2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Inferring initial state of the ancestral network of cellular fate decision: a case study of phage lambda

Yiyu Pang [1,2] and Jie Liang[1]

***Abstract*— Gene regulatory networks (GRNs) describe how gene expression is controlled by interactions among DNA and proteins. The decision network controlling prophage induction in phage lambda has served as a paradigm for studying decision control of cellular fate, which has broad implications for understanding phenomena such as embryo development, tissue regeneration, and tumorigenesis. The phage-lambda GRN dictates whether the phage enters the lytic mode or the lysogenic mode. In this work, we study the evolutionary origin of this GRN and explore the initial architecture of the proto-GRN, from which the modern GRN is evolved. Specifically, we examined the model of proto-GRN of phage-lambda containing one operator, from which the modern GRN with three operators evolved. We constructed 9 network architectures of the proto-GRNs by different combinations of the three operators O_R3, O_R2, O_R1 and the three different genomic locations. We quantified the full stochastic behavior of each of these networks through exact computation of their steady-state probability landscapes using the Accurate Chemical Master Equation(ACME) algorithm. We further analyzed changes in the copy numbers of the two key proteins CI and Cro during prophage induction upon UV irradiation at different dosages. By examining the dynamic changes of the protein copy numbers upon different UV irradiations, our results show that the network in which O_R1 located at the second site is the most probable architecture for the ancestral phage-lambda network. Our work can be extended for further analysis of the evolutionary trajectories of this cellular fate decision network.**

I. INTRODUCTION

Biochemical reactions occur among interacting molecules in cells. The interactions between transcription factors and specific binding sites on DNA are the basis of the regulation of gene expression. Gene regulatory networks (GRNs) describe details of the molecular species involved and how they interact to control gene expression. While GRNs have evolved into highly complex machineries, deciphering their evolutionary origins and evolutionary trajectories has important implications. First, understanding how GRNs respond to different environmental perturbations may afford us a global view on how environmental changes affect the GRNs. Conversely, an evolutionary view of GRNs may help to gain understanding on how the architecture and properties of GRNs evolve to adapt to environmental changes. For example, studies of cancer in the evolutionary context suggested that neoplasms recapitulate ancestral phenotypes [1], [2], [3]. These ancient traits are found to reappear in cancer cells as the regulation of suppress them are disrupted [3]. By identifying their evolutionary origins and trajectories, analysis of the behavior of evolving GNRs can shed light on tumorigenesis and offer further opportunities to design therapeutic regimens [2].

The lysogeny-lysis switching network of phage λ is a prototypical model system that provides a general framework for studying mechanism of gene regulation and cellular fate decision. The dynamics of this network is responsible for phage to chose between two distinct life cycles, namely, lysogeny and lysis. This network has been studied extensively [4], [5], [6], [7], [8]. However, the evolutionary origin of this GRN remains unclear.

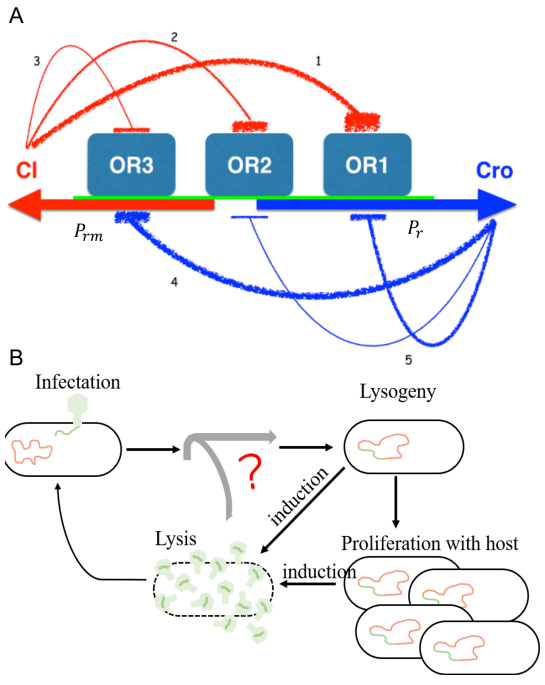

Fig. 1. Simplified diagram of the core components of the gene regulatory network of phage-λ and schematic of the fate decision process of phage λ. (A) The core regulatory network of lambda switch; (B) the decision making process of phage after infecting *E. Coli*.

