Cooperativity leads to temporally-correlated fluctuations in the bacteriophage lambda genetic switch

Jacob Q. Shenker[1] and Milo M. Lin[2]*

[1] *Department of Physics, California Institute of Technology, Pasadena, CA, USA,* [2] *Pitzer Center for Theoretical Chemistry, University of California, Berkeley, CA, USA*

Cooperative interactions are widespread in biochemical networks, providing the nonlinear response that underlies behavior such as ultrasensitivity and robust switching. We introduce a temporal correlation function — the *conditional activity* — to study the behavior of these phenomena. Applying it to the bistable genetic switch in bacteriophage lambda, we find that cooperative binding between binding sites on the prophage DNA lead to non-Markovian behavior, as quantified by the conditional activity. Previously, the conditional activity has been used to predict allosteric pathways in proteins; here, we show that it identifies the rare unbinding events which underlie induction from lysogeny to lysis.

Keywords: gene regulatory networks, markov state models, phage lambda, mutual information, information theory, conditional activity

1. Introduction

Cells use biochemical networks to sense, process information, and respond to their environments. Many cellular behaviors have been found to be controlled by genetic switches, in which the expression levels of a set of genes form a stable memory of a transient stimulus, allowing the cell to make a decision and remember it. These networks range from a simple bistable switch to complicated networks involving dozens of genes and many stable states of the switch (fixed points). In bacteria and viruses, these switches enable phenotypic switching to optimize the fitness of the organism in response to environmental conditions. In temperate bacteriophages, a switch selects between dormancy and virulence, and in *Escherichia coli*, bistable switches are known to regulate the lactose (Ozbudak et al., 2004) and arabinose utilization systems (Fritz et al., 2014). More complex gene regulatory networks are used in plants and animals to coordinate development and determine cell fates. In *Arabidopsis thaliana*, a fifteen-gene network was identified whose fixed points correspond to the ten flower cell types (Espinosa-Soto et al., 2004). This is typical of many gene regulatory networks, where the phenotype corresponds not to expression of any single gene, but rather to the collective state of the system. These systems are challenging to study theoretically and experimentally, since the effects of a given gene on phenotype is difficult to disentangle from the behavior of the other genes in the network. We propose a theoretical technique which identifies the functional connectivity between different elements in a biochemical network. This map elucidates the often non-intuitive connection between genotype and phenotype in these networks, and may be used to design experimental interventions which most effectively modify or disrupt this collective behavior, and hence most directly affect phenotype.

The multistability which characterizes these switch-like gene networks often results from the interplay of feedback and ultra-sensitivity. These dynamics arise from individual binding interactions between the nucleic acids, proteins, and small ligands that comprise these networks. Information propagates between these species through macromolecular complexes containing multiple binding sites. The presence of a ligand at one binding site modulates the activity or binding strength of another binding site through cooperativity or conformational change (allostery). Allosteric regulation has been widely and successfully described by the Monod-Wyman-Changeux (MWC) model (Monod et al., 1963), while cooperativity and other forms of indirect regulation have been described by generalizations thereof (Marzen et al., 2013). These models describe the thermodynamic equilibrium of these systems. For the two-site MWC model, cooperative binding energies can be fit to experimental binding curves. However, the number of possible cooperative interactions increases rapidly with the number of binding sites, and for complex systems these interactions are often too numerous to constrain experimentally. In the cases for which it is not possible to fit a model, thermodynamic correlations within the system can still be inferred by observing the system and calculating the mutual information between its binding sites. However, for systems stabilized by cooperative binding, important behavior may be invisible to thermodynamic correlations and evident only from the timing of binding events.

Inspired by the physics of glasses, we introduce a new quantity, the conditional activity, to measure temporal correlations of binding activity in a biochemical system. The conditional activity can be calculated from direct experimental measurements or from a stochastic model. Recently, we used the conditional activity to measure temporal correlations between different regions of proteins as calculated from molecular dynamics simulations (Lin, under review). For these proteins, the conditional activity correctly distinguished functional modules and identified the allosteric connections between sites within a protein. This type of intramolecular communication was not detectable using mutual information or other equilibrium-based correlation functions.

In this contribution, we go beyond the molecular scale and apply the conditional activity to study temporally-correlated binding activity in a gene regulatory network. In these networks, complex system-level behaviors—such as adaptation, switching, and oscillation—arise from macromolecular binding of transcription factors to DNA and the resulting modulation of gene transcription, and this binding is often subject to cooperativity and competition between transcription factors. We consider the genetic switch in *E. coli* infected with bacteriophage lambda, whose bistability emerges from cooperativity between six binding sites, including long-range DNA looping. The switch reliably maintains its initial pathway—lysogeny—until it is flipped into another pathway—lysis—by an external trigger (the bacterium's SOS response).

