Epigenetics as a First Exit Problem

E. Aurell[1,2,3] and K. Sneppen[3]
[1]*SICS, Box 1263, SE-164 29 Kista, Sweden*
[2]*SANS/NADA/KTH, SE-100 44 Stockholm, Sweden*
[3]*NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*
(Received 3 March 2001; published 8 January 2002)

We develop a framework to discuss the stability of epigenetic states as first exit problems in dynamical systems with noise. We consider in particular the stability of the lysogenic state of the λ prophage. The formalism defines a quantitative measure of robustness of inherited states.

DOI: 10.1103/PhysRevLett.88.048101 PACS numbers: 87.16.–b

Epigenetics means inherited states in living systems, which are not encoded as genes, but as the (inherited) patterns of expressions of genes. Modulation of gene expression underlies a wide number of biological phenomena, from response to a changing nutrient supply in bacteria, to cell differentiation in multicellular organisms. Some of the simplest examples of inherited gene expression are found among bacteriophages, DNA viruses growing on bacterial hosts. The classical example is the lysogenic state of phage λ in *Escherichia coli* [1–3].

Upon infection of an *E. coli* cell, either λ enters a pathway leading to lysis, multiplying and killing the host, or it integrates its DNA into that of the host. In both cases different sets of genes are expressed. The latter, known as lysogeny, can be passively replicated for very long times. Indeed, the wild-type rate of spontaneous loss of λ lysogeny is only about 10^{-5} per cell and generation [3], a lifetime of the order of 5 yr. Moreover, this number is but a consequence of random activation of another part of the genetic system, the bacterial DNA repair response involving RecA. The intrinsic loss rate has in several independent experiments been found to be less than 10^{-7} per cell and generation [4–6]. The rate of mutations in the part of the lambda genome involved in lysogeny is between 10^{-6} and 10^{-7} per generation [5,6]. Epigenetics is therefore in this system actually more stable than the genome itself.

A stable state can be likened to a control switch that is on. For λ the analogy is quite direct [1,3]: lysogeny is maintained by regulatory proteins CI and Cro, and λ DNA, around an operator O_{R}, which consists of three binding sites $O_{\mathrm{R}}1$, $O_{\mathrm{R}}2$, and $O_{\mathrm{R}}3$, overlapping with two promoter sites P_{RM} and P_{R}; see Fig. 1. At the binding sites either CI or Cro can bind. CI has highest affinity for $O_{\mathrm{R}}1$, thus blocking RNA polymerase, the enzyme which catalyses the production of mRNA transcripts from DNA, from binding to promoter P_{R}, and initiating transcription of cro. Cro conversely binds primarily to $O_{\mathrm{R}}3$, thereby blocking promoter P_{RM} and subsequent production of CI. Finally, the rate of initiation of transcription of *cI* from P_{RM} depends on whether CI is bound at $O_{\mathrm{R}}2$. In lysogeny, CI production balances dilution from bacterial growth at a steady state with 200–350 CI per bacterial cell [3]. This is functionally a control switch, because if CI concentration becomes sufficiently low, increased activation of cro increases Cro concentration and decreases *cI* activation, so that lysogeny is ended and lysis follows.

The simplest mathematical model which embodies Fig. 1 is a set of coupled equations for the time rate of change of numbers of CI and Cro in a cell [7]:

$$\begin{aligned} \dot{N}_{\mathrm{CI}} &= \phi_{\mathrm{CI}}(N_{\mathrm{CI}}, N_{\mathrm{Cro}}), \\ \dot{N}_{\mathrm{Cro}} &= \phi_{\mathrm{Cro}}(N_{\mathrm{CI}}, N_{\mathrm{Cro}}), \end{aligned} \tag{1}$$

where the net production rates are

$$\begin{aligned} \phi_{\mathrm{CI}} &= S_{cI} f_{\mathrm{CI}}(N_{\mathrm{CI}}, N_{\mathrm{Cro}}) - N_{\mathrm{CI}}/\tau_{\mathrm{CI}}, \\ \phi_{\mathrm{Cro}} &= S_{\mathrm{cro}} f_{\mathrm{Cro}}(N_{\mathrm{CI}}, N_{\mathrm{Cro}}) - N_{\mathrm{Cro}}/\tau_{\mathrm{Cro}}. \end{aligned} \tag{2}$$