Upon infecting its host $Escherichia\ coli$, phage lambda has two alternative propagation modes, namely, the lysogenic (prophage) mode and the lytic mode (Fig 1). In lysogeny, all of the phage genes responsible for developmental functions are turned off, and phage lambda integrates its genetic material into the genome of the host $E.\ coli$ and replicates its DNA at each subsequent cell division for many generations.

*This work is supported by NIH grant R35 GM127084.

[1]Yiyu Pang is with Department of Bioengineering and Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA ypang22@uic.edu

[2]Jie Liang is with Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA jliang@uic.edu

When triggered by a cellular signal such as threatening DNA damage by UV irradiation, the lysogenic mode switches into the lytic mode and the $E.\ coli$ host cell produces a large number of phage particles until it is destroyed. This switching process is called *prophage induction* [6].

To investigate the evolution of a GRN, one can investigate how it is locally rewired and how bindings of proteins are altered under evolutionary selection pressure [9]. As transcription factors are usually highly conserved, evolution of the sequences of their binding sites (such as operators) is thought to be the primary cause for the evolution of GRNs [10], [11], [12]. In this study, we investigate the evolution of the switching network of phage lambda by examining the evolution of the sequences of the right operators (O_Rs). We assume that the extant core network of lambda switch, which includes 3 operators of O_R3, O_R2 and O_R1, is evolved from an initial ancient network containing only a single operator. Due to environmental pressure, this primitive network evolves incrementally through duplication of the operator(s), followed by subsequent mutations. Here the environmental pressure is taken as different levels of UV irradiation. By observing changes in the global behavior of the network during prophage induction, our goal is to reconstruct the ancestral network containing a single operator. With the recent development of the accurate chemical master equation (ACME) [13], [8], [14], [15], we analyze the stochastic behavior of different possible candidate ancestral networks under different levels of UV irradiation. By comparing behavior of these different networks, we can infer the most probable network structure as the ancient network.

II. Models and Methods

A. Computing probability landscape for a biochemical reaction network

We assume that the reaction network system is well-mixed with constant volume and temperature. The network system has n molecular species $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ and m reactions with reaction constants $\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_m$. The k-th reaction is denoted as

$$c_{1,k}x_1(t) + c_{2,k}x_2(t) + \ldots + c_{n,k}x_n(t) \rightarrow^{r_k} c_{1,k}\prime x_1(t) + c_{2,k}\prime x_2(t) + \ldots + c_{n,k}\prime x_n(t), \tag{1}$$

where the microstate of the system at time t is $x(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \in \mathbb{Z}^n$. The rate of reaction R_k that brings the microstate from x_i to x_j is denoted as

$$A_k(x_i, x_j) = r_k \prod_{l=1}^{n} \binom{x_l}{c_{l,k}}, \tag{2}$$

where r_k is the intrinsic reaction constant and $c_{l,k}$ is the stoichiometry constant of the relevant reactants x_l in reaction R_k

The discrete chemical master equation can be written using above definitions to describe the probability change of every microstate over time:

$$\frac{dp(\mathbf{x}, t)}{dt} = \sum A(\mathbf{x}, \mathbf{x}')p(\mathbf{x}', t) - A(\mathbf{x}', \mathbf{x})p(\mathbf{x}, t). \tag{3}$$

where $p(\mathbf{x}, t)$ is probability of microstate $\mathbf{x}$ at time t. $A(\mathbf{x}, \mathbf{x}')$ is the transition rate from state $\mathbf{x}'$ to state $\mathbf{x}$. Using above equations, the probability landscape $p(\mathbf{x}, t)$ can be computed using the Accurate Chemical Master Equation (ACME) method [14].

B. Computing CI and Cro protein level

Once the steady state probability landscape is constructed, we compute the protein levels of CI and Cro proteins by summing up the properly weighted probabilities of the copy number of CI and Cro from each microstate.

III. Results

The gene regulatory module of phage-lambda contains the molecular species of protein CI and Cro, operator O_R3, O_R2 and O_R1, and protein-DNA complexes where CI or Cro binding to these three operator sites. Details of the reactions and their rates are listed in Table I. The levels of CI and Cro protein determine whether phage lambda is in the lysogenic or the lytic mode. Specifically, the lysogenic mode corresponding to microstates with high CI protein copy numbers and low Cro protein copy numbers, while the lytic mode corresponding to microstates with low CI protein copy numbers and high Cro protein copy numbers.

