Stony Brook University

Potential Landscape of Adiabatic and Non-Adiabatic Non-Equilibrium Networks

A Dissertation Presented

by

Bo Han

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2011

Stony Brook University

The Graduate School

Bo Han

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Jin Wang - Advisor
Associate Professor, Department of Chemistry

Xiaolin Li - Committee Chair
Professor, Department of Applied Mathematics and Statistics

Wei Zhu
Professor, Department of Applied Mathematics and Statistics

Dale G. Drueckhammer
Professor, Department of Chemistry,
Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Potential Landscape of Adiabatic and Non-Adiabatic Non-Equilibrium Networks

by

Bo Han

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2011

We study the origin of robustness of yeast cell cycle cellular network through uncovering its underlying energy landscape. This is realized from the information of the steady steady probabilities by solving a discrete set of kinetic master equations for the network. We discovered that the potential landscape of yeast cell cycle network is funnelled towards the global minimum, G1 state. The ratio of the energy gap between G1 and average versus roughness of the landscape termed as robustness ratio (RR) becomes a quantitative measure of the robustness and stability for the network. The funnelled landscape is quite robust against random perturbations from the inherent wiring or connections of the network. There exists a global phase transition between the more sensitive response or less self degradation phase leading to underlying funneled global landscape with large RR, and insensitive response or more self degradation phase leading to shallower underlying landscape of the network with small RR. Furthermore, we show the more robust landscape also leads to less dissipation cost of the network. We

quantify the cost in terms of the dissipation or heat loss characterized through the steady state properties: the underlying landscape and the associated flux. We found that the dissipation cost is intimately related to the stability and robustness of the network. With least dissipation cost, the network becomes most stable and robust under mutations and perturbations on sharpness of the response from input to output as well as self degradations. The least dissipation cost may provide a general design principle for the cellular network to survive from the evolution and realize the biological function.

We explore the stochastic dynamics of self regulative genes from fluctuations of molecular numbers and of on and off switching of gene states due to regulatory protein binding/unbinding to the genes. We found when the binding/unbinding is relatively fast (slow) compared with the synthesis/degradation of proteins in adiabatic (non-adiabatic) case, the self regulators can exhibit one or two peak (two peak) distributions in protein concentrations. This phenomena can also be quantified through Fano factors. This shows even with the same architecture (topology of wiring), networks can have quite different functions (phenotypes), consistent with recent single molecule single gene experiments. We further found the inhibition and activation curves to be consistent with previous results in adiabatic regime, but show significantly different behaviors in non-adiabatic regimes with previous predictions with monomer binding. Such difference is due to the dimer effect and never reported before. We derive the non-equilibrium phase diagrams of mono-stability and bi-stability in adiabatic and non-adiabatic regimes. We study the dynamical trajectories of the self regulating genes on the underlying landscapes from non-adiabatic to adiabatic limit, provide a global picture of understanding and show an analogy to the electron transfer problem. We study the stability and robustness of the systems through mean first passage time (MFPT) from one peak (basin of attraction) to another and found both monotonic and non-monotonic turnover behavior from adiabatic to non-adiabatic regimes. For the first time, we explore global dissipation by entropy production and the relation with binding/unbinding processes. Our theoretical pre-

dictions for steady state peaks, fano factors, inhibition/activation curves and MFPT can be probed and tested from experiments.

We established a landscape framework to explore the global stability and robustness of the dynamical systems and networks. We explore in particular a gene network motif appeared in the experimental synthetic biology studies of two genes mutually repress and activate each other with self activation and repression. We found that coherent limit cycle oscillations emerge in two regimes: adiabatic and non-adiabatic regimes, with two mechanisms of producing the stable oscillations: nonlinear cooperative interactions in the adiabatic regime and time delays due to the slow binding to the promoters in the non-adiabatic regime. In both regimes, the landscape has a topological shape of Mexican hat in protein concentrations. The shape of the Mexican hat provides the quantitative description of the capability of the system to communicate with each other in concentration space and time. Therefore, the topology of the landscape quantitatively determines the global stability and robustness of the dynamical systems and networks. The coherence of the oscillations are shown to be correlated with the shape of Mexican hat characterize by the height from the top to the ring of the hat.

Contents

List of Figures

Chapter 1

Quantifying Robustness of Yeast Cell Cycle Network: The Funneled Energy Landscape Perspectives

1.1 Energy Landscape and Cellular Network

To understand the biological function and robustness of the cellular network, it is crucial to uncover the underlying global principle [16, 39, 48]. The natures of the cellular network have been explored by the many experimental techniques [41]. It is found that the cellular networks are in general quite robust against genetic and environmental perturbations. There are increasing number of studies on the global topological structures of the networks recently [43, 8, 58, 61]. However, there are so far very few studies of why the network should be robust and perform the biological function from the physical point of view [89, 63, 88, 10, 76, 54, 3, 71, 72, 36, 98, 50, 30].

Theoretical models of the cellular networks have often been formulated with a set of deterministic chemical rate equations. These dynamical descriptions are inherently local. To probe the global properties, one often has to change the parameters. The parameter space is huge. The global robustness therefore is hard to see from this approach.

Here we will explore the nature of the network from another angle: formulate the problem in terms of the potential function or potential landscape. If the potential landscape of the cellular network is known, the global properties can be explored

[88, 76, 54, 3, 71, 36, 98, 50, 30, 23]. This is in analogy with the fact that the global thermodynamic properties can be explored when knowing the inherent interaction potentials in the system. In the cell, statistical fluctuations coming from the finite number of molecules (typically on the order of 1-1000) provide the source of intrinsic internal noise and the fluctuations from highly dynamical and inhomogeneous environments of the interior of the cell provide the source of the external noise for the networks [59, 20, 83, 85, 92, 65]. Both the internal and external noise play important roles in determining the properties of the network.

In general, one should study the chemical reaction network equations in the noisy conditions to model more realistically the cellular environments. In other words, instead of following the deterministic evolution of the concentrations of proteins in the network by the normal chemical rate equations, one should describe the dynamics of protein concentrations probabilistically. We can realize this through the kinetic master equations. We can study the steady state probability distributions of these chemical concentrations under noisy environments. The generalized potential function for steady state of the network is closely associated with the steady state probability [88, 76, 54, 3, 71, 36, 98, 50, 30, 98]. Once the network problem is formulated in terms of the generalized potential function or potential landscape, the issue of the global stability or robustness is much easier to address. In fact, an explicit illustration of energy landscape and robustness for MAP Kinase signal transduction network has been given recently [98].

It is the purpose of this chapter to study the global robustness problem directly from the properties of the potential landscape for the budding yeast cell cycle network. Furthermore, cellular network is an open non-equilibrium system due to the interactions with the environments. There is often a dissipation cost associated with the network. It will also be interesting to see for our model system how the dissipation cost is related to the features of the landscape reflecting the stability and robustness of the network.

1.2 Budding Yeast Cell Cycle

To explore the nature of the underlying potential landscape of the cellular network, we will study budding yeast cell cycle network. One of the most important function of the cell is the reproduction and growth. It is therefore crucial to understand

the cell cycle and its underlying process. The cell cycles during the development are usually divided in several phases: G1 phase in which cell starts to grow under appropriate conditions; S phase in which DNA synthesis and chromosome replication occurs, G2 phase where the cell is in the stage of preparation for mitosis, and M phase in which chromosome separation and cell division occurs. After passing through the M phase, the cell enters back to G1 phase and thus completes a cell cycle. In most of the eucaryotic cells, the elaborate control mechanisms over DNA synthesis and mitosis make sure the crucial events in the cell cycle are carried out properly and precisely. Physiologically, there are usually several check points (where cells are in the quiescent phase waiting for the signal and suitable conditions for further progress in the cell cycle) for controlling and coordination: G1 before the new round of division; G2 before the mitotic process begins; and M before segregation.

Recently, many of the underlying controlling mechanisms are revealed by the genetic techniques such as mutations or gene knock outs. It is found that control has been centered around cyclin-dependent protein kinases (CDKs) which trigger the major events of the eukaryotic cell cycle. For example, the activation of cyclin/CDK dimer drives the cells at both G1 and G2 check points for further progress. During other phases, check points CDK/cyclin are activated. Although molecular interactions regulating the CDK activities are known, the mechanisms of the check point controls are still uncertain [89, 63, 88, 10].

The cell cycle process has been studied in details in the budding yeast Saccharomyces cerevisiae [40, 89, 63, 88, 10, 54]. There are many genes involved in controlling the cell cycle processes. But the number of the crucial regulators is much less. A network wiring diagram based on the crucial regulators can be constructed [89, 63, 88, 10, 54] as shown in Figure 3.1.

Under the rich nutrient conditions and when the cell size grows large enough, a cyclin Cln3 will be turned on. Thus the cell-cycle sequence starts when the cell commits to division through the activation of Cln3 (the START). The Cln3/Cdc28 will be activated. This in turn activates through phosphorylation a pair of transcription factor groups, SBF and MBF, which activate the genes of the cyclins Cln1 and Cln2 and Clb5 and Clb6, respectively. The subsequent activity of Clb5 drives the cell into the S phase where DNA replication begins. The entry into the M phase for segregation is controlled by the activation of Clb2 through the transcription factor

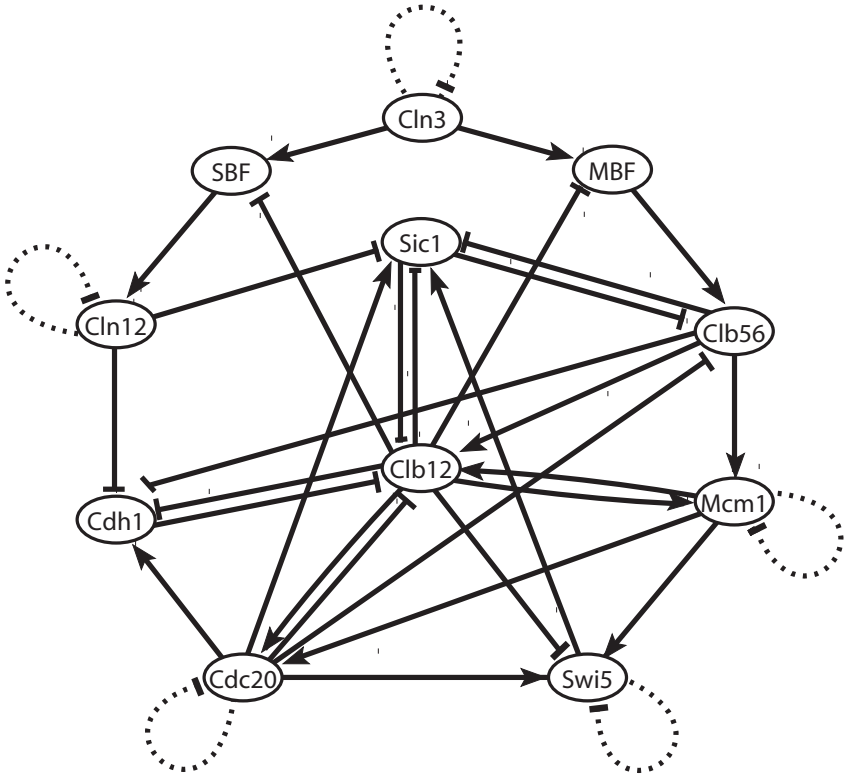


Figure 1.1: The yeast cell cycle network scheme: wiring diagram, the arrow sing(⟶) represent positive activating regulations(1); the inhibition sign(⊣) represents negative suppressing regulations(-1); and the loop sign(–|) represent self-degradation.

MCM1/SFF activation. The exit of the M phase is controlled by the inhibition and degradation of Clb2 through the Sic1, Cdh1 and Cdc20. Clb2 phosphorylates Swi5 to prevent its entry into the nucleus. After the M phase, the cell comes back to the stationary G1 phase, waiting for the signal for another round of division. Thus the cell-cycle process starts with the excitation from the stationary G1 state by the cell-size signal and evolves back to the stationary G1 state through a well defined sequence of states.