S_{cI} and S_{cro} the number of CI and Cro protein molecules produced from one mRNA transcript of the respective gene. The transcription rates f_{CI} and f_{Cro} are assumed the following functions of CI and Cro concentrations:

$$\begin{aligned} f_{\mathrm{CI}} &= R_{\mathrm{RM}}(P_{010} + P_{011} + P_{012}) + R^{u}_{\mathrm{RM}}(P_{000} + P_{001} + P_{002} + P_{020} + P_{021} + P_{022}), \\ f_{\mathrm{Cro}} &= R_{\mathrm{R}}(P_{000} + P_{100} + P_{200}), \end{aligned} \tag{3}$$

where P_s is the probability of a state s, encoding whether the operator sites in Fig. 1 are free, occupied by CI, or occupied by Cro. R^{u}_{RM} is the base rate of transcription of *cI*, R_{RM} is the stimulated rate if $O_{\mathrm{R}}2$ is occupied by a CI dimer, and R_{R} is the rate of cro transcription. We use a grand canonical formula

$$P_s = \mathcal{N}^{-1}[\mathrm{CI}]^{i_s}[\mathrm{Cro}]^{j_s} e^{-G(s)/RT}, \tag{4}$$

where [CI] and [Cro] are the concentrations of CI and Cro dimers in cytoplasm, and i_s and j_s are the numbers of these molecules bound to operator sites in state s. By dimerization balance we solve for the number of free dimers as a function of the total molecule numbers, using (4) to subtract the number of these molecules bound to operator sites, and association balance to subtract the numbers bound

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. CI dimers bind cooperatively to O_R1 and O_R2.

unspecifically to DNA [2]. We further use a formula analogous to (4) to compute and subtract the expected number of CI and Cro bound to a second operator site O_L, and finally assume that the bacterial and viral genome are, on the average, present in three copies in the cell [8]. The reference values for the binding free energies are shown in Table I, and the other numerical constants discussed above in the caption to that table. We note that f_{CI} has been directly measured in the absence of Cro [16], and agrees well with the parametric representation given by (3). The model is conveniently visualized by the phase space plot in Fig. 2.

TABLE I. Binding free energies of CI and Cro dimers to operator sites. The bindings at O_R are from [9–11]. Cro is assumed to bind without cooperativity, and, e.g., the binding energy G_{211} is taken to be $G_{200} + G_{011}$. The single-site binding energies at O_L are taken from [11], the cooperativities have been assumed the same as at O_R. The value of RT is 0.617 kcal/mol. The bacterial volume is taken $2.0 \times 10^{-15} l$ [12]. The reference value of S_{cI} is 1, which is in the range given by the comparison to lac*Z* in [13], while that for S_{cro} is 20, as deduced from [14] and [15]. Relative values of the rate constants R_{RM}, R^u_{RM}, and R_R have been reported in [16], and can be deduced from, e.g., the number of resulting CI molecules in lysogeny, as discussed in [6]. We here use $R_{RM} = 0.115$, $R^u_{RM} = 0.01045$, and $R_R = 0.30$, all three in units s^{-1}. The decay constants have been taken $\tau_{CI} = 2943$ s to match the generation time in the strains used in [5], and $\tau_{Cro} = 5194$ s, see [17]. The dimerization constants of CI and Cro have been taken -11.1 [18] and -7.0 kcal/mol [19], respectively. Cro dimers bind to unspecific DNA with -6.5 kcal/mol [9]

State	O_R free energy [kcal/mol]	O_L free energy [kcal/mol]
000	0.0	0.0
001	−12.5	−11.5
010	−10.5	−9.7
100	−9.5	−9.7
011	−25.7	−23.9
101	−22.0	−21.2
110	−22.9	−22.3
111	−35.4	−33.8
002	−14.4	−14.5
020	−13.1	−12.6
200	−15.5	−14.5

