Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 231 (2004) 525–533
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Nonspecific binding of the O_R repressors CI and Cro of bacteriophage λ

Audun Bakk*, Ralf Metzler

NORDITA, Nordic Institute for Theoretical Physics, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Received 26 March 2004; received in revised form 8 July 2004; accepted 12 July 2004
Available online 27 August 2004

Abstract

We estimate the Gibbs free energy for nonspecific binding (ΔG_{NSB}) to the *Escherichia coli* DNA for two regulatory proteins of the λ phage, CI and Cro. By means of a statistical–mechanical approach, we calculate the *c*I and *cro* activities associated with the operator O_R of an introduced λ phage genome (prophage). In this statistical model we apply in vitro-measured binding free energies to fit in vivo experimental data for *c*I and *cro* activities, respectively, where ΔG_{NSB} is introduced as a free (fitting) parameter. Without nonspecific binding included in the model, the quality of the description is fairly poor, whereas data are nicely correlating with our model with nonspecific binding included over the entire data range. The obtained values of ΔG_{NSB} are -4.1 ± 0.9 kcal/mol, for CI, and -4.2 ± 0.8 kcal/mol, for Cro. In particular, in a lysogen ($\approx$250 CI monomers per cell) we conclude that 86% of the total CI in the cell is nonspecifically bound, leaving on average around 10 CI dimers freely available in the *E. coli* cytoplasma. These findings corroborate the view that due to low free cellular particle numbers a dynamical analysis of genetic regulation at O_R and comparable systems should include a stochastic component. In addition, we perform a stability analysis of the O_R system in the presence of nonspecific binding.

Keywords: Genetic regulation; Lambda phage; Nonspecific binding; Statistical mechanics

1. Introduction

About 50 years ago, Lwoff observed that a colony of *Escherichia coli* cells infected by bacteriophage lambda (λ phage) lysed when irradiated by UV-light (Lwoff, 1953). Since then, numerous details of the λ phage and its pathways have been revealed. Today, we know the genome and the protein products of the bacteriophage. Furthermore, an understanding of the genetic network seems to be attained (Ptashne, 1992).

The fate of a λ phage infected bacterium bifurcates in two directions. Either the λ genome is introduced into the host genome (prophage), whereupon it silently becomes replicated for generations of *E. coli* life cycles, called the lysogenic pathway, or the phage becomes massively reproduced, on the time-scale of a cell generation, until the bacterium bursts (lyses) such that 50–100 new phages are released, called the lytic pathway (see Ptashne, l992 for details).

The right operator (O_R) of the λ phage genome and the binding of the regulatory proteins CI and Cro to O_R are essential to understand the life cycle of the infected *E. coli* bacterium. In particular, O_R is believed to be central for the stability of the switch and for the turnover to lytic growth (Brooks and Clark, 1967). Two regulatory proteins, CI and Cro, are able to bind as dimers to three different binding sites of O_R (Fig. 1). Depending on the specific binding pattern of these proteins, RNA polymerase (RNAP) may bind and transcribe in one of two directions from the O_R-associated promoters P_{RM} and P_R such that either *c*I or *cro* is transcribed (divergent transcription),

*Corresponding author. Fax: +45-35-38-9157.
E-mail addresses: audunba@nordita.dk (A. Bakk), metz@nordita.dk (R. Metzler).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.07.007

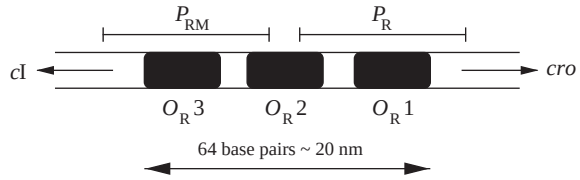

Fig. 1. Schematic illustration of the O_R of the λ phage genome. The operator has three binding sites, O_R1, O_R2, and O_R3, where CI and Cro dimers are able to bind. P_{RM} and P_R indicate the promoter regions where RNA polymerase binds to initiate transcription of *c*I and *cro* genes. The arrows associated with *c*I and *cro* indicate the transcription direction of these genes, respectively.

respectively. In a lysogen, O_R1 and O_R2 are usually occupied by one CI dimer each, exhibiting a cooperative interaction, and P_{RM} is occupied by RNAP such that CI is continuously expressed, thereby maintaining repression of *cro*. This is an example of an auto-regulating system, i.e. CI regulates its own synthesis. Auto-regulation is also the case when Cro is abundant, e.g. after an induction event, that will block for RNAP binding at P_{RM} such that only *cro* transcription is possible.

In order to quantitatively understand genetic networks one needs detailed information about the system parameters (rate constants, free energies of binding, etc.), in addition to detailed structural information of its compartments (Bukau and Horwich, 1998). The past 25 years have provided a growing amount of quantitative data of the kinetics (Hawley and McClure, 1982; Fong et al., 1994) and energetics (Takeda et al., 1992; Burz and Ackers, 1996; Darling et al., 2000b) associated with O_R of the λ phage, i.e. there are sufficient and reliable data to quantitatively model the O_R regulatory network.

Whereas in vitro affinity experiments are based on short DNA sequences containing O_R, in vivo experiments usually include the entire *E. coli* genome (4.6×10^6 base pairs). The system in vivo therefore features an additional effect, namely that CI and Cro may also bind outside O_R, called nonspecific binding (NSB) (von Hippel et al., 1974). Furthermore, in vivo experiments typically provide an estimate of the total number of proteins per cell (Reinitz and Vaisnys, 1990; Dodd et al., 2001). However, in a thermodynamical description it is the free concentration of CI, Cro, and RNAP that is assumed to be important for specific binding to O_R (Ackers et al., 1982) (see also Eq. (1)). Thus, in order to provide reliable quantitative information about O_R in vivo it is essential to obtain a more quantitative understanding of NSB.

In this work we study the impact of NSB of CI and Cro repressors on the two activities associated with P_{RM} and P_R measured in vivo. The activities are calculated by means of a modified version of the statistical-mechanical approach proposed by Ackers et al. (1982), such that in addition to the specific binding to O_R we also quantify the NSB of CI and Cro. Based upon experimentally determined specific protein-DNA binding affinities of O_R, we obtain the free energy upon NSB (ΔG_{NSB}) as a free parameter. By fitting our model to experimental data of CI activity from Dodd et al. (2001) and Cro activity from Pakula et al. (1986)[1] we find ΔG_{NSB} of -4.1 ± 0.9 and -4.2 ± 0.8 kcal/mol, respectively (Bakk and Metzler, 2004). The significant NSB strength implies that a large fraction of the proteins in vivo are nonspecifically bound. Given this quantification of NSB, we perform a perturbation analysis of the activities due to the experimental error of the parameters involved in the model (Bakk et al., 2004a, b). We also put the protein–DNA binding strength in context with the *c12* and *r1* mutants of Dodd et al. (2001).