A. Model justification

In our network, dimers of CI and Cro are not explicitly modeled for consideration of computational costs. To ensure that the simplification will not inadvertently alter the dynamics of prophage induction, we compared variations in the CI and Cro protein level during induction. We modeled UV irradiation doses using the degradation rate of CI: the stronger the UV dose is, the larger the degradation rate of CI. Our results show that without explicit dimers, our network without dimers has similar changes in CI and Cro protein level at different UV irradiation as the model where the dimers are explicitly modeled. Specifically, the switching threshold, which corresponds to the cross-point of CI and Cro protein levels at different UV irradiation (Fig 2), is at CI degradation rate of $\sim$0.0023/s for both models. Therefore, we conclude that the dynamics of proteins during prophage induction are not affected by the simplification of the dimers.

B. Ancient regulatory network inference

The extant gene regulatory network of lambda switch includes three operator sites, which can be bound by dimers of protein CI and Cro at differential affinities. CI has the lowest affinity for O_R3 compared to binding to O_R2 and O_R1, while Cro has a higher affinity for O_R3 and lower affinities for O_R2 and O_R1. For convenience, we name these sites in the order from left to right as site 3, site 2, and site 1 (Fig 1A). Since the three operators have sequence identities near 50% but are not 100% identical (Fig 3A), they have different binding affinities with protein CI and Cro. We also assume that the ancient regulatory network contains only one operator site. During the long process of

TABLE I

REACTIONS AND THEIR PARAMETERS [14]

Reactions	reaction rate(/s)
Synthesis reactions	
$\phi + (OR3 + OR2) \rightarrow CI + (OR3 + OR2)$	$k_s\_CI = 0.0069/s$
$\phi + (OR3 + ROR2) \rightarrow CI + (OR3 + ROR2)$	$k_{s1}\_CI = 0.0069/s$
$\phi + (OR1 + OR2) \rightarrow Cro + (OR1 + OR2)$	$k_s\_CI = 0.066/s$
$\phi + (OR1 + OR2) \rightarrow Cro + (OR1 + OR2)$	$k_s\_Cro = 0.0929/s$
Degradation reactions	
$CI \rightarrow \phi$	$k_d\_CI = 0.0027/s$
$Cro \rightarrow \phi$	$k_d\_Cro = 0.0025/s$
DNA-protein binding reactions	
$2CI + OR1 \rightarrow ROR1$	$k_b\_CI = 0.0021/nM^2 \cdot s$
$2CI + OR2 \rightarrow ROR2$	$k_b\_CI = 0.0021/nM^2 \cdot s$
$2CI + OR3 \rightarrow ROR3$	$k_b\_CI = 0.0021/nM^2 \cdot s$
$2Cro + OR1 \rightarrow COR1$	$k_b\_Cro = 0.001289/nM^2 \cdot s$
$2Cro + OR2 \rightarrow COR2$	$k_b\_Cro = 0.001289/nM^2 \cdot s$
$2Cro + OR3 \rightarrow COR3$	$k_b\_Cro = 0.001289/nM^2 \cdot s$
Dissociation reactions	
$ROR1 + (OR2) \rightarrow 2CI + OR1 + (OR2)$	$k_u\_CI2OR1 = 0.03998/s$
$ROR1 + (ROR2 + OR3) \rightarrow 2CI + OR1 + (ROR2 + OR3)$	$k_{u1}\_CI2OR1 = 0.0005/s$
$ROR1 + (ROR2 + ROR3) \rightarrow 2CI + OR1 + (ROR2 + ROR3)$	$k_{u1}\_CI2OR1 = 0.05531/s$
$ROR1 + (ROR2 + COR3) \rightarrow 2CI + OR1 + (ROR2 + COR3)$	$k_{u2}\_CI2OR1 = 0.0005/s$
$ROR1 + (COR2) \rightarrow 2CI + OR1 + (COR2)$	$k_u\_CI2OR1 = 0.03998/s$
$ROR1 + (ROR2 + OR3) \rightarrow 2CI + OR1 + (ROR2 + OR3)$	$k_{u1}\_CI2OR1 = 0.0005/s$
$ROR1 + (ROR2 + ROR3) \rightarrow 2CI + OR1 + (ROR2 + ROR3)$	$k_{u1}\_CI2OR1 = 0.05531/s$
$ROR1 + (ROR2 + COR3) \rightarrow 2CI + OR1 + (ROR2 + COR3)$	$k_{u2}\_CI2OR1 = 0.0005/s$
$ROR1 + (COR2) \rightarrow 2CI + OR1 + (COR2)$	$k_u\_CI2OR1 = 0.03998/s$
$ROR2 + (OR1 + OR3) \rightarrow 2CI + OR2 + (OR1 + OR3)$	$k_u\_CI2OR2 = 1.026/s$
$ROR2 + (ROR1 + OR3) \rightarrow 2CI + OR2 + (ROR1 + OR3)$	$k_{u1}\_CI2OR2 = 0.01284/s$
$ROR2 + (OR1 + ROR3) \rightarrow 2CI + OR2 + (OR1 + ROR3)$	$k_{u2}\_CI2OR2 = 0.00928/s$
$ROR2 + (ROR1 + ROR3) \rightarrow 2CI + OR2 + (ROR1 + ROR3)$	$k_{u1}\_CI2OR2 = 0.01284/s$
$ROR2 + (COR1 + OR3) \rightarrow 2CI + OR2 + (COR1 + OR3)$	$k_u\_CI2OR2 = 1.026/s$
$ROR2 + (OR1 + COR3) \rightarrow 2CI + OR2 + (OR1 + COR3)$	$k_u\_CI2OR2 = 1.026/s$
$ROR2 + (COR1 + COR3) \rightarrow 2CI + OR2 + (COR1 + COR3)$	$k_u\_CI2OR2 = 1.026/s$
$ROR2 + (ROR1 + COR3) \rightarrow 2CI + OR2 + (ROR1 + COR3)$	$k_{u1}\_CI2OR2 = 0.01284/s$
$ROR2 + (COR1 + ROR3) \rightarrow 2CI + OR2 + (COR1 + ROR3)$	$k_{u2}\_CI2OR2 = 0.00928/s$
$ROR3 + (OR2) \rightarrow 2CI + OR3 + (OR2)$	$k_u\_CI2OR3 = 5.19753/S$
$ROR3 + (ROR1 + OR2) \rightarrow 2CI + OR3 + (OR1 + OR2)$	$k_{u1}\_CI2OR3 = 0.04702/S$
Dissociation reactions of Cro with $OR1, OR2, OR3$ is not showing here, please find in [14]	

