021, USA
[2] Center for Molecular Modeling and Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104, USA

Corresponding author: Vilar, Jose MG (vilar@cbio.mskcc.org)

Current Opinion in Genetics & Development 2005, **15**:136–144

This review comes from a themed issue on
Chromosomes and expression mechanisms
Edited by Barbara Meyer and Jonathan Widom

Available online 5th March 2005

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2005.02.005

Introduction

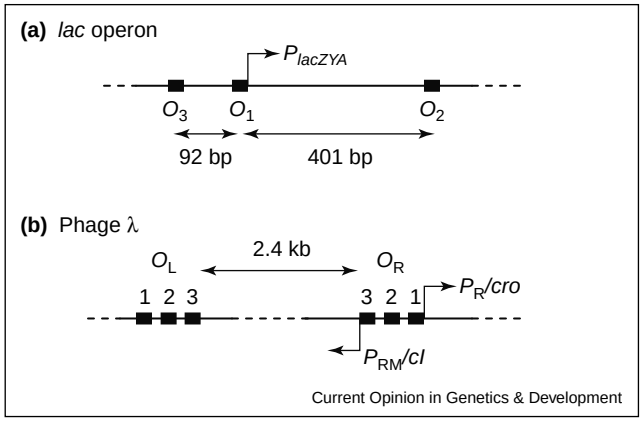
Gene regulation relies to a great extent on proteins that bind to DNA, not only at sites proximal to the genes they regulate but also at distal DNA sites that can be brought to the initiation of transcription region by looping the intervening DNA. This process, known as DNA looping, is widely used in gene regulation. It was first discovered in the *ara* operon of *Escherichia coli* [1] — although it was already suspected to be present in eukaryotic enhancers [2] — and since then it has been found in many other systems, such as the *gal*, *lac* and *deo* operons in *E. Coli* [3–5], the lysogenic-to-lytic switch in phage λ [6••], and the human *β-goblin* locus [7], to name just a few. In eukaryotes, multiple DNA-binding sites that are spread over long distances are involved in controlling the same localized DNA events. Thus, DNA looping is crucial to enable multiple proteins to affect the RNA polymerase in the promoter region. Enhancers, silencers or mediators bound at distal DNA sites are then brought to form part of, affect, or interfere with the transcriptional complex. In prokaryotes, notwithstanding some exceptions [8,9], such complexity in the signal integration machinery is absent and the role of DNA looping appears to be subtler, more similar to a 'fine-tuning' of the underlying biochemistry.

In this review we will, on the one hand, present the key thermodynamic concepts needed to understand the formation of the looped DNA–protein complexes and, on the other hand, explore the effects that the formation of DNA loops has on gene regulation. We consider explicitly as prototype systems the *lac* operon [10] and phage λ [6••], the two systems that led to the discovery of gene regulation. These systems share many characteristics, but also exhibit crucial differences. In the *lac* operon, the lac repressor, a single divalent protein, can bind simultaneously to two DNA sites and induce the formation of a loop. In phage λ, the loop is not formed by a single protein but by a protein complex that is assembled on DNA as the loop forms. The situation with phage λ is, therefore, similar to that of eukaryotic transcription, in which multiple proteins are assembled on the DNA to form the transcriptional complex with RNA polymerase.

Increase in the local concentration and its paradoxes

The *E. coli lac* operon is a clear example that illustrates the elegance of the functioning of biological systems and the subtlety on which they are built. It consists of a regulatory domain and three genes required for the uptake and catabolism of lactose (Figure 1a) [11]. The lac repressor can bind to the main operator O_1, thereby preventing the RNA polymerase from binding to the promoter and transcribing the genes. Furthermore, there are two auxiliary operators, O_2 and O_3, to which the repressor can also bind but not prevent transcription. Interestingly, elimination of either auxiliary operator has only minor effects; yet simultaneous elimination of both reduces the repression level ~100-fold. The reason for this effect is that the lac repressor can bind simultaneously to two operators and loop the intervening DNA. Thus, the presence of O_1 and at least one auxiliary operator is enough to form a DNA loop that substantially increases the ability of the repressor to bind O_1. Binding to an auxiliary operator keeps the repressor close to O_1 and effectively increases its concentration around this site. In this way, the efficiency of the binding to O_1 is greatly enhanced. However, the affinity of the repressor for each of both auxiliary operators is much lower than for O_1 — as much as 10- and 300-fold lower for O_2 and O_3, respec-

Figure 1



Scheme of the different operator positions on the DNA for the *lac* operon and phage λ. **(a)** Location of the main, O_1, and auxiliary operators, O_2 and O_3, of the *lac* operon. Binding of the lac repressor to O_1 represses transcription of the *lacZ*, *lacY* and *lacA* genes. **(b)** Location of the right (O_R) and left (O_L) operators in phage λ. Binding of the λ cI dimer to O_R2 activates transcription of its own gene. Binding of cI dimers to O_R1 and to O_R3 prevents transcription of *cro* and *cI* genes, respectively.

tively. For the increase in the local concentration to be a valid argument to explain DNA looping, the repressor needs to be bound to the auxiliary site in the first place. How is it then possible that a weaker operator can help the binding to O_1?

A related paradox is observed in phage λ. The lysogenic-to-lytic switch is controlled in the phage DNA at two operators, known as the left (O_L) and right (O_R) operators, which are 2.4 kb away from one another (Figure 1b) [6••,12]. Each has a tandem of three DNA motifs to which the λ cI repressor dimers can bind: O_R1, O_R2 and O_R3 for the right; and O_L1, O_L2 and O_L3 for the left operator. It was found recently that two cI dimers bound to O_R1 and O_R2 can form an octamer with two cI dimers bound to O_L1 and O_L2 by looping the intervening DNA [6••,12,13,14••]. The striking fact was not only that this fundamental result came more than 40 years after the discovery of gene regulation in phage λ but also that the increase of the local concentration argument alone cannot be applied to explain looping in this case. Current theories predict that the local concentration for such long loops would be increased, at the most, by less than a factor of 10 [15], which is well below the factor of 1000 that is required for the formation of the octamers in solution [16]. Therefore, the loop is too long to increase substantially the local concentration. How is it then possible that the loop forms when the octamer that ties it would not exist in such low concentrations?

