Stability puzzles in phage λ

Erik Aurell,[1] Stanley Brown,[2] Johan Johanson,[3] and Kim Sneppen[4]
[1]*SICS, Box 1263, SE-14 29 Kista, Sweden*
and Nordita, Blegdamvej 17, DK-2100 Copenhagen, Denmark
[2]*Department of Molecular Cell Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen, Denmark*
[3]*Department of Microelectronics and Nanoscience, CTH, SE-412 96 Göteborg, Sweden*
[4]*Department of Physics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway*
(Received 9 January 2002; published 16 May 2002)

In the absence of RecA-mediated cleavage of the repressor, the λ prophage is exceptionally stable. We develop a stochastic model that predicts the stability of such epigenetic states from affinities of the molecular components. We find that the stability, in particular, depends on the maximum possible cI protein production, and on the number of translated cro proteins per transcribed mRNA. We apply the model to the behavior of recently published mutants of O_R and find, in particular, that a mutant that overexpress cro behaves in a different way than what was predicted, thus suggesting that the current view of the O_R switch is incomplete. The approach described here should be generally applicable to the stability of expressed states.

DOI: 10.1103/PhysRevE.65.051914 PACS number(s): 87.10.+e, 87.16.−b

I. INTRODUCTION

The lysogeny maintenance switch in phage λ can be viewed as one of the simplest examples of the molecular level of computation, command, and control in a living system. If, following infection of the bacterium *Escherichia coli*, the virus enters the lysogenic pathway, it represses its developmental functions, in response to a small set of sensory inputs, and integrates its DNA into the host chromosome. In this state the prophage may be passively replicated for many generations of *E. coli*.

Established lysogeny is maintained by the protein cI which blocks operators O_R and O_L on the λ DNA, and thereby prevents transcription of all lytic genes including cro [1]. In lysogeny the cI concentration functions as a sensor of the state of the bacterium: if DNA is damaged the protease activity of RecA is activated, leading to degradation of cI. A low cI concentration allows for transcription of the lytic genes, starting with *cro*, the product of which is the protein cro.

According to the picture so beautifully conveyed by Ptashne [1], the λ lysogeny maintenance switch is centered around operator O_R, which consists of three binding sites O_R1, O_R2, and O_R3, each of which can be occupied by either a cro dimer or a cI dimer. As illustrated in Fig. 1 these three binding sites control the activity of two promoters P_{RM} and P_R for respectively, cI and *cro* transcription. Transcription of *cro* starts at P_R, which partly overlaps O_R1 and O_R2. Transcription of cI starts at P_{RM}, which partly overlaps O_R2 and O_R3. The affinity of RNA polymerase for the two promoters, and subsequent production of the two proteins, depends on the pattern of cro and cI bound to the three operator sites and thereby establishes lysogeny with 180–350 cI molecules per bacterium [2]. If, however, cI concentration becomes sufficiently low, the increased production of cro throws the switch to lysis.

There does not seem to be a quantitative theoretical understanding of how stable the lysogenic state is to spontaneous switching. We will demonstrate that the large stability observed experimentally puts constraints on the mechanism of the switch, constraints that we explore in a quantitative model. We will show that in order to span the time scales from single molecule on-off binding times of order about 10 s, to the stability of order 10^{12} s, a simple model, where lysis occurs every time cro is transcribed, is inadequate. On the basis of the transition state theory for lysogeny maintenance [3] we here elaborate a dynamical model that takes into account all known mechanisms of O_R. This allow us to discuss key elements in the stability maintenance, and to falsify whether the currently known molecular mechanisms are adequate. We will see that the recently reported robustness of the switch against mutational changes of the operators [4] suggests that an additional mechanism of stability must be present.

II. EXPERIMENTS ON LYSOGEN STABILITY

Repression of bacteriophage λ can be eliminated either by cleavage of the repressor by the RecA protease (reviewed by [5]) or by the absence of repressor at the operators. The frequency of spontaneous induction in strains deleted for the recA gene has recently been reported. Rozanov *et al.* [6] found the frequency of infectious centers to be 1.1×10^{-7}. Little *et al.* [4] measured the ratio of the released phage per lysogen and the burst size. They estimate the frequency of one cell to switch to the lytic state to be 4×10^{-7} per gen-

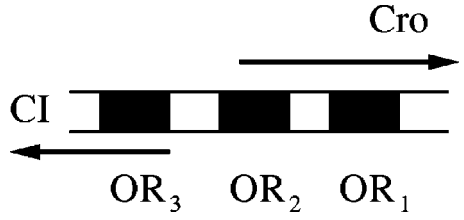

FIG. 1. Right operator complex, O_R, consisting of the three operators O_R1, O_R2, and O_R3. cI is transcribed when O_R3 is free and O_R2 is occupied by cI. cro is transcribed when both O_R2 and O_R1 are free.

eration, assuming the switch occurred 1 h before lysis. Little *et al.* verified that the lysogens contained only a single prophage by polymeraze chain reaction (PCR). Reference [6] does not report testing the number of prophages in the lysogens they used. We repeated these measurements using single lysogens [7] of MG1655 [8] containing the del(srl-recA)306::Tn10 allele [9]. We measured the frequency of switching to the lytic state by measuring the appearance of turbid plaques in a limiting dilution experiment at 37° in YT broth [10]. We find lysed cells to appear with frequencies between 1.0 and 1.8×10^{-8}, which, if we include the same correction factor as [4] to account for the increase in cell numbers during the period of phage development, implies a rate in the range of 3.5×10^{-8} and 6.0×10^{-8} per generation and cell. Thus, all three sets of measurements fall within a factor of 10, in spite of the use of different strain backgrounds.

Recent results of Little (Ref. [42]) show that at 37 °C about 99% of the turbid plaques are not wild type, but instead a marginally stable $P_{RM}240$ mutant. As a consequence, Little estimates that wild-type λ phages inside *E. coli* cells grown on nutrient rich medium have a frequency of spontaneous lysis that does not exceed 2×10^{-9} per generation and cell.