Mathematical models of the cell cycle controls have been formulated with a set of ordinary first order (in time) differential equations mimicking the underlined biochemical processes [89, 63, 88, 10, 54]. The models have been applied to budding yeast cycle and explained many qualitative physiological behavior. The check points can be viewed as the steady states or stationary fixed points. Since the intracellular and intercellular signals are transduced into the changes in the regulatory networks, the cell cycle becomes the dynamics in and out of the fixed points. Although detailed simulations give some insights towards the issues, due to the limitation of the parameter space search, it is difficult to perceive the global or universal properties of the cycle networks (for example, for different species). It is the purpose of the current study to address this issue.

We will study the global stability by exploring the underlying potential landscape for yeast cell cycle network.

1.3 Methods and Materials

The average dynamics of the network can be usually described by a set of chemical rate equations for concentrations where both the concentrations and the links among them through binding rates with typically quite different time scales are treated in a continuous fashion. In the cycle cycle, most of the biological functions seem to be from the on and off properties of the network components. Further more, the global properties of the network might depend less sensitively on the details of the model. Therefore, a simplified representation [54] can be proposed with each node i has only two states $S_i = 1$ and $S_i = 0$, representing the active and the inactive state of the protein, or high concentration and low concentration of proteins, respectively. As illustrated in Figure 3.1, we have 11 protein nodes in the network wiring diagram, we have all together 2^{11} states, each state represented by S

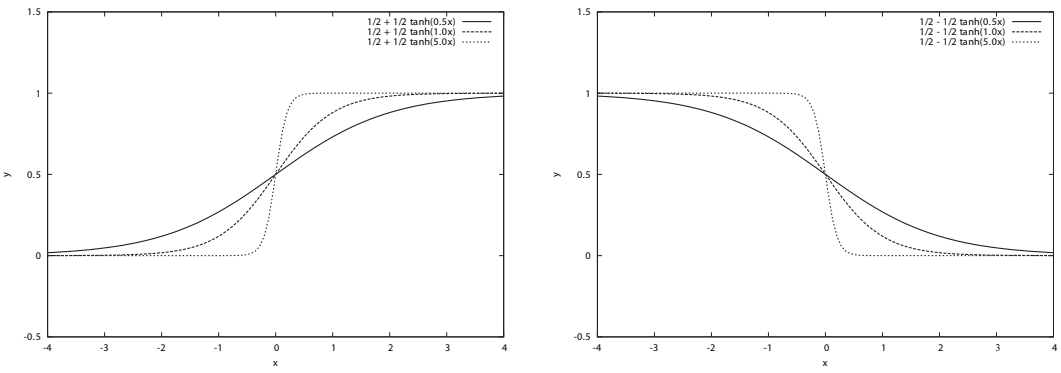
with a distinct combination of the on and off of the 11 protein nodes of Cln3, MBF, SBF, Cln1-2, Cdh1, Swi5, Cdc20, Clb5-6, Sic1, Clb1-2, Mcm1 represented by $\{S_1, S_2, S_3, ...S_{11}\} = S$. Green arrows represent positive regulations or activations (1). Red arrows represent negative regulations or repressions (-1). The yellow loop represents self degradations to the nodes which are not regulated by others. We can then define some rules to follow the subsequent dynamics of the network. Therefore the evolution of the network is deterministic.

As mentioned, in the cell the average dynamics of the cellular network might not give a good description of the system. This is due to the intrinsic fluctuations from the limited number of the proteins in the cell and extrinsic fluctuations from the environments in the interior of the cell. It is then more appropriate to approach the network dynamics based on statistical description. In other words, we should replace the deterministic or average description of the dynamics of states in cellular network to a probabilistic description of the evolution of the cellular network dynamics. So instead of following the on and off state switching in the network, we follow the probability of on and off of each state in the network.

In order to follow the evolution of the states in the cellular network, we need to first figure out the transition probability from one state S_1 at present time to another state S_2 at the next moment. This is difficult to solve and in general almost impossible. We therefore will make some simplifications so that we can handle the case without the loss of the generality by assuming that the transition probability T from one state to another can be split into the product of the transition probability for each individual flip (or no flip) of the on or off state from this moment to the next moment. The transition probability from one state at current state to another at next moment will be assumed not to depend on the earlier times (no memory). This leads to the Markovian process [24, 18, 107]. The transition matrix T can thus be written as:

$$T_{\{S_1(t'),S_2(t'),...,S_{11}(t')|S_1(t),S_2(t),...,S_{11}(t)\}} = \Pi_{i=1}^{11} T_{\{S_i(t')|S_1(t),S_2(t),...S_{11}(t)\}} \tag{1.1}$$

where t is the current time and t' is the next moment. So the whole transition probability from current state to the next is split into the product of the transition probability of each individual flip (or no flip) of the node i. For each individual flip, the transition probability for a particular node can be modeled as a nonlinear switching function as shown in Fig. 1.2A and Fig. 1.2B from the input through the



A B

Figure 1.2: Nonlinear Response Function versus Inputs: (A) is for $y = 1/2 + 1/2tanh(\mu x)$ when $\mu = 0.5, 1, 5$ and (B) is for $y = 1/2 - 1/2tanh(\mu x)$ when $\mu = 0.5, 1, 5$

interactions to the output which is often used in neural science [35]:

$$T_{\{S_i(t')|S_1(t),S_2(t),...S_{11}(t)\}} = \frac{1}{2} \pm \frac{1}{2}tanh[\mu \sum_{j=1}^{11} a_{ij}S_j(t)] \tag{1.2}$$

When the input $\sum_{j=1}^{11} a_{ij}S_j(t) > 0$ is positive (activation), the transition probability to the on state is higher (close to 1). When the input is negative (repression) the transition probability to the on state is lower (close to zero). Furthermore

$$T_{S_i(t')|S_1(t),S_2(t),...S_{11}(t)} = 1 - c \tag{1.3}$$

when there is no input of activation or repression ($\sum_{j=1}^{11} a_{ij}S_j(t) = 0$), c is a small number mimicking the effect of self degradation. Here a_{ij} is the arrow or link representing the activating (+1) or suppressing (-1) interactions between ith and jth protein node in the network which is explicitly shown in the wiring diagram of Fig 3.1. μ is a parameter controlling the width of the switching function from the input to the output. The physical meaning is clear. If the inputs through the interactions among proteins to a specific protein node in the network is large enough, then the state will flip, otherwise the state will stay without the flip. The positive (negative) sign in the T expression gives probability of flipping from 0(1) to 1(0) state. If μ is small (large), the transition width is large (small), the transition is smooth (sharp or sensitive) from the original state to the output state. Therefore we have an analytical expression of the transition probability.

With the transition probability among different states specified, finally we can write down the master equation for each of the 2^{11} states as:

$$dP_i/dt = -\sum_j T_{ij}P_i + \sum_j T_{ji}P_j \tag{1.4}$$

where T_{ij} (T_{ji}) represents the transition probability from state i(j) to state j(i) specified in details above. Here i and j are from 1 to $2^{11} = 2048$ states and $\sum_{i=1}^{i=2^{11}} P_i = 1$.

We solved the $2^{11} = 2048$ master equations numerically of the yeast cell cycle (by using iterative method) to follow the evolution of the probability distribution of each state, with the initial condition of equal small probability of all the cell states ($P_i = 1/2048$). Both the time dependent evolution and the steady state probability distribution for each state are obtained.

Let us focus on the steady state probability distribution. For each state, there is a probability associate with it. One can write the probability distribution for a particular state as $P_i = exp[-U_i]$ ($\sum_{i=1}^{i=2^{11}} P_i = 1$) or $U_i = -lnP_i$. One can immediately see that U_i acquires the meaning of generalized potential energy (from Boltzman distribution). This is the key point: although there is no potential energy function directly from the normal deterministic averaged chemical reaction rate equations for the network, a generalized potential energy function does exist and can be constructed from the probabilistic description of the network instead of the deterministic averaged one. This generalized potential energy function is inversely related to the steady state probability. When the probability is large, the potential energy is lower and when probability is small, the potential energy is higher. The dynamics of the cell cycle thus can be visualized as passing through mountains and ridges of the energy landscape in state space of the cell cycle network to the final destiny. The advantage of introducing the concept of energy is that once we have the potential landscape, we can discuss the global stability of the protein cellular networks. Otherwise, it is almost impossible to address the global issues without going through the parameter space locally which is often cosmologically big.

The network is an open system in non-equilibrium state. Even at steady state, the system is not necessarily in equilibrium. This is clear from the fact that although we can obtain the steady state probability and can define an equilibrium like quantity such as steady state probability, the flux is not necessarily equal to zero ($F_{ij_{steady-state}} = -T_{ij}P_{i\,steady-state} + T_{ji}P_{j\,steady-state}$). This is different from the equilib-

rium situation where the local flux is equal to zero (detailed balance condition). The flux defines a generalized force for the non-equilibrium state along with the associated generalized chemical potential [75, 78]. The non-equilibrium state dissipates energy. In the steady state, the heat loss rate is equivalent to entropy production rate, where entropy S_0 is defined as $S_0 = -\sum_i P_i ln P_i$ and entropy production rate (per unit time) S is given by:

$$S = \sum_{ij} T_{ji} P_j ln(\frac{T_{ji} P_j}{T_{ij} P_i}) \tag{1.5}$$

Entropy production rate is a characterization of the global properties of the network. We can study how the entropy production rate or dissipation cost of the network varies with the changes of internal and external perturbations. We can explore the global natures of the network such as stability, robustness and dissipation cost and their interrelationships.

In each of the simulations, we study the robustness of the network by exploring different values of switching and self degradation parameters μ and c, as well as the mutations of the links or interactions in the network.

1.4 Results and Discussions

Since the potential energy is a multidimensional function in protein states, it is difficult to visualize U. So we directly look at the energy spectrum (Figure 1.3) and explore the nature of the underlying potential landscape U.

Figure 1.3A shows the spectrum as well as the histogram of U. We can see that the distribution is approximately Gaussian. The lowest potential U is the global minimum of the potential landscape. It is important to notice this global minimum of U is found to be the same state as the steady state or fixed point (the stationary G1 state=(0; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0)) of the deterministic averaged chemical reaction network equations for yeast cell cycle. It is clear that the global minimum of the potential is significantly separated from the average of the potential spectrum or distribution.

To quantify this, we define the robustness ratio RR for the network as the ratio of the gap δU, the difference between this global minimum of G1 state $U_{global-minimum}$ and the average of U, $< U >$ versus the spread or the half width of the distribution of

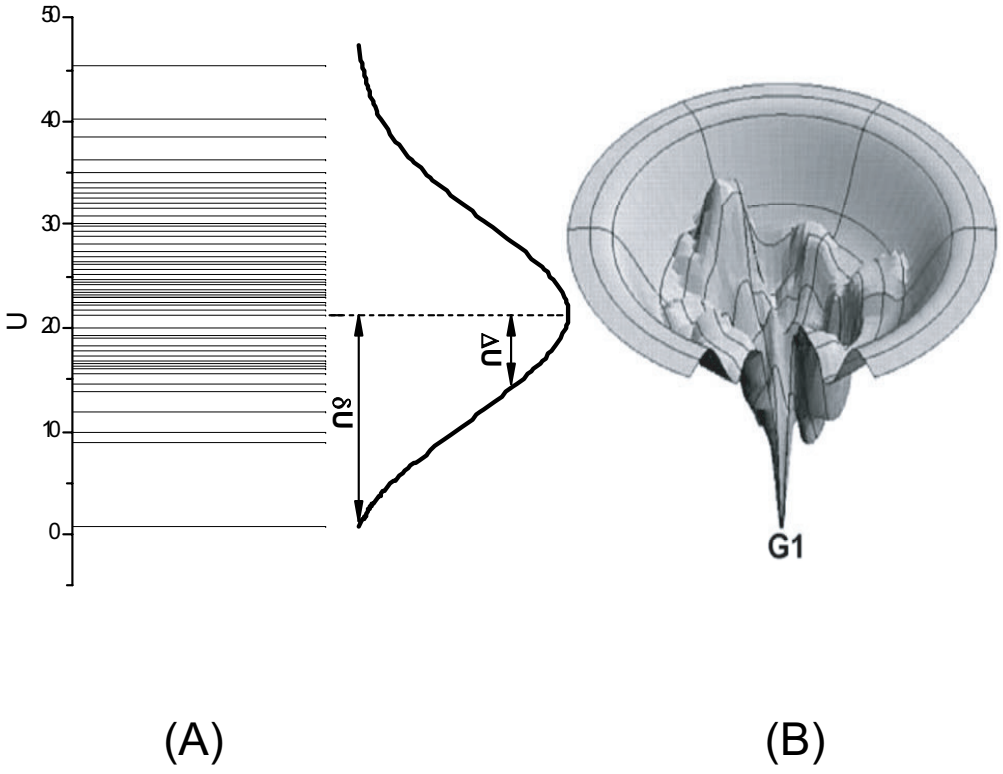


Figure 1.3: The global structures and properties of the underlying potential landscape of the yeast cell cycle network. (A) The spectrum and the histogram or the distribution of the potential energy U. (B) An illustration of the funneled landscape of the yeast cell cycle network. The global minimum of the energy is at G1 state.