The novelty of this work is the calculation of the NSB strength of CI and Cro based upon the most recent and accurate values of the parameters involved. Furthermore, to our knowledge this is the first attempt to systematically estimate the error of NSB for CI and Cro. The previous error-analyses of the O_R-system (Bakk et al., 2004a, b) were performed without NSB, in contrast to the present work. Compared to the work of Bakk and Metzler (2004), we now perform a thorough and more physical derivation of the model, in addition to the comprehensive error analysis of the NSB and the promoters activities. The fitting procedure is described in detail, in addition to a more extensive discussion of the biological implications of the results (noise, small free particle numbers, etc.) The discussion of the two *c*I mutants in the context of NSB is also new. Moreover, we present interesting results for the detailed functional dependence of NSB on the total protein concentration.

2. Modeling the system

To study the binding of proteins to the DNA, both specifically to O_R and nonspecifically, we assume that the protein associations are in thermodynamical equilibrium (Ackers et al., 1982). For the specific protein binding to O_R of λ phage we apply the statistical–mechanical (equilibrium) approach of Ackers et al. (1982). It is possible to show that the binding of CI dimers (CI_2), Cro dimers (Cro_2), and RNAP to O_R of phage λ occur in 40 different combinations s (Shea and Ackers, 1985). The associated probability f_s for finding the system in one of these 40 states s is

$$f_s = \frac{\exp(-\Delta G(s)/(RT))[CI_2]^{i_s}[Cro_2]^{j_s}[RNAP]^{k_s}}{\sum_s \exp(-\Delta G(s)/(RT))[CI_2]^{i_s}[Cro_2]^{j_s}[RNAP]^{k_s}}, \quad (1)$$

[1]The activity data in absolute amounts are presented in Reinitz and Vaisnys (1990) with reference to Pakula et al. (1986) and personal communication between Reinitz/Vaisnys and Pakula.

Table 1
Parameters with error estimates applied in the model (Eqs. (1)–(3))

	Parameter	Wild type[a]	Error[b]
CI[c]	ΔG_1	−12.5	0.3
	ΔG_2	−10.5	0.2
	ΔG_3	−9.5	0.2
	ΔG_{12}	−2.7	0.3
	ΔG_{23}	−2.9	0.5
Cro[d]	$\Delta G_{1'}$	−12.0	0.1
	$\Delta G_{2'}$	−10.8	0.1
	$\Delta G_{3'}$	−13.4	0.1
	$\Delta G_{12'}$	−1.0	0.2
	$\Delta G_{23'}$	−0.6	0.2
	$\Delta G_{123'}$	−0.9	0.2
RNAP[e]	ΔG_{RM}	−11.5	0.5
	ΔG_R	−12.5	0.5
	P[f]	30	6
	N_{DNA}	2.3/2.7[g]	10%

[a]Experimental (wild-type) value. CI, Cro, and RNAP data in units of kcal/mol, P in units of nM, and N_{DNA} is dimensionless.

[b]Experimental error that corresponds to 67% confidence intervals in the same unit as the respective parameter.

[c]Protein–DNA GFEs for CI dimers from Koblan and Ackers (1992) measured at 37 °C . ΔG_1 is the GFE associated with the binding between CI and operator site O_R1, etc. ΔG_{12} is the GFE associated with cooperative binding between CI and O_R1 and O_R2, etc. Experimental data are obtained in vitro in 200 mM KCl, resembling "physiological" conditions (Kao-Huang et al., 1977; Ackers et al., 1982) (also for Cro and RNAP). Free energy of dimerization is −11.1 kcal/mol (Koblan and Ackers, 1991).

[d]Protein–DNA binding energies for Cro dimers from Darling et al. (2000b) measured at 20 °C. Same notation as for the CI with a prime (′) in the subscript to indicate Cro data. Free energy of dimerization is −8.7 kcal/mol (Darling et al., 2000a).

[e]RNAP–DNA GFEs from Shea and Ackers (1985).

[f]Free RNAP concentration is usually assumed to be constant 30 nM (Ackers et al., 1982; Shea and Ackers, 1985; Aurell et al., 2002), however, we here investigate the effect of perturbing this parameter ±20%.

[g]The average number of DNA-chromosomes in an *E. coli* bacterium is 2.3 in the experiment of Dodd et al. (2001) (doubling time of 40 min) and 2.7 in the experiment of Pakula et al. (1986) (doubling time 33 min). The experimental error is assumed to be within 10% (Bremer and Dennis, 1996).

where $R = 8.31$ J/(mol K) is the gas constant, $T = 310$ K is the absolute temperature corresponding to physiological temperature 37 °C, and $\Delta G(s)$ is the Gibbs free energy difference (binding energy) between state s and the unoccupied state ($s = 1$). Both individual binding affinities and cooperative interactions contribute to the different $\Delta G(s)$ (see Table 1). $[CI_2]$, $[Cro_2]$, and [RNAP] are the free (unbound) concentrations of CI dimers, Cro dimers, and RNAP, respectively. $i_s \in \{0,1,2,3\}$, $j_s \in \{0,1,2,3\}$, and $k_s \in \{0,1,2\}$ are the number of CI dimers, Cro dimers, and RNAP bound to O_R in the state s.

Free monomers and dimers of CI and Cro are supposed to be in equilibrium with standard free energies of association[2] constants of −11.1 kcal/mol, for CI (Koblan and Ackers, 1991), and −8.7 kcal/mol, for Cro (Darling et al., 2000a). The former is measured at 37 °C, which is the temperature in the experiments we want to compare with, while the latter value is measured at 20 °C. We apply a Boltzmann activation factor to recalculate the equilibrium constant for 37 °C.