III. ELEMENTS OF A MODEL

The considerations in this section are in part closely similar to the model of Refs. [11,12]. The main difference is that our interest is in longer time scales, and in the stability of the switch against spontaneous stochastic fluctuations. We will, therefore, begin with a discussion of characteristic time scales, then discuss the chemistry, production, and decay. The stochastic element is included in a definite model, see the following section.

A. Time scales

The measured half-lives of cI and cro dissociation from O_R1 at 22 °C are, respectively, 34 s for cI [13], and varies from 1200 s for cro [14] dissociation from O_R alone down to 20 s [15] from OR flanked by 1 kb of λ-DNA. At equilibrium dissociation balances association, and thus the time it takes a cI or cro dimer to bind to the operator is

$$\tau_1=\tau_{off}V[1M]\exp(\Delta G/RT), \tag{1}$$

where ΔG are the free energies of binding, which for cI and cro are tabulated in the reference by Darling *et al.* [16] and $RT=0.617$ kcal/mol. Assuming a bacterial volume of $V=2\times10^{-15}$ l [17] we have $V[1M]=12\times10^{8}$, and can then estimate the association rates of single molecules upto orders of tens of seconds, for both cI and cro. This is compatible with diffusion-limited association into a region of size $L=5$ nm, at a rate determined by the Smoluchowski equation [18]: $\tau_1=V/(4\pi DL)$, assuming a diffusion constant $D\sim5$ $\mu m^2/s$, similar to that of other proteins in bacterial cytoplasma [19]. The association rate from N molecules is proportional to N, and thus for a bacterium with more than ten molecules of either cI or cro, association events will occur at a frequency of fractions of a second.

The shortest time scale in which we are actually interested is that which it takes for a switching event to take place. This cannot be much less than those of significant changes in cI and cro concentrations, which are of order one bacterial generation, 34 min in the strains used in [4]. Hence, all binding-unbinding events of cro and cI can be assumed to take place in homeostatic equilibrium. In addition, we need to characterize RNA polymerase (RNAp) binding and initiation of transcription from promotor sites P_{RM} and P_R. As discussed below (see Sec. III C), these events also have to occur on time scales much shorter than one bacterial generation, and can, therefore, be described by their overall rates. We note that a more detailed description of RNAp action may be necessary to resolve the shorter time scales associated to the lysis-lysogeny entry decision, as discussed in modeling by [20,21].

B. Chemistry

We assume that cI and cro molecules in the cell are in homeostatic equilibrium. This does not mean that there is always the same number of cI dimers bound to the operators at any particular time. These numbers are fluctuating, and the equilibrium assumption gives the size of these fluctuations. The key inputs are cI and cro dimerization constants and the Gibbs free energies for their bindings to the three operator sites O_R1, O_R2, and O_R3, for which we here use the latest available experimental values, Table 4 in Darling *et al.* [16]. We also assume an affinity of RNAp binding of -11.5 kcal/mol to P_R and of -12.5 kcal/mol to P_{RM} [11]. Since P_R overlaps with O_R3, and shuts out cI and cro to bind there if present, this means that, for our purposes, RNAp binds to O_R3 with affinity -11.5 kcal/mol. Similarly, when bound to P_{RM}, which overlaps with O_R1 and O_R2, RNAp can be thought to bind to these two operator sites together, with affinity -12.5 kcal/mol. We assume a RNAp concentration in the cell of 30 nM.

We remark that these free energies are taken from *in vitro* studies, and that the *in vivo* conditions could be different, e.g., the measured protein-DNA affinities could depend sensitively on the ions present in the buffer solutions. On the other hand, *in vivo* the effects of such changes should be compensated for, as, e.g., changed KCl concentrations are by putrescine and other ions and crowding effects [22]. The data quoted in [16] was obtained at a KCl concentration of 200 mM, which most closely resembles *in vivo* conditions. Parameter sensitivity is analyzed separately below.

Following [16] we encode a state s of cI and/or cro bound to O_R by three numbers (i,j,k) referring, respectively, to O_R3, O_R2, and O_R1. The coding is 0 if the corresponding site is free, 1 if the site is occupied by a cI dimer, 2 if the site is occupied by a cro dimer, and 3 if the site is occupied by RNAp. As described above, RNAp can be considered to effectively bind to O_R3, or to O_R1 and O_R2 together. At this point we would like to stress that although RNAp occasionally initiates transcription when bound to P_R or P_{RM}, the *de facto* rates in a lysogen are sufficiently low to allow us to make an equilibrium ansatz. In Appendix A we discuss how to self-consistently include the nonequilibrium occupation of

operators when RNAp is initiating transcription at a rate significantly higher than the rate at which it spontaneously leaves DNA.

The probability of a state s at O_{R}, with i_s cI, j_s cro dimers and k_s RNAp molecules bound, is in the grand canonical approach of Shea and Ackers [11]

$$\mathrm{Prob}_R(s)=\mathcal{N}^{-1}[\mathrm{cI}]^{i_s}[\mathrm{cro}]^{j_s}[\mathrm{RNAp}]^{k_s}e^{-G(s)/RT}. \quad (2)$$

The normalization constant $\mathcal{N}^{-1}$ is determined by summing over s. The free dimer concentrations ([cI] and [cro]) are equal to the numbers of free dimers (n_D) per volume. The free dimer number of cI is given by

$$2n_D=N-n_M-2\mathcal{N}_{chr}n_U-2\mathcal{N}_{\lambda}\times\left(\sum_s i_s\mathrm{Prob}_R(s)+\sum_{s'} i_{s'}\mathrm{Prob}_L(s')\right), \quad (3)$$

where Prob_L is the analog of Prob_R for operator site O_{L}, to which both cI and cro can also bind, N is the total number of cI, n_U the number of cI dimers bound nonspecifically to the DNA of one bacterial chromosome, n_M the number of free cI monomers, and $\mathcal{N}_{\lambda}$ and $\mathcal{N}_{chr}$ the average number of copies present of the λ genome and of the full *E. coli* chromosome. The numbers n_M and n_U are in the homeostatic equilibrium ansatz given by n_D, bacterial volume, and equilibrium in the dimerization and DNA association reactions. The free cro dimer number is obtained from a formula similar to Eq. (3), but with weights $j(s)$ of Eq. (2) substituted for $i(s)$.