If the numbers of CI and Cro were macroscopically large, then (1) would be an entirely accurate description of the dynamics. The numbers are, however, only in the range of hundreds. The actual production process is influenced by many chance events, such as the time it takes for a CI or a Cro in solution to find a free operator site, or the time it takes a RNA polymerase molecule to find and attach itself to an available promoter. If in a time interval Δt the expected number of transcriptions is $f\Delta t$, then an actual realization has scatter $\sqrt{f\Delta t}$. As a minimal model of the switch with finite-N noise, we therefore consider the following system of two coupled stochastic differential equations, with two independent standard Wiener noise sources $(d\omega_t^{CI}, d\omega_t^{Cro})$:

$$\begin{aligned} dN_{CI} &= \phi_{CI} dt + g_{CI} d\omega_t^{CI}, \\ dN_{Cro} &= \phi_{Cro} dt + g_{Cro} d\omega_t^{Cro}. \end{aligned} \tag{5}$$

We assume that there is an equal amount of finite-N noise in decay as in production, and the two noise amplitudes are hence

$$\begin{aligned} g_{CI} &= \sqrt{S_{cI}^2 f_{CI} + N_{CI}/\tau_{CI}}\,, \\ g_{Cro} &= \sqrt{S_{cro}^2 f_{Cro} + N_{Cro}/\tau_{Cro}}\,. \end{aligned} \tag{6}$$

The problem of escape from a stable equilibrium point like S under a dynamics like (5) is a first-exit problem in the theory of stochastic processes [20]. The probability of a given realization of the noise in time $[0, T]$ is

$$\begin{aligned} &\mathrm{Prob}(\{\omega_t^{CI}, \omega_t^{Cro}\}_0^T) \\ &\propto \exp\left[-\frac{1}{2} \int_0^T (\dot\omega_t^{CI})^2 + (\dot\omega_t^{Cro})^2 \, dt \right] \\ &= \exp\left[-\int_0^T \frac{(\dot N_{CI} - \phi_{CI})^2}{\Gamma_{CI}} + \frac{(\dot N_{Cro} - \phi_{Cro})^2}{\Gamma_{Cro}} \, dt \right], \end{aligned} \tag{7}$$

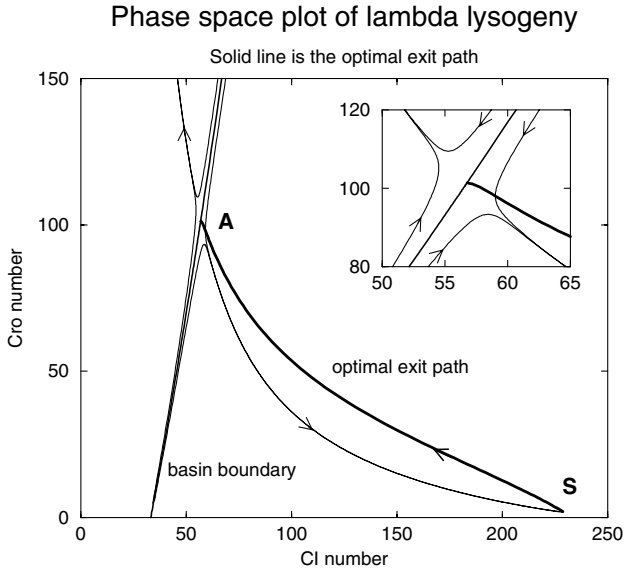

FIG. 2. The phase space plot of the dynamical system (1), and the optimal exit path in the stochastic dynamical system (5).

where we have introduced the diagonal elements of the diffusion matrix, $\Gamma_{\rm CI} = g_{\rm CI}^2$ and $\Gamma_{\rm Cro} = g_{\rm Cro}^2$.