Lambda, a bacterial virus, infects an *E. coli* cell, and depending on the environment inside the *E. coli* proceeds along one of two pathways (**Figure 1** top). In the more-common lytic pathway, the virus DNA enters free-floating into the bacterium. The host bacterium's own machinery replicates and synthesizes protein from the viral DNA, which in turn self-assemble into new bacteriophages. When approximately one hundred progeny phage have been produced, the viral DNA produces proteins that rupture the bacterial cell and the newly-made phages are released into the environment. In some cases, however, the viral DNA integrates itself into the host DNA and lies dormant—this is the lysogenic pathway. Once there, it is replicated along with the host DNA and is passed on to daughter cells upon division. The dormant state is exceedingly stable, and the virus may lie dormant for hundreds of millions of generations without activation (Little and Michalowski, 2010). However, the bacterium's SOS response, triggered by DNA damage, reliably induces the activation of the virus genes, and the virus switches over to the lysis pathway, replicating itself before it lyses its host.
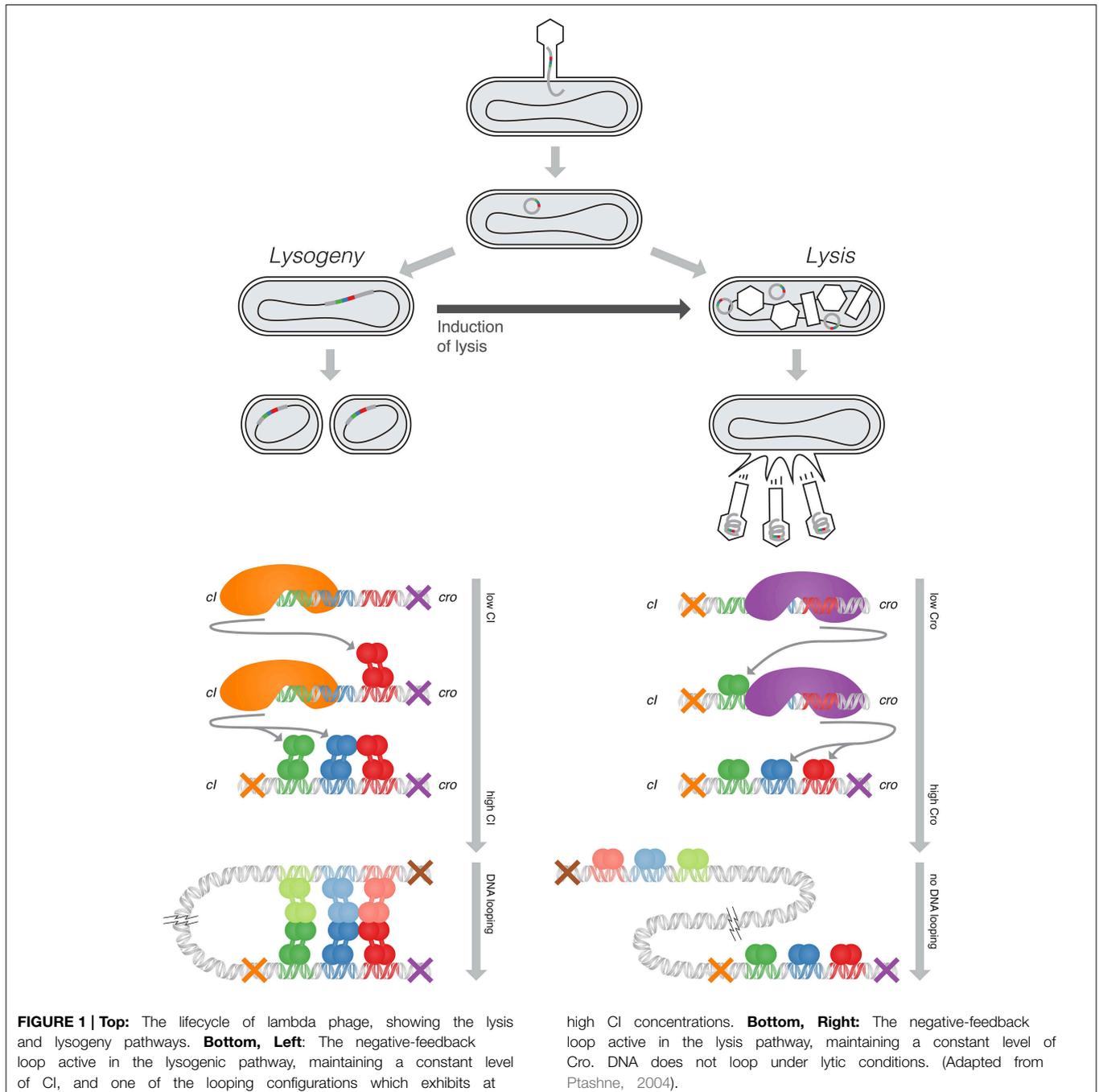
Upon infecting its host bacterium, the lambda phage chooses one of these two pathways. This choice is maintained by a bistable genetic switch composed of the *cI* gene, the *cro* gene, and an operator that modulates the rate of their transcription. The CI and Cro proteins, upon dimerizing, become transcription factors which bind to this operator region. The lysogenic pathway is characterized by expression of *cI* and inactivation of *cro*, whereas the lytic pathway is initiated by expression of *cro*, which in turn inactivates *cI*. CI and Cro each negatively autoregulate, maintaining stable levels of protein when their gene is turned on (**Figure 1** bottom). The cooperativity and competition between the binding of CI and Cro to the operator couples these negative-feedback loops, resulting in switch-like behavior.

We analyze this system looking for correlations in binding state between different binding sites, using the mutual information, and correlations in binding times (i.e., non-Markovian behavior), using the conditional activity. Since cooperative binding leads to stability, we find that the mutual information fails to describe any nontrivial relationships between binding sites because mutual information is small for stable systems. The conditional activity, however, is sensitive to rare collective fluctuations of the system and uses this information to infer a rich network of interactions between binding sites.

2. Materials and Methods

2.1. Model

We adopt a standard model for the phage lambda switch (Ackers et al., 1982; Shea and Ackers, 1985; Santillán and Mackey, 2004; Tian and Burrage, 2004; Gedeon et al., 2008). The region of the lambda phage DNA relevant to this switch contains the right operator with three binding sites (O_{R1}, O_{R2}, O_{R3}) sandwiched between two promoter regions (P_R and P_{RM}, governing transcription of *cI* and *cro*, respectively), and the left operator with three binding sites (O_{L1}, O_{L2}, O_{L3}) and a promoter region (P_L, whose function is immaterial here). Each operator site can be bound by a CI dimer, a Cro dimer, or left unbound. Each promoter region may be bound by RNA polymerase (RNAP) or left unbound. The promoter regions physically overlap the operator: RNAP bound to P_R blocks anything from binding to O_{R3}; RNAP bound to P_{RM} blocks binding at both O_{R1} and O_{R2}; and RNAP bound to P_L block binding at O_{L1} and O_{L2}. All possible binding configurations of these nine sites give 1200 microstates.

FIGURE 1 | Top: The lifecycle of lambda phage, showing the lysis and lysogeny pathways. **Bottom, Left**: The negative-feedback loop active in the lysogenic pathway, maintaining a constant level of CI, and one of the looping configurations which exhibits at high CI concentrations. **Bottom, Right:** The negative-feedback loop active in the lysis pathway, maintaining a constant level of Cro. DNA does not loop under lytic conditions. (Adapted from Ptashne, 2004).