U, ΔU, $RR = \frac{\delta U}{\Delta U}$ as shown in Figure 1.3A. δU is a measure of the bias or the slope towards the global minimum (G1 state) of the potential landscape. ΔU is a measure of the averaged roughness or the local trapping of the potential landscape. When RR is significantly larger than 1, the gap is significantly larger than the roughness or local trapping of the underlying landscape, then the global minimum (G1 state) is well separated and distinct from the average of the network potential spectrum. Since $P_i = exp\{-U(x)\}$, the weight or population of the global minimum (G1 state) will be dominated by the one with large RR. The populations of the other possible states are much less significant. This leads to the global stability or robustness discriminating against others. The RR value for the yeast cell cycle network is $RR = 3$ (for $\mu = 5$ and $c = 0.001$) as shown in Figure 1.3A, significantly larger than 1. This shows a funnel picture of energy going downhill towards G1 state in the evolution of network states, as illustrated in Figure 1.3B. So RR gives a quantitative measure of the property of the underlying landscape spectrum.

We found the typical values for random networks are close to 2 (RR can not be less than 1). A typical random network with RR 2 is illustrated in Figure 1.3C for a random network. The ground state is not necessarily the G1 state any more. The probability of G1 is smaller for random network compared with the biological one and therefore less stable. Thus, only the cellular network landscape with large value of RR will be able to form a stable global minimum G1 state, be robust, perform biological function and survive the natural evolution.

We identified the preferential global pathway towards the global minimum G1 by following the most probable trajectory in each step of the kinetic moves from the kinetic master equations towards G1. The protein can be either 1 or 0 representing active or inactive. The 11 proteins are arranged in a vector form to represent the state of the system as (Cln3; MBF; SBF; Cln1,2; Cdh1; Swi5; Cdc20; Clb5, 6; Sic1; Clb1,2; Mcm1). The most probable global path follows the states 1→13 sequentially towards G1 from the start signal. Where start signal is in state sequence 1 given by: (1; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0). Three excited G1 states are in sequence 2, 3, 4, given respectively by: (0; 1; 1; 0; 1; 0; 0; 0; 1; 0; 0), (0; 1; 1; 1; 1; 0; 0; 0; 1; 0; 0), (0; 1; 1; 1; 0; 0; 0; 0; 0; 0; 0). The S phase is in state with sequence 5 given by: (0; 1; 1; 1; 0; 0; 0; 1; 0; 0; 0). The G2 phase is in state with sequence 6 given by (0; 1; 1; 1; 0; 0; 0; 1; 0; 1; 1). The M phase is in states with sequence 7, 8, 9, 10, 11, given respectively by: (0; 0; 0; 1; 0; 0; 1; 1; 0; 1; 1), (0; 0; 0; 0; 0; 1; 1; 0; 0; 1;

1), (0; 0; 0; 0; 0; 1; 1; 0; 1; 1; 1), (0; 0; 0; 0; 0; 1; 1; 0; 1; 0; 1), (0; 0; 0; 0; 1; 1; 1; 0; 1; 0; 0). The another excited G1 state is with sequence 12 given by (0; 0; 0; 0; 1; 1; 0; 0; 1; 0; 0). Finally stationary G1 phase is in state sequence 13 given by (0; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0). The most probable path turns out to be the biological path going through G1→S→G2→M→G1.

We arranged the state space into the two dimensional grids with the constraints of minimal overlapping or crossings of the state connectivity for clear visualization purpose. Each point on the two dimensional grid represents a state (one of 2048 states). The energy landscape on the two dimensional grids is shown in Figure 1.4. The lowest energy state corresponds to the stationary G1 state. The global biological path is represented by the narrow green band on the projected two dimensional state space plane. It is sequentially from state 1 to 13 as mentioned in the above text (sequences 1→13). As we can see, the global biological path is in the low energy valley of the landscape towards G1. In addition, we can also see some other off pathway traps (states with low energies).

Figure 1.5A shows robustness ratio, RR of the cell cycle network versus the steady state probability of the G1 (with $\mu = 5$ and $c = 0.001$) against various perturbations through deleting an interaction arrow, adding an activating or repressing arrow between the nodes that are not yet connected in the network wiring diagram in Figure 1, or switching an activating arrow to a repressing arrow or vice versa, and deleting an individual node. There is a monotonic relationship between the G1 probability and robustness ratio RR. When RR is larger (smaller), the landscape is more (less) robust, the network is more (less) stable with G1 state dominating (less significant). Therefore RR is indeed a robustness measure for the network.

Figure 1.5B shows the robustness ratio RR versus steady state probability of the global biological path with important biological states including G1 [54]. We see again that network with large RR characterizing the funneled landscape leads to higher steady state probability and therefore more stable biological path. Random networks typically have smaller RR and smaller probability of G1 compared with the biological one. They are less stable. The biological functioning network is quite different from the random ones in terms of the underlying energy landscape and stability.

Figure 1.6A shows the the robustness ratio of the underlying energy landscape versus different switching parameters μ ($c = 0.001$). We see that when μ is large

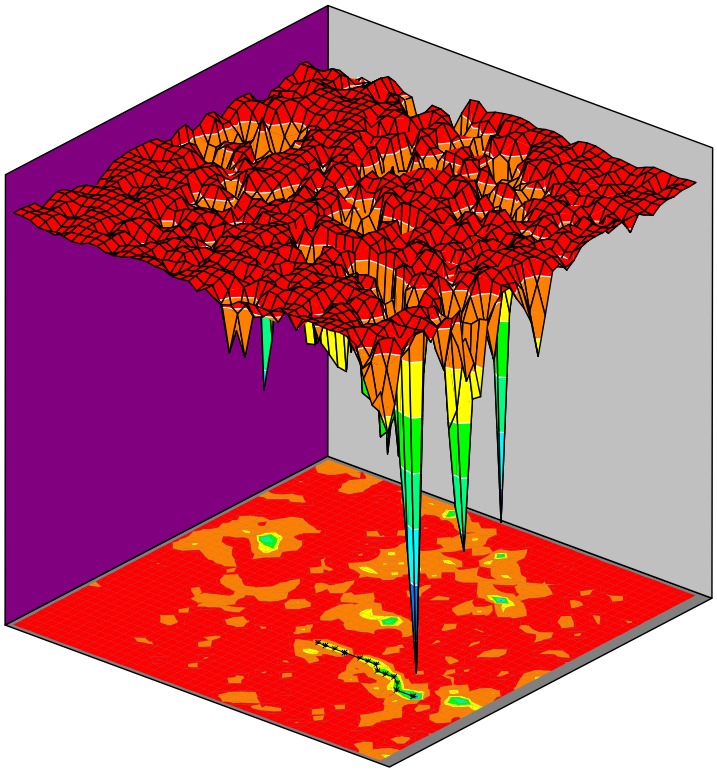


Figure 1.4: The potential energy landscape of the yeast cell cycle network and biological path to stationary G1: The lowest energy state corresponds to stationary G1 state. The green band with arrows corresponds to the biological path.

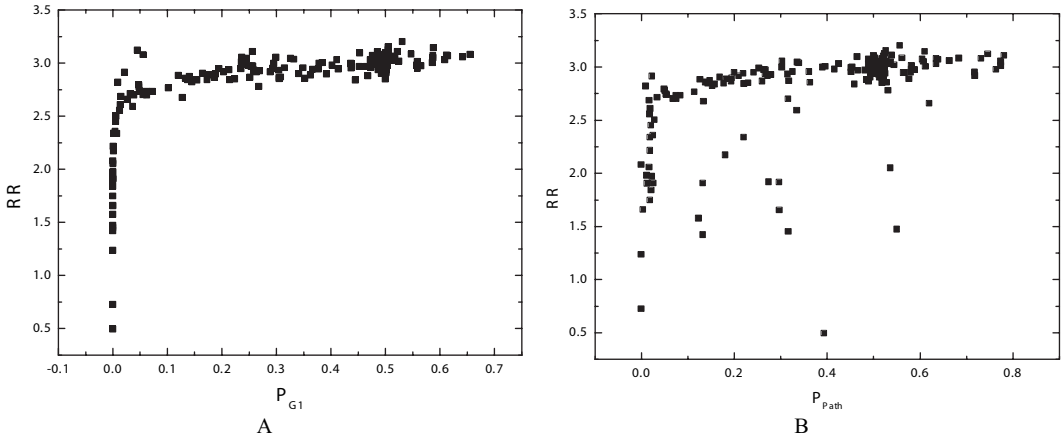


Figure 1.5: Robustness against mutation perturbations. (A) Robustness ratio versus steady-state probability of G1, P_{G1} for different mutations of the links. (B) Robustness ratio RR versus steady-state probability of biological path, P_{path} for different mutations of the links.

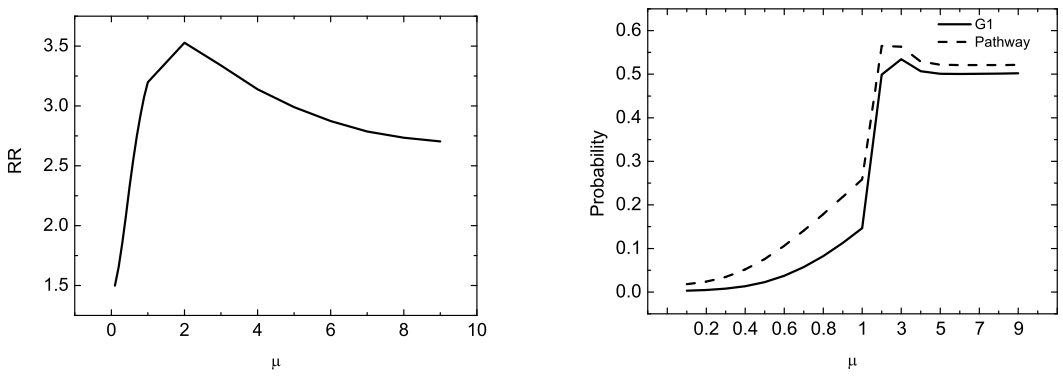


Figure 1.6: Robustness against the sharpness of the response or the inverse noise level. (A) Robustness ratio versus sharpness of the response or inverse of noise level μ. (B) Steady-state probability of stationary G1, P_{G1} and biological path, P_{path} versus m.

(small) indicating a sharp (smooth) transition or response from input to output for a single flip of the protein states, the robustness ratio increases with μ increases. This means, a sharper transition or response from input to output gives more robust network compared with the smoother transition or response. μ can also be seen as a measure or characterization of the strengths of the noise from the intrinsic or extrinsic statistical fluctuations in the cellular environments [107]. The μ could then be related to the inverse of the "temperature" (temperature here is a measure of the strength of the noise level). The energy U we defined in this chapter is in units of μ. So U is a dimensionless quantity. When μ is not changing then the two definitions of U ($U = -\mu logP$ and $U = -logP$) are only different by a constant. The RR is not influenced by the above two definitions of U since it involves the ratio of the U's.

When μ is large, the transition is sharp. This corresponds to all or none deterministic behavior for the response or transition (0 or 1). This is the situation when the underlying statistical fluctuations are small. When μ is small, the response or the transition is no longer all or none (1 or 0) but a smooth function in between 0 and 1. This is due to the fact the statistical fluctuations lead to the states more distributed and with less sharp response. Therefore, the associated probability of distributed states has more chances being between 0 and 1. In other words, less (more) statistical fluctuations or shaper response with larger μ (less sensitive response with smaller μ) leads to more (less) robust network characterized by large (small) RR. Then, there exist two phases for the network: a robust phase with RR

is significantly larger than 2, where the network is stable and the underlying energy landscape is funneled towards G1; and a fragile phase with RR drops to around or below 2, where the network is less stable and the underlying energy landscape is more shallower towards G1.