In this paper we take $\mathcal{N}_{\lambda}$ and $\mathcal{N}_{chr}$ both equal to 3 [17]. In our analysis we estimate that cI binds nonspecifically by -2.0 kcal/(mol bp), and that cro binds nonspecifically by -3.0 kcal/(mol bp). The effect of a stronger putative nonspecific cI binding to DNA is analyzed separately in Appendix B. Note that we also include occupancy of O_{L} in order to account for finite depletion of cI, which has some significance for the fractional binding at O_{R} when the total number of cI molecules is small. Parameters for O_{L} may be taken to be identical to O_{R} eventually with the $\Delta\Delta G$ changes measured by [23,24].

In Fig. 2(a) we show $P(0,3,3)+P(1,3,3)+P(2,3,3)$ as function of N_{cI} at $N_{Cro}=0$. We remark that Fig. 2(a) corresponds to the scaling of P_{R} promoter strength in the in vitro experiment of [25].

C. Production

cI and cro are produced from mRNA transcripts of cI and cro, which are initiated from promotor sites P_{RM} and P_{R}. The rate of initiation of transcription from P_{RM} when stimulated by cI bound to $O_{\mathrm{R}}2$ is denoted R_{RM}, and when not stimulated R^u_{RM}. The number of cI molecules produced per transcript is S_{cI}, and overall expected rate of production of cI is

$$\mathcal{R}_{\mathrm{cI}}=R_{\mathrm{RM}}S_{\mathrm{cI}}[P(310)+P(311)+P(312)]+R^u_{\mathrm{RM}}S_{\mathrm{cI}}[P(300)+P(301)+P(302)+P(320)+P(321)+P(322)+P(333)]. \quad (4)$$

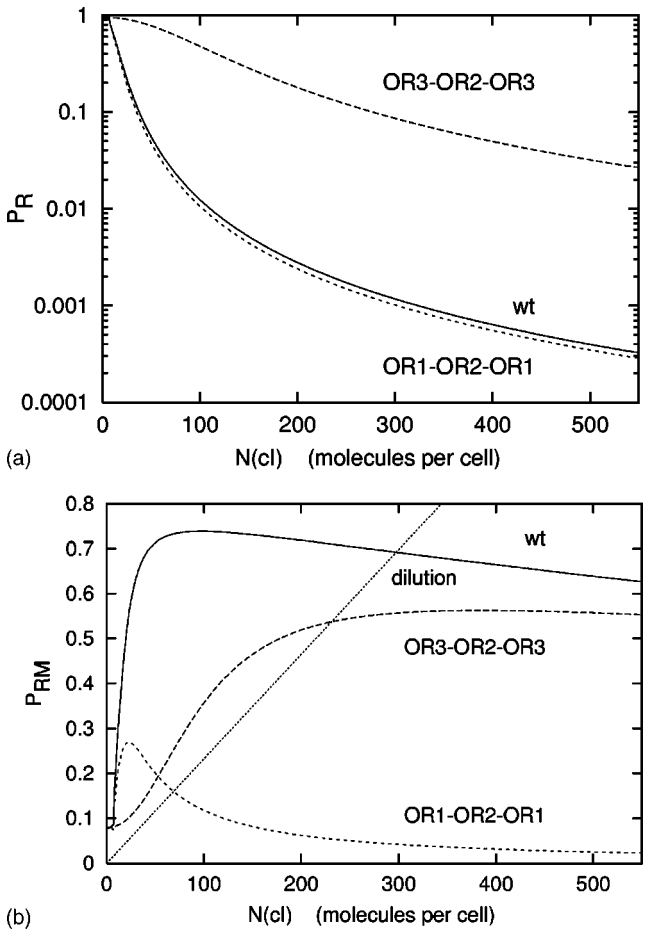

FIG. 2. (a) Probability of open cro promoter, $P(033)=P(033)+P(133)+P(233)$, as a function of cI number in an *E.coli* cell of volume 2 $\mu\mathrm{m}^3$. The dashed lines marked $O_{\mathrm{R}}1O_{\mathrm{R}}2O_{\mathrm{R}}1$ and $O_{\mathrm{R}}3O_{\mathrm{R}}2O_{\mathrm{R}}3$ show repression for the two O_{R} mutants investigated by Little (Ref. [4]); (b) activity of P_{RM} in units of activity when $O_{\mathrm{R}}3$ is free and $O_{\mathrm{R}}2$ is occupied by cI. Same notation as in (a). For relative activity of nonstimulated P_{RM} promoter we use the data of [25].

For the relative values of the coefficients, [25] report R^u_{RM} to be equal to $R_{\mathrm{RM}}/11$.

The above equations can be used to estimate *in vivo* nonspecific bindings by fitting the overall activity profile for O_{R} complexes with reported genes. Thus cro unspecific binding can be found by fitting data on P_{R} activity as a function of cro concentration in an *E. coli* where there is no cI. Such data have been reported by Reinitz and Vaisnys [26], with reference to Pakula *et al.* [27], and demonstrate that about 300 nM of cro inside a cell leads to half-suppression of P_{R}. This we can reproduce in the model with a -3.0 kcal/(mol bp) nonspecific cro binding. Similarly, a cI nonspecific binding of -2.0 kcal/(mol bp) is the minimal strength of nonspecific binding that is compatible with Johnson *et al.* [28] data on P_{RM} activity as a function of cI concentration in the cell. For weaker nonspecific binding the decrease of P_{RM} activity at high cI concentration is incompatible with data.

The absolute *in vivo* values of R_{R} and R_{RM} are inferred from data. We discuss how we determine these below. Ac-

cording to Shean and Gottesman [29] the number of cI molecules produced from one transcript of cI is a factor 20–70 smaller compared to lacZ. Thus S_{cI} is small, between 1 and 5, and we take S_{cI} to be 1.

As there are about 180–350 cI in a lysogenic cell, and as each transcription only results in a small number of cI molecules, transcription has to be initiated from P_{RM} many times per generation in lysogeny. We can, therefore, safely assume that the time scales for RNA polymerase to find the operator sites is short. Figure 2(b) shows average cI production according to Eq. (4) as a function of the cI number in the cell, at constant, volume. The small gap in cI number before full activation of P_{RM} at low cI in Fig. 2(b) originates in part from finite system filling of O_{R} and O_{L}. We note that the overall behavior resembles the *in vitro* measurements of [25].