The energies of binding and cooperativity of CI, Cro, and RNAP have been measured experimentally (Reinitz and Vaisnys, 1990; Darling et al., 2000a,b), and from these we may calculate the free energies ΔG_i of each microstate i. The probabilities that the system will be found in a microstate i is given by

$$K_i = \frac{1}{Z} \exp(-\Delta G_i/RT)[CI_2]^{\alpha_i}[Cro_2]^{\beta_i}[RNAP]^{\gamma_i} \quad (1)$$

where α_i, β_i, γ_i are the numbers of bound CI, Cro, and RNAP in each microstate i, $[x]$ indicates the concentration of x, R is the universal gas constant and $T = 310\ K$ is the temperature. The partition function Z is a normalization chosen to ensure that the system lies in one of the microstates at all times, i.e., that the probabilities sum to one: $\sum_{i=1}^{1200} K_i = 1$.

As we show below, this model exhibits behavior under lysogenic conditions that is qualitatively different from that under lytic conditions (see Results). This difference arises because lysogeny and lysis occur at different concentrations of CI and Cro, yielding different equilibrium probabilities K_i. The concentrations of CI, Cro, and RNAP that correspond to the lysogenic

and lytic pathways are obtained using the model and parameters from Santillán and Mackey (2004) (see Supplementary Material for details and parameter values). We choose a repressor degredation rate $\gamma_{cI} = 0.015$ min^{-1}, which lies in the range consistent with bistability.

Note that CI bound to O_{R2} increases the transcription rate of CI elevenfold when RNAP is bound to P_{RM}. This cooperativity between CI and RNAP is a key interaction in phage lambda. However, our analysis encompasses only the behavior of the system on the fast timescale of macromolecular binding and unbinding, and not the far slower timescale of protein production. As such, in this work we consider cooperativity only in the sense of cooperative binding. Furthermore, because our model does not model the dynamics of protein production and degradation, our model cannot transition between lysogenic and lytic pathways by varying the protein concentrations, as the protein concentrations are held fixed at either lysogenic or lytic conditions. However, we examine the tendency for the system to switch from lysogeny to lysis by considering the binding of RNAP to P_R. The more often P_R is bound by RNAP, the faster Cro is being produced, pushing the system toward lysis.

The probabilities K_i describe the occupancies of a system in thermodynamic equilibrium. For each of the two pathways, we now construct a continuous-time Markov chain which adds kinetic information—namely, the rates of transitions between microstates—to this thermodynamic model. In these Markov models, each transition corresponds to a single binding or unbinding event. We consider the substitution of one ligand for another as two distinct transitions: an unbinding of the first ligand followed by a binding of the second ligand. Since the binding and unbinding rates have not been directly determined experimentally, we infer approximate kinetics from the thermodynamics. Assuming detailed balance fixes the ratio between the forward (q_{ij}) and backwards (q_{ji}) rates for each transition, $q_{ij}/q_{ji} = K_j/K_i$. We are free to choose the q_{ij} so that binding events have the same rate D:

$$q_{ij} = \begin{cases} D & i \to j \text{ is a binding transition} \\ D\frac{K_j}{K_i} & i \to j \text{ is an unbinding transition} \end{cases} \quad (2)$$

Here we have assumed detailed balance and equal binding rates D to derive the kinetics q_{ij} from the equilibrium occupancies K_i. This approximation is reasonable because all binding rates represent the physical process of diffusion-limited binding of macromolecules to DNA. Furthermore, not only was our choice of rates the simplest and most natural consistent with detailed balance, but we also have found that our results are highly insensitive to large deviations in these rates.

In **Figure 2** we show the Markov model by illustrating each microstate and the transitions between them, where the size of each microstate (gray circles) is proportional to the logarithm of its occupancy probability K_i and the width of each arrow is proportional to the logarithms of the flux ($K_i q_{ij}$) between two microstates. Even with this simple model of nine binding sites and three species of ligands, the resulting graph is complex and the functionality of the system is not evident: from the graph, it is difficult to discern how different binding sites couple to one
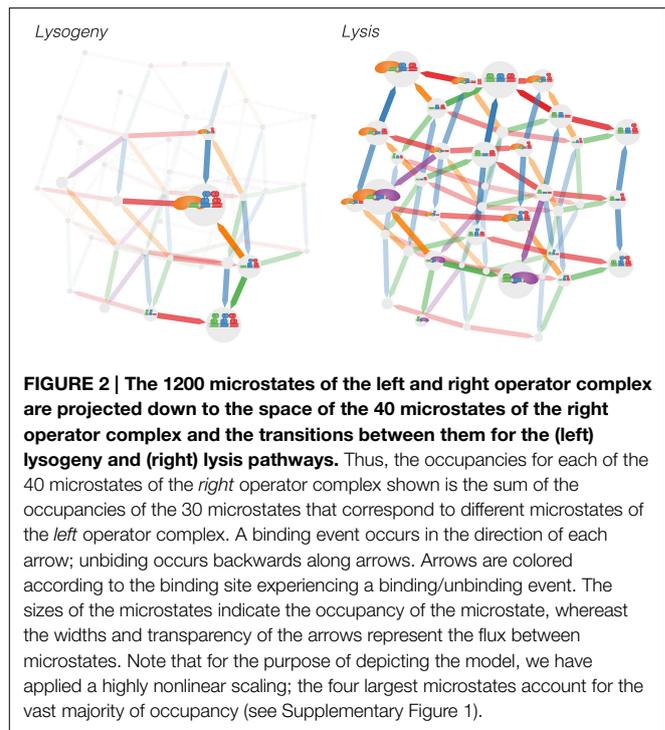