The total concentration of CI molecules in the cell, in monomeric equivalents, yields

$$[CI_t] = [CI_1] + 2[CI_2] + 2N_{DNA}[V_{CI}^{-1}] \times \left(\sum_s i_s f_s + N_{bp}[CI_2] \exp(-\Delta G_{NSB}^{CI}/(RT)) \right), \quad (2)$$

where the first and second term on the right-hand side count the free monomeric and dimeric concentrations, respectively, and the last term accounts for both the average number of specifically bound CI dimers at O_R (first part of the last term) and the average number of nonspecifically bound CI dimers (second part of the last term). The derivation of the NSB term in Eq. (2) is performed in Appendix A where Eq. (A.3) is converted to cellular concentration by dividing by the cellular volume to take the form expressed in Eq. (2). The culture in the experiment of Dodd et al. (2001) has a doubling time of 40 min (I.B. Dodd, pers. comm.) that corresponds to a cellular volume $V_{CI} = 1.13\,\mu m^3$ (Donachie and Robinson, 1987). $[V_{CI}^{-1}] = 1.47$ nM is then the molar concentration of one particle in the cells of Dodd et al. (2001). The average number of DNA molecules (N_{DNA}) present in the *E. coli* cell is 2.3 (Bremer and Dennis, 1996) and the number of base pairs of the *E. coli* genome (N_{bp}) is 4.6×10^6 (Blattner et al., 1997).

Similarly, Eq. (2) yields for Cro

$$[Cro_t] = [Cro_1] + 2[Cro_2] + 2N_{DNA}[V_{Cro}^{-1}] \times \left(\sum_s j_s f_s + N_{bp}[Cro_2] \exp(-\Delta G_{NSB}^{Cro}/(RT)) \right). \quad (3)$$

The doubling time of the culture of Pakula et al. (1986) is 33 min corresponding to a cellular volume $V_{Cro} = 1.41\,\mu m$ ($[V_{Cro}^{-1}] = 1.18$ nM) and N_{DNA} equals 2.7 (Bremer and Dennis, 1996).

3. Results and discussion

3.1. CI

To estimate the free energy change upon NSB for CI dimers (ΔG_{NSB}^{CI}) we compare our model to experimental

[2]For example, the standard free energy of CI association (ΔG_a) for the monomer–dimer equilibrium $2CI_1 \longleftrightarrow CI_2$ is defined as $\exp(-\Delta G_a/(RT)) = [CI_2]/[CI_1]^2$.

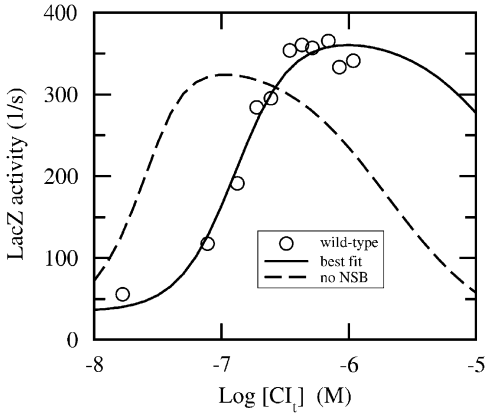

Fig. 2. Activity of P_{RM} in LacZ units (Dodd et al., 2001) versus total CI concentration (logarithmic scale). The Cro concentration is zero, "wild-type": experimental data from Dodd et al. (2001); "best fit": best-fit curve from theoretical expressions (Eqs. (1) and (2)) yielding $\Delta G^{CI}_{NSB} = -4.1$ kcal/mol; (no NSB) theoretical prediction without nonspecific binding ($\Delta G^{CI}_{NSB} \to \infty$).

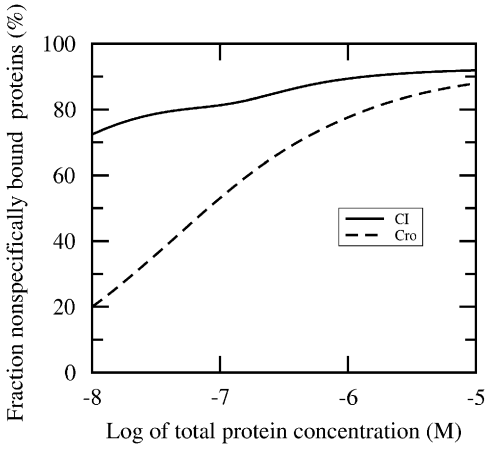

Fig. 3. Fraction of nonspecifically bound CI and Cro proteins versus total protein concentration (in monomeric units), respectively. The curves are based on $\Delta G^{CI}_{NSB} = -4.1$ kcal/mol and $\Delta G^{Cro}_{NSB} = -4.2$ kcal/mol, respectively.

P_{RM} activity data from Dodd et al. (2001) (see Fig. 2). These data are obtained in absence of the left operator (O_L) that excludes the possibility of long-range loop formation between O_R and O_L (Révet et al., 1999) and thereby violating Eq. (1). The best-fit procedure is performed as described in Appendix B and we obtain $\Delta G^{CI}_{NSB} = -4.1$ kcal/mol.[3] This value is comparable with the in vitro result of Senear and Batey (1991) who found a ΔG^{CI}_{NSB} of -3.7 kcal/mol for a 1.7 kbp fragment at 200 nM as KCl and 20 °C. Despite the relatively short DNA fragments, the latter optimized fit was not improved by including cooperativity for neighboring dimers, justifying the neglect of cooperative NSB as we have done in the present work for the entire genome in vivo. Koblan and Ackers (1992) obtained for similar conditions as Senear and Batey (1991) an approximate ΔG^{CI}_{NSB} around -3.5 kcal/mol. We also note that Aurell et al., 2002 estimate a minimal $\Delta G^{CI}_{NSB}$of -2.0 kcal/mol by comparing their model (similar to the model in this work although some parameter values are different) to data of Johnson et al. (1981).

In Fig. 2 we show the activity data obtained in vivo. The dashed line corresponds to the original model in which NSB for CI is ignored ($\Delta G^{CI}_{NSB} \to \infty$), clearly showing that without NSB the data cannot be reasonably reproduced as then P_{RM} model data are significantly shifted to the left in the activity–concentration plot. In contrast, with NSB in the model the data are nicely described. The fraction of bound CI versus the total number of CI is displayed in Fig. 3. For a lysogen ($\approx$ 250 CI monomer equivalents per cell) we find that 86% of the total number of CI are nonspecifically bound. Effectively, this observation leads to a significant dilution of the free cellular volume. We note that for *E. coli lac* repressors in vivo Kao-Huang et al. (1977) estimated that less than 10% of the repressors are free in solution.