The two counterintuitive examples that we have just discussed — namely, a weak site helping a strong site, and a protein complex that would not exist in solution fastening a DNA loop — have a straightforward explanation when formulated in terms of the appropriate thermodynamic quantities.

Free energies and the thermodynamic basis of regulated recruitment

It is often assumed that cellular processes are similar to chemical reactions occurring in an ideal well-stirred macroscopic reactor. The cell, however, is a small and crowded environment in which many events take place at the same time. At the cellular level, the problem is not so much how to make two proteins interact but, rather, how to prevent them from interacting with all the other proteins they are not supposed to. Concentrations of the different molecular species are, therefore, kept low. To achieve specificity and affinity at the same time, cells have evolved mechanisms to bring molecules close to their interaction sites. It is becoming increasingly evident that this idea, referred to as regulated recruitment, is one of the unifying principles of the molecular functioning of living systems [17••]; DNA looping is just one fundamental example. In this section we present the basic thermodynamic ideas that will enable us to implement regulated recruitment in a quantitative manner and apply it to DNA looping.

One key point — perhaps the most important piece of information — is that the free energy of binding, ΔG_{bind}, can be decomposed into two main contributions:

$$\Delta G_{bind} = \Delta G_{pos} + \Delta G_{int}$$

The interaction free energy, ΔG_{int}, arises from the interactions between the two molecules — interactions such as electrostatic, hydrophobic and Van der Waals forces. The positional free energy, ΔG_{pos}, results from positioning the molecules in the right place and orientation so that they can interact, and it accounts, among other potential contributions, for the loss of translational and rotational entropy upon binding.

Let us consider in more detail the meaning of the positional free energy. If two molecules are to be bound, they have to be positioned within a small volume of the order of the interaction forces. The probability that one molecule is in this volume just by chance is given by the ratio of the volume of interaction, V_{int}, to the volume in which the reaction takes place, V_{reac}. If there are N molecules instead of one, the probability is scaled up accordingly, leading to the simple expression $P_{pos} = NV_{int}/V_{reac}$. Statistical thermodynamics [18] (see also Bintu *et al.*, this issue) this issue links this probability with its corresponding free energy, ΔG_{pos}, through the relationship $P_{pos} \approx e^{-\Delta G_{pos}/RT}$ where R is the gas constant and T is the absolute temperature ($RT \approx 0.6$ kcal/mol for typical experimental conditions). Equating both expressions for the positional probability and taking logarithms leads to

$\Delta G_{pos} = -RT\ln(NV_{int}/V_{reac})$, which links the positional free energy with the concentration, N/V_{reac}, and a microscopic parameter of the binding, V_{int}.

For practical purposes, the expression for the positional free energy can be rewritten as:

$$\Delta G_{pos} = \Delta G^{o}_{pos} - RT\ln[N]$$

by normalizing both volumes by the volume (V_{mol}) associated with one molecule at 1M concentration, $V_{mol} = 1.7\ \mathrm{nm}^3$, and by using the property that the logarithm of the product of two factors is the sum of the logarithm of each factor. The quantity $\Delta G^{o}_{pos} = -RT\ln(V_{int}/V_{mol})$ is the molar positional free energy and $[N] = NV_{mol}/V_{reac}$ is the concentration expressed in moles. In general, the free energy of binding depends on the concentrations of the different components through the positional free energy. Once it is known for a given concentration, ΔG^{o}_{pos} can be easily determined and the preceding expression gives the free energy values for any concentration. Free energies at 1M concentration are known as standard free energies and are usually labelled by the superscript ^{o}.

Typical values of the positional free energy are $\Delta G^{o}_{pos} \approx 15\,\mathrm{kcal/mol}$ [19]. This value indicates that if the free energy of interaction is zero, the probability that two molecules are as close as if they were bound is extremely low. Interaction forces are the ones that provide stability to the bound state. Even low values of binding free energies, such as $\Delta G_{bind} \approx -2$ kcal/mol, would imply considerably high interaction free energies, such as $\Delta G_{int} \approx -17$ kcal/mol. Now imagine that a univalent binding is transformed into a divalent one; that is to say that molecules originally having an interaction site now have two. The positional free energy would remain basically the same, because it is associated with the molecule itself. The free energy of interaction, however, would be twice as much, leading to $\Delta G_{bind} \approx -19$ kcal/mol in the preceding example. Similarly, a trivalent binding would lead to $\Delta G_{bind} \approx -36$ kcal/mol.

This large increase in affinity for multivalent binding is known in inorganic chemistry as the chelate effect [20], and is crucial for the assembly of the macromolecular complexes that implement regulated recruitment. Multiple binding domains also introduce additional complexity. In general, one should also consider the conformational free energy, ΔG_{conf}, which accounts for the structural changes needed to accommodate multiple simultaneous interactions. The free energy of binding is, thus, the sum of all the contributions: $\Delta G_{bind} = \Delta G_{pos} + \sum \Delta G_{int} + \Delta G_{conf}$. In Figure 2a we show illustrative examples of the effects that different contributions have on the free energy of binding.

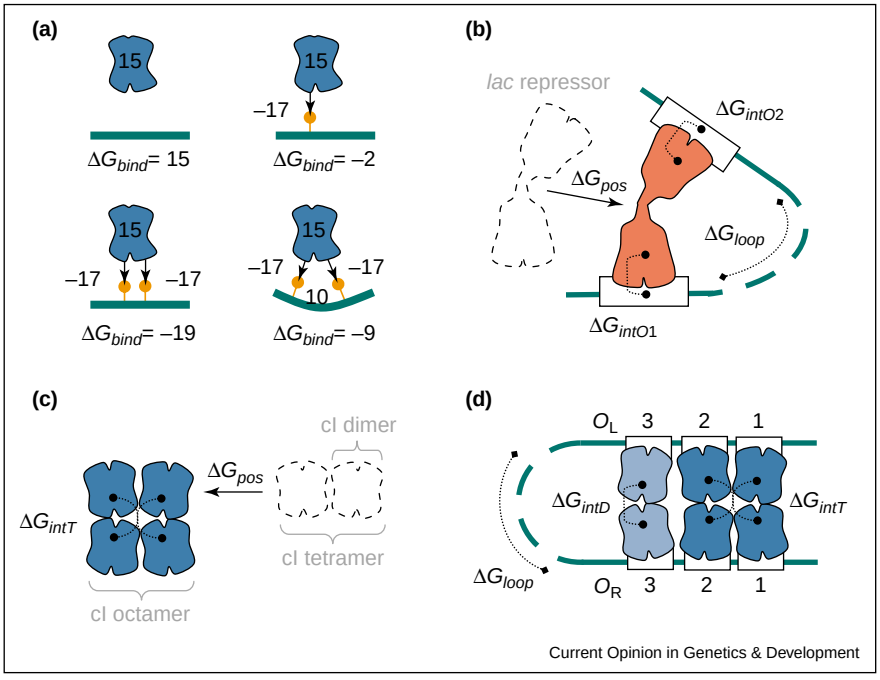
Stability of the loop in the *lac* operon