Of all the realizations that move the system from S to A, the most probable is the one that minimizes the action functional

$$\mathcal{A} = \frac{1}{2}\int_{0^T}^{T}\left[\frac{(\dot{N}_{\rm CI} - \phi_{\rm CI})^2}{\Gamma_{\rm CI}} + \frac{(\dot{N}_{\rm Cro} - \phi_{\rm Cro})^2}{\Gamma_{\rm Cro}}\right]dt\,, \tag{8}$$

where the initial position is S, the final position A, and the minimization is taken over all paths that go from S to A in time T. If $\mathcal{A} \gg 1$ it can be proved, see [21], that the most probable exit point from the basin of attraction of S is indeed A, and the rate of exit is

$$\text{Rate(exit)} \propto \exp(-\mathcal{A}^{\rm min})\,. \tag{9}$$

The leading correction to (9) is the appropriate fluctuation determinant, which has dimension one over time. In our case it sets a scale of the order of once per bacterial generation.

Since the Lagrangian in (8) is not explicitly time dependent, the Hamiltonian

$$\mathcal{H} = \frac{1}{2}(\Gamma_{\rm CI}p_{\rm CI}^2 + \Gamma_{\rm Cro}p_{\rm Cro}^2) + p_{\rm CI}\phi_{\rm CI} + p_{\rm Cro}\phi_{\rm Cro} \tag{10}$$

is conserved along the path. The momenta $(p_{\rm CI}, p_{\rm Cro})$ are conjugate to the generalized coordinates $(N_{\rm CI}, N_{\rm Cro})$, and the path is given by a solution of Hamilton's equations $\dot{\vec{N}} = \frac{\partial\mathcal{H}}{\partial\vec{p}}$ and $\dot{\vec{p}} = -\frac{\partial\mathcal{H}}{\partial\vec{N}}$. We note that the energy of the optimal exit path, the value of $\mathcal{H}$ in the auxiliary mechanical system, must be non-negative, since the drift field vanishes at the two end points. On the other hand, we have in general $\partial\mathcal{A}/\partial T = -E$, where E is the energy and T is the transit time. It hence follows that the optimal exit path is a zero-energy path from S to A under the Hamiltonian in (10).

Conceptually, the general problem of exit from an equilibrium through a basin boundary is similar to thermal exit from a potential well, see [22,23], with (9) playing the role of the Arrhenius factor. Indeed, the integrand in Eq. (8) can be rewritten $\frac{1}{2}(\dot{\vec{N}} + \vec{\phi})\cdot\Gamma^{-1}(\dot{\vec{N}} + \vec{\phi}) - 2\dot{\vec{N}}\cdot(\Gamma^{-1}\vec{\phi})$. If $\Gamma^{-1}\vec{\phi} = -\vec{\nabla}V$ for some V, the second term gives the same value for all paths, and therefore drops out of the minimization, while the quadratic term can be decreased to zero [20]. For a recent derivation in the context noise in switches, see [24]. The action is hence then always minimized by $\dot{\vec{N}} = -\vec{\phi}$, i.e., a path going against the drift field, and equal to twice the difference in V. In the general case no similar analytic prescription exists, and the optimal path must be computed by numerical procedure, see below.

An important feature of the general exit problem is, however, that the action (9) is only defined locally around each equilibrium, that is, from S the bottom of the effective well, up to A, the saddle point. One can therefore, in contrast to the potential case, have a series of minima, $S_1, S_2, \ldots, S_N$, such that the system jumps preferentially between them in a definite order, e.g., as $S_1 \to S_2 \to \cdots \to S_N \to S_1 \to \cdots$. This might be a possible model of cyclic processes, for instance the cell cycle with checkpoints [25], as recently remarked in [24].

In a general exit problem the optimal path can be computed by the following numerical procedure, using the relaxation method of computing solutions to 2-point boundary problems in an ODE [26,27]. We first find a natural parameter in the system, and vary that to get close to a bifurcation where the stable and unstable equilibria (S and A) coalesce. The diffusion matrix Γ is then practically constant in a neighborhood around both points, while the drift field is small. Hence, we can compute a path between the two points at high energy starting from a straight path at constant speed. The energy and the parameters are then changed incrementally, using the previous solution as the trial solution. The zero-energy paths in the intermediate neighborhoods of the two points always need to be taken care of by a local calculation [21]. The optimal path is shown in Fig. 2.