The rate of initiation of transcription from promoter P_{R} is denoted R_{R} and the number of cro molecules per transcript S_{cro}. From [30] one can estimate S_{cro} to be 51% of ideal lacZ and [31] thus allow a lower estimate of S_{cro} to be 20. Thus, in contrast to the many small production events of cI, cro production is intermittent. The ratio of R_{R} to R_{RM} has been reported to be between 1.3 and 20 [25] In practice, we determine R_{RM} by balancing cI production and decay, as discussed below, and mostly treat R_{R} as a free parameter. We end this part by stressing that R_{R}, R_{RM}, and R^{u}_{RM} parametrize the total production rates from all copies of λ-DNA in the cell. The real rates of transcription per O_{R} complex will generally be smaller, e.g., about a factor $\mathcal{N}_{\lambda}\sim 3$ smaller at high growth rates.

D. Decay

Concentrations decay due to dilution and degradation. For cI we only take into account dilution from cell growth and division, while for cro we include, in addition, an *in vivo* half-life time t_{cro} of about 2600 sec at 37° [27], which is significant compared to, e.g., generation time of 34 min in the experiments of [4].

E. The steady state

Balance between production and decay of cI over one generation, where production is given by Eq. (4) and decay by N_{cI}/t_{life}, gives one equation for the parameter R_{RM}, if the total numbers of cI and cro in lysogeny are assumed to be known. Figure 2(b) displays the steady state as the intersection between the linear dilution and the P_{RM} activity curve, having, in this case, about 300 cI and no cro molecules in the system.

The average number of cro molecules in lysogens has not been measured, but in the models considered here this number is not zero. Equation (5) gives the average rate of cro production. If we assume that R_{R} is not very different from R_{RM}, we find the rate of cro transcription in lysogeny to be once per five to ten generations. This production is balanced by cro decay to give an average number of cro. Hence, the balance of only cI production and decay does not completely determine R_{RM}, since the probabilities depend parametrically on the cro concentrations, which in turn depend on R_{R}. However, the implied dependence of R_{RM} on R_{R} is not very pronounced. In the model presented below we fit R_{RM} and R_{R} to data, where the value of 0.14 for R_{RM} is not very sensitive to changes in model assumptions.

IV. THE MODEL

The basic time scale in our model is one bacterial generation, where we resolve shorter times around cro transcription events when necessary. A simulation goes as follows. Each bacterial generation starts with N_{cI} total number of cI, and N_{cro} total number of cro. From these one computes by Eq. (2) the occupation probabilities P_s of the states s of cI and cro dimers bound to O_{R}, and from these the average cI and cro production rates f_{cI} and f_{cro}. The first is computed by Eq. (4), and the second by the overall cro rate

$$\mathcal{R}_{\mathrm{cro}}=S_{\mathrm{cro}}R_{\mathrm{R}}(P_{033}+P_{133}+P_{233}), \tag{5}$$

cI production is assumed continuous, since S_{cI} is about 1. cro production is on the other hand treated as a discrete event. In each generation the bacterial volume V is assumed to start at $V_{start}=\frac{2}{3}V_{ave}$, and then to grow linearly until it has doubled. $V_{ave}=2\times 10^{-15}$ l is the average volume over one generation. The growth of bacterial DNA is treated similarly, with an average of three copies of bacterial DNA ($\sim 15\times 10^{6}$ bp per cell).

(1) The rates f_{cI} and f_{cro} are computed with the current values of N_{cI}, N_{cro} and V. cro transcription events are assumed Poisson distributed with mean waiting time $S_{\mathrm{cro}}/f_{\mathrm{cro}}$. A random time t_r is drawn from this distribution. If this time is less than the remaining time of the current bacterial generation, one cro transcription will occur, and we continue the simulation under point (2b) below. Else, no more cro transcription will take place in this generation, and we continue under (2a).

(2a) The time since beginning of the generation, or since last cro transcription in the generation, is t. We update cI numbers by the expected number produced in the interval t, $f_{\mathrm{cI}}t$, plus ξ, a random number drawn from a Gaussian distribution with zero mean and variance $f_{\mathrm{cI}}t$. We update cro number by removing each cro molecule with a probability $p=p(t)=2^{-t/t_{\mathrm{cro}}}$, and then proceed to cell division under point (3) below.

(2b) We update cI number and degrade cro number as above, (2a), but under time t_r. The number of cro molecules produced from this cro transcription is a random variable drawn from a Poisson distribution of mean S_{cro}, and is added to the remaining cro. We now return back to point (1) above for the remaining time of the current generation.

(3) This is the stage of cell division. The volume of a daughter cell is restarted at V_{start}, and will contain each cI and cro molecule of the mother cell with probability one-half. If cI number has shrunk below a threshold, a switch to the lytic pathway has occurred, and we leave the loop. Otherwise we return to (1) with the new values of cI and cro.

The expected number of generations before lysis is calculated by an average over at least ten independent trajectories. The value of the threshold in cI number is not important,