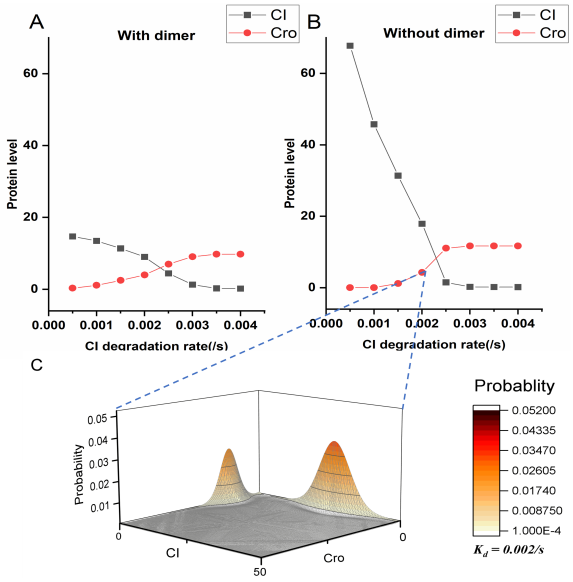

Fig. 2. CI and Cro protein levels during prophage induction at different CI degradation rates, which models the dosage of the UV irradiation. (A) Protein titration by UV irradiation in a detailed model with dimer explicitly included. (B) Protein titration by UV irradiation in a simplified model without explicit dimer. (C) The probability landscape projected to 2D subspace when CI degradation rate K_d is 0.002/s. The red data point in (B) connected to (C) by blue dashed lines is the protein level of Cro when K_d is 0.002/s, which is the integration of products of weighted probabilities and copy number at each microstate.

evolution, the operator site was duplicated and to form new operators, with subsequent mutations following duplications.

In ancient times, UV irradiation was much stronger than that of present days [16]. The only way that bacteriophage lambda could survive is to have strong resistance to UV irradiation during prophage induction. Once phage lambda switched to the lytic state, it would replicate itself as much as possible until lysis of the host occurs. Therefore, if prophage induction occurs at a low level of UV irradiation, phage lambda will stay persistently switched to the lytic state and eventually exhaust all available hosts nearby. In the end, phage lambda would become extinct once there is no host left to infect. On the other hand, phage lambda cannot be too resistant to UV irradiation that it never switches into the lytic mode. If so, as the lysogenic mode only replicates itself with the host life cycle, there would be little chance for the phage to have enough posterity to survive. In other words, phage lambda needs to stay in a lysogenic mode at low UV irradiation dosage and be induced to the lytic mode only at a relatively high-level UV in ancient times. As the CI degradation rate reflects the intensity of UV irradiation [17], in our model, the higher the CI degradation rate, the stronger the UV irradiation dosage is.