These ideas can be applied to study the formation of the DNA loop between the main operator and the auxiliary operator in the *lac* operon [21••]. The free energy of the looped state, $\Delta G_{O_1 \sim O_2} = \Delta G_{pos} + \Delta G_{intO_1} + \Delta G_{intO_2} + \Delta G_{loop}$,includes the positional free energy of one repressor, the interaction of each of the two repressor domains with O_1 and O_2, and the conformational free energy of forming the loop (Figure 2b). All these contributions must add up to the experimentally observed free energy (Table 1a), a substantially smaller quantity than the free energy of binding to just O_1, $\Delta G_{O_1} = \Delta G_{pos} + \Delta G_{intO_1}$ (Table 1a). Therefore, the extra interaction of the repressor with O_2 compensates for the conformational free energy of forming a 401 bp DNA loop. This value is estimated as $\Delta G_{loop} = \Delta G^{o}_{pos} + 8.4\,\mathrm{kcal/mol}$ from the values of Table 1a and by combining different expressions for the free energy, $\Delta G_{loop} = \Delta G_{pos} + \Delta G_{O_1 \sim O_2} - (\Delta G_{O_1} + \Delta G_{O_2})$. Considering, from this point of view, the apparent paradox of a weak site helping a strong one, one realizes that the weak site is indeed weak with respect to the binding of a free repressor but strong compared with the conformational free energy of looping the DNA. In fact, the looped complex will be more stable than the repressor bound to O_1 when $\Delta G_{O_2} < \Delta G_{pos} - \Delta G_{loop}$, which happens for standard free energies of binding to O_2 smaller than -8.4 kcal/mol.

It is interesting to compare the energetics of the looped state with the free energy of the simultaneous binding of two repressors to O_1 and O_2, $\Delta G_{O_1/O_2} = \Delta G_{pos} + \Delta G_{intO_1} + \Delta G_{intO_2} + \Delta G_{pos}$. There is just one, yet crucial, difference: the conformational free energy of closing the loop has been replaced by one positional free energy. At physiological lac repressor concentrations (1.5×10^{-8} M; about 10 repressors per cell), the positional free energy is $\Delta G_{pos} = \Delta G^{o}_{pos} + 10.8\,\mathrm{kcal/mol}$ (recall $\Delta G_{pos} = \Delta G^{o}_{pos} - RT\ln[N]$), which indicates that the looped state is more stable. Only at concentrations higher than 8.3×10^{-7} M (about 500 repressors per cell) would the simultaneous binding of two repressors dominate over the looped state.

The deconstruction procedure we have followed for the free energy has the advantage that the resulting contributions can be related to each other by combining them with the available experimental data for different experimental setups. From the measured activity of the *lac* operon *in vivo* [22] and its mathematical expression in terms of free energies [21••], we have inferred the free energy of looping for different lengths of the DNA between operators (Table 1b). For intermediate lengths, within the range 150 bp to 1.5 kb, the conformational free energy of looping nicely fits the theoretically predicted expression for an ideal flexible polymer $\Delta G_{loop}(l) = \Delta G_{loop}(l_0) + \alpha RT\ln(l/l_0)$, where l is the length of the loop, l_0 is a reference length and α is a constant. Intriguingly, theoretical estimates give $\alpha \sim 2.25$ [15,23], which is

Figure 2



Examples of different contributions to the free energy of binding. **(a)** Illustrative imaginary situations for zero, one and two interaction domains. The different contributions to the free energy of binding, ΔG_{bind}, are a positional free energy of 15 kcal/mol, an interaction free energy of −17 kcal/mol for each interaction site, and a conformational free energy of 10 kcal/mol. **(b)** Energetics of the formation of the DNA loop–lac repressor complex. Contributions to the free energy of the complex formed by the divalent lac repressor (shown in orange) when its two DNA binding domains interact simultaneously with the two operators O_1 and O_2 by looping the intervening DNA: $\Delta G_{O_1 \sim O_2} = \Delta G_{pos} + \Delta G_{intO_1} + \Delta G_{intO_2} + \Delta G_{loop}$. The positional free energy term, ΔG_{pos}, accounts for the free energy necessary to bring the lac repressor (shown in white with dashed contour lines) to the appropriate position and orientation (shown in orange) so that it can bind the operators. Two interaction free energy terms account for the interaction between the two domains of the lac repressor with the O_1 and O_2 operators: ΔG_{intO_1} and ΔG_{intO_2}, respectively. The last contribution to the free energy of the DNA loop–lac repressor complex accounts for the conformational free energy cost of looping the DNA between the two operators (ΔG_{loop}). **(c)** Energetics of the formation of cI octamers from tetramers in solution. Under physiological conditions, cI exists in solution as a dimer (shown in white with dashed contour lines) and as a monomer (not shown). Dimers can oligomerize to form tetramers, and two tetramers can form an octamer. The free energy of forming cI octamers (shown in blue as the assembly of four dimers) from the tetramers in solution (shown in white with dashed contours as the assembly of two dimers) is given by $\Delta G_{4\rightarrow 8} = \Delta G_{pos} + \Delta G_{intT}$. ΔG_{pos} is the positional free energy of bringing the tetramer from solution (shown in white with dashed contour lines) to form part of the octameric complex (shown in blue with continuous contour lines) and ΔG_{intT} is the interaction free energy between the two tetramers that form the octamer. **(d)** Energetics of the cI–DNA loop formation. ΔG_{intT} is the interaction free energy between cI tetramers bound to O_R1 and O_R2, and to O_L1 and O_L2; ΔG_{intD} is the interaction free energy between dimers bound to O_R3 and O_L3 (shown in light blue); and ΔG_{loop} is the conformational free energy of looping the intervening DNA. The free energy of forming the cI–DNA loop complex from a non-looped conformation with cI dimers bound to all the operator sites is given by $\Delta G_{R123\sim L123} - \Delta G_{R123/L123} = \Delta G_{intT} + \Delta G_{intD} + \Delta G_{loop}$.

significantly different from the inferred *in vivo* value of $\alpha \sim 1.25$ (Table 1b).