FIGURE 2 | The 1200 microstates of the left and right operator complex are projected down to the space of the 40 microstates of the right operator complex and the transitions between them for the (left) lysogeny and (right) lysis pathways. Thus, the occupancies for each of the 40 microstates of the *right* operator complex shown is the sum of the occupancies of the 30 microstates that correspond to different microstates of the *left* operator complex. A binding event occurs in the direction of each arrow; unbinding occurs backwards along arrows. Arrows are colored according to the binding site experiencing a binding/unbinding event. The sizes of the microstates indicate the occupancy of the microstate, whereast the widths and transparency of the arrows represent the flux between microstates. Note that for the purpose of depicting the model, we have applied a highly nonlinear scaling; the four largest microstates account for the vast majority of occupancy (see Supplementary Figure 1).

another. To learn about the connections between binding sites, we may use a correlation function to project the connections between the 1200 microstates into the smaller space of nine binding sites. Note that the transitions between the 1200 microstates are Markovian; the non-Markovian dynamics arise when the model is projected into the space of the nine binding sites. In most experimental contexts it is not possible to observe all binding sites simultaneously, so the binding dynamics of the subset of binding sites which is observed appears non-Markovian. The conditional activity uses these non-Markovian dynamics to infer functional connectivity between the binding sites.

We now define two correlation functions with the aim of demonstrating that they capture orthogonal and complimentary information about a molecular system. In Section 2.2, we review the mutual information, a widely used correlation function which is useful for characterizing the dominant behavior of a system. We then define the conditional activity, which we introduced in Lin (under review), and is sensitive to fluctuations away from the dominant behavior. In Section 2.3, we demonstrate how the conditional activity may be directly calculated from the transition rate matrix for a Markov model of a molecular system; these correlation functions can also be calculated from simulating the Markov model, as described in Section 2.4.

2.2. Correlation Functions
The mutual information is commonly used to infer connectivity in networks (Margolin et al., 2006). Here we use it to deduce the coupling between different binding sites. The mutual information (Shannon and Weaver, 1949; Cover and Thomas, 2006) between

two binding sites a, b is defined as

$$
\begin{aligned}
MI(a, b) = & \sum_{x \in X} P(a = x) \log_2 P(a = x) \\
& + \sum_{y \in Y} P(b = y) \log_2 P(b = y) \quad (3) \\
& - \sum_{x \in X} \sum_{y \in Y} P(a = x, b = y) \log_2 P(a = x, b = y)
\end{aligned}
$$

where site a takes on binding configurations X, site b takes on binding configurations Y, and P is the probability function. Note that P can be written in terms of the K_i: $P(a = x, b = y) = \sum_{i=1}^{1200} K_i \chi_{a,x}^i \chi_{b,y}^i$, where $\chi_{a,x}^i$ is 1 if site a is in binding configuration x in system microstate i and 0 otherwise. We see that the mutual information is entirely a thermodynamic quantity, since it depends only on the equilibrium occupation probabilities K_i.

The conditional activity is defined for systems with degrees-of-freedom that take on discrete configurations (or may be sensibly discretized), and captures correlations in the times between the times at which different degrees-of-freedom transition between configurations. As such, the CA is sensitive to the kinetics of the system, not only the thermodynamics. In this case, the degrees-of-freedom are the different binding sites, and each binding/unbinding event is considered a transition. Let $T(a, t)$ be the t-th transition time of binding site a and $\tau \equiv T(a, N(a))$ the duration of the observation period, with $N(i) \gg 1$ being the number of recorded transitions. $W(a, T)$ is the time between time T and the next transition of a after T [e.g., $W(a, T(a, t)) = T(a, t + 1) - T(a, t)$, the time between a given binding event and the next unbinding event, or the time between a given unbinding event and the next binding event].

We define the mean observed persistence time to be half the mean squared waiting time between transitions for a given binding site a:

$$
\tau_p[a] \equiv \frac{1}{2\tau} \sum_{t=1}^{N(a)} W(a, T(a, t))^2 \quad (4)
$$

For each transition time of binding site b, there is a previous transition of b and a next transition of a. These three times define two adjacent time periods, and the mean observed exchange time for a following b is:

$$
\tau_x[a \leftarrow b] \equiv \frac{1}{\tau} \sum_{t=1}^{N(b)-1} W(a, T(b, t + 1)) W(b, T(b, t)) \quad (5)
$$

The persistence time $\tau_p[a]$ and exchange time $\tau_x[a \leftarrow a]$ correspond to standard notions in glassy physics (Jung et al., 2005; Hedges et al., 2007); here we extend the exchange time to the case when $a \neq b$. We now define the conditional activity of a following b:

$$
CA[a \leftarrow b] \equiv -\log_{10} \frac{\tau_x[a \leftarrow b]}{\tau_p[a]} \quad (6)
$$

If a and b are independent, then $CA[a \leftarrow b] = 0$; if the transitions of a are Markovian, $CA[a \leftarrow a] = 0$. Note that the conditional activity, unlike the mutual information, is asymmetric: in general, $CA[a \leftarrow b] \neq CA[b \leftarrow a]$, because the conditional activity depends on the time-ordering of transitions.

For a real system, the occupancy probabilities K_i may be measured experimentally; for the Markov model, they are given in Equation (1). From these, the mutual information may be calculated using Equation (3). Similarly, using Equation (6), the conditional activity may be obtained from a time-series of configuration changes for each binding site. This time-series may be obtained experimentally for a real system. For the Markov model, we may generate this time-series data by simulating the continuous-time Markov chain using the Gillespie algorithm (Gillespie, 1977). However, the conditional activity is sensitive to rare events, so a long simulation is required to accurately calculate the conditional activity. For our simple Markov model, this simulation is still very tractable, but would pose a problem for more complicated models. We now use techniques of Markov chain theory to express the conditional activity directly in terms of the transition rates q_{ij} given above in Equation (2).