Figure 1.6B shows probability of stationary G1 state as well as the probability of the global path towards G1 versus μ. We can see a global transition phase transition at $\mu \sim 1$ below which P_{G1} and P_{path} significantly drops. From Figure 1.6A and 6B, we see when P_{G1} and P_{path} is small, RR is also small implying the system is less stable. Therefore the network looses the stability below $\mu \sim 1$. Significantly above $\mu \sim 1$, the network becomes stable. We can interpret this as the phase transition from the weak noise limit where the underlying landscape and the associated global path are not influenced much by the noise level to the limit where underlying landscape and associated global path are disturbed significantly or disrupted by the strong noise. We can also interpret this as the transition from hyper sensitive response leading to the robustness of the landscape and the associated global path, to the inert or insensitive response leading to the fragile landscape and associated global path to G1. We can see a sharper response or more sensitivity of the individual protein nodes to the rest of the protein network through interactions usually leads to more robustness of the network with stable G1 and biological path.

The low μ corresponds to strong noise limit or insensitive response for the node to the input. The landscape has low RR and is less stable or robust. The landscape is more flat and less biased towards G1. When μ increases, the noise level decreases, the response to the input is more sensitive for each node. This results in a more funneled towards G1 and more robust landscape. The maximal funnel is found around $\mu = 2$. There is a sharp change of the shape of the landscape near $\mu = 2$ from $\mu < 2$ side. When $\mu > 2$, the RR value is slightly lower and quickly approach to a constant as μ becomes larger corresponding to smaller noise, and more sensitive response from a node to the input. The landscape becomes stabilized with a definite robustness ratio and probability of P_{G1}. The peak value of the RR, P_{G1} as well as P_{path} implies there might exist traps in the landscape (deep energy states other than G1 and not on biological path). Large noise will destroy landscape which leads to low RR, P_{G1} as well as P_{path}. Zero noise leads to relatively stable network with relatively large RR, P_{G1} as well as P_{path}. In the presence of traps, adding a small amount of noise helps the system to to reach the global minimum without getting

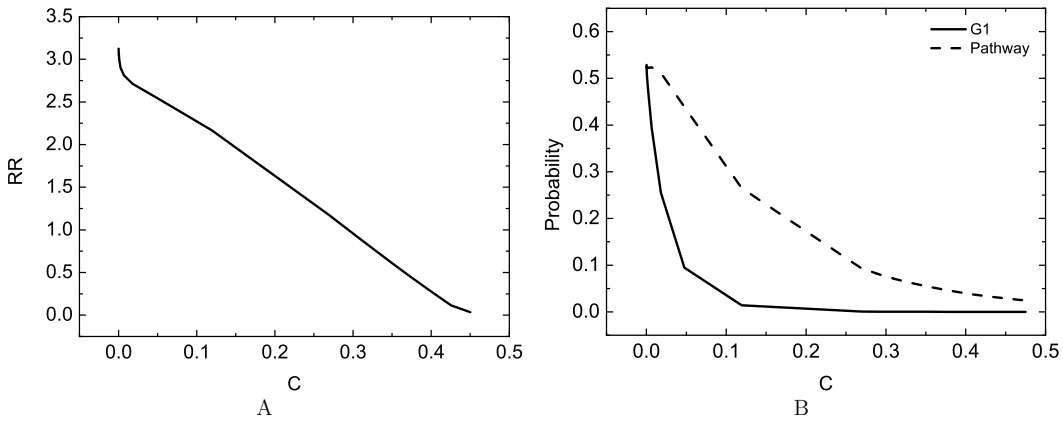


Figure 1.7: Robustness against self-degradation. (A) Robustness ratio RR versus degree of self-degradation, c. (B) Steady-state probability of stationary P_{G1} and biological path P_{path} versus c.

caught or trapped in the intermediate off pathway trapping states. This increases the probability and enhances the stability of G1 and biological path. Therefore the presence of the peak of RR, P_{G1} as well as P_{path} is an indication of the existence of traps in the landscape. We found 6 major off pathway traps responsible for the peak in RR, P_{G1} as well as P_{path} (Some are shown in Figure 1.4).

Figure 1.7A shows the robustness ratio of the underlying energy landscape versus different self degradation parameters c (at $\mu = 5$). We see that when c is large (small) indicating a large (small) self degradation, the robustness ratio increases with c decreases. This means, a less degradation gives more robust network. Figure 1.7B shows the the probability of stationary G1 as well as biological path versus different self degradation parameters c (at $\mu = 5$). We see that when c is large (small) indicating a large (small) self degradation, the probability of stationary G1 phase and biological path increases with c decreases. This means, a less degradation gives more probable and stable stationary G1 phase and biological path and therefore more robust network.

In Figure 1.8, we plotted the entropy production (per unit time) or the dissipation cost of the network, S, versus RR for different μ. We can see the entropy production rate decreases as RR increases. This implies the more robust the network is, the less entropy production or heat loss the network is. This can be very important for the network design. The nature might evolves such that the network is robust against internal (intrinsic) and environmental perturbations, and perform specific biological

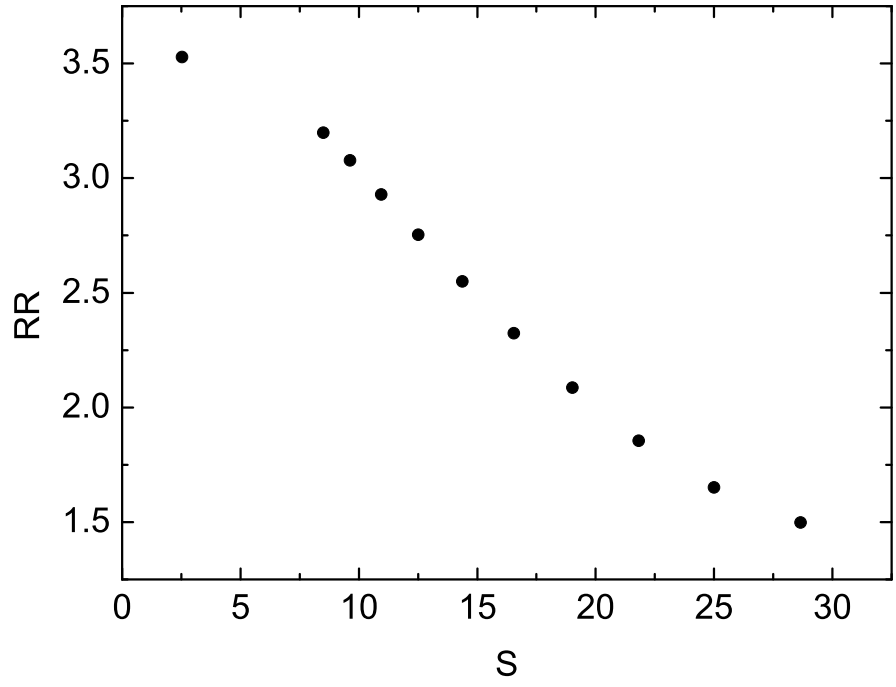


Figure 1.8: Dissipation Cost Versus Robustness of the Network: Entropy Production Rate S versus Robustness Ratio RR.

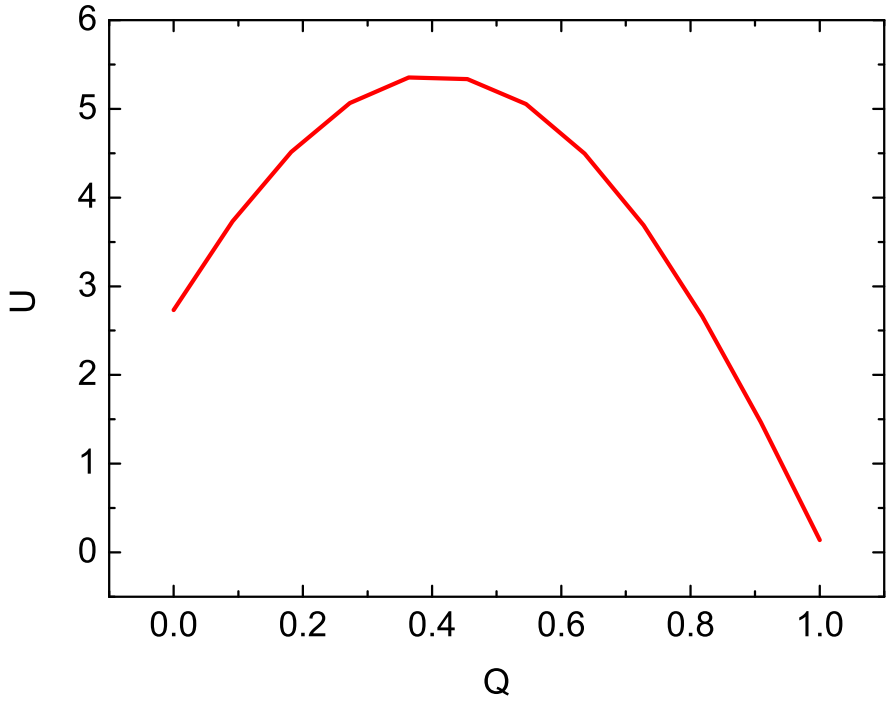


Figure 1.9: One Dimensional Projection of the Energy Landscape: Potential U versus Fraction of Protein Nodes Consistent with the Stationary G1 Phase, Q

functions with minimum dissipation cost. The fact that robustness is linked with the entropy production rate may reflect the fact that less fluctuations and perturbations leads to more robust and stable network, also more energy saving, and therefore less costs in the mean time. This might provide us a design principle of optimizing the connections of the network with minimum dissipation cost for the network. In our study here this is also equivalent of optimizing the robustness or stability of the network.

In Figure 1.9, we define an order parameter Q for this model, based on how many nodes there are in the same state relative to the stationary G1 phase, normalized to 1. So $Q = 1$ when the network is in stationary G1 phase, and $Q = 0$ when the network is in a state which is completely uncorrelated with the stationary G1 phase. We plotted the projection of energy U to the order parameter Q. We can see that there are two minimum or two basins of attractions. One is at Q=1. It is the

global minimum corresponding to the global stationary G1 phase. The other is near Q=0, corresponding to the G2 phase. The existence of different basins of attraction is reasonable in the cell cycle network with several check points. One of the major check points in the experiments turns out to be G2. So check points could be seen as on pathway "trapping". The robustness network will be able to pull out itself from the on pathway "trapping" to proceed the normal cell cycle function to G1.

1.5 Conclusions

The energy landscape is a statistical based approach which is good in two folds: It is a approach capturing the global properties. On the other hand, the statistical approach can be very useful and informative when the data are rapidly accumulating. In this picture, there are many possible energy states of the network corresponding to different patterns of combinations of activation and inhibition of the protein states. Each check point can be viewed as basins of attractions of globally low energy states. The G1 phase states has the lowest global energy since it is the end of the cycle. We think it might be possible to describe the cell cycle as the dynamic motion in the energy landscape state space from one basin to another. This kinetic search can not be entirely random but directed since the random search takes cosmological time. The direction or gradient of the landscape is provided from the bias in terms of the energy gap towards the G1 phase. So the landscape picture becomes that there is a funnel towards the G1 state (the bottom of the funnel, we can call that native state). At the end of G1 phase, the network is pumped upon receiving the new start signal or nutrition (with those, the system will stay at G1 and network can not continue the cycling process) to high energy excited states at the top of the funnel (cycling). Then the cell cycle follows as it cascades through the configurational state space (or energy landscape) in a directed way passing several check points (basin of attractions) and finally reach the bottom of the funnel-G1 phase before being pumped again for another cycle (Figure 1.4).