Assembly of macromolecular complexes in phage λ

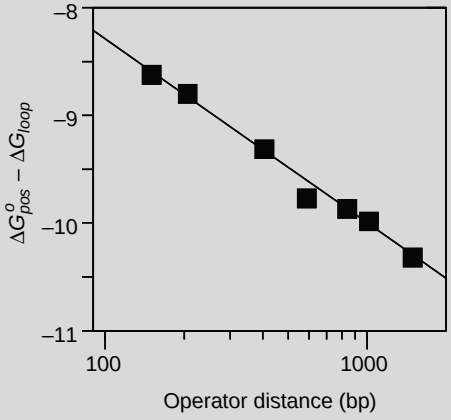
Phage λ represents a step forward in complexity. The loop is formed not by a single protein but by a protein complex that is assembled on the DNA as the loop forms. The free energy of the looped state with an assembled λ cI octamer consists of different contributions, $\Delta G_{OR12\sim OL12} = \Delta G_{intT} + \Delta G_{loop} + \Delta G_{OR12/OL12}$, which account for the interaction free energy between tetramers, the conformational free energy of forming the DNA loop, and the free energy of the tetramers bound to the right and left operators (Figure 2d). We can compare these contributions with those of the free energy of the octamer in solution, $\Delta G_{2\rightarrow 8} = \Delta G_{intT} + \Delta G_{pos} + 2\Delta G_{2\rightarrow 4}$, which account for the interaction free energy between tetramers, the positional free energy, and the free energy of the tetramers in solution, respectively (Figure 2c). The main difference between the formation of the octamer on DNA and the formation in solution results from the formation of the constituent pair of tetramers, whose formation free energies at 0.5×10^{-9} M are −4.2 and 10.6 kcal/mol, respectively (Table 1c,d). The conformational free energy of forming the DNA loop, $\Delta G_{loop} = \Delta G^{o}_{pos} + 6.8\,\mathrm{kcal/mol}$, can be obtained from the preceding equations for $\Delta G_{OR12\sim OL12}$ and $\Delta G_{2\rightarrow 8}$,

Table 1

Representative free energies of binding and looping states.

(a) lacR/state *i*	ΔG_t^o	$\Delta G_i^{1.5\times10^{-8}\mathrm{M}}$
O_1	−13.7	−2.9
O_2	−11.8	−1.0
O_1/O_2	−25.5	−4.0
$O_1{\sim}O_2$	−17.1	−6.3

(b)



$$\Delta G_{loop} = \Delta G^o_{pos} + 4.9 + 1.24RT\ln(l)$$

(c) cI/state *i*	ΔG_i^o	$\Delta G_i^{0.5\times10^{-9}\mathrm{M}}$
O_R1	−13.2	−0.4
O_R2	−10.7	2.1
O_R3	−10.2	2.6
O_R1O_R2	−26.9	−1.3
$O_R1O_R2O_R3$	−37.1	1.2
O_L1	−13.8	−1.0
O_L2	−12.1	0.7
O_L3	−12.4	0.4
O_L1O_L2	−28.4	−2.8
$O_L1O_L2O_L3$	−40.8	−2.5
*R*12∼*L*12	−55.8	−4.7
*R*12/*L*12	−55.3	−4.2
*R*123∼*L*123	−81.4	−4.7
*R*123/*L*123	−77.9	−1.2

(d) cI/process *i*	ΔG_i^o	$\Delta G_i^{0.5\times10^{-9}\mathrm{M}}$
2→4	−7.5	5.3
2→8	−22.3	16.0
4→8	−8.0	4.8

Representative free energies (in kcal/mol) of binding and looping states for the (a,b) *lac* operon and (c,d) phage λ at 1M (standard conditions) and at the physiological concentrations of 1.5×10^{-8} M (*lac* operon) and 0.5×10^{-9} M (phage λ). Whenever DNA looping can be present, the tilde (∼) and solidus (/) symbols in the labels of the states indicate looped and non-looped DNA, respectively. **(a)** Binding of the lac repressor to different operators. The free energy of binding was inferred [21••] from previous experimental data [22]. **(b)** Free energy of looping as a function of the distance between operators for the *lac* operon. The symbols correspond to the estimation obtained from the experimental data [22] for the repression level, *R*, and the mathematical expression that links it to the free energy: $R = 1 + e^{-\Delta G_{O1}/RT}([N] + e^{-(\Delta G_{loop}(l) - \Delta G^o_{pos})/RT})$ [21••]. The continuous line corresponds to the displayed equation. **(c)** Binding of λ cI dimers to different right operators (top section), left operators (middle section), and left and right operators at the same time with looped and non-looped DNA conformations (bottom section), as inferred in reference [14••] from diverse experimental sources. **(d)** Free energies of the formation of cI octamers (2→8) and tetramers (2→4) from dimers in solution; and free energy of the formation of cI octamers (4→8) from the tetramers in solution [16].

and from Table 1c,d. It is smaller than the positional free energy required to bring the tetramers together in solution, $\Delta G_{pos} = \Delta G^o_{pos} + 12.8\,\mathrm{kcal/mol}$. Therefore, DNA acts as a scaffold for the formation of tetramers and also helps the tetramers to encounter each other, which explains why an octamer that does not exist in solution is able to fasten a DNA loop in phage λ.