2.3. Direct Calculation of Conditional Activity From the Transition Rates

Given the transition rates q_{ij}, for each binding site a define the transition rate matrix Q^a where transitions that change the binding configuration of site a are disallowed:

$$
(Q^a)_{ij} = \begin{cases} q_{ij} & i \neq j \text{ and microstates } i, j \text{ have site } a \text{ in} \\ & \quad \text{the same binding configuration} \\ 0 & i \neq j \text{ and microstates } i, j \text{ have site } a \text{ in} \\ & \quad \text{different binding configurations} \\ -\sum_{k \neq i} q_{ik} & i = j \end{cases}
$$

The state space of all 1200 microstates may partitioned into subspaces, where each subspace is a connected component of the transition graph where transitions that change the binding configuration of site a are disallowed. The Q^a matrices are block diagonal (under a relabeling of the indices), and each block represents the transitions within a given subspace. In the language of Markov chain theory, each of these blocks represents a Markov chain where the transitions that change the binding configuration of site a have been made absorbing. Using techniques for analyzing absorbing Markov chains (Kemeny and Snell, 1961; Tavare, 1979; Syski, 1992), we may express the exchange and persistence times in terms of the fundamental matrix for a regular Markov chain $N^a = -(Q^a)^{-1}$.

We now define

$$
B^a = N^a R^a \quad (7)
$$

$$
K_i^a = \frac{\sum_j K_j (R^a)_{ji}}{\sum_j \sum_k K_j (R^a)_{jk}} \quad (8)
$$

where the matrices R^a encode the transitions between subspaces:

$$
(R^a)_{ij} = \begin{cases} q_{ij} & i \neq j \text{ and microstates } i, j \text{ have site } a \text{ in different} \\ & \quad \text{binding configurations} \\ 0 & \text{otherwise.} \end{cases}
$$

$(B^a)_{ij}$ is the probability that j is the first microstate in which the binding configuration of site a has changed, given that the system has started in microstate i. K_i^a represents the probability of finding the system in microstate i just after the binding configuration of a has changed.

We may now define the persistence time

$$\tau_p[a] = \frac{1}{\tau^a} \sum_i \sum_j \sum_k K_i^a (N^a)_{ij} (N^a)_{jk} \quad (9)$$

and exchange time

$$\tau_x[a \leftarrow b] = \frac{1}{\tau^b} \sum_i \sum_j \sum_k \sum_l K_i^b (N^b)_{ij} (B^b)_{jk} K_k^a (N^a)_{kl} \quad (10)$$

where $\tau^a = \sum_i \sum_j K_i^a (N^a)_{ij}$ is a normalization. The conditional activity then follows from Equation (6).

(See the Supplemental Material for a derivation).

2.4. Simulation

The conditional activity and mutual information may be directly calculated from the Markov model, as described above, or from simulated or experimental time-series data. For a reduced five binding site model, the analytic calculations presented above agreed with the same quantities calculated from simulated time-series data (Supplementary Figure 3) simulated using the Gillespie algorithm (Gillespie, 1977).

3. Results

The lysogenic and lytic pathways exhibit differing levels of fluctuations, as can be seen qualitatively in **Figure 2**. In the lysogenic pathway, the system is stabilized by long-range DNA looping—cooperativity between the left and right operators—and cooperative interactions between adjacent sites on each operator (Santillán and Mackey, 2004; Anderson and Yang, 2008; Zurla et al., 2009; Norregaard et al., 2013). With more than 99% probability the system exists in one of three microstates exhibiting an octameric configuration of CI bound to O_{R1}, O_{R2}, O_{L1}, and O_{L2} (Supplementary Figure 1).

This high degree of stability is quantified by the entropies of each binding site (**Figure 3** top left). Most binding sites are fixed in one microstate with high probability. Only the P_{RM}, O_{R3}, and O_{L3} sites are fluctuating, as shown by their non-negligible entropy. The high mutual information between P_{RM} and O_{R3} represents the near perfect anticorrelation between the two sites: one is almost always bound and no more than one may be bound at a time because they physically block each other.

Similarly, in the lytic pathway, the mutual information (**Figure 3** top right) clearly indicates that P_R overlaps with, and hence is anticorrelated with, both O_{R1} and O_{R2}; P_{RM} overlaps with O_{R3}; and P_L overlaps with O_{L1} and O_{L2}. The mutual information between O_{R1} and O_{R2}, and that between O_{L1} and O_{L2}, arise because when those sites are unbound by their overlapping RNAP, the high concentration of Cro leads to both sites being bound simultaneously with Cro. Note that this simultaneous binding is highly correlated, but does not represent any