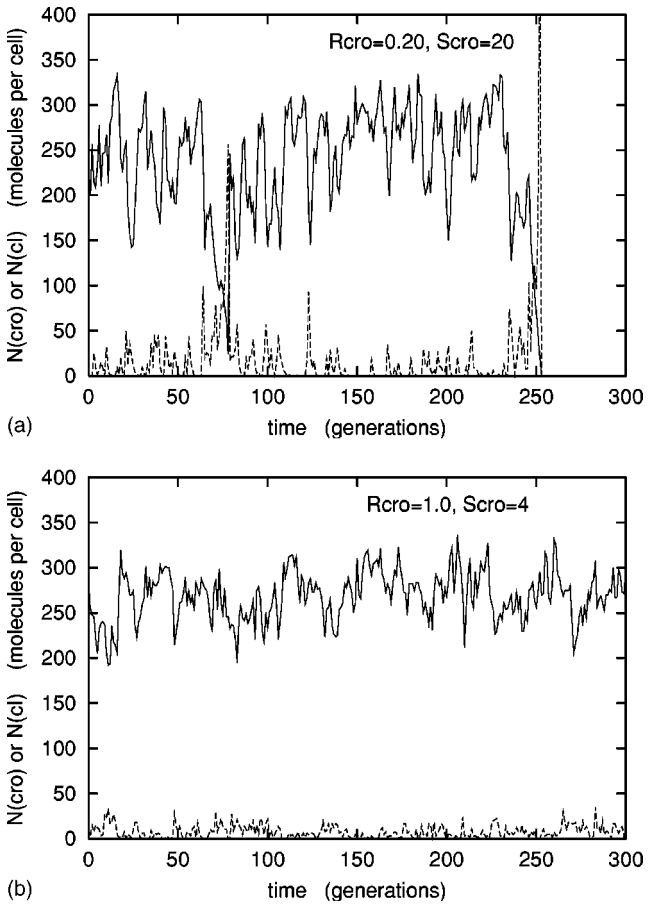

FIG. 3. Time course of the last few hundreds of generations before lysis in two simulations with the same high value of average cro production: i.e., $R_R S_{cro}=4\ \ s^1$ in both cases. The upper figure (a) refers to a situation where $S_{cro}=20$, the lower (b) to a case where $S_{cro}=4$ and thus represents a situation where cro production is steady. Thus for given average production, a main source of instability is therefore the ''burstiness'' of the protein production.

provided it is chosen sufficiently low. In the model presented here, lysis practically always follows if N_{cI} goes below 20.

V. RESULTS

A. Standard parameter behavior

The two parameters R_{RM} and R_R have been adjusted such that the average number of cI molecules in lysogeny is 300, and the mean rate of spontaneous lysis is 2×10^{-9} generations. R_R can alternatively be estimated by adjusting the density of cro in the anti-immune state measured by Reinitz and Vaisnys [26], suggesting that R_R is about 0.2 s^{-1}. However, this value then predicts a wild-type stability with a too high lysis frequency of about 0.01 per generation, which recalls the findings of [26] of a completely unstable switch when fitting anti-immune cro levels. Note that the anti-immune state is an artificially constructed state of defect phage where lysis is prevented, and a lysislike state is maintained by a high cro number and a low cI number in the cell.

Fine adjusting R_R from the value as discussed amounts to a minimal parametrization of additional unknown levels of repression of P_R, not included in the above standard parametrization. The other parameter values are discussed above in Sec. III. The values thus obtained are $R_{RM}=0.14\ \ sec^{-1}$ and $R_R=0.025\ sec^{-1}$. The ratio between these two rates is hence smaller than the limits set by [25]. We stress that any estimate on R_R based on lysis frequency is very uncertain, as R_R act in concordance with binding of cro to O_R3: a 1 kcal/mol change in $\Delta G(2,\ldots)$ can offset a tenfold change in R_R.

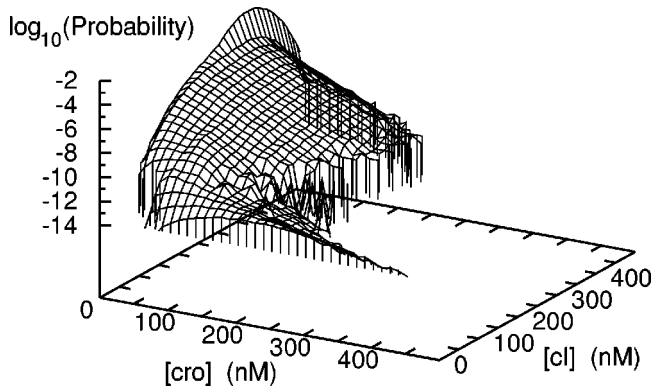

FIG. 4. Probability to be at various values of cI and cro number in the cell during lysogeny. Please notice the extended distribution at low cI corresponding to the rare fluctuations that leads to lysis.

For illustrative purpose we show in Fig. 3 a typical simulation with R_R equal to 0.2. The numbers of cI and cro fluctuate for a long time around the metastable lysogenic state. A particularly large fluctuation (of both cI and cro) then builds up over a few generations, leading to lysis. This kind of switching event is similar to the escape over an activation barrier as in the well-known diffusion model of chemical reactions. Thus, spontaneous lysis is the result of a number of rare events happening in a given short time interval. If each event has probability $p<1$, and lysis depends on n events, the resulting frequency of lysis is $\sim p^n$. This frequency is, therefore, very sensitive to p, which means to the parameters in the model.

In Fig. 3(b) we show another simulation, with same average cro production as in Fig. 3(a), i.e., $S_{cro}\times R_R=4\ \ s^{-1}$ is kept fixed, but where number of cro per transcript changed from 20 to 4. One observes less fluctuations in lysogen, and a much stabilized lysogen, corresponding to the much smaller probability of building up many unlikely events within a short time interval.

In Fig. 4 we explore the activation barrier picture in a more detailed way by computing the probability that the system over time visits various values of N_{cI} and N_{cro}. The distribution has a maximum around the lysogenic state, as it should. From there to the origin, along the N_{cI} axis, the probability becomes almost vanishingly small, which reflects the fact that a fluctuation of cI alone, without cro, cannot induce lysis. The switch instead happens along a ridge in the (N_{cI},N_{cro}) plane, where probability does not go down so quickly.

B. Parameter sensitivity

A main objective to build a model for stability is to address the quantities that determines the stability. As we have

TABLE I. Switching stabilities and cI and cro numbers as functions of model parameters for wild type and the Little mutants. We observe that the 323 mutants fail to stabilize unless R_R is decreased by a factor 30. nd means not determined.