There is an emerging consensus in molecular biology and biological physics that chemical networks in living cells have to be robust [28,29]. For the λ phage, robustness of lysogeny has been experimentally established for several large modifications of the $O_{\rm R}$ complex [5]. The present work allows us to quantify robustness of epigenetic states. A state only exists at all if deterministic equations like (1) have a stable equilibrium with the corresponding properties. This state is stable for long times, even if the number of molecules involved is small, if the action $\mathcal{A}$ (9) is much larger than unity. The state is finally robust with respect to variation of a parameter, if the state still exists and is stable after the variation.

In Figs. 3(a) and 3(b) we examine lysogenic stability as function of one parameter, the binding of Cro to $O_{\rm R}3$. If we first disregard the noise, we see that a change of

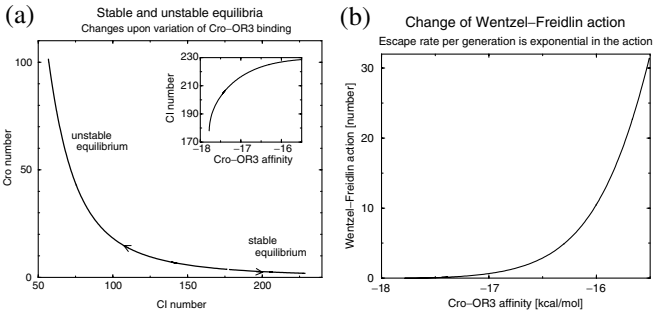

FIG. 3. Systemic changes due to changes in affinity of Cro to operator site $O_{\rm R}3$. The standard value is -15.5 kcal/mol. Stronger binding energies are investigated for use in numerical procedure (see main text), and to explore robustness of lysogeny to parameter changes. (a) Location of stable equilibrium (S) and unstable equilibrium (A) as affinity is varied. (insert) CI number in lysogenic equilibrium as function of affinity (b) Wentzel-Freidlin action as a function of affinity.

affinity by 2.25 kcal/mol brings the stable and unstable equilibria together, such that the lysogenic state disappears altogether. We also observe a sensitive dependence of the position of the unstable equilibrium, while the number of CI in the stable equilibrium (lysogenic state) only changes by 30%. The lysogenic state therefore looks qualitatively similar over this range of parameters. These are features of the model embodied by Eq. (1) only. If we then bring in our model of the noise, Eq. (5), we see that the action $\mathcal{A}$ changes from more than 30 to less than 3 when affinity changes by 1 kcal/mol, the approximate change of binding energy under a single point mutation. Such a change hence suffices to destabilize the switch over biologically relevant time scales. The model is therefore not robust to such changes, in contradiction to recent experimental data [5], which implies the presence of some additional mechanism, in order for robustness to prevail.

In conclusion, we have examined the general problem of escape from a stable equilibrium in more than one dimension, and demonstrated how this determines the stability of states of genetic networks. In contrast to Kramers' escape from a potential well, the stability of inherited states in such networks is not a mathematically, or computationally, trivial problem. Indeed, in an earlier separate contribution we studied lysogenic stability in phage λ in a purely numerical model [6]. The semianalytic model introduced here, where the most likely exit path is computed from an auxiliary classical mechanical system, has the advantages of being conceptually simpler, of extending standard theoretical tools in condensed matter physics, and of facilitating investigations of robustness properties. We note, however, that a more detailed modeling of the noise, as in [6], may be important for phenomena on shorter time scales, as in the lysis/lysogeny entry decision [30]. The overall lesson of this study is that an examination of equilibria and their bifurcations with changing parameter values allows us to quantify both the stability and the robustness of states of a genetic control system.