Let us estimate the error bars for the ΔG^{CI}_{NSB} due to the experimental error underlying the parameters entering the fit. In Table 2 we present the results from systematic perturbations of the different parameters involved in the model, where the parameters, one by

Table 2
Best fit of ΔG_{NSB} in kcal/mol for CI and Cro due to systematical perturbations of the parameters corresponding to the standard deviation of the experimental error as listed in Table 1

CI			Cro		
Parameter	+Error[a]	−Error[b]	Parameter	+Error	−Error
ΔG_1	−4.0	−4.3	$\Delta G_{1'}$	∼	∼
ΔG_2	−4.0	−4.2	$\Delta G_{2'}$	∼	∼
ΔG_3	∼[c]	∼	$\Delta G_{3'}$	∼	∼
ΔG_{12}	−4.0	−4.3	$\Delta G_{12'}$	∼	∼
ΔG_{23}	∼	∼	$\Delta G_{23'}$	∼	∼
			$\Delta G_{123'}$	∼	∼
ΔG_{RM}	∼	∼	ΔG_{RM}	∼	∼
ΔG_R	−4.4	−3.9	ΔG_R	−4.5	−3.8
P	∼	∼	P	∼	∼
N_{DNA}	∼	∼	N_{DNA}	∼	∼

[a]Best fit of ΔG^{CI}_{NSB} when the corresponding parameter is *added* the experimental error listed in Table 1.

[b]Best fit of ΔG^{CI}_{NSB} when the corresponding parameter is *subtracted* the experimental error listed in Table 1.

[c]A change within ± 0.1 kcal/mol of ΔG_{NSB} compared to the unperturbed value of ΔG_{NSB} (i.e., no error included in the involved parameters) is termed "∼" in this table. The best fit of ΔG_{NSB} values of the unperturbed parameter set is -4.1 kcal/mol for CI and -4.2 kcal/mol for Cro.

[3]The square root of the least quadratic error for the activity (between theory and experiment) divided by the number of experimental data points is 5.0 s^{-1} (units as in Fig. 2).

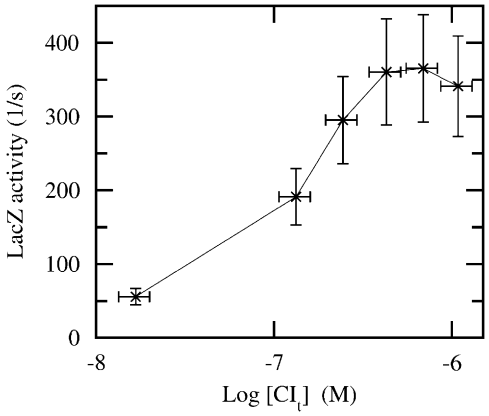

Fig. 4. Experimental LacZ activity of P_{RM} (also shown in Fig. 2). Vertical error bars indicate 20% relative uncertainty in the activity while horizontal error bars indicate 20% relative uncertainty in the concentration. Only every second data point of Dodd et al. (2001) is shown to make the figure more clear.

one, are added to and subtracted from the experimental error, respectively, as listed in Table 1, for example when we add the experimental error 0.3 kcal/mol to the wild-type GFE of ΔG_1, the optimized (best fit) ΔG^{CI}_{NSB} changes to -4.0 kcal/mol as stated in Table 2. ΔG^{CI}_{NSB} is most sensitive to the experimental error of ΔG_R. The experimental error in the GFE for the CI monomer–dimer equilibrium is ± 0.3 kcal/mol (Koblan and Ackers, 1991) and leads to negligible change in the best fit estimate of ΔG^{CI}_{NSB}.

In order to estimate the total error of ΔG^{CI}_{NSB} we search for the largest deviation of this parameter compared to the mean (unperturbed) value (-4.1 kcal/mol) due to *simultaneous* perturbations of the parameters in Table 1 within the experimental error. This results in an error estimate of ± 0.9 kcal/mol for ΔG^{CI}_{NSB}. The experimental error of the activity of Dodd et al. (2001) is unknown to the authors' knowledge. However, assuming a relative uncertainty of the activity of 20% (vertical error bars in Fig. 4), with the concentration unchanged, does not change the ΔG^{CI}_{NSB}. Conversely, allowing a relative uncertainty of 20% in the protein levels (horizontal error bars in Fig. 4), leads to a change in $\Delta G^{CI}_{NSB} < 0.2$ kcal/mol.

It is reasonable to assume that O_R mutations will not affect ΔG^{CI}_{NSB}. By applying the ΔG^{CI}_{NSB} value obtained above, we tried to fit activity data from Dodd et al. (2001) corresponding to the two different single base substitutions of O_R3 termed *r1* and *c12*. These mutations are supposed to change the ΔG_3 free energy (i.e. $\Delta\Delta G_3$) by $+2.9$ and -0.8 kcal/mol, respectively (Sarai and Takeda, 1989), but one should note that they are measured in vitro and at different temperature (0 °C) compared to the in vivo experiment by Dodd et al. (2001). Furthermore, since P_{RM} is overlapping O_R3, these mutations are likely to perturb ΔG_{RM} as well. By

Table 3
Relative change in activity at lysogenic concentration ([CI$_t$] ≈ 370 nM and [Cro$_t$] ≈ 0) compared to wild-type activity at promoters P_{RM} and P_R due to perturbations of ±1 kcal/mol of the different affinities of CI

	+1 kcal/mol		−1 kcal/mol	
	P_{RM}	P_R	P_{RM}	P_R
ΔG_1	−0.2[a]	2.7	0.1	−0.8
ΔG_2	−0.3	2.4	0.1	−0.8
ΔG_3	0[b]	0	−0.1	0
ΔG_{12}	−0.3	2.3	0.1	−0.8
ΔG_{23}	0	0	0	0
ΔG_{RM}	−0.5	0	0.2	0
ΔG_R	0.1	−0.8	−0.2	2.8

[a] −0.2 corresponds to a 20% reduction of the activity.
[b] 0 corresponds to a relative change in activity less than ±5%.

applying the $\Delta\Delta G_3$ results from Sarai and Takeda (1989) we are able to fit the activity of the mutants of Dodd et al. (2001) by $\Delta\Delta G_{RM}$ values of -0.3 kcal/mol, for *r1*, and $+0.5$ kcal/mol, for *r12*. This may indicate that single base substitutions of O_R3 influence the GFE for RNAP association to P_{RM}.