cooperative binding effects: there is cooperativity between binding at O_{R1} and O_{R2}, and between O_{R2} and O_{R3}, yet there is only mutual information between the former pair, that which overlaps P_R. Similarly, there is cooperativity between O_{L1} and O_{L2}, and between O_{L2} and O_{L3}, and again we only see mutual information between the former pair, that which overlaps P_L. This indicates that correlations indicated by high mutual information are due solely to high concentrations of Cro, leading to simultaneous binding as soon as the blocking RNAP unbinds.
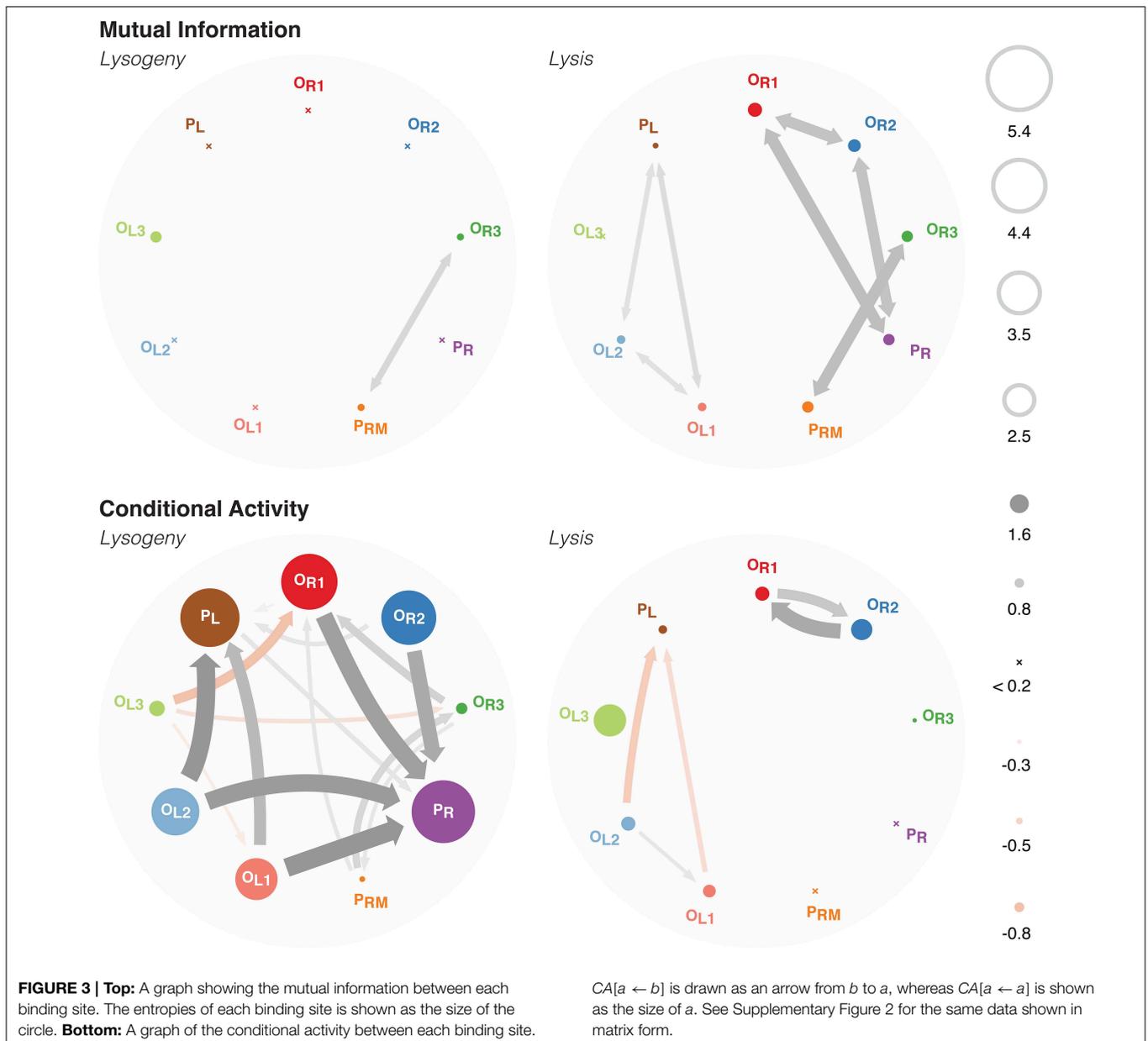
We see that the mutual information is sensitive only to correlations that occur between physically-overlapping binding sites. Since these overlaps were explicitly introduced when constructing the model, the mutual information does not give any new information. The conditional activity, however, shows a rich network of interactions between binding sites that is not evident directly from the model. Furthermore, unlike the mutual information, the conditional activity is highly informative even when the system is in a highly stable regime (as is the lysogenic pathway), because it is sensitive to fluctuations away from equilibrium behavior even when they are rare.

In the lysogenic pathway (**Figure 3** lower left), the conditional activity maps out the interactions which can lead the system to switch to lysis. Lysogeny is characterized by the maintenence of a stable level of CI by RNAP binding to P_{RM}, and strong repression of Cro production. Thus, RNAP is bound to P_{RM} with 86% probability, whereas RNAP is bound to P_R with probability $\leq 10^{-5}$. Because RNAP binding to P_R leads to production of Cro, and high levels of Cro lead to lysis, those events represent fluctuations of the system toward the lysis pathway.

P_R physically overlaps with O_{R1} and O_{R2}, so both must be unbound before RNAP may bind to P_R. However, O_{R1} and O_{R2} are kept bound with CI by cooperative binding with each other and with O_{L1}, O_{L2}. An unbinding event of O_{R1}, O_{R2}, O_{L1}, or O_{L2} therefore represents a deviation from this highly-stable octameric configuration, incurring the associated penalty of breaking multiple cooperative binding interactions. Once the octamer is disrupted, it is more likely that O_{R1} and O_{R2} will unbind, thus allowing RNAP to bind to P_R. This is indicated by the large CA arrows originating at O_{R1}, O_{R2}, O_{L1}, and O_{L2} and pointing at P_R.

The binding behavior of P_R is highly non-Markovian, as quantified by $CA[P_R \leftarrow P_R] \approx 5.4$, indicating that the binding state of P_R changes rarely, but just after it changes it will change again $\sim 10^5$ times sooner than it would have otherwise. This large conditional activity is due to a long persistence time. Since RNAP binding at P_R and the resulting production of Cro serves to move the system from lysogeny to lysis, this large persistence time provides a measure of the stability of the lysogenic pathway. The much shorter exchange time indicates that RNAP binding at P_R is controlled by other cooperative binding processes, which are revealed by the off-diagonal elements of the CA matrix (see below).

Similarly, O_{L1} and O_{L2} have CA arrows pointing at P_L since they physically overlap P_L, and hence their unbinding makes RNAP more likely to bind at P_L. The octameric configuration with O_{R3} either bound by CI or unbound is the dominant behavior in the lysogenic pathway. O_{L3} is slightly more likely to be bound with CI than unbound, whereas is O_{R3} significantly more

FIGURE 3 | Top: A graph showing the mutual information between each binding site. The entropies of each binding site is shown as the size of the circle. **Bottom:** A graph of the conditional activity between each binding site.

$CA[a \leftarrow b]$ is drawn as an arrow from b to a, whereas $CA[a \leftarrow a]$ is shown as the size of a. See Supplementary Figure 2 for the same data shown in matrix form.

likely unbound than bound with CI. Thus, when the binding configuration of O_{L3} changes, it is more likely to be an unbinding event, which eliminates the cooperativity between CI at O_{L3} and O_{R3} which could recruit CI to bind to O_{R3}. Hence, we see a negative CA arrow from O_{L3} to O_{R3}.