O_R type	R_R (s^{-1})	S_{cro} number	cI (s)	cro number	Lysis frequency $(generations)^{-1}$
321	1.00	4	270	7	6.3×10^{-5}
321	0.20	20	234	20	0.01
321	0.10	20	275	7	6.6×10^{-4}
321	0.05	20	280	2	3.3×10^{-6}
321	0.03	20	287	1	1.4×10^{-8}
321	0.025	20	291	1	1.4×10^{-9}
121	0.10	20	nd	nd	<0.06
121	0.05	20	64	56	$<2\times10^{-4}$
323	0.10	20	nd	nd	>0.1
323	0.01	20	nd	nd	>0.1
323	0.001	20	171	7	5×10^{-3}
323	0.0005	20	205	3	4×10^{-5}

already discussed, stability is governed by near simultaneous occurrences of many rare events. Consequently, the stability will be very sensitive to the probability of these events, which in turn depend sensitively on parameters, in particular, of the parameters describing cro production in the cell. We now examine parameter sensitivity, i.e., lysis frequency as a function of production rates and binding affinities. The least uncertain combination is the product $S_{cI}R_{RM}$, which is determined, to good accuracy, by the number of cI in lysogeny, known to be between 180 and 350 for growth in a rich medium. The individual terms in this product are less well known, but only influence the dynamics through the relative amplitude of the noise in cI production, which will be proportional to the square root of S_{cI}, all else equal. This uncertainty is much less than what stems from cro, as we describe next, and will therefore be ignored.

By simulations we find that the key uncertain quantity that determines stability is total cro affinity for O_R3. Changes in cro affinities to O_R1 and O_R2 do not influence the stability significantly. cro affinity for O_R3 can be changed either directly, by changing the value of cro nonspecific DNA binding, or by changing the amount of cro produced and degraded in the cell. The production is controlled by R_R and by cI repression of P_R. Varying R_R amounts to a simplistic scan of effects due to variation of cI repression due to effects not explicitly taken into account. In Table I we show the dependence of stability with R_R.

cI influences the stability of a lysogen by binding to O_R1 and O_R2, which determines the fraction of time for which P_R is open, see Fig. 1. Increasing for instance binding to these sites by 0.5 kcal/mol means that the probability of P_R activation is reduced by a factor 5. cI also indirectly influences the stability of a lysogen through the binding to O_R3, which determines the fraction of time for which P_{RM} is open. To balance decay through dilution at given total number of cI, such a change must be compensated by an increase in the product $R_{RM}S_{cI}$. Further, if one takes the measured stability for granted, any increase in specific cI can be compensated for by a corresponding increase in R_R.

The conclusion so far is that with present parameter information, R_R needs to be rather small to be compatible with observed rate of spontaneous lysis in recA-*E. coli* strains. This indicates that there are additional levels of repression of P_R. Allowing adjustment of absolute level of PR we now examine the switch mechanisms by comparing mutants.

C. The Little mutants

In Table I we show the mean rate of switching in variants of the model corresponding to the recently reported lysogenic state in λ mutants, where either operator O_R1 is replaced by a copy of O_R3, or vice versa Little *et al.* [4]. Thus, while wild-type (321) has an O_R site made up of $O_R3O_R2O_R1$, the mutant labeled mutant 121 has $O_R1O_R2O_R1$ at O_R, and mutant 323 has $O_R3O_R2O_R3$. From the table we see that the mutant 121 is well fitted by the model, both with regard to the cI level in lysogen and the stability, i.e., Ref. [4] reported a cI level of 25–30 % of wild type, and a stability of 3×10^{-6}. However, we find it impossible to reproduce simultaneously the measured stability of wild-type recA lysogens and the existence of stable lysogens in the mutant 323. For 323 the cause of decreased stability is demonstrated in Fig. 2(a), which shows that P_R is activated much more than in wild type. For 121, on the other hand, the changed stability is associated to a much lower cI level in lysogens, as demonstrated in Fig. 2(b). Thus the two mutants demonstrate two mechanisms of destabilization, one through enhanced cro, the other through depleted cI. The prime lesson is that the current molecular mechanisms fail to reproduce the effect of increased cro in the cell.

When adjusting R_R down by a factor 30 we can obtain stable lysogens for 323. This parameter change allows us to analyze what lysogens in the 323 mutant would look like, as in the last rows of the table. We see that the number of cI for 323 is 70% of wild type (WT). Thus cI levels for both WT and 323 are compatible with the cI levels reported by [4], indicating that the promoter P_{RM} is not changed by these mutations. This is consistent with the conventional view that promoted activity is controlled by the respective -10 and -35 regions, which are unaffected by mutations [4]. Note that for low cro the cI levels in the mutants can also be found by balancing the production and decay of cI only, as in Fig. 2(b). In further qualitative agreement with experiment we observe that in spite of having less cI, the 121 mutant is markedly stabler than 323. This is mainly caused by a larger probability for open P_R in 323, as seen from Fig. 2(a).

Finally we report that the above conclusions remain valid also if cI binds significantly to nonspecific DNA, see Appendix: In that case we may simultaneously fit the stability of wild type and 121, but not wild type and 323.

VI. DISCUSSION

The switch to lysis in our model is essentially a first exit time problem [3], in a system influenced by a combination of deterministic and stochastic forces. A well-known analogy is

the thermal escape of a particle from a potential [32,33], a model of chemical reactions with activation energies. There the rate of escape depends exponentially on the height of the potential barrier, and is therefore very sensitive to small changes in the potential. Our case is not exactly in Kramers' form, but the analogy is nevertheless illuminating. We also find that the rate of spontaneous switching depends very sensitively on model parameters. In fact, the negative of the logarithm of the probability from Fig. 4 can be identified with the Wentzel-Freidlin quasipotential [34,35], which plays the same role in this more general exit time problem as the potential in Kramers' problem. An analogy to temperature is, on the other hand, the noises associated to typical fluctuations of cI and cro numbers in a lysogen [3].

The large stability of the lysogenic state of the λ puts constraints on working mechanisms of the switch. First, we have shown that a simple model, where every transcription of cro leads to lysis, only provides stability for a few generations. A mechanism must therefore exist that stabilizes the switch against most transcriptions of cro. We have explored a straightforward model, where lysogens are stabilized against spontaneous lysis by each cro transcription, which occurs about once per 3–5 bacterial generations. In this way, many cro transcriptions are necessary for lysis. The probability of this event is essentially the probability of a single cro transcription to a high power, implying a very rare spontaneous breakdown of lysogeny. The need for multiple cro transcriptions for breakdown of lysogen stability is associated to the following mechanisms.