To obtain a more systematic picture of the sensitivity of the activity upon mutations in the presence of NSB, we perform a similar analysis as pursued by Bakk et al. (2004a) (and further expanded in Bakk et al., 2004b), however, NSB is included this time. In light of the exceptional stability of the lysogenic state (Brooks and Clark, 1967; Aurell et al., 2002) we study the system in this concentration regime (≈370 nM). Due to the (assumed) zero Cro concentration in a lysogen the relevant affinities are the ones related to CI and RNAP. These are, one by one, perturbed ±1 kcal/mol, in order to mimic mutations, whereupon the relative shift of the activity (sensitivity) is calculated from our thermodynamic model (Table 3). The sensitivity of a lysogen is larger for P_{RM} in the present work compared to Bakk et al. (2004a), a natural consequence of the lowered CI concentration. Although we applied a different CI affinity set in the latter case, the most important difference between the two works is that the free concentration of CI dimers was 79 nM in Bakk et al. (2004a), while in the present work due to the NSB, this concentration is only 12 nM. A striking feature of a lysogen, except for the parameter ΔG_{RM}, is that the sensitivity of P_{RM} upon perturbations of the affinities is much smaller compared to the corresponding sensitivity of P_R.

3.2. Cro

In order to estimate the NSB of Cro, we apply P_R activity data from Pakula et al. (1986) where the cellular CI concentration is zero (see Fig. 5). The analysis for Cro is performed in a similar manner as we did for CI.

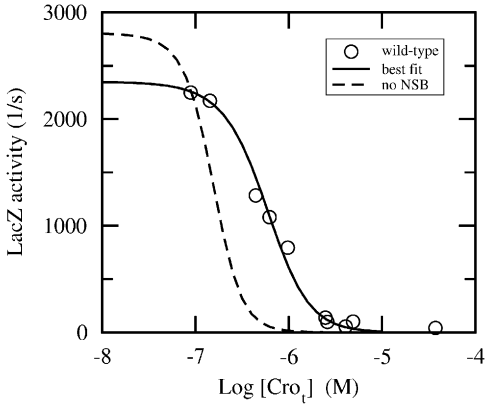

Fig. 5. Activity of P_R in LacZ units (Pakula et al., 1986) versus total Cro concentration (logarithmic scale). The CI concentration is zero. "wild-type": experimental data from Pakula et al. (1986). Due to the logarithmic concentration scale the zero cro data point is not shown (LacZ activity 2411 s^{-1}); "best fit": best-fit curve from theoretical expressions (Eqs. (1) and (3)) yielding $\Delta G^{Cro}_{NSB} = -4.2$ kcal/mol; "no NSB": theoretical prediction without nonspecific binding ($\Delta G^{Cro}_{NSB} \to \infty$).

The best-fit value of ΔG^{Cro}_{NSB} is -4.2 kcal/mol.[4] The experimental data and the theoretical prediction are shown in Fig. 5. As for CI (Fig. 2) we see that the fit without the NSB term is unsatisfactory. This further emphasizes the need for a term that incorporates the NSB for models describing similar systems, as we have performed in this work.

From the results of the error analysis presented in Table 2 it is only the error in ΔG_R that is able to perturb the ΔG^{Cro}_{NSB} value more than 0.1 kcal/mol. However, it should be noted that the experimental errors associated with Cro-DNA binding are small ($\leqslant 0.2$ kcal/mol) compared to that of CI. ΔG_{NSB} is sensitive within the experimental error of ΔG_R for both CI and Cro. The error in the GFE for the Cro monomer–dimer equilibrium is ± 0.2 kcal/mol (Darling et al., 2000a) and leads to negligible change in ΔG^{CI}_{NSB} when included.

The total error of ΔG^{Cro}_{NSB}, resulting from the combination of the experimental errors in the different parameters as listed in Table 1 is ± 0.8 kcal/mol. Thus, best fit values of ΔG_{NSB} and the error estimates for CI and Cro dimers are similar. When we systematically include 20% error in the wild-type activities of Cro or 20% error in the Cro concentrations, as we did for CI (corresponds to vertical and horizontal error bars as shown for CI data in Fig. 4), and fit data to the limits of the error bars we find that ΔG^{Cro}_{NSB} changes $< \pm 0.2$ kcal/mol. This analysis is performed to estimate the impact of experimental errors on ΔG^{Cro}_{NSB}, because the error estimates of the experimental activities are unknown to the authors. However, note in the case where the experimental error turns out to alter the shape of the activity curve the effect may be more significant. We note that Reinitz and Vaisnys (1990) were not able to improve their fit to the P_R data of Pakula et al. (1986) by including NSB. However, one should bear in mind that these authors did not convert the Cro concentrations reported by Pakula et al. (1986) to absolute amounts, as we do here. Furthermore, the nonspecific binding itself does not alter the shape of the activity curve much, and consequently, Reinitz and Vaisnys (1990) were not able to detect any difference between the fits with and without NSB included in their model. Furthermore, Aurell et al. (2002) applied a similar model compared to the model we apply in this work for Cro, although with different parameter values, and they concluded by comparing their model to the concentration corresponding to half repression in the data of Pakula et al. (1986) that NSB is of order -3.0 kcal/mol.

As for CI, a significant amount of Cro is nonspecifically bound, however, this fraction for Cro is lower than for CI (Fig. 3). The main reason for that is the weak dimerization constant of Cro, relative the one for CI. At total Cro concentrations $> 10^{-7}$ nM more than 50% of Cro is bound at nonspecific sites.