We thank G. Dahlquist, B. Altshuler, S. Brown, and P. Muratore-Ginanneschi for discussions. E. A. thanks the Swedish Natural Research Council for support under Grant No. M-AA/FU/MA 01778-333. K. S. thanks the ITP (U.C. Santa Barbara) for hospitality, and consequently the National Science Foundation for financial support under Grant No. PHY99-07949.

[1] A. D. Johnson, A. R. Poteete, G. Lauer, R. T. Sauer, G. K. Ackers, and M. Ptashne, Nature (London) **294**, 217–223 (1981).
[2] M. A. Shea and G. K. Ackers, J. Mol. Biol. **181**, 211–230 (1985).
[3] M. Ptashne, *A Genetic Switch: Phage λ and Higher Organisms* (Blackwell Scientific Publications and Cell Press, Cambridge, MA, 1992).
[4] D. V. Rozanov, R. D'Ari, and S. P. Sineoky, J. Bacteriol. **180**, 6306–6315 (1998).
[5] J. Little, D. P. Shepley, and D. W. Wert, EMBO J. **18**, 4299–4307 (1999).
[6] E. Aurell, S. Brown, J. Johansson, and K. Sneppen, cond-mat/0010286 [Biophys. J. (to be published)].
[7] J. Reinitz and J. R. Vaisnys, J. Theor. Biol. **145**, 295–318 (1990).
[8] H. Bremmer and P. P. Dennis, in *Escherichia coli and Salmonella,* edited by F. C. Neidhardt (ASM Press, Washington, DC, 1996), pp. 1553–1569.
[9] J. G. Kim, Y. Takeda, B. W. Matthews, and W. F. Anderson, J. Mol. B **196**, 149–158 (1987).
[10] S. K. Koblan and G. K. Ackers, Biochemistry **31**, 57–65 (1992).
[11] Y. Takeda, P. D. Ross, and C. P. Mudd, Proc. Natl. Acad. Sci. U.S.A. **89**, 8180–8184 (1992).
[12] F. C. Neidhardt *et al.,* in *Escherichia coli and Salmonella* (Ref. [8]), pp. 13–16.
[13] C. S. Shean and M. E. Gottesman, Cell **70**, 513–522 (1992).
[14] S. Ringquist, S. Shinedling, D. Barrick, L. Green, J. Binkley, G. D. Stormo, and L. Gold, Mol. Microbiol. **6**, 1219–1229 (1992).
[15] D. Kennell and H. Riezman, J. Mol. B. **114**, 1–21 (1977).
[16] D. K. Hawley and W. R. McClure, J. Mol. Biol. **157**, 493–525 (1982).
[17] A. A. Pakula, V. B. Young, and R. T. Sauer, Proc. Natl. Acad. Sci. U.S.A. **83**, 8829–8833 (1986).
[18] S. K. Koblan and G. K. Ackers, Biochemistry **30**, 7817–7821 (1991).
[19] R. Jana, T. R. Hazbun, A. K. M. M. Mollah, and M. C. Mossing, J. Mol. Biol. **273**, 402–416 (1997).
[20] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems* (Springer-Verlag, New York, 1984).
[21] R. S. Maier and D. S. Stein, SIAM J. Appl. Math. **57**, 752–790 (1997).
[22] H. A. Kramers, Physica (Utrecht) **7**, 284–304 (1940).
[23] P. Hänggi, P. Talkner, and M. Borkevic, Rev. Mod. Phys. **62**, 251–341 (1990).
[24] W. Bialek, cond-mat/0005235.
[25] A. W. Murray, Nature (London) **359**, 599–604 (1992).
[26] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipies* (Cambridge University Press, Cambridge, U.K., 1988).
[27] P. P. Eggleton, Mon. Not. R. Astron. Soc. **151**, 351 (1971).
[28] M. A. Savageau, Nature (London) **229**, 855–857 (1971).
[29] U. Alon, M. G. Surette, N. Barkai, and S. Leibler, Nature (London) **397**, 168–171 (1999).
[30] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633–1648 (1998).