We can see from the above discussions that maximizing the ratio of the potential gap (or the slope) versus the roughness of the underlying potential landscape is the criterion for the global stability or robustness of the network. Only the cellular network landscape satisfying this criterion will be able to form a thermodynamically stable global steady state, be robust (Figure 1.5, 1.6, 1.7), perform the biological

functions with minimal dissipation cost (Figure 1.8) and survive the nature evolution. Similar to protein folding and binding problem [108, 96], a funnelled potential landscape of cellular network emerges. The landscape biases towards the global minimum G1 state and dominates the fluctuations or wiggles in the configurational space. From this picture, at the initial stage of the yeast cell cycle network process, there could be multiple parallel paths leading towards the global minimum G1 state. As the kinetic process progresses, the discrete paths might emerge and give dominant contributions (biological path) when the roughness of the underlying landscape becomes significant (Figure 1.4).

The cellular network with too rough underlying potential landscape can neither guarantee the global robustness nor perform specific biological function. They are more likely to phase out from evolution. The funneled landscape therefore is a realization of Darwinian principle of natural selection at the cellular network level. As we see, the funneled landscape provides an optimal criterion to select the suitable parameter subspace of cellular networks, guarantee the robustness and perform specific biological function with less dissipation cost. This will lead to an optimal way for the network connections and is potentially useful for the network design.

It is worth pointing out that the approach described here is general and can be applied to many cellular networks such as signaling transduction network [46], metabolic network [87], and gene regulatory network [76, 71].

Chapter 2

Least Dissipation Cost as a Design Principle for Robustness and Function of the Cellular Networks

2.1 Introduction

Understanding the function and stability of the cellular network requires a global characterization of the system. The natures of the network have been explored through various experimental techniques [16, 39]. It is found that the cellular networks are often quite stable and robust against intrinsic and environmental perturbations. There are increasing number of bioinformatic studies on the global topological structures of the networks [58], as well as some studies from the physical perspectives through the underlying chemical reactions on the network robustness [39, 89, 54, 37, 53]. Recently, efforts have been made in understanding biological function from the energy landscape perspectives [109, 96, 113, 74, 99, 97, 31, 50]. The advantage of this approach is that both global and local properties of the network can be explored in fluctuating environments [59, 92]. In fact, explicit illustrations of the underlying energy landscape and robustness for MAP Kinase signal transduction, yeast cell cycle and gene regulatory networks have been given recently [98, 97, 31, 50].

These studies provided us insights towards the understanding of the robustness of the network with finite number of deep basins of attractions either through a funnel (one basin of attraction, for MAPK or yeast cell cycle) or multiple funnels

(several basins of attractions, for gene regulatory network). The deep basins of attractions of the landscape might be the result of the evolution selection to perform the biological function and maintain the robustness. The question is then how the networks realize that? Cellular networks are open non-equilibrium systems due to the interactions and exchanges with the environments. At steady state, both the steady state probability or the corresponding landscape and the local flux are needed to characterize the non-equilibrium network. Contrary to the situation for the equilibrium case where the steady state probability can characterize the whole system, the local flux is not necessarily zero in the non-equilibrium case because the detailed balance may not be satisfied. For open non-equilibrium network, there are dissipation costs from the exchange with the environments which can be described using the underlying landscape and the flux for the system. Here we provide a possible evolution scenario: The network may have evolved to minimize the dissipation cost to realize the biological function, robustness and structurally stability against genetic and environmental perturbations. Minimizing the dissipation cost might provide a design principle for evolution selection of biological function and robust network.

2.2 Budding Yeast Cell Cycle

To explore the dissipation cost of the cellular network, we will use the budding yeast cell cycle as an example. We will summarize the previous robustness and landscape investigations [31] in Figure 2.2, to serve as a basis for the current study. A network wiring diagram based on the crucial regulators was constructed [89, 54] as shown on the in Figure 2.2A.

In Figure 2.2A, each protein node [54] is assumed to have only two states $S_i = 1$ and $S_i = 0$, representing the active and the inactive state of the protein. There are 11 protein nodes in the network wiring diagram, and all together 2^{11} states. Each state can be represented by S with a distinct combination of the on and off of the 11 protein nodes of Cln3, MBF, SBF, Cln1-2, Cdh1, Swi5, Cdc20, Clb5-6, Sic1, Clb1-2, Mcm1 represented by $\{S_1, S_2, S_3, ...S_{11}\} = S$. $\rightarrow$ arrows represent positive regulations or activations (1). $\dashv$ arrows represent negative regulations or repressions (-1). The loop represents self degradations to the nodes which are not regulated by others.

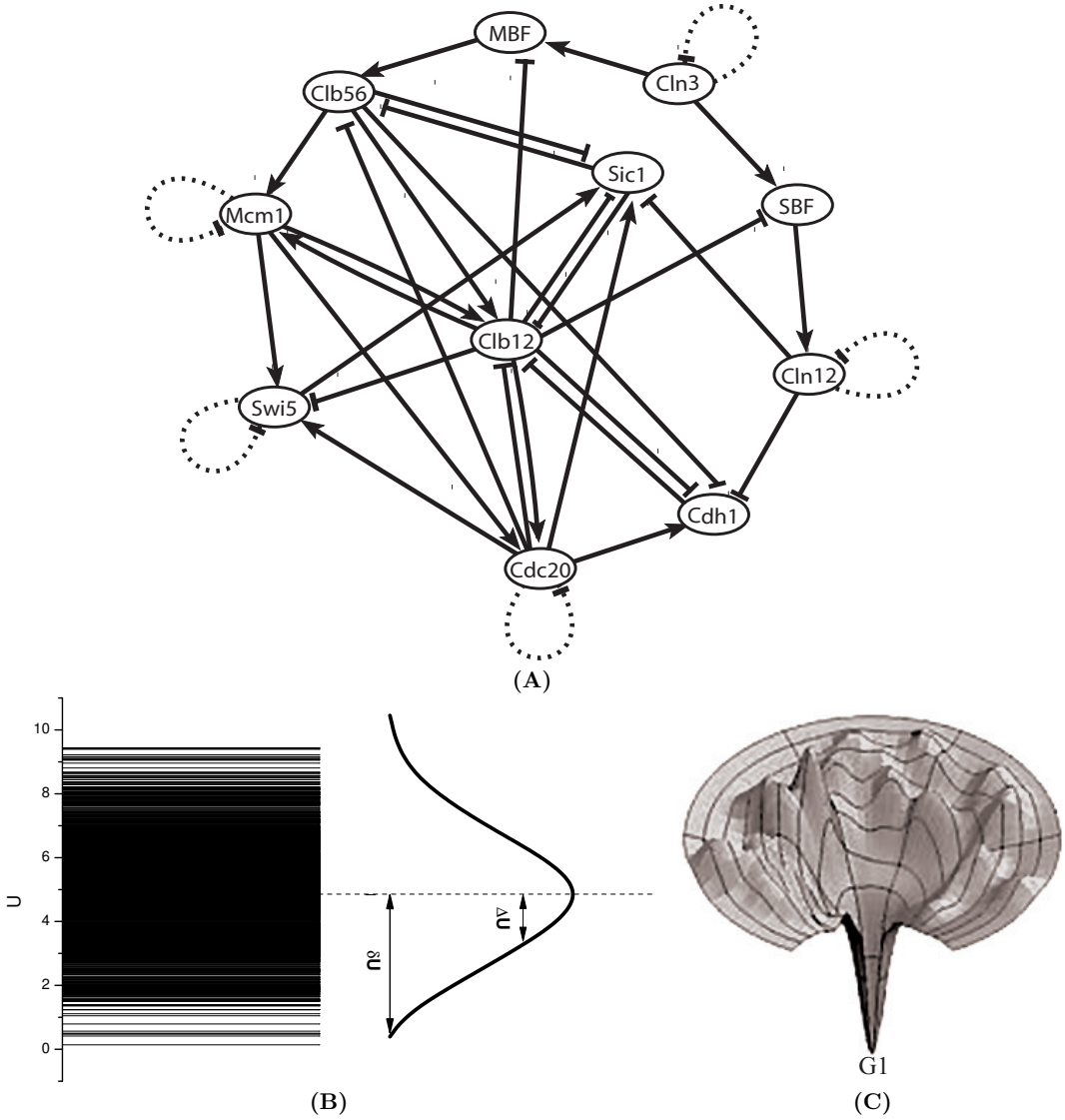


Figure 2.1: A: The Yeast Cell Cycle Network Scheme: Wiring Diagram, → arrow represents positive activating regulations (1); ⊣ arrow represents negative suppressing regulations (-1); loop represents self degradation. B: The spectrum and the histogram or the distribution of the potential energy U. C: An illustration of the funneled landscape of the yeast cell cycle network. The global minimum of the energy is at G1 state.

The significant intrinsic and extrinsic fluctuations within the cell imply that we should follow the probability evolution rather than deterministic dynamics in the network. The transition matrix T can be simplified by assuming the Markovian process, [112, 31]. The Markovian approximations is introduced here for simplicity. There can be time delays for example due to the presence of translation step in addition to transcription. This can be rendered by the introduction of more variables such as mRNA. Therefore we introduce the transition matrix: $T_{\{S_1(t'),S_2(t'),....,S_{11}(t')|S_1(t),S_2(t),....,S_{11}(t)\}} = \Pi_{i=1}^{11} T_{\{S_i(t')|S_1(t),S_2(t),....S_{11}(t)\}}$ where t is the current time and t' is the next moment. The input-output switching response function has a similar form often seen in neural science [34]. The form of the response function although similar to the neural science can actually approximate the nonlinear increase of the protein number production upon regulations of others. It has been widely used in the literature [39, 89]. The transition matrix can be defined as (with nonzero input): $T_{\{S_i(t')|S_1(t),S_2(t),....S_{11}(t)\}} = \frac{1}{2} \pm \frac{1}{2} tanh[\mu \sum_{j=1}^{11} a_{ij} S_j(t)]$. Furthermore $T_{S_i(t')|S_1(t),S_2(t),....S_{11}(t)} = 1 - c$ when there is no input of activation or repression ($\sum_{j=1}^{11} a_{ij} S_j(t) = 0$), c is a small number mimicking the effect of self degradation. Here a_{ij} is the arrow or link representing the activating (+1) or suppressing (-1) interactions between ith and jth protein node in the network which is explicitly shown in the wiring diagram of Figure 2.2. μ is a parameter controlling the sharpness or sensitivity of the response from input to output. $\frac{1}{\mu}$ can also be a measure of the fluctuation strength (for example, mimicking the effects of temperature) [31]. The fluctuation referred in this chapter is a measure of the environmental noise or external noise, not directly the intrinsic noise from protein number fluctuations. When the environmental noise is low, then the response will be sharper.

With the transition probability among different states specified, we can write down the master equation for each of the 2^{11} states as: $dP_i/dt = -\sum_j T_{ij} P_i + \sum_j T_{ji} P_j$ where T_{ij} (T_{ji}) represents the transition probability from state i(j) to state j(i) specified in details above. Here i and j are from 1 to $2^{11} = 2048$ states and $\sum_{i=1}^{i=2^{11}} P_i = 1$.

The steady state probability distribution can be solved numerically [31]. One can link the steady state probability distribution with the generalized potential energy as $U_i = -lnP_i$. [96, 113, 98, 97, 31, 50]

Figure 2.2B shows the spectrum as well as the histogram or distribution of U ($U = -lnP_{steady-state}$). We can see that the distribution is approximately Gaussian.

The global minimum of U was found to be the same state as the fixed point (the stationary G1 state=(0; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0)) for yeast cell cycle [31]. In quantifying the stability or robustness, we previously defined the robustness ratio RR [98, 97, 31] for the network as the ratio of the gap δU, the difference between this global minimum of G1 state $U_{global-minimum}$ and the average of U (mimicking the slope of the landscape), $< U >$ versus the spread or the half width of the distribution of U (mimicking the roughness or trapping of the landscape), When RR is significantly larger than 1, the global minimum (G1 state) is well separated and distinct from the average of the network potential spectrum. Since $P = exp\{-U(x)\}$, the weight or population of the global minimum (G1 state) will be dominant. This leads to the global stability or robustness discriminating against others. It shows a funnel picture of energy going downhill towards G1 state in the evolution of network states, as illustrated in Figure 2.2C. So RR gives a quantitative measure of the shape or topography of the underlying landscape.