In order to further pursue the error analysis we performed for CI (Table 3) we apply the method performed by Bakk et al. (2004b), which they applied to the same system without NSB. Shortly explained, the DNA binding free energies of CI, Cro, and RNAP are perturbed within the experimental error to check the uncertainty of the activity. The perturbations are performed such that the ΔGs are simultaneously chosen at random (Gaussian distribution) around the literature (mean) values, with a width corresponding to the experimental error. Within this procedure we also assume that the Cro production is in equilibrium, thus, for a given CI concentration the Cro concentration is implicitly given (see Fig. 6a). In the thermodynamic description as explained above, with the randomly drawn protein-DNA binding energies as input and the Cro concentration determined self-consistently, we calculate the corresponding activities of the promoters P_{RM} and P_R. By repeating this procedure we obtain a mean value ($\times$ in Fig. 6b) of the activity that roughly corresponds to unperturbed activity (fully drawn lines in Fig. 6b). The standard deviation emerging from this scheme, a measure of the sensitivity due to experimental error, is significant (typically $> 20\%$ relative to wild-type activity). However, we note that the in vivo experiments of Bailone et al. (1979) showed that the P_{RM} activity may be significantly reduced, and still the lysogenic state is maintained. Thus, within the error estimates showed in Fig. 6b the promoter activities are probably sufficiently separated to make the λ-switch feasible.

[4]The square root of the least quadratic error for the activity (between theory and experiment) divided by the number of experimental data points is 19.8 s^{-1} (units as in Fig. 5).

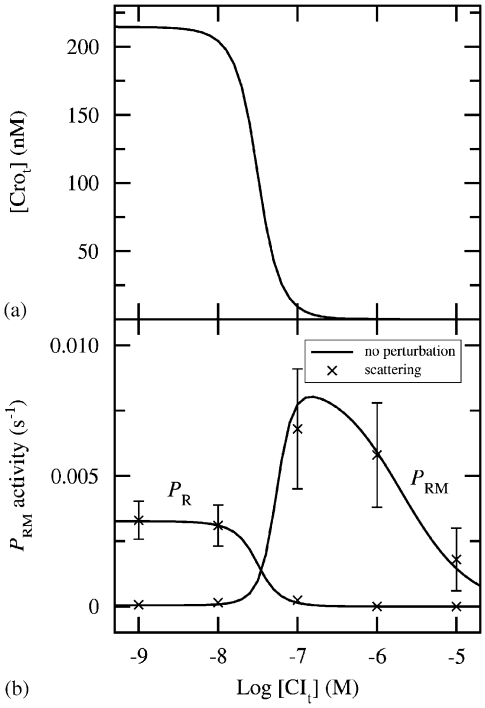

Fig. 6. (a) Total Cro concentration versus total repressor concentration (logarithmic scale) where the Cro concentration is determined self-consistently as described by Bakk et al. (2004b). (b) Promotor activity versus total CI concentration where Cro concentration is determined self-consistently. The scaling of the activities are similar to Shea and Ackers (1985). "no perturbation": P_{RM} activity based on the literature (mean) values of the parameters; "scattering" P_{RM} activity calculated from the perturbation scheme (within experimental error) described by Bakk et al. (2004b), yielding a mean value (×) and corresponding error bars (error bars corresponding to absolute errors $>3.4\times10^{-4}\,s^{-1}$ are omitted). All plots are based upon the ΔG_{NSB} values −4.1 kcal/mol for CI and −4.2 kcal/mol for Cro.

4. Summary and conclusion

We extended the thermodynamic model of Ackers et al. (1982) that describes specific protein-DNA association of the regulatory proteins CI and Cro to O_R, by introducing nonspecific binding (NSB), i.e. we incorporate the effect of protein-DNA binding outside O_R. By fitting our model, where the Gibbs free energy upon NSB (ΔG_{NSB}) is a free parameter, to experimental in vivo data we obtain the ΔG_{NSB} values -4.1 ± 0.9 kcal/mol for CI and -4.2 ± 0.8 kcal/mol for Cro. These ΔG_{NSB} values seem reasonable keeping in mind that in vitro experiments give NSB estimates for CI around −3.6 kcal/mol (Senear and Batey, 1991; Koblan and Ackers, 1992). We note that our model including NSB shows excellent agreement with both experimental data sets for CI and Cro, in contrast to the model without NSB where we are not able to reasonably reproduce the in vivo data.

By applying the ΔG_{NSB} obtained from the optimized fit to the in vivo activity data, we estimate at a typical lysogenic concentration (≈ 250 CI molecules per cell) that 86% of the total number of CI repressors are nonspecifically bound, i.e. there remain around 10 free dimers per cell. A significant NSB was also seen in vivo experiment by Kao-Huang et al. (1977) with *E. coli lac* repressors, who concluded that less than 10% of the *E. coli lac* repressors in vivo are free in solution, i.e. most regulatory proteins are nonspecifically bound.

The finding in our work that most of the cellular CI in a lysogen is bound to DNA outside O_R should also be included in future modeling (equilibrium as well as non-equilibrium studies) of the O_R-system. We note that the majority of the modeling attempts of the system has only considered the free (unbound) and O_R-bound states of CI and Cro (Ackers et al., 1982; Shea and Ackers, 1985; Darling et al., 2000b; Santillán and Mackey, 2004). If the large fraction of nonspecifically bound CI and Cro proteins is representative for other *E. coli* intracellular proteins as well, one may think that this has some impact on the stiffness (and local melting properties) of the DNA. Thus, the sum of all NSB-bound proteins associated with the λ phage genome may have some impact on the energetics associated with the O_R-O_L DNA-loop (this loop was first described by Révet et al. (1999)).

We also investigate two single base mutants of O_R3, *r1* and *c12*, that Dodd et al. (2001) have provided P_{RM} LacZ activity data on. We apply the $\Delta\Delta G$ estimates for CI of Sarai and Takeda (1989) and predict a $\Delta\Delta G$ for the RNAP affinity to P_{RM} of order −0.3 and +0.5 kcal/mol for *r1* and *c12*.

Due to the significant NSB in the cell that leads to low concentrations of the regulatory proteins, it might be relevant to model the system with spatial and/or time-dependent protein distributions (noise) (Metzler, 2001; Ozbudak et al., 2002; Elowitz et al., 2002; Aurell and Sneppen, 2002; Blake et al., 2003; Raser and O'Shea, 2004). We also note that Isaacs et al. (2003) performed a partly theoretical and partly experimental in vivo study of an isolated module of O_R. They concluded that noise is a significant factor in genetic regulation, and, in particular in systems of low particle numbers. There are indications that noise is significant also in the naturally occurring O_R-system (Bæk et al., 2003). This stochasticity and low free regulatory protein-concentrations might actually make the switch from the lysogenic to the lytic state more efficient, which adds to the effect of the O_R-O_L looping (the latter reduces the CI concentration compared to the situation when the O_L is absent (Dodd et al., 2001)). Arkin et al. (1998) also point out that the stochastic character, due to low free cellular protein concentrations, such that stochasticity becomes a necessary ingredient to predict the dynamics of λ phage infected *E. coli* cells and related systems. However, one should keep in mind that fluctuations in the free particle numbers may be damped significantly

due to the large number of nonspecifically bound species that effectively act as a buffer in this respect.