Interestingly, once the tetramers are formed, the interaction free energy between them is barely able to compensate for the conformational free energy cost of closing the loop, $\Delta G_{intT} + \Delta G_{loop} = -0.5$ kcal/mol. The octamer is formed on looped DNA but it is present, at the most, only 70% of the time. To further stabilize the loop, binding of λ cI dimers to O_R3 and O_L3 is required. In this case, the free energy of forming the loop from the non-looped complex is $\Delta G_{intD} + \Delta G_{intT} + \Delta G_{loop} = -3.5$ kcal/mol, where ΔG_{intD} is the interaction free energy between dimers bound to O_R3 and O_L3 (Figure 2d). Such a contribution to the free energy enables the loop to be present up to 99.7% of the time.

Effects of DNA looping in gene regulation

The thermodynamic approach we have presented provides a straightforward method to obtain the free energy of the assembly of macromolecular complexes from the different contributions of their components. Notably, there are potentially many more complexes than components. Therefore, measuring the free energy of a selected set of complexes can be sufficient to thermodynamically characterize all the components and their mutual interactions. The resulting information can, in turn, be used to obtain the free energies of all the possible complexes. Knowing the free energies of all the complexes or, equivalently, all the possible configurations in which the components can be arranged, enables statistical thermodynamics to make quantitative predictions about the probability of finding different configurations [18] and the resulting effects on gene regulation (See Bintu *et al.*, this issue).

The widespread view in the field of gene regulation is that DNA looping is just a mechanism to increase the

binding of regulatory molecules to their corresponding DNA binding sites [3]. The thermodynamic approach we have discussed shows how such an increase is achieved in representative instances. DNA looping, however, is intrinsically different from other common mechanisms that could be used to increase the affinity of a molecule for its cognate site. Below, we discuss three crucial effects that DNA looping can have on gene regulation — namely, the suppression of cell-to-cell variability, the control of transcriptional noise, and the supply of cooperativity on demand.

Cell-to-cell variability

The numbers of different molecular species are expected to differ from cell to cell, especially when they are as low as tens or hundreds. Such differences can even make genetically identical cells behave differently under the very same conditions [24,25]. Therefore, it is important to understand how the molecular and behavioral variability are connected through the different cellular processes.

Let us focus again on the *lac* operon. A useful quantity to measure the strength of a repressor is the repression level, which is defined as the maximum level of transcription divided by the actual level of transcription. Because transcription takes place when the repressor is not bound to the main operator, O_1, the actual transcription rate is the maximum rate multiplied by the probability that the main operator is free. Thus, 0%, 95% and 99.95% occupancy of O_1 by the lac repressor result in repression levels of 1, 21 and 2001, and in the production of 6000, 300 and 3 β-galactosidase molecules (the product of the *lacZ* gene) per cell each hour, respectively [10].

The parameters that characterize different mechanisms of transcription regulation can be chosen so that the repression level is the same for all mechanisms at a given number of repressors per cell. However, differences are expected to arise when the number of repressors changes. We consider explicitly the repression level for three alternative mechanisms: two operators and DNA looping, as in the *lac* operon lacking O_3; a tandem of two identical O_1 operators without DNA looping, so that both operators have to be free for transcription to occur; and a single, but stronger, operator so that the repression level is the same as for the looping case when the number of the repressors per cell is 10. Strikingly, the repression level for the looping case is a convex function of the number of repressors (Figure 3a), which leads to a low sensitivity to changes in the number of repressors. The other two cases exhibit linear and quadratic relationships between the change in the repression level and the number of repressors. The implications of these dependences extend up to the cell-population level. The low sensitivity obtained for DNA looping can be used to achieve fairly constant repression levels among cells in a population (Figure 3c), irrespective of the fluctuations of the numbers of regulatory molecules (Figure 3b). By contrast, using a single operator just propagates the fluctuations proportionally, and two operators without DNA looping can lead to an amplification of the underlying molecular variability (Figure 3b and c).