(i) cro bound to operators is in homeostatic equilibrium, and cI production is therefore possible with cro present.

(ii) cro is degraded and diluted over time.

(iii) cro binds also to nonspecific DNA, with significant affinity. Measured lysogenic stability of wild-type recA-*E. coli* can then be reproduced in the model, if one assumes P_R activity that is a factor of about 5–10 lower than the one measured by [26], or alternately if one assumes that cro binding to O_R3 is between 0.5 and 1 kcal/mol weaker *in vivo* than *in vitro*.

The actual stability depends sensitively on the parameters in the model, and we thus predict a stability that changes dramatically with changes in, e.g., growth conditions. The model allows for quantitative examination of mutants, by changing parameters in the model, and we have in this sense investigated the O_R mutants reported by [4]. Although the model can reproduce the ratios of cI numbers in mutant and wild-type lysogens, and of the stabilities of each one of wild-type, 121, and 323 mutants, it fails to reproduce simultaneously the stability of wild type and 323. This all in all suggests that some additional mechanism outside the O_R complex contributes to the stability. Origins of such a mechanism could be either an increased repression of cro production by cI relative to our model, or, conversely, a decreased repression of cI production by cro. Examples of such mechanisms are as follows: (i) additional cI transcription directed from P_{RE}; (ii) an unknown interaction between cro and cI that allows cro to stabilize lysogens; (iii) cI mediated binding between O_L and O_R if that would repress P_R in a way that differs significantly between 323 and wild type; (iv) cro is unable to block P_{RM} completely, even when bound to O_R3; and (v) the strength of P_R is actually changed in the 323 mutant.

The first mechanism (i) would give a second role to P_{RE} in repressor maintenance, while mechanism (ii) was suggested already by [36] based on a study of the hyper immune (Hyp) phenotype. Also one might imagine that 323 often switches to the anti-immune state, but then mostly switches back. However, if this happens with a frequency of order one (which it would for WT parameters, see the table), one should expect a much lower average cI level in the measured lysogen, than that found in our model. An argument for mechanism (iii) is the recent report by Révet *et al.* [37], which shows a (at least) fourfold increase in repression of P_R in a plasmid construction, involving two operator complexes, both of which can bind cI dimers, and one of which overlaps with P_R as in wild-type λ-O_R. However, as we already treat R_R as a free parameter, and as 323 binds cI weaker that WT the overall effect of O_L-O_R coupling is likely to destabilize 323 relative to WT. The last possibility (v) is the most conservative, as concerns our view of the λ switch, but demands that the activity of promotor P_R depends on the exact DNA sequence at a location, which is usually not thought to be important in determining promoter strengths. Also note that the promoter strengths of P_{RM} is not affected by any of the mutations. In any case, the above mechanisms are, of course, not exclusive, and do not constitute an exhaustive list.

In summary, we have developed a quantitative model for the stability of lysogens. The model builds a direct connection between processes, their affinities, and the resulting stability. The model allows for quantitative numerical tests of genetic feedback mechanisms and identifies key parameters, which determine this stability. In particular we have seen that the maximum firing rate of derepressed P_{RM} is very important for stability. Further we have seen that the multiple copies of cro proteins per cro transcript should have a pronounced effect on stability. Finally we have demonstrated that the model points to possible shortcomings in the standard molecular mechanisms assumed to be important for the λ-lysogeny maintenance switch. Thus the model suggest a series of experiments, to test, in particular, cro levels in various lysogens, as well fluctuations in cI and cro levels between lysogenic cells. Finally, we wish to point out again that the model developed here speaks of the stability of data storage in biocomputing [38].

ACKNOWLEDGMENTS

We thank D. Court, I. Dodd, B. Egan, H. Eisen, J. Little, M. Mossing, and P. Muratore-Ginanneschi for discussions and valuable comments. We thank I. Dodd and B. Egan for suggesting that the repression by cro may be incomplete. We thank the Swedish Natural Research Council for support under Grant No. M-AA/FU/MA 01778-333 (E.A.), and the Danish research council for financial support (S.B.). We also thank K.F. Jensen and B.E. Uhlin for gifts of MG1655 and the recA allele, respectively.

APPENDIX A: NONEQUILIBRIUM EFFECTS DUE TO RNAP ACTIVATION

Above we have assumed that RNAp binds and unbinds to its potential target with equilibrium kinetics. In principle, this is not entirely true, RNAp may initiate transcription and thus leave the target in a directed way. This is expressed through the reaction path

$$\mathrm{RNAp}+\mathrm{O}\rightleftharpoons\mathrm{RNApO}\rightarrow\mathrm{O}\ \ldots\ \mathrm{RNAp}, \tag{A1}$$

where the last step is transcription initiation with rate constant k. Reaction (A1), of course, takes place at both P_{RM} and P_{R}, with rates either R_{RM} or R_{R}. The first reaction is an equilibrium kinetic equation with an on-rate k_{on} (in $\mathrm{M}^{-1}\ \mathrm{s}^{-1}$) and an off-rate $k_{off}=k_{on}\cdot\exp(G/RT)$ (in s^{-1}). Setting operator concentration to be $[\mathrm{O}]=1/V$ we get the change in RNApO concentration as

$$\frac{d}{dt}[\mathrm{RNApO}]=k_{on}[\mathrm{RNAp}]\left(\frac{1}{V}-[\mathrm{RNApO}]\right)-k_{off}[\mathrm{RNApO}]-k[\mathrm{RNAp\ O}] \tag{A2}$$

from which steady state condition gives occupancy

$$[\mathrm{RNApO}]=\frac{[\mathrm{RNAp}]}{1+V\exp(G'/RT)} \tag{A3}$$

with a pseudoaffinity G' given by

$$\exp(G'/RT)=\exp(G/RT)+\frac{kV}{k_{on}}, \tag{A4}$$

where the last term is the nonequilibrium mean rate at which RNAp clears the promoter and starts transcription, in units of the mean rate at which it arrives to the promotor from solution. k_{on} can be estimated from a diffusion-limited reaction to be about $0.1\ \mathrm{s}^{-1}$ per promoter, which is of same order as R_{RM} and R_{R} in the present study, giving a correction of only 0.1 kcal/mol in G', compared with for G, which is -11.5 kcal/mol. This is insignificant for the present setup, but would be easy to include, if necessary. For cases with weaker RNAp-promoter bindings, or stronger transcription initiation rates, this effect may be important, and eventually one may have situations where RNAPp activity is simply proportional to the probability for having a vacant promoter binding site.