Another interesting issue, in light of the significant NSB reported herein, is that specific binding of proteins may profit from the nonspecifically bound portion of the proteins, i.e. a combination of a one-dimensional diffusion process along the DNA and protein interchange between close DNA segments as a source for specific binding that then comes in addition to the pool of freely accessible proteins in solution (Berg et al., 1981; von Hippel and Berg, 1989; Gerland et al., 2002). This diffusion process corresponds to a specific protein translocation on the DNA that opens the possibilities for sliding, hopping, and intersegment/interdomain protein transfer.

Acknowledgements

It is a pleasure to acknowledge discussions with J.S. Høye.

Appendix A. Expression for NSB

Consider the binding of CI dimers to the *E. coli* genome of $N_{bp} = 4.6 \times 10^6$ base pairs. As for the specific binding to O_R we assume that only dimers bind nonspecifically in significant amounts (Ptashne, 1992). We assume that each dimer covers λ base pairs (bp). Then, the number of combinations for binding q dimers to the DNA is $[N_{bp} - q(\lambda - 1)]!/q![N_{bp} - \lambda q]!$ (McQuistan, 1968). The grand canonical partition function for protein–DNA binding is then

$$\Xi = \sum_{q=0}^{N_{bp}/\lambda} \frac{[N_{bp} - q(\lambda - 1)]!}{q![N_{bp} - \lambda q]!}\, r^i, \tag{A.1}$$

where

$$r = [\mathrm{CI}_2] \exp(-\Delta G^{CI}_{NSB}/(RT)). \tag{A.2}$$

$[\mathrm{CI}_2]$ is the free CI dimeric concentration (in molar) and ΔG^{CI}_{NSB} is the free energy change for NSB upon DNA-binding of CI dimers. We note in the experimental "window" we compare our model to leads to $\Delta G_{NSB} \approx -4\,\mathrm{kcal/mol}$ and $[\mathrm{CI}_2] < 1.1\,\mu\mathrm{M}$, thus, $r \ll 1$.

We consider the model in the low density limit where the mean number of NSB-bound dimers (N_{NSB}) yields $N_{NSB} \ll N_{bp}$. Using Stirling's formula[5] the most probable term in the partition sum of Eq. (A.1) is approximately $N_{bp}r$. Furthermore, Ξ is nearly symmetric around $N_{bp}r$ and has a variance $\sim 1/\sqrt{N_{bp}r}$. This means that $N_{NSB} \approx N_{bp}r$ and by inserting r from Eq. (A.2) we obtain

$$N_{NSB} \approx N_{bp}[\mathrm{CI}_2] \exp(-\Delta G^{CI}_{NSB}/(RT)). \tag{A.3}$$

That is in this binding limit without overlaps, each protein binds independently. The result in Eq. (A.3) is also supported by numerical evaluation leading to a relative error less than 0.02 between the approximate and the exact value of N_{NSB} for the parameters within the experimental range we study. Eq. (A.3) is the same expression that von Hippel et al. (1974) proposed for NSB. Finally, we note that the value $\lambda = 1$ in Eq. (A.1) leads to the binomial distribution that has the exact expectation value $N_{bp}r$.

Appendix B. Fitting procedure

The activity at P_{RM} is

$$Activity_{RM} = a\,\mathrm{Prob}_{RM}(\Delta G^{CI}_{NSB}), \tag{B.1}$$

where $\mathrm{Prob}_{RM}(\Delta G^{CI}_{NSB})$ is the probability for binding of RNAP at P_{RM} calculated from Eq. (1), and indicates that Prob_{RM} depends upon ΔG^{CI}_{NSB}. Note that according to Hawley and McClure (1982), a CI dimer bound to O_R2 is estimated to increase the transcription rate by a factor 11 (Shea and Ackers, 1985) that is implicitly included in the activity in Eq. (B.1) (see Eq. (4) of Bakk et al. (2004a)). a is a proportionality constant that scales $\mathrm{Prob}(\Delta G^{CI}_{NSB})$ to the experimental activity (unit: s^{-1}). The model in Eq. (B.1) has two free parameters, a and ΔG^{CI}_{NSB}. The best-fit procedure of the activity is performed as a two-dimensional grid search over a and ΔG^{CI}_{NSB}, where the least quadratic error (between theory and experimental data points) is pursued (Bevington and Robinson, 1992). The procedure is repeated for finer grids until an accuracy of ΔG^{CI}_{NSB} within 0.01 kcal/mol is obtained.

The activity for P_R is defined in a similar manner as Eq. (B.1):

$$Activity_R = \tilde{a}\,\mathrm{Prob}_R(\Delta G^{Cro}_{NSB}), \tag{B.2}$$

where $\mathrm{Prob}_R(\Delta G^{Cro}_{NSB})$ is the probability for binding of RNAP at P_R. The fitting procedure is performed as described for the P_{RM} activity above where $\tilde{a}$ and ΔG^{Cro}_{NSB} are free (fitting) parameters in the model.

References

Ackers, G.K., Johnson, A.D., Shea, M.A., 1982. Quantitative model for gene regulation by λ phage repressor. Proc. Natl Acad. Sci. USA 79, 1129–1133.

Arkin, A., Ross, J., McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. Genetics 149, 1633–1648.

Aurell, E., Sneppen, K., 2002. Epigenics as a first exit problem. Phys. Rev. Lett. 88, 048101-1–048101-4.

[5] $n! \approx \sqrt{2\pi n}(n/e)^n$, where the relative error is less than 0.01 for $n \geqslant 10$.

Aurell, E., Brown, S., Johanson, J., Sneppen, K., 2002. Stability puzzles in phage λ. Phys. Rev. E 65 (2002) 051914-1–051914-9.

Bæk, K., Svenningsen, S., Eisen, H., Sneppen, K., Brown, S., 2003. Single-cell analysis of λ immunity regulation. J. Mol. Biol. 334, 363–372.

Bailone, A., Levine, A., Devoret, R., 1979. Inactivation of prophage λ repressor in vivo. J. Mol. Biol. 131, 553–572.