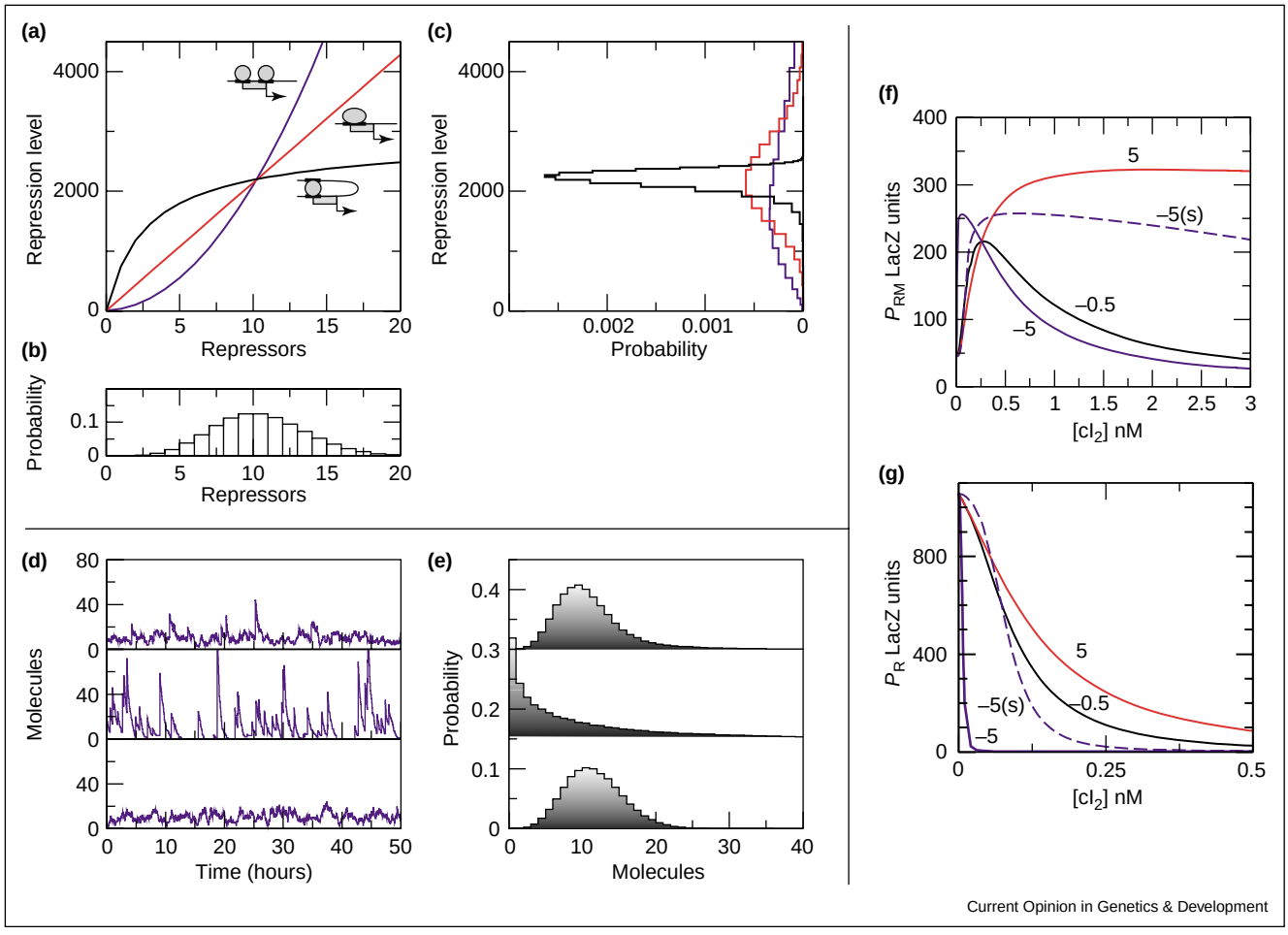
Transcriptional noise

Differences in the numbers of regulatory molecules are not the only source of cellular variability. The intrinsic stochasticity of cellular processes, usually referred to as noise, also plays an important role [25]. When the average production of proteins is as low as 10 per hour, stochastic effects can become relevant for the behavior of the system [26–31]. Figure 3d and e show the typical computed time courses and the histograms of the number of molecules produced from operons regulated with DNA looping and with just a single binding site. The parameters of the systems are such that the single operator cases have the same repression level and average transcription rate as those for DNA looping. Nevertheless, as the figure clearly illustrates, these mechanisms are not equivalent, not even when the numbers of regulatory molecules are kept constant. In particular, if a lower repressor–operator dissociation rate constant is used to increase the repression level up to that obtained with DNA looping, fluctuations are greatly enhanced. By contrast, using a higher association rate constant for the same purpose will keep the fluctuations as low as for the looping case.

The reason for these differences is a matter of time scales. If transcription switches slowly between active and inactive, there are long periods of time in which proteins are produced constantly and long periods without any production. Therefore, the number of molecules fluctuates severely between high and low values. By contrast, if the switching is very fast, the production is in the form of short and frequent bursts. This absence of long periods of time with either full or null production gives a narrower distribution of the number of molecules. DNA looping naturally introduces a fast time scale: the typical time for the repressor to be recaptured by the main operator before unbinding the auxiliary operator is much shorter than the time needed by a repressor in solution to find the main operator [21••]. DNA properties are therefore important in controlling transcriptional noise. Conspicuously, it has been observed recently that DNA can cyclize at unusually high rates [32••,33•,34•].

Looping and a higher association rate constant might seem to provide equivalent mechanisms with regards to fluctuations. There are, however, certain limits for the values that the rate constants can achieve. The theoretical limit for the association rate constant of diffusion-limited reactions is $k_a \approx 10^9\ \mathrm{M}^{-1}\mathrm{s}^{-1}$ [35]. To reduce

Figure 3



Effects of DNA looping in gene regulation. (a–c) Cell-to-cell variability **(a)** Computed repression level as a function of the number of repressors per cell, *N*, for three mechanisms: two operators and DNA looping, as in the *lac* operon without O_3 (black line) [21••]; a single operator, as in the *lac* operon without O_3 and O_2 (red line) [21••]; and two identical operators that overlap with the promoter (blue line). The different mechanisms are schematically represented next to each line. The curves corresponding to the continuous red and blue lines were computed for parameters chosen to obtain the same repression level as for the looping case for $N = 10$. This value of *N* corresponds to the wild type average number of lac repressors per cell. Assuming a Poisson distribution for fluctuations in the number of repressors, such as that shown in **(b)**, the corresponding distribution of repression levels obtained for the three cases considered (looping in black, single stronger operator in red, and two proximal operators controlling RNA polymerase binding in blue) are plotted in **(c)**. (d,e) Transcriptional noise **(d)** Time courses and **(e)** histograms of the number of molecules produced from operons regulated with (top) and without (center and bottom) looping. When the repressor is not bound to O_1, β-galactosidase molecules, encoded by the *lacZ* gene, are randomly produced at a rate of 13 per second and degraded with a characteristic life-time of 30 min. (top) Regulation with looping (rates are the same as in [21••]). (center) Regulation with a single (but stronger) operator with a dissociation rate constant 0.021 times smaller, chosen so that the repression level is the same as for the looping case. (bottom) Same situation as in the center but with the same dissociation rate constant of the top example and an association rate constant 47 times larger, which is beyond the limit of diffusion-limited reactions. (f,g) Cooperativity on demand: production of LacZ from the **(f)** P_{RM} promoter that transcribes the λ *cI* gene and from the **(g)** P_R promoter that transcribes the *cro* gene as a function of the concentration of cI dimers, $[cI_2]$, computed similarly to Dodd *et al*. [14••]. Regulation by looping (black continuous lines) is compared with regulation when the looped DNA conformation cannot be formed because of a too high free energy of looping (red continuous lines) and when the DNA is trapped in a looped conformation because of a too low free energy of looping (blue continuous lines). The free energies of looping for each case are $\Delta G_{loop} = -\Delta G_{intT} - 0.5$kcal/mol, $\Delta G_{loop} = -\Delta G_{intT} + 5$kcal/mol, and $\Delta G_{loop} = -\Delta G_{intT} - 5$kcal/mol, respectively. The latter case can be adequately scaled for comparison with the other two cases, as shown by the blue dashed lines, by plotting the production of LacZ as a function of $[cI_2]e^{0.5\Delta G_{intD}/RT}$ instead of $[cI_2]$. Each curve is labeled with its corresponding value of $\Delta G_{loop} + \Delta G_{intT}$. The symbol −5(s) labels the scaled curves.

the fluctuations by increasing the association rate constant, the diffusion limit would have to be surpassed, which does not seem to be the case for the lac repressor [36,37]. DNA looping consequently provides the cell with a mechanism to circumvent the physical constraints imposed by diffusion-limited reactions.