The conditional activity in the lysis pathway (**Figure 3** bottom right) shows a much smaller degree of non-Markovian behavior than that seen in the lysogenic pathway. The small, positive CA arrows between O_{R1} and O_{R2}, and between O_{L1} and O_{L2}, represent the same correlated binding shown by the mutual information. The small, negative CA originating at O_{L1} and O_{L2} and pointing at P_L represent the same correlations due to overlap as were detected by the mutual information. (There are also small, negative CAs originating at P_R and pointing at O_{R1} and O_{R2}, too

small to be plotted in this figure. They are visible in Supplementary Figure 2). The conditional activity is evidently diminished for systems in which fluctuations are common; this is the regime correspond to large entropies and mutual information values.

The mutual information and conditional activity thus provide orthogonal, and complimentary, representations of the behavior of the system. Because cooperativity should lead to correlated binding, we expected the mutual information to recover cooperative interactions between binding sites. However, we find that mutual information fails to do so. Cooperativity does lead to correlations between binding activity at different sites. However, cooperativity is also a strongly stabilizing effect, leading to the system residing in one of a few microstates which maximize cooperative interactions, and hence to small entropies. Because

the mutual information between two binding sites is bounded above by the entropy of each binding site, this implies that systems exhibiting strong cooperative interactions are also likely to show little mutual information. By looking for correlations in the binding times instead of correlations between equilibrium binding states, the conditional activity is sensitive to the rare fluctuations which reveal the effects of cooperativity.

4. Discussion

We designed the conditional activity to be a quantitative measure of the deviation from Markovian dynamics. While the Markov model of transitions between the 1200 microstates is Markovian by construction, the cooperative binding and overlap of binding sites encoded in the transition rate matrix lead to highly non-Markovian behavior when the space of 1200 microstates is projected down to the space of the nine binding sites. Because non-Markovian dynamics are a general feature of systems exhibiting cooperativity, the conditional activity can be used to study the interactions and flow of information in such systems. Furthermore, since the conditional activity is an efficiently-computable low-dimensional representation of a high-dimensional stochastic model that preserves important information about the dynamical behavior of the system, it is an ideal tool for model inference and model reduction of biochemical systems. Since simulating the Markov model is computationally expensive or intractable for large models with widely-separated time scales, a major contribution of this paper is deriving an analytic expression for the conditional activity in terms of the Markov transition matrix for the microstates of the system, bypassing the need for simulations.

Cooperativity in biochemical systems often results in highly stable configurations of macromolecular complexes. In this work, we showed that rare fluctuations away from these stable configurations contain information about important interactions between binding sites and mechanisms of switching between lysogeny and lysis in phage lambda. We find that the conditional activity is sensitive to these fluctuations, whereas the mutual information is not. Our results illustrate the limited utility of the mutual information in characterizing systems with strong cooperative interactions. In Lin (under review), we showed that proteins exhibit non-Markovian dynamics on the intramolecular scale, and that the conditional activity effectively characterized the flow of information through these molecular systems. In this work, we find cooperativity leads to non-Markovian dynamics at the intermolecular level, and motivate the conditional activity as a tool for studying information flow through gene regulatory networks.

For the model of phage lambda, the energies of each of the 1200 microstates were known experimentally, allowing for the construction of a full Markov model involving the rates of transition between microstates. Therefore, we could calculate the correlation functions between the different binding sites directly from the transition rate matrix as well as from explicitly simulating the Markov model. By recording the times of binding events for pairs of binding sites, the conditional activity may be measured experimentally. Since cooperativity and the resulting

stability are typical features of biochemical systems, the limitations of mutual information we encounter here will apply equally to those systems, and we expect the methods we present here to be useful in analyzing a wide variety of such systems.

The *lac* operon in *E. coli* is a bistable switch with an almost identical architecture to that of the phage lambda switch we have considered here (except its most stable configuration is a DNA loop bound by a repressor tetramer; it lacks the octomeric configuration of the phage lambda switch). Just as our analysis of the phage lambda switch found that Cro production resulted from unbinding of CI repressor from the highly-stable octomer in the looped-DNA configuration, Choi et al. (2008) reach the same conclusion experimentally in the *lac* system. Their single-molecule experiment observed bursts of protein production following unbinding of the repressor from the tetrameric configuration.

Our results demonstrate that one may infer such behavior without directly observing it experimentally, as it is revealed in temporally-correlated binding activity when the system is fluctuating around a stable configuration—the conditional activity elucidates the architecture of the bistable switch without actually observing it switch. Furthermore, our analysis suggest that in some cases it may not be necessary to perform a difficult single-molecule time-series experiment to directly measure the conditional activity in gene regulatory networks. All our results were calculated using thermodynamic parameters alone (binding energies), which in some cases may be obtained *in vitro* and are often easier to measure than real-time single-molecule binding activity. We suspect that the strategy of inferring kinetics from thermodynamic parameters by assuming detailed balance and equal binding rates, and then calculating dynamic quantities from these kinetics, is valid for a large class of biochemical systems. It remains for future work to explore the limits of this approximation (see e.g., Daniels et al., 2008).

Colquhoun and Hawkes (1981) and subsequent work on ion channels compellingly demonstrated how non-Markovian behavior could bridge scales in biological systems: by measuring non-Markovian dynamics on observable scales one could place strong constraints on microscopic behavior which may not be directly accessible experimentally. This line of work leverages explicit stochastic models of ion channels to relate observable dynamics with microscopic parameters, so that the latter may be inferred from the former. The conditional activity, however, was designed to detect interactions between parts of the system in a model-independent way.

Because the *lac* system is well-studied and relatively simple, Choi and coauthors were able to hypothesize that the dissociation of the repressor tetramer would lead to bursts of protein production, and could design an experiment to observe this phenomenon. As one considers more complex and less well-understood biochemical networks, it becomes highly non-intuitive to design an appropriate experiment to observe the desired phenomena, since the interactions between the components in the network may not be known. By identifying the functional connectivity between the elements in a biochemical network, the conditional activity could serve to guide the design of such experiments. Developmental gene regulatory networks are a

particularly interesting application of this approach. For the gene network which controls cell-fate determination in *Arabidopsis*, the conditional activity would suggest the perturbations to which the system is insensitive and also those experimental interventions which would most efficiently lead to specific cell fates.