The network is an open system in non-equilibrium state. Although we can obtain the steady state probability and can define an equilibrium like quantity, the local flux from j to i ($F_{ji_{steady-state}} = -T_{ij}P_{i\,steady-state} + T_{ji}P_{j_{steady-state}}$) is not necessarily equal to zero (no detailed balance). The flux defines a generalized force for the non-equilibrium state along with the associated generalized chemical potential (from j to i) $A_{ji} = ln(\frac{T_{ji}P_j}{T_{ij}P_i})$ [75, 78, 68]. There is a mapping between the cellular networks and electric circuits. The flux F_{ij} corresponds to current I and chemical potential A_{ij} corresponds to voltage V. The non-equilibrium cell network dissipates energy just as the electric circuits. In the steady state, the heat loss rate is related to the entropy production rate. The entropy production or dissipation characterizes "time irreversibility" and provides a lower bound for the actual heat loss in Boolean network. [75, 78, 74, 68]. The total entropy change is equal to the part from the system or source plus the part from the bath or sink (dissipation). Since in steady state the entropy change of the system is equal to zero, thus the total entropy change (source) is equal to the entropy change of the sink (dissipation). The total entropy change (source) = $\sum F_i j A_i j$ is the entropy production and the sink term is dissipation. Therefore in steady state, knowing the entropy production, we know the dissipation quantitatively. The entropy S from the system part is defined as $S = -\sum_i P_i ln P_i$ and entropy production rate (per unit time) $\frac{dS_{tot}}{dt}$ (system plus bath) is given by: $\frac{dS_{tot}}{dt} = \sum F_{ji}A_{ji} = \sum_{ij} T_{ji}P_j ln(\frac{T_{ji}P_j}{T_{ij}P_i})$. Entropy production rate is a

characterization of the global properties of the network.

2.3 Results and Discussions

Equipped with the quantification of landscape and flux from the previous work [31] as illustrated in Figure 2.2, we are now ready to explore the global nature of the network and the interrelationships between the robustness and stability with the dissipation cost in terms of entropy production under genetic and external perturbations (under different responses μ, self degradations C and mutations).

Figure 2.3A shows the self degradation versus entropy production rate $\frac{dS}{dt}$ (at $\mu = 5$). In [31], we show a less degradation c gives more robust network with large RR. Here we see that less self degradation leads to less dissipation costs and less dissipation cost $\frac{dS}{dt}$ leads to more robust network with large RR and larger stability of the biological path towards G1 phase.

Figure 2.3B shows the robustness ratio RR of the underlying energy landscape as well as the steady state probability of the biological path towards G1 phase versus entropy production $\frac{dS}{dt}$ with different self degradation parameters c (at $\mu = 5$).

We observed in Figure 2.3B a relatively sharp decrease of the robustness through the robustness ratio RR and path probability P_{path} upon changing the self degradation rate c. We noticed from previous studies that both RR and P_{path} drops with the increase of self degradation c (Figure 7 in [31]). So the landscape changes significantly with the self degradation. The entropy production rate is the accumulated effects from the combination of both landscape and flux. Therefore the entropy production rate is in general a nonlinear function of the accumulated effects of landscape and flux. The sharper transition of entropy production rate with respect to the stability upon changing the self degradation might be from the more sensitive dependence on the accumulated landscape and flux.

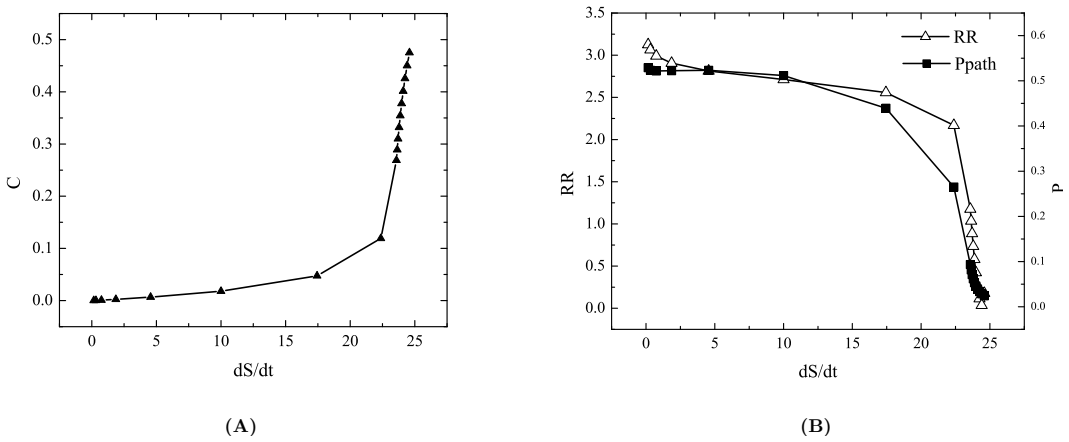
We identified the preferential global pathway towards the global minimum G1 by following the most probable trajectory in each step of the kinetic moves from the deterministic equations of the corresponding to the master equations towards G1. Therefore the global path is referring to the deterministic path without the perturbations and noises. We explored the sum of the probabilities passing through this path in various conditions. When the perturbation or noise is small we expect this path is similar to the actual path. When the perturbation or noise is larger, we

expect the actual path starts to deviate from this path. We expect to see less partition of the probabilities on this path. Obtaining the actual most probable path under various conditions is an important and challenging issue. We plan to investigate that in the future study.

The protein can be either 1 or 0 representing active or inactive. The 11 proteins are arranged in a vector form to represent the state of the system as (Cln3; MBF; SBF; Cln1,2; Cdh1; Swi5; Cdc20; Clb5, 6; Sic1; Clb1,2; Mcm1). The most probable global path follows the states $1 \rightarrow 13$ sequentially towards G1 from the start signal. Where start signal is in state sequence 1 given by: (1; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0). Three excited G1 states are in sequence 2, 3, 4, given respectively by: (0; 1; 1; 0; 1; 0; 0; 0; 1; 0; 0), (0; 1; 1; 1; 1; 0; 0; 0; 1; 0; 0), (0; 1; 1; 1; 0; 0; 0; 0; 0; 0; 0). The S phase is in state with sequence 5 given by: (0; 1; 1; 1; 0; 0; 0; 1; 0; 0; 0). The G2 phase is in state with sequence 6 given by (0; 1; 1; 1; 0; 0; 0; 1; 0; 1; 1). The M phase is in states with sequence 7, 8, 9, 10, 11, given respectively by: (0; 0; 0; 1; 0; 0; 1; 1; 0; 1; 1), (0; 0; 0; 0; 0; 1; 1; 0; 0; 1; 1), (0; 0; 0; 0; 0; 1; 1; 0; 1; 1; 1), (0; 0; 0; 0; 0; 1; 1; 0; 1; 0; 1), (0; 0; 0; 0; 1; 1; 1; 0; 1; 0; 0). The another excited G1 state is with sequence 12 given by (0; 0; 0; 0; 1; 1; 0; 0; 1; 0; 0). Finally stationary G1 phase is in state sequence 13 given by (0; 0; 0; 0; 1; 0; 0; 0; 1; 0; 0). The most probable path turns out to be the biological path going through $G1 \rightarrow S \rightarrow G2 \rightarrow M \rightarrow G1$. [31].

The rational of considering not only the G1 phase but also the whole biological path of the cell cycle is as follows: As the cell cycle progresses, cells have to visit many different states on the biological path, not just stay in one state. The state of G1 phase can be stable, but this may not imply other states of the cell cycle are also stable. To create a robust cell cycle network all important states (local minimums on the potential) have to be investigated not just a stationary G1 phase (the global minimum). Therefore we include the probabilities of all the states on the biological path. In [31], we show a less degradation leads to more weight or stability for G1 phase. Here we see that less self degradation leads to less entropy cost and when entropy production $\frac{dS}{dt}$ decreases, the probability of the whole biological path including the stationary G1 phase also increases. This means, a less dissipation cost gives more probable and stable biological path to stationary G1 phase and therefore more stable and robust network.

As mentioned before, robustness ratio RR here is a measure of the stability of



(A) (B)

Figure 2.2: Robustness and Stability versus Dissipation Cost of the Yeast Cell Cycle Network for Various Protein Self Degradation Rate. A: Self Degradation Rate versus Entropy Production Rate $\frac{dS}{dt}$. B: Robustness Ratio RR and Probability of the Biological Path towards G1 phase versus Entropy Production $\frac{dS}{dt}$ for Various Protein Self Degradation Rate.

the global minimum G1. If we consider the whole biological path, then the 13 local minimum states visited by the biological path should be grouped together. We expect a positive correlation between the stability of the global minimum G1 measured by RR and that of 13 local minimum constituting the whole biological path.

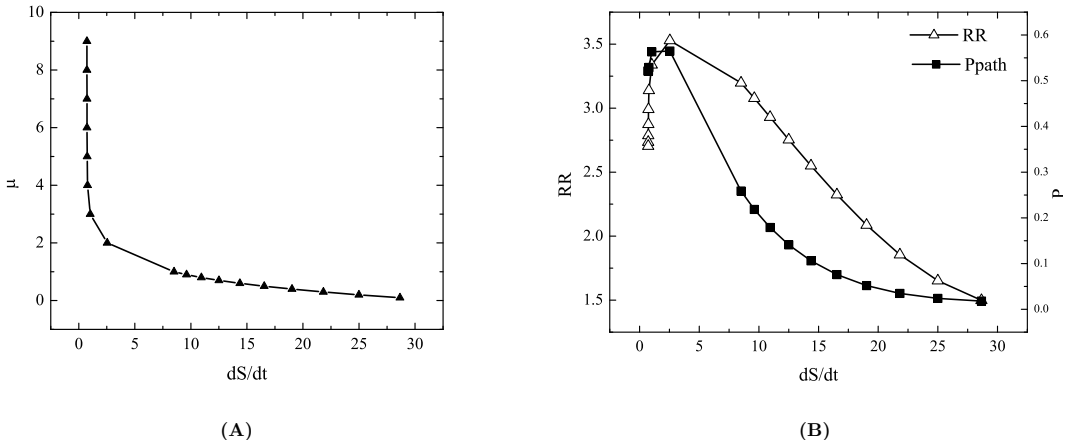
In Figure 2.3A, we plotted the μ versus entropy production (per unit time) or the dissipation cost of the network, $\frac{dS}{dt}$. In [31], we show sharper response or less noise (larger μ) leads to more weight or stability for biological path including G1 phase. Here we see that sharper response and less noise leads to less entropy cost, and when entropy production $\frac{dS}{dt}$ decreases, the probability of biological path including stationary G1 phase also increases.

In Fiure 2.3B, we plotted the robustness of the network, RR as well as the steady state probability of the biological path towards G1 phase versus entropy production (per unit time) or the dissipation cost of the network, $\frac{dS}{dt}$, for different μ (fixed $c = 0.001$). In [31], we show sharper response or less noise (larger μ) in general leads to more robust network with larger RR. We also show in that the presence of the peak of RR is due to traps in the landscape [31]. On the right hand side of the peak of RR and steady state probability of the biological path towards G1 phase in Figure 2.3B, as the noise increases (μ decreases in Figure 2.3A), the network quickly becomes unstable (smaller RR below 2.5 and probability of the biological

path is below 0.3) and dissipation cost increases significantly. Here, we can see when the entropy production rate decreases, the robustness RR and probability of the biological path towards G1 increases. The left hand side of the peak of RR and steady state probability of the biological path towards G1 in Figure 2.3B corresponds to a stable regime. The peak represents the most stable state. At zero noise limit, there are no fluctuations, so the system can easily get trapped in the local minimum of energy (local maximum of the steady state probability). Increasing the noise level slightly from zero (larger μ) can help to adjust the system by overcoming the local traps to reach to the global minimum of energy [31]. Above all, we can see, a sharper response or less noise environments normally leads to less dissipation, the less dissipation cost gives more probable and stable biological path and stationary G1 phase and therefore more stable and robust network in general.

We also in Figure 2.3B observed a relatively smooth decrease of the robustness through the robustness ratio RR (from high values to be below 2) and path probability P_{path} (from high values to be below 0.2 upon changing the response μ (from high values to be below 1). We noticed from previous studies that both RR and P_{path} drops sharply with the decrease of response μ (Figure 6 in [31]). So the landscape changes significantly with the response. As mentioned above, the entropy production rate is the accumulated effects from the combination of both landscape and flux. So the entropy production rate is in general a nonlinear function of the accumulated effects of landscape and flux. The smoother transition of entropy production rate with respect to the stability upon changing the response might be from the less sensitive dependence on the the accumulated landscape and flux.