Cooperativity on demand

The logic of λ cI regulation is to activate its own transcription at the P_{RM} promoter and repress the *cro* gene at the P_R promoter (Figure 1b) [6••]. cI dimers bind cooperatively to O_R1 and O_R2 to form a tetramer on the DNA. Once the concentration of cI is sufficiently high, O_R3 becomes occupied and *cI* transcription from the P_{RM} promoter is turned off. For a long time, one of the main puzzles in the regulation of phage λ was that the strength of O_R3 was too weak for it to be occupied at physiological concentrations of cI [38].

It is now clear that O_R3 is occupied at physiological concentrations because of the effects of DNA looping [6••,12,13]. In Figure 3f and g we display the activity of the P_R and P_{RM} promoters as a function of cI dimer concentration, $[cI_2]$, for different values of the free energy of looping [14••]. As the free energy of looping decreases, the probability for O_R3 to be occupied by cI dimers increases (Figure 3f). Thus DNA looping provides the cooperativity needed for the occupation of O_R3.

Unexpectedly, if the free energy of looping is decreased further, beyond wild type levels, the qualitative behavior resembles that of the case in which there is no DNA looping at all. Let us consider this key point in more detail. For the phage λ switch to function properly, repression of the *cro* gene and activation of the *cI* gene should occur simultaneously. This is accomplished by the formation of cI tetramers at the O_R operator. In addition, as soon as cI concentration is sufficiently high, its production must be turned off to enable an effective RecA-mediated switch to the lytic state [6••]. If the free energy of looping is too high, DNA cannot loop and *cI* is repressed only for concentrations significantly higher than those of *cI* activation and *cro* repression. If the free energy of looping is too small, DNA readily loops and *cI* activation and *cro* repression occur at a much lower concentration than that required for *cI* repression. Therefore, both high and low conformational free energies of looping lead to the same qualitative behavior. The main difference is just a shift in concentrations at which different regulatory events happen (Figure 3f and g). Only an intermediate range of free energies of looping in the order of the *in vivo* value will give the adequate behavior. To date, all these properties of DNA looping have not been considered in quantitative approaches [39] to analyze the remarkable stability of the switch [40].

DNA looping is, thus, crucial for the high sensitivity of the activation and repression of the P_{RM} promoter. The ability of the DNA loop to open and close upon small changes of cI dimer concentration enables DNA looping to sharply switch on the cooperativity that it provides to the binding of the cI dimers to O_R3 and O_L3. Thus, it maintains a tightly regulated concentration of cI dimers in the lysogenic state. Therefore, cooperativity between O_R3 and O_L3 binding is only present when the loop is formed by the cI octamer. A similar pattern of induced cooperativity is also observed in the retinoid X receptor (RXR), a nuclear hormone receptor [41]. In its tetrameric form, RXR has two DNA binding domains and can loop DNA to bring transcription factors close to the promoter region. Retinoic acid controls whether or not the loop is formed, by its ability to prevent the assembly of the tetrameric complex from the constituent dimers, which also bind DNA.

Conclusions

Regulation systems have evolved constrained by the intrinsic molecular nature of the cell. Cells are densely packed with thousands of different molecular species and their function is built on molecular events that are inherently stochastic. The integration of such a multiplicity of components into a functional unit requires balance among a series of factors, all of which might not be attainable at the same time. If concentrations of the different molecular species are kept low to prevent non-specific interactions, not only is the binding to the specific sites decreased but also fluctuations are expected to become important. DNA looping is a mechanism that can be used to increase specificity and affinity simultaneously and, at the same time, to control the intrinsic stochasticity of cellular processes. In particular, it can buffer molecular variability to produce phenotypically homogeneous populations, decrease the transcriptional noise and enable cooperative interactions to take place on demand, as required from the cellular context.

It is clear that the effects of DNA looping in gene regulation cannot be fully understood just in terms of increased local concentrations of the regulatory molecules. DNA looping relies on key thermodynamic quantities that extend beyond the macroscopic theory of chemical reactions. Explicitly, the free energy of each binding molecule can be decomposed into a single unfavorable positional free energy and multiple, potentially favorable interaction free energies (one per interacting binding domain). Such deconstruction provides a starting point to characterize and predict the collective properties of macromolecular complexes, such as looped DNA–protein complexes, in terms of the properties of their constituent elements.

Uncovering how different molecular mechanisms determine cellular behavior is of fundamental importance for understanding both naturally occurring [17,42] and artificially designed cellular systems [43].

Acknowledgements

We are indebted to Nicolas Buchler, Calin Guet, Stanislas Leibler, Mark Ptashne, Miguel Rubi, Nikolaus Schultz, Wenying Shou and Jon Widom for comments and discussions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dunn TM, Hahn S, Ogden S, Schleif RF: **An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression**. *Proc Natl Acad Sci USA* 1984, **81**:5017-5020.

2. Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P: **The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants**. *Nucleic Acids Res* 1981, **9**:6047-6068.

3. Schleif R: **DNA looping**. *Annu Rev Biochem* 1992, **61**:199-223.

4. Adhya S: **Multipartite genetic control elements: communication by DNA loop**. *Annu Rev Genet* 1989, **23**:227-250.

5. Matthews KS: **DNA looping**. *Microbiol Rev* 1992, **56**:123-136.

6. •• Ptashne M: *A Genetic Switch: Phage Lambda Revisited*. edn 3. Cold Spring Harbor Laboratory Press; 2004.

This book is one of most illuminating expositions of gene regulation. It explains in simple molecular terms how proteins bind to DNA and turn genes on and off as the phage λ virus grows in bacteria.

7. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: **Looping and interaction between hypersensitive sites in the active β-globin locus**. *Mol Cell* 2002, **10**:1453-1465.

8. Goosen N, van de Putte P: **The regulation of transcription initiation by integration host factor**. *Mol Microbiol* 1995, **16**:1-7.

9. Wade JT, Belyaeva TA, Hyde EI, Busby SJ: **A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter**. *EMBO J* 2001, **20**:7160-7167.