In future work, we also aim to move beyond systems with established models and seek to demonstrate the efficacy of the conditional activity in characterizing systems for which a full Markov model is not yet known. This approach has promise in neural systems, where almost all of the information in the network is encoded by the times of spike events and effective experimental protocols exist for recording spike times for large networks of neurons. Previous work has shown how correlations between subsequent waiting times (Farkhooi et al., 2009; Schwalger and Lindner, 2010) reveal the dynamics of a single neuron; our results suggest that correlations of waiting times between different neurons may reveal the functional connectivity of the network.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2015.00214/abstract

References

Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. U.S.A.* 79, 1129–1133. doi: 10.1073/pnas.79.4.1129

Anderson, L. M., and Yang, H. (2008). DNA looping can enhance lysogenic CI transcription in phage lambda. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5827–5832. doi: 10.1073/pnas.0705570105

Choi, P. J., Cai, L., Frieda, K., and Xie, X. S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322, 442–446. doi: 10.1126/science.1161427

Colquhoun, D., and Hawkes, A. G. (1981). On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 211, 205–235. doi: 10.1098/rspb.1981.0003

Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory, 2nd Edn.* Hoboken, NJ: Wiley-Interscience.

Daniels, B. C., Chen, Y.-J., Sethna, J. P., Gutenkunst, R. N., and Myers, C. R. (2008). Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* 19, 389–395. doi: 10.1016/j.copbio.2008.06.008

Darling, P. J., Holt, J. M., and 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* 39, 11500–11507. doi: 10.1021/bi000935s

Darling, P. J., Holt, J. M., and 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. doi: 10.1006/jmbi.2000.4050

Espinosa-Soto, C., Padilla-Longoria, P., and Alvarez-Buylla, E. R. (2004). A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell Online* 16, 2923–2939. doi: 10.1105/tpc.104.021725

Farkhooi, F., Strube-Bloss, M. F., and Nawrot, M. P. (2009). Serial correlation in neural spike trains: experimental evidence, stochastic modeling, and single neuron variability. *Phys. Rev. E* 79:021905. doi: 10.1103/PhysRevE.79.021905

Fritz, G., Megerle, J. A., Westermayer, S. A., Brick, D., Heermann, R., Jung, K., et al. (2014). Single cell kinetics of phenotypic switching in the arabinose utilization system of *E. coli*. *PLoS ONE* 9:e89532. doi: 10.1371/journal.pone.0089532

Gedeon, T., Mischaikow, K., Patterson, K., and Traldi, E. (2008). Binding cooperativity in phage λ is not sufficient to produce an effective switch. *Biophys. J.* 94, 3384–3392. doi: 10.1529/biophysj.107.121756

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008

Hedges, L. O., Maibaum, L., Chandler, D., and Garrahan, J. P. (2007). Decoupling of exchange and persistence times in atomistic models of glass formers. *J. Chem. Phys.* 127, 211101. doi: 10.1063/1.2803062

Jung, Y., Garrahan, J. P., and Chandler, D. (2005). Dynamical exchanges in facilitated models of supercooled liquids. *J. Chem. Phys.* 123, 084509. doi: 10.1063/1.2001629

Kemeny, J. G., and Snell, J. L. (1961). Finite continuous time Markov chains. *Theory Probab. Appl.* 6, 101–105. doi: 10.1137/1106012

Little, J. W., and Michalowski, C. B. (2010). Stability and instability in the lysogenic state of phage lambda. *J. Bacteriol.* 192, 6064–6076. doi: 10.1128/JB.00726-10

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7

Marzen, S., Garcia, H. G., and Phillips, R. (2013). Statistical mechanics of Monod–Wyman–Changeux (MWC) models. *J. Mol. Biol.* 425, 1433–1460. doi: 10.1016/j.jmb.2013.03.013

Monod, J., Changeux, J.-P., and Jacob, F. (1963). Allosteric proteins and cellular control systems. *J. Mol. Biol.* 6, 306–329. doi: 10.1016/S0022-2836(63)80091-1

Norregaard, K., Andersson, M., Sneppen, K., Nielsen, P. E., Brown, S., and Oddershede, L. B. (2013). DNA supercoiling enhances cooperativity and efficiency of an epigenetic switch. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17386–17391. doi: 10.1073/pnas.1215907110

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., and van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 737–740. doi: 10.1038/nature02298

Ptashne, M. (2004). *A Genetic Switch: Phage Lambda Revisited, 3rd Edn*. New York, NY: Cold Spring Harbor Laboratory Press Cold Spring Harbor.

Reinitz, J., and Vaisnys, J. R. (1990). Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of cooperativity. *J. Theor. Biol.* 145, 295–318. doi: 10.1016/S0022-5193(05)80111-0

Santillán, M., and Mackey, M. C. (2004). Why the lysogenic state of phage λ is so stable: a mathematical modeling approach. *Biophys. J.* 86, 75–84. doi: 10.1016/S0006-3495(04)74085-0

Schwalger, T., and Lindner, B. (2010). Theory for serial correlations of interevent intervals. *Eur. Phys. J. Spec. Top.* 187, 211–221. doi: 10.1140/epjst/e2010-01286-y

Shannon, C. E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Shea, M. A., and 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. doi: 10.1016/0022-2836(85)90086-5

Syski, R. (1992). *Passage Times for Markov Chains*. Amsterdam: IOS Press.

Tavare, S. (1979). A note on finite homogeneous continuous-time Markov chains. *Biometrics* 35, 831–834. doi: 10.2307/2530117

Tian, T., and Burrage, K. (2004). Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage λ. *J. Theor. Biol.* 227, 229–237. doi: 10.1016/j.jtbi.2003.11.003

Zurla, C., Manzo, C., Dunlap, D., Lewis, D. E., Adhya, S., and Finzi, L. (2009). Direct demonstration and quantification of long-range DNA looping by the λ bacteriophage repressor. *Nucleic Acids Res.* 37, 2789–2795. doi: 10.1093/nar/gkp134