Figure 2.3A shows the steady state probability of the G1 (with $\mu = 5$ and $c = 0.001$) versus dissipation cost of the network against various mutations or perturbations through deleting an interaction arrow, adding an activating or repressing arrow between the nodes that are not yet connected in the network wiring diagram in Figure 2.2, or switching an activating arrow to a repressing arrow or vice versa, and deleting an individual node. Upon mutations, when the entropy production is smaller (larger), the G1 state tends to be more (less) stable and dominating. This is the regime where the underlying energy landscape is a funnel. We noticed that there are steady states with low entropy and low probability. Those are the outliers. They correspond to perturbed underlying energy landscape which are either not very stable and unable to perform biological functions, or possibly become sig-



(A) (B)

Figure 2.3: Robustness and Stability versus Dissipation Cost of the Yeast Cell Cycle Network for Different Responses or Noises. A: Response μ versus Entropy Production Rate $\frac{dS}{dt}$. B: Robustness Ratio RR as well as Steady State Probability of the Biological Path Towards the G1 Phase versus Entropy Production $\frac{dS}{dt}$ for Different Response or Noise.

nificantly perturbed cell cycles (i.e. cycles without stable G1 phase as in wild type fission yeast cells). This implies that dissipation cost and stability of the network through G1 state might be more correlated in relatively high stability region while less correlated in low stability region. In other words, since low stability region often corresponds to more fluctuating region, the dissipation cost is a less reliable measure of the network property. Therefore in this outlier regime, we may need to use both dimensions to explore the network, one for stability and one for function through dissipation cost.

Figure 2.3B shows the P_{G1} versus robustness ratio under various of mutations mentioned above. We see that larger (smaller) RR corresponds to larger (smaller) P_{G1}. Since less entropy production leads to more stable network (larger steady state probability of G1, P_{G1}), therefore less entropy dissipation also leads to larger RR and therefore more robust network. Random networks typically have smaller RR and smaller probability of G1 compared with the biological one, corresponding to rough underlying energy landscape. They are less stable and robust. The biological functioning network is quite different from the random ones in terms of the underlying energy landscape and stability. In the low RR region, the robustness and entropy dissipation is less correlated. These networks are less of biological relevance.

The exploration of the relationship among statistical fluctuations, stability, ro-

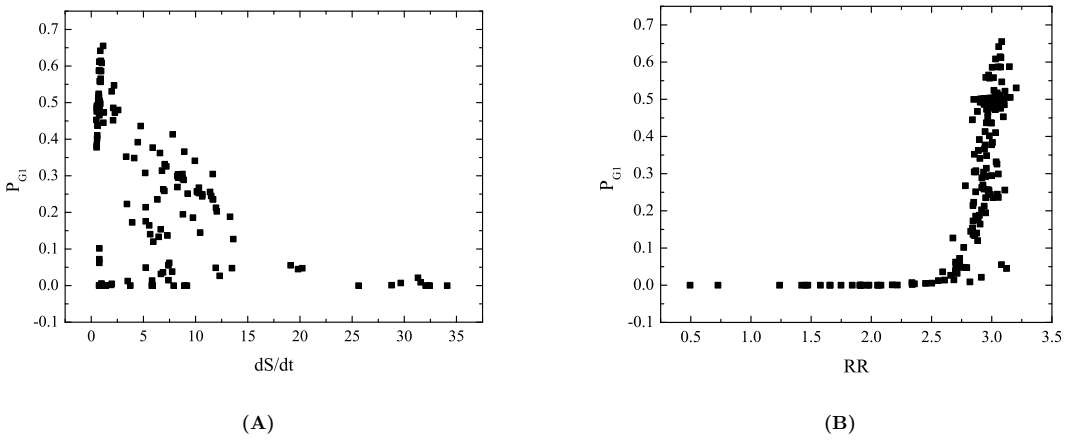


Figure 2.4: Stability and Robustness versus Dissipation Cost of the Yeast Cell Cycle Network for Different Perturbations through Mutations. A: Steady State Probability of G1, P_{G1} versus Entropy Production Rate $\frac{dS}{dt}$ for Different Mutations. B: Robustness Ratio RR versus Steady State Probability of G1, P_{G1} for Different Mutations.

bustness and dissipation cost of the network here can be important for the network design. The nature might evolves such that the network is robust against internal (intrinsic) and environmental perturbations, and perform specific biological functions with minimum dissipation cost. From evolution point of view, the fact that robustness and stability are often correlated with the entropy production rate may reflect the fact that more cost saving requires the system to have less fluctuations and perturbations, leading to more robust and stable network. This may provide us a design principle of optimizing the connections of the network with minimum dissipation cost. In our study here this is also equivalent of optimizing the robustness or stability of the network. The less dissipation cost or robust landscape therefore might be a quantitative realization of Darwinian principle of natural selection at the cellular network level. The nature might evolve such that the biological networks become robust against perturbations, and perform specific biological functions with minimum dissipation cost. The dissipation criterion was also used in the context of constrained-based modeling of metabolic network.

Probing a non-equilibrium network is crucial for uncovering the mechanisms. We believe the dissipation cost is a quantitative measure of non-equilibrium property which can be used to effectively show the noise level of the inherent system as seen in Figure 2.3A.

It is worthwhile to mention there is a difference between cellular networks and

protein folding/binding problem [109, 96]. In protein folding and binding, people typically assume quasi-equilibrium condition and system obeys the usual detailed balance conditions [109, 96]. Therefore one can define the usual energy and potential. On the contrary, cellular networks are in non-equilibrium state. There is no apparent energy or potential to use. We pointed out that we could still define the landscape as the $-logP_{steady-state}$, but we also need to take into account of the flux in addition to characterize the whole non-equilibrium network. The entropy production rate which combines both the information of steady state probability P and local flux can be used to globally quantify the non-equilibrium networks. An interesting and challenging question is to study the dissipation along the biological path. We will address this in a separate study.

Chapter 3

Funneled Flux Landscape Determines the Stability and Robustness of the Oscillation Networks with an Example of Budding Yeast Cell Cycle

3.1 Introduction

The global stability and robustness are crucial for uncovering underlying mechanisms of networks. However, it is very difficult to quantify them for dynamic systems and networks. This presents a great challenge for the dynamical systems and the field of systems biology. [42] [23] [109] [96] [29], [62] [76] [3] [5][69, 70] [98, 50, 30, 32, 52, 101].

In equilibrium systems, the global nature of the system is characterized by the underlying potential landscape U which is usually known (the energy function) and given a priori. The potential landscape U is directly linked with the probability P through the Boltzmann equilibrium distribution law $P \sim exp[-\beta U]$. The local dynamics is determined by the gradient of the potential landscape.

The dynamical systems however, does not typically have a gradient potential as in the equilibrium case. Furthermore, it is not clear how the dynamics is determined. So global natures are hard to address. The dynamic systems however

are not in isolations. They constantly interact with the environments. The intrinsic fluctuations also might emerge for mesoscopic systems [84]. Stochastic description is therefore more appropriate. Instead of following the deterministic dynamical trajectories, one needs to follow the evolution of the probabilistic distributions, which is inherently global. The probabilistic evolution is governed by master equations for discrete space (more general) and Fokker-Planck equations for continuous space.

It turns out the steady state distribution of the probability evolution in long time limit can give a global quantification of the dynamical systems . [23] [109] [96] [29][62] [76] [3] [5][69, 70] [98, 50, 30, 32, 52, 101]. This defines a probability landscape for characterizing the system. On the other hand, the dynamics of the systems can be decomposed to gradient of the potential landscape related to the steady state probability distribution and a curl probability flux. The existence of a non-zero curl flux correlates with the breakdown of the detailed balance. It quantifies the degree of the non-equilibriumness. [101]. While this decomposition is shown explicitly in continuous space through Fokker-Planck equation description of the stochastic dynamics, the corresponding decomposition of stochastic dynamics in discrete space from the master equation still needs further explorations [73, 26].

In this study, we study the more general stochastic dynamics in discrete space of the non-equilibrium networks (Markov chains) governed by the probabilistic master equations. We found the network dynamics and global properties are determined by two features: the potential landscape and the probability flux landscape. While potential landscape quantifies the probabilities of different states forming hills and valleys, the probability flux landscape is composed of many flux loops, with each loop going through some states and carrying a probability flux value going through the loop. The flux landscape quantifies the probability fluxes of different loops through states. These two landscapes can be quantitatively constructed through the decomposition of the dynamics into the detailed balance part and non-detailed balance part. These two landscapes can be quantitatively constructed through the decomposition of the dynamics into the detailed balance part and non-detailed balance part.

We found that while funneled landscape [109] [96] is crucial for the stability of the single attractor networks, the funneled flux landscape is crucial for stable oscillation networks. The stability and the robustness of the networks can be quantified through a dimensionless ratio of the steepness against the averaged variations or

roughness of the landscape (which measures the degree of funnelness, we termed as robustness ratio RR) versus the changes of the network topologies and stochastic fluctuations.

This flux landscape picture may provide a new interpretation of the origin of the limit cycle oscillations: There are always many cycles and loops forming the probability flux landscapes, each with a probability flux value going through the loop. With the homogeneous landscape, each loop carries similar values of probability flux. No global oscillations will be seen, because each loop carries equal and small probability flux and no individual loops are highly probable. The global oscillation only emerges when one specific loop stands out and carries much more probability flux and therefore more probable than the rest of the others. This happens when the nonlinearity of the inherent dynamical systems increases.

We specific studied the budding yeast cell cycle as an example to illustrate the idea. We found the flux landscape of the budding yeast cell cycle oscillations funneled, guarantees its stability and the robustness of the oscillations. This is quantified by the robustness ratio RR of the flux funnel landscape with respect to the changes in topology of the network (wirings) and stochastic fluctuations. The landscape analysis here allows us to identify the key factors and structure elements of the networks in determining the global robustness and stability of the budding yeast cell cycle oscillations.

Theory and Methodology

3.1.1 Decompositions of the dynamics into detailed balance part with potential and non-detailed balance part with probability flux loops

The master equation describes the probability evolution of the stochastic dynamics of non-equilibrium systems in discrete state space (Markov chains) [85, 25]. So the state variables take discrete instead of continuous values. The corresponding master equation for the probability evolution is:

$$\frac{dP_i}{dt} = -\sum_j T_{ij}P_i + \sum_j T_{ji}P_j, \tag{3.1}$$

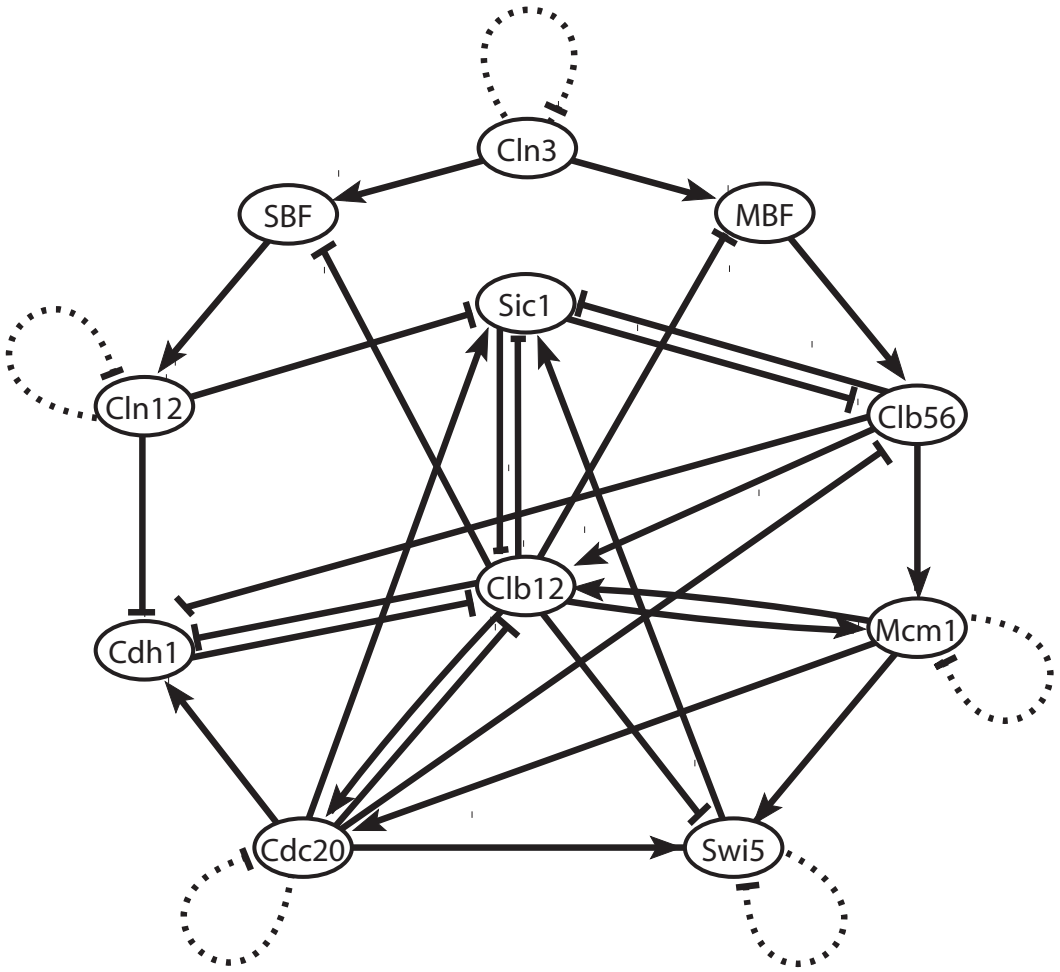


Figure 3.1: The yeast cell cycle network wiring diagram: wiring diagram, the arrow sing(⟶) represent positive activating regulations(1); the inhibition sign(⊣) represents negative suppressing regulations(-1); and the loop sign(–|) represent self-degradation.