To infer which operator most likely occurs in the ancestral regulatory network, we constructed all possible molecular networks with 9 possible combinations of three different operator sequences and three different sites on the phage DNA. We then explored the dynamics of changes in protein level at prophage induction.

The three different operators are O_R3, O_R2 and O_R1, and there are three different sites: the site located within in CI's promoter region (where O_R3 is located in the extant GRN), and two other sites located within Cro's promoter region (where O_R2 and O_R1 are located in the extant GRN), as shown in Fig 1. The CI and Cro protein levels in the 9 different networks are computed from the corresponding steady state probability landscapes of these networks. By varying CI degradation rate from 0.0005/s to 0.006/s, the protein levels are calculated to obtain the protein titration curve at different UV irradiation.

Our results show that O_R1 at site 2 is the most probable scenario for the ancestral regulatory network. When O_R3, O_R2, or O_R1 is at site 3, the copy number of protein Cro is always greater than protein CI at any CI degradation rate, indicating that the system will be at the lytic mode under any circumstances. Other networks, except that of O_R1 at site 2, exhibit the behavior that Cro copy number is greater than that of CI if CI degradation rate is $>$ 0.002/s. This indicates that these systems would switch into the lytic mode at low UV irradiation during prophage induction. In contrast, the network with O_R1 at site 2 exhibits the behavior that only when CI degradation rate is greater than a larger threshold $\sim$0.0041/s, Cro copy number is greater than that of CI. That is, phage lambda switches from lysogeny to lysis only at a high UV irradiation. We therefore can conclude that O_R1 at site 2 is the most probable scenario to have occurred in the ancestral regulatory network, as it endows phage with strong resistance against UV irradiation, which was required for phage lambda to survive and thrive.

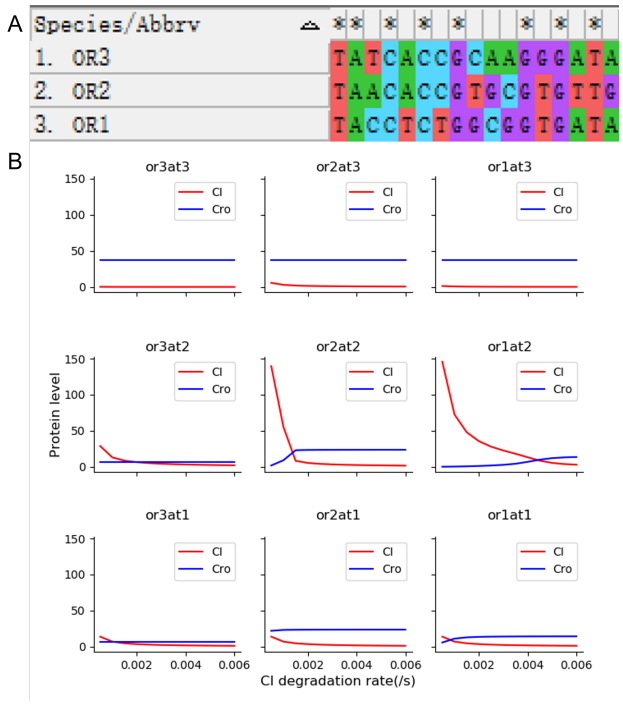

Fig. 3. Operator sequences and protein level variation of possible ancestral networks. (A) sequences of O_R3, O_R2 and O_R1. (B) Protein levels at different UV irradiation modeled by CI degradation in different models of ancient network structures. The label in each plot specifies the operator and its location in the ancestral network. As an example, $or3at3$ indicates this network has O_R3 at site 3.

IV. CONCLUSIONS

In this study, we have inferred the most probable operator and binding site in the ancestral regulatory network of phage lambda switch. By computing the exact steady state probability landscape through the direct solution of the chemical master equations of the networks, we explored the switching behavior of the network of phage lambda under different UV irradiation, which is modeled by CI degradation rate. The exact computation of the steady state probability landscape is carried out using the finite-buffer ACME method. Our results show that O_R1 at site 2 is the most probable scenario of the ancient regulatory network, as phage lambda is switched to the lytic state at a very high CI degradation rate, ensuring that phage lambda can survive the challenging environment of high UV irradiation and avoid the fate of extinction.