APPENDIX B: CONSEQUENCES OF NONSPECIFIC cI BINDING

Koblan and Ackers [39,40] report that binding per base pair between cI dimers and nonspecific DNA is at least 9 kcal/mol weaker than binding between cI and $O_{\mathrm{R}}3$. This sets a limit on nonspecific binding to be at most -3.5 kcal/(mol bp). Here we consider the effects of a -3.0 kcal/(mol bp) nonspecific binding between cI dimers and DNA, see also Table I. The 2×10^{-9} stability of lysogens with $N_{\mathrm{cI}}=200$ is then obtained with $R_{\mathrm{RM}}=0.085\ \mathrm{s}^{-1}$ and $R_{\mathrm{R}}=0.02\ \mathrm{s}^{-1}$. In this state $O_{\mathrm{R}}3$ is 20% occupied in lysogeny, in agreement with [41]. When we also apply these parameters to the 121 and 323 mutants we find that 121 can form stable lysogens [with $N_{cI}(121)\approx40$ and lysis frequency in the range 10^{-3}–10^{-5}], whereas 323 fails to do so. This lack of stability of 323 in the model fails to reproduce the stability observed by Little *et al.* [4].

[1] M. Ptashne, *A Genetic Switch; Phage λ and Higher Organisms* (Blackwell, Oxford, 1992).

[2] L. Reichardt and A. D. Kaiser, Proc. Natl. Acad. Sci. U.S.A. **68**, 2185 (1971).

[3] E. Aurell and K. Sneppen, Phys. Rev. Lett. **88**, 048101 (2002).

[4] J. W. Little, D. P. Shepley, and D. W. Wert, EMBO J. **18**, 4299 (1999).

[5] J. W. Roberts and R. Devoret, in *Lambda II*, edited by R. W. Hendrix *et al.* (Cold Spring Harbor Laboratory, New York, 1983), p. 123.

[6] D. V. Rozanov, R. D'Ari R, and S. P. Sineoky, J. Bacteriol. **180**, 6306 (1998).

[7] B. S. Powell *et al.*, Nucleic Acids Res. **22**, 5765 (1994).

[8] K. F. Jensen, J. Bacteriol. **175**, 3401 (1993).

[9] L. N. Csonka and A. J. Clark, Genetics **93**, 321 (1979).

[10] J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, New York, 1972).

[11] M. A. Shea and G. K. Ackers, J. Mol. Biol. **181**, 211 (1985).

[12] G. Ackers, A. Johnson, and M. Shea, Proc. Natl. Acad. Sci. U.S.A. **79**, 1129 (1982).

[13] H. M. Nelson and R. T. Sauer, Cell **42**, 549 (1985).

[14] R. Jana *et al.*, Biochemistry **37**, 6446 (1998).

[15] J. G. Kim *et al.*, J. Mol. Biol. **196**, 149 (1987).

[16] P. J. Darling, J. M. Holt, and G. K. Ackers, J. Mol. Biol. **302**, 625 (2000).

[17] H. Bremmer and P. P. Dennis, in *Escherichia coli and Salmonella*, edited by F. C. Neidhardt (ASM Press, Herndon, VA, 1996), p. 1553. At doubling time of 30 min one *E. coli* cell has dry weight 6.4×10^{-13} g and consists of ~70% water (Neidhardt and Umbarger, *ibid.*, pp. 13–16), giving a volume of 2.1×10^{-15} l. At the same conditions *E. coli* DNA is on average present in three copies, the part belonging to λ in 2.4 copies.

[18] O. G. Berg, R. B. Winter, and P. H. von Hippel, Trends Biochem. Sci. **7**, 52 (1982).

[19] M. B. Elowitz *et al.*, J. Bacteriol. **181**, 197 (1999).

[20] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997).

[21] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).

[22] M. T. Record *et al.*, Trends Biochem. Sci. **23**, 190 (1998).

[23] Y. Takeda, A. Sarai, and V. M. Rivera, Proc. Natl. Acad. Sci. U.S.A. **86**, 439 (1989).

[24] Y. Takeda, P. D. Ross, and C. P. Mudd, Proc. Natl. Acad. Sci. U.S.A. **89**, 8180 (1992).

[25] D. K. Hawley and W. R. McClure, J. Mol. Biol. **157**, 493 (1982); Cell **32**, 327 (1983).

[26] J. Reinitz and J. R. Vaisnys, J. Theor. Biol. **145**, 295 (1990).
[27] A. A. Pakula, V. B. Young, and R. T. Sauer, Proc. Natl. Acad. Sci. U.S.A. **83**, 8829 (1986).
[28] A. D. Johnson *et al.*, Nature (London) **294**, 217 (1981).
[29] C. S. Shean and M. E. Gottesman, Cell **70**, 513 (1992).
[30] S. Ringquist *et al.*, Mol. Microbiol. **6**, 1219 (1992).
[31] D. Kennell and H. Riezman, J. Mol. Biol. **114**, 1 (1977).
[32] H. A. Kramers, Physica (Amsterdam) **7**, 284 (1940).
[33] P. Hänggi, P. Talkner, and M. Borkevic, Rev. Mod. Phys. **62**, 251 (1990).
[34] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems* (Springer-Verlag, New York, 1984).
[35] R. S. Maier and D. S. Stein, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **57**, 752 (1997).
[36] H. Eisen *et al.*, Gene **20**, 71 (1982); **20**, 83 (1982).
[37] B. Révet *et al.*, Curr. Biol. **9**, 151 (1999).
[38] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[39] S. K. Koblan and G. K. Ackers, Biochemistry **30**, 7817 (1991).
[40] S. K. Koblan and G. K. Ackers, Biochemistry **31**, 57 (1992).
[41] R. Maurer, B. J. Meyer, and M. Ptashne, J. Mol. Biol. **139**, 147 (1980).
[42] J. Little (private communication).