Bakk, A., Metzler, R., 2004. In vivo non-specific binding of λ CI and Cro repressors is significant. FEBS Lett. 563, 66–68.

Bakk, A., Metzler, R., Sneppen, K., 2004a. Sensitivity of O_R in phage λ. Biophys. J. 86, 58–66.

Bakk, A., Metzler, R., Sneppen, K., 2004b. Sensitivity of phage lambda upon variations of the Gibbs free energy. Israel J. Chem., in press.

Berg, O.G., Winter, R.B., von Hippel, P.H., 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids: 1. Models and theory. Biochemistry US 20, 6929–6948.

Bevington, P.R., Robinson, D.K., 1992. Data Reduction and Error Analysis for the Physical Sciences, second ed. McGraw-Hill, Boston.

Blake, W.J., Kærn, M., Cantor, C.R., Collins, J.J., 2003. Noise in eucaryotic gene expression. Nature 422, 633–637.

Blattner, F.R., et al., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1462.

Bremer, H., Dennis, P.P., 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella*, vol. 2, second ed. ASM Press, Washington, pp. 1553–1569.

Brooks, K., Clark, A.J., 1967. Behavior of λ bacteriophage in a recombination deficient strain of *Escherichia coli*. J. Virol. 1, 283–293.

Bukau, B., Horwich, A.L., 1998. The Hsp70 and Hsp60 chaperone machines. Cell 92, 351–366.

Burz, D.S., Ackers, G.K., 1996. Cooperativity mutants of bacteriophage λ cI repressor: temperature dependence of self-assembly. Biochemistry US 35, 3341–3350.

Darling, P.J., Holt, J.M., Ackers, G.K., 2000a. Coupled energetics of λ *cro* repressor self-assembly and site-specific DNA operator binding I: analysis of *cro* dimerization from nanomolar to micromolar concentrations. Biochemistry US 39, 11500–11507.

Darling, P.J., Holt, J.M., Ackers, G.K., 2000b. Coupled energetics of λ *cro* repressor self-assembly and site-specific DNA operator binding II: cooperative interactions of *cro* dimers. J. Mol. Biol. 302, 625–638.

Dodd, I.B., Perkins, A.J., Tsemitsidis, D., Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of P_{RM} and efficient switching from lysogeny. Gene. Dev. 15, 3013–3022.

Donachie, W.D., Robinson, A.C., 1987. Cell division: parameter values and the process. In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella Typhimurium*, vol. 2, second ed. American Society of Microbiology, Washington, pp. 1578–1593

Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186.

Fong, R.S.-C., Woody, S., Gussin, G.N., 1994. Direct and indirect effects of mutations in λP_{RM} on open complex formation at the divergent P_R promoter. J. Mol. Biol. 240, 119–126.

Gerland, U., Moroz, J.D., Hwa, T., 2002. Physical constraints and functional characteristics of transcription factor–DNA interaction. Proc. Natl Acad. Sci. USA 99, 12015–12020.

Hawley, D.K., McClure, W.R., 1982. Mechanism of activation of transcription initiation from the λP_{RM} promoter. J. Mol. Biol. 157, 493–525.

Isaacs, F.J., Hasty, J., Cantor, C.R., Collins, J.J., 2003. Prediction and measurement of an autoregulatory module. Proc. Natl Acad. Sci. USA 100, 7714–7719.

Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K., Ptashne, M., 1981. λ repressor and cro-components of an efficient molecular switch. Nature 294, 217–223.

Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W., von Hippel, P.H., 1977. Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli lac* repressor in vivo. Proc. Natl Acad. Sci. USA 74, 4228–4232.

Koblan, K.S., Ackers, G.K., 1991. Energetics of subunit dimerization in bacteriophage λ *c*I repressor: linkage to protons, temperature, and KCl. Biochemistry US 30, 7817–7821.

Koblan, K.S., Ackers, G.K., 1992. Site-specific enthalpic regulation of DNA transcription at bacteriophage λ O_R. Biochemistry US 31, 57–65.

Lwoff, A., 1953. Lysogeny. Bacteriol. Rev. 17, 269–337.

McQuistan, R.B., 1968. Exact occupation statistics for one-dimensional arrays of λ-bells. Nuovo Ciment. B 58, 86–92.

Metzler, R., 2001. The future is noisy: the role of spatial fluctuations in genetic switching. Phys. Rev. Lett. 87, 068103-1–068103-4.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., van Oudenaarden, A., 2002. Regulation of noise in the expression of a single gene. Nat. Genet. 31, 69–73.

Pakula, A.A., Young, V.B., Sauer, R.T., 1986. Bacteriophage λ *Cro* mutations: effects on activity and intracellular degradation. Proc. Natl Acad. Sci. USA 83, 8829–8833.

Ptashne, M., 1992. A Genetic Switch: Phage λ and Higher Organisms, second ed. Cell Press & Blackwell, Cambridge, MA.

Raser, J.M., O'Shea, E.K., 2004. Control of stochasticity in eucaryotic gene expression. Science 304, 1811–1814.

Reinitz, J., Vaisnys, J.R., 1990. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. J. Theor. Biol. 145, 295–318.

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

Santillán, M., Mackey, M.C., 2004. Why the lysogenic state of phage λ is so stable: a mathematical modeling approach. Biophys. J. 86, 75–84.

Sarai, A., Takeda, Y., 1989. λ repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. Proc. Natl Acad. Sci. USA 86, 6513–6517.

Senear, D.F., Batey, R., 1991. Comparison of operator-specific and nonspecific DNA binding of the λ *c*I repressor: [KCl] and pH effects. Biochemistry US 30, 6677–6688.

Shea, M.A., Ackers, G.K., 1985. The O_R control system of bacteriophage lambda: a physical–chemical model for gene regulation. J. Mol. Biol. 181, 211–230.

Takeda, Y., Ross, P.D., Mudd, C.P., 1992. Thermodynamics of Cro protein-DNA interactions. Proc. Natl Acad. Sci. USA 89, 8180–8184.

von Hippel, P.H., Berg, O.G., 1989. Facilitated target location in biological systems. J. Biol. Chem. 264, 675–678.

von Hippel, P.H., Revzin, A., Gross, C.A., Wang, A.C., 1974. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: 1. The *lac* operon: equilibrium aspects. Proc. Natl Acad. Sci. USA 71, 4808–4812.