Work reported here is the first systematic examination of the global stochastic behavior of a GRN in the evolutionary context according to our knowledge. An important advantage of our approach is the ability to study how environment elicits different responses of GRNs by directly observing the behavior of the network under different environmental perturbations. Such analysis is different from traditional phylogenetic analysis of molecules, which are building-block components of GRNs. Network behavior such as phenotype switching is a system emerging property that requires analysis of the global behavior of the probability landscape of each hypothetical ancestral network.

It is expected that experiments of synthetic biology can be constructed to test the inferred behavior of candidate ancestral networks to gain further insights into the evolution of GRNs. Additional future work can be extended from our findings reported here, so further analysis of likely evolutionary pathways can be inferred. Our work has broad implications in understanding the origin of networks controlling cellular fate decisions, as similar models can be constructed to study the emergence of phenotype switching in tumor cells that result from somatic mutations.

REFERENCES

[1] Lucien Israel. Tumour progression: random mutations or an integrated survival response to cellular stress conserved from unicellular organisms? *Journal of theoretical biology*, 178(4):375–380, 1996.

[2] Amy Wu, Qiucen Zhang, Guillaume Lambert, Zayar Khin, Robert A Gatenby, Hyunsung John Kim, Nader Pourmand, Kimberly Bussey, Paul CW Davies, James C Sturm, et al. Ancient hot and cold genes and chemotherapy resistance emergence. *Proceedings of the National Academy of Sciences*, 112(33):10467–10472, 2015.

[3] Anna S Trigos, Richard B Pearson, Anthony T Papenfuss, and David L Goode. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proceedings of the National Academy of Sciences*, 114(24):6406–6411, 2017.

[4] M Ptashne, K Backman, MZ Humayun, A Jeffrey, R Maurer, B Meyer, and RT Sauer. Autoregulation and function of a repressor in bacteriophage lambda. *Science*, 194(4261):156–161, 1976.

[5] John W Little, Donald P Shepley, and David W Wert. Robustness of a gene regulatory circuit. *The EMBO journal*, 18(15):4299–4307, 1999.

[6] Mark Ptashne. *A genetic switch: phage lambda revisited*. CSHL press, 2004.

[7] Ian B Dodd, Keith E Shearwin, and J Barry Egan. Revisited gene regulation in bacteriophage λ. *Current opinion in genetics & development*, 15(2):145–152, 2005.

[8] Youfang Cao, Hsiao-Mei Lu, and Jie Liang. Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. *Proceedings of the National Academy of Sciences*, 107(43):18445–18450, 2010.

[9] M Madan Babu, Sarah A Teichmann, and L Aravind. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *Journal of molecular biology*, 358(2):614–633, 2006.

[10] Michalis Averof, Rachel Dawes, and David Ferrier. Diversification of arthropod hox genes as a paradigm for the evolution of gene functions. In *Seminars in Cell & Developmental Biology*, volume 7, pages 539–551. Elsevier, 1996.

[11] Michael Z Ludwig, Nipam H Patel, and Martin Kreitman. Functional analysis of eve stripe 2 enhancer evolution in drosophila: rules governing conservation and change. *Development*, 125(5):949–958, 1998.

[12] David N Arnosti. Analysis and function of transcriptional regulatory elements: insights from drosophila. *Annual review of entomology*, 48(1):579–602, 2003.

[13] Youfang Cao and Jie Liang. Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Systems Biology*, 2(1):1–13, 2008.

[14] Youfang Cao, Anna Terebus, and Jie Liang. Accurate chemical master equation solution using multi-finite buffers. *Multiscale Modeling & Simulation*, 14(2):923–963, 2016.

[15] Youfang Cao, Anna Terebus, and Jie Liang. State space truncation with quantified errors for accurate solutions to discrete chemical master equation. *Bulletin of mathematical biology*, 78(4):617–661, 2016.

[16] Charles S Cockell and Gerda Horneck. The history of the uv radiation climate of the earth—theoretical and space-based observations¶. *Photochemistry and Photobiology*, 73(4):447–451, 2001.

[17] Moisés Santillán and Michael C Mackey. Why the lysogenic state of phage λ is so stable: a mathematical modeling approach. *Biophysical journal*, 86(1):75–84, 2004.