10. Müller-Hill B: *The lac Operon: A Short History of a Genetic Paradigm*: Walter de Gruyter; 1996.

11. Jacob F, Monod J: **Genetic regulatory mechanisms in the synthesis of proteins**. *J Mol Biol* 1961, **3**:318-356.

12. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB: **Octamerization of lambda CI repressor is needed for effective repression of P_{RM} and efficient switching from lysogeny**. *Genes Dev* 2001, **15**:3013-3022.

13. Revet B, von Wilcken-Bergmann B, Bessert H, Barker A, Müller-Hill B: **Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances**. *Curr Biol* 1999, **9**:151-154.

14. •• Dodd IB, Shearwin KE, Perkins AJ, Burr T, Hochschild A, Egan JB: **Cooperativity in long-range gene regulation by the lambda CI repressor**. *Genes Dev* 2004, **18**:344-354.

The authors discuss the experiments and thermodynamic basis of the formation of the DNA loop between the right and the left operators of phage λ.

15. Rippe K: **Making contacts on a nucleic acid polymer**. *Trends Biochem Sci* 2001, **26**:733-740.

16. Burz DS, Ackers GK: **Cooperativity mutants of bacteriophage λ cI repressor: temperature dependence of self-assembly**. *Biochemistry* 1996, **35**:3341-3350.

17. •• Ptashne M, Gann A: *Genes & Signals*.Cold Spring Harbor Laboratory Press; 2002.

This book describes the essence of the molecular mechanisms used by cells to regulate their function.

18. Hill TL: *An Introduction to Statistical Thermodynamics*. Addison-Wesley; 1960.

19. Finkelstein AV, Janin J: **The price of lost freedom: entropy of bimolecular complex formation**. *Protein Eng* 1989, **3**:1-3.

20. Page MI, Jencks WP: **Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect**. *Proc Natl Acad Sci USA* 1971, **68**:1678-1683.

21. •• Vilar JMG, Leibler S: **DNA looping and physical constraints on transcription regulation**. *J Mol Biol* 2003, **331**:981-989.

The authors use thermodynamic models of the *lac* operon in *E. coli* to explore the effects of DNA looping in gene regulation. They demonstrate that DNA looping can provide a mechanism to regulate not only the average properties of transcription but also its fluctuations.

22. Müller J, Oehler S, Müller-Hill B: **Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator**. *J Mol Biol* 1996, **257**:21-29.

23. Hanke A, Metzler R: **Entropy loss in long-distance DNA looping**. *Biophys J* 2003, **85**:167-173.

24. Spudich JL, Koshland DE Jr: **Non-genetic individuality: chance in the single cell**. *Nature* 1976, **262**:467-471.

25. Korobkova E, Emonet T, Vilar JM, Shimizu TS, Cluzel P: **From molecular noise to behavioural variability in a single bacterium**. *Nature* 2004, **428**:574-578.

26. Rao CV, Wolf DM, Arkin AP: **Control, exploitation and tolerance of intracellular noise**. *Nature* 2002, **420**:231-237.

27. Vilar JMG, Guet CC, Leibler S: **Modeling network dynamics: the *lac* operon, a case study**. *J Cell Biol* 2003, **161**:471-476.

28. Maloney PC, Rotman B: **Distribution of suboptimally induces –D-galactosidase in *Escherichia coli*. The enzyme content of individual cells**. *J Mol Biol* 1973, **73**:77-91.

29. Paulsson J: **Summing up the noise in gene networks**. *Nature* 2004, **427**:415-418.

30. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell**. *Science* 2002, **297**:1183-1186.

31. Raser JM, O'Shea EK: **Control of stochasticity in eukaryotic gene expression**. *Science* 2004, **304**:1811-1814.

32. •• Cloutier TE, Widom J: **Spontaneous sharp bending of double-stranded DNA**. *Mol Cell* 2004, **14**:355-362.

The authors demonstrate that DNA molecules that are 94 bp in length — comparable to the length of sharply looped DNAs *in vivo* — spontaneously bend into circles. Random DNA sequences cyclize 10^2–10^4 times more easily than predicted from current theories of DNA bending, and DNA sequences that position nucleosomes cyclize up to 10^5 times more easily.

33. • Yan J, Marko JF: **Localized single-stranded bubble mechanism for cyclization of short double helix DNA**. *Phys Rev Lett* 2004, **93**:108108.

See annotation [34•].

34. • Wiggins PA, Phillips R, Nelson PC: **Exact theory of kinkable elastic polymers**.*Phys Rev E* 2005, in press.

This article and that by Yan and Marko [33•] provide a theoretical explanation of the experiments of Cloutier and Widom [32••].

35. Berg OG, von Hippel PH: **Diffusion-controlled macromolecular interactions**. *Annu Rev Biophys Biophys Chem* 1985, **14**:131-160.

36. Fickert R, Müller-Hill B: **How Lac repressor finds *lac* operator *in vitro***. *J Mol Biol* 1992, **226**:59-68.

37. Hsieh WT, Whitson PA, Matthews KS, Wells RD: **Influence of sequence and distance between two operators on interaction with the lac repressor**. *J Biol Chem* 1987, **262**:14583-14591.

38. Reinitz J, Vaisnys JR: **Theoretical and experimental analysis of the phage λ genetic switch implies missing levels of co-operativity**. *J Theor Biol* 1990, **145**:295-318.

39. Aurell E, Brown S, Johanson J, Sneppen K: **Stability puzzles in phage lambda**. *Phys Rev E* 2002, **65**:051914.

40. Little JW, Shepley DP, Wert DW: **Robustness of a gene regulatory circuit**. *EMBO J* 1999, **18**:4299-4307.

41. Yasmin R, Yeung KT, Chung RH, Gaczynska ME, Osmulski PA, Noy N: **DNA-looping by RXR tetramers permits transcriptional regulation "at a distance"**. *J Mol Biol* 2004, **343**:327-338.

42. Jacob F: *The Possible and the Actual*.Pantheon; 1982.

43. Guet CC, Elowitz MB, Hsing W, Leibler S: **Combinatorial synthesis of genetic networks**. *Science* 2002, **296**:1466-1470.