Where P_i represents the probability of state i, and T_{ij} represents the transition probability from state i to state j. The physical meaning of the master equation is the conservation law of probability: the local change of the probability of a particular state i in time is equal to the probability flow (flux) from the other states to this state i given by $\sum_j T_{ji}P_j$ subtracting the probability flow (flux) from the state i to other states $\sum_j T_{ij}P_i$.

We define the steady state flux between state i and j as: $F_{ij_{steady-state}} = -T_{ij}P_i^{(ss)} + T_{ji}P_j^{(ss)}$, where $P_i^{(ss)}$ is the solution of 3.4 in the long time limit (when $\frac{dP_i}{dt}$). If for any i, j, $F_{ij_{steady-state}} = 0$, this Markov chain is called to be detail-balanced, and the steady state of the system becomes the equilibrium state (no flux), since $dp_i/dt = \sum_j F_{ij_{steady-state}} = 0$. However, in general the steady state probability can be obtained, but it does not have to satisfy the detailed balance condition($F_{ij_{steady-state}} \neq 0$),and the system is in non-equilibrium steady state. In other words, the local steady state probability flux is not necessarily zero. Although the steady-state distribution is fixed and does not change in time, there can be an internal probability flow among states.

In order to study the non-equilibrium steady states and characterize the global properties, one can separate the dynamical process into two parts, a detailed balance part and a pure irreversible non-detailed balance flux part by decomposing the transition probability matrix M [73]. The master equation can be rewritten as $dP/dt = M^T P$, where P is the vector of probability of all the discrete states, M is the transition probability matrix (or rate matrix) with $M_{ij} = T_{ij}, i \neq j$ and $M_{ii} = (-1)\sum_j T_{ij}$. We define a matrix C such that the ith row and jth column of it is $C_{ij} = max\{T_{ij}P_i^{(ss)} - T_{ji}P_j^{(ss)}, 0\}/P_i^{(ss)}, i \neq j$ and $C_{ii} = (-1)\sum_j C_{ij}$, and matrix D whose ith row and jth column is $D_{ij} = min\{T_{ij}P_i^{(ss)}, T_{ji}P_j^{(ss)}\}/P_i^{(ss)}, i \neq j$ and $D_{ii} = (-1)\sum_j D_{ij}$. It follows that $M = C + D$ and $D^T P^{(ss)} = 0$. Since $M^T P = (C + D)^T P^{(ss)} = 0$, $C^T P^{(ss)} = 0$. By separating the transition probability matrix this way, two Markov processes are obtained [73]. So the probability transition matrix (or rate matrix) M for characterizing the dynamics can be decomposed into two terms: C and D. Both C and D have the same steady-state(stationary) probability distribution, and one of the processes D satisfies detailed balance($D_{ij}{P_i}^{ss} = D_{ji}{P_j}^{ss}$), while the other C is non-detailed balanced and irreversible (if $C_{ij}{P_i}^{ss} > 0$,$C_{ji}{P_j}^{ss} = 0$). In this way, the dynamics is decomposed to detailed balance part and non-detailed balanced part.

The non-equilibrium irreversible part is usually called circulation or flux part, since it can be further decomposed into flux circles or loops with a flux value on each circle [73]. The prove of the circulation also provides a way to obtain all the circles and their corresponding flux values for the dynamic part. By definition of flux, we have $F_{ij} = -C_{ij}P_i^{(ss)} + C_{ji}P_j^{(ss)} = -F_{ji}$. Now define $J_{ij} = C_{ij}P_i^{(ss)}, i \neq j; J_{ii} = 0$, we have $\sum_i J_{ij} = \sum_i J_{ji}$. Since $Jii = 0$, suppose $J_{k_0k_1} > 0$, from the summation equation just mentioned, we can find a $k_2 \neq k_0, k_1$ such that $J_{k_1k_2} > 0$. We can keep on doing this, until a repeat is found: $k_n \in \{k_0, k_1, ..., k_{n-2}\}$. Suppose $k_n = k_{n_0}$, let $i_1 = k_{n_0}, i_2 = k_{n_0} + 1, ..., i_{n_1} = k_{n-1}$ and $i_{n_1+1} = i_1$, we now construct a circle or closed loop with $i_1, ..., i_{n_1}$. Let $r_1 = min_{k=1,2,...,n_1}\{J_{i_k i_{k+1}}\}$, define r_1 as the flux value of this circle. Then subtract their flux value from the whole J matrix, J, thus $J_{ij}^{(1)} = \begin{cases} J_{ij} - r_1, i \in \{i_1, ..., i_{n_1}\}, \\ J_{ij}, otherwise. \end{cases}$ If $J^{(1)} \neq 0$ (all the elements of $J^{(1)}$, repeat what we did above to find another circle as well as its flux value, then subtract those fluxes from $J^{(1)}$ to get $J^{(2)}$. Since the number of non-zero elements in $J^{(i)}$ is at least one less than that in $J^{(i-1)}$, there exit an integer N such that $J^{(N+1)} = 0$ (all the elements of the J matrix are zero). Therefore the non-detailed balanced part of the dynamics, the flux can be decomposed of finite number of circles or closed loops, each with a flux value, $J = \sum_{i=1}^{i=N} J^i$ [73].

Therefore we have two quantitative features to characterize the system, one is the steady state probability and the other is the non-zero flux which can be further decomposed into loops. The steady state probability obeys the evolution equation of the transition probability matrix (or rate matrix) with only the detailed balance part. The detailed balance condition allows one to identify the path independent probability measures [114]. This naturally leads to the potential. We can see how both potential and flux landscape influence the dynamics and stability of the system through an example on budding yeast cell cycle.

3.1.2 The stochastic probabilistic model of budding yeast cell cycle

The average dynamics of the biological network dynamics can be usually described by a set of chemical rate equations for concentrations where both the concentrations and the links among them through binding rates with typically quite different timescales are treated in a continuous fashion. For some network systems (for

example, some gene regulation networks), the biological functions can be approximately described by the on- and off-properties of the network components. Further more, the global properties of the network might depend less sensitively on the details of the model. Therefore, a simplified Boolean representation [44] can be proposed with each node i having only two states, $s_i = 1$ and $s_i = 0$, representing the active and the inactive state of the protein, or high concentration and low concentration of proteins, respectively. As illustrated in Fig. 3.1, we have 11 protein nodes in the network of budding yeast cell cycle wiring diagram [54]. Altogether, we have 2^{11} states. Each state S with a distinct combination of the on and off of the 11 protein nodes (the key regulators for the underlying cell cycle process) of Cln3, MBF, SBF, Cln1-2, Cdh1, Swi5, Cdc20, Clb5-6, Sic1, Clb1-2, and Mcm1 is represented by $\{s_1, s_2, s_3, ..., s_{11}\} = S$. The arrows ($\longrightarrow$) represent positive regulations or activations (1). Inhibition sign (——|) represent negative regulations or repressions (-1). The loop ($- - |$) represents self-degradations to the nodes which are not regulated by others. We can then define some rules to follow the subsequent dynamics of the network. Therefore, the evolution of the network is deterministic.

Due to the intrinsic fluctuations from the limited number of the proteins in the cell and extrinsic fluctuations from the environments in the interior of the cell [84], it is then more appropriate to approach the network dynamics based on statistical description. In other words, we should use a probabilistic description of the evolution of the cellular network dynamics rather than the deterministic or average description of the dynamics of states in cellular networks. Therefore, instead of following the on- and off-state switching in the network, we follow the probability of on and off for each state in the network [30, 32].

To follow the evolution of the states in the cellular network, we need to first figure out the transition probability from one state S at present time to another state S' at the next moment. This is difficult to study and in general almost impossible. We therefore will make some simplifications, so that we can handle the case without the loss of the generality, by assuming that the transition probability T from one state to another, can be split into the product of the transition probability for each individual flip (or no flip) of the on- or off-state from this moment to the next moment. The transition probability from one state at current state to another at next moment will be assumed not to depend on the earlier times (no memory). This leads to the

Markovian process [24, 30, 32]. The transition matrix T can thus be written as

$$T\Big(S(t') = \{s_1(t'), s_2(t'), ..., s_{11}(t')\}\Big|S(t)\Big) = \prod_{i=1}^{11} T\Big(s_i(t')\Big|S(t)\Big), \tag{3.2}$$

where t is the current time and t' is the next moment. So the whole transition probability from current state to the next is split into the product of the transition probability of each individual flip (or no flip) of the node i. For each individual flip, the transition probability for a particular node can be modeled as a nonlinear switching function, mapping the input through the interactions to the output, which is often used in neural science[35]. The input is defined as $I = \sum_{j=1}^{11} a_{ij} S_j(t)$, where a_{ij} is the arrow or link representing the activating (+1) or suppressing (-1) interactions between i^{th} and j^{th} protein node in the network, which is explicitly shown in the wiring diagram of Fig. 3.1. The transition probability of a single node is:

$$T(s_i(t') = s_i(t)|S(t)) = \begin{cases} \frac{1}{2} + \frac{1}{2}tanh[\mu I], & \text{if } I \neq 0 \text{ and } s_i(t') = 1, \\ \frac{1}{2} - \frac{1}{2}tanh[\mu I], & \text{if } I \neq 0 \text{ and } s_i(t') = 0, \\ \frac{1}{1-e^{-\gamma}}, & \text{if } S(t) = G1 \text{ and } i = 0, \\ 1 - c, & \text{otherwise,} \end{cases} \tag{3.3}$$

where c is is a small number mimicking the effect of self-degradations when the total input to a node is zero and γ is the number to control the positive feedback or the excitation (finite transition probability) from the global checkpoint G1 to the start signal of the cycle cln3 when G1 is reached. This means we add specifically a kinetic excitation once the global G1 check point is reached. The rational for doing so is from the biology that if there is no nutrition supply constantly pumping into the system, the cell cycle will stop and not continue. So the origin of finite transition probability from G1 to the start of the cycle state of Cln3 mimicking the excitation to maintain the cycle, is from nutrition supply. The value μ is a parameter controlling the width of the switching function from the input to the output. The physical meaning is clear. If the inputs through the interactions among proteins to a specific protein node in the network are large enough, then the state will flip, otherwise the state will stay without the flip. If μ is small (large), the transition width is large (small), the transition is smooth and linear (sharp or sensitive and nonlinear) from the original state to the output state. Therefore, we have an analytical expression of the transition probability.

With the transition probability among different states specified, finally we can write down the master equation for each of the 2^{11} states as

$$\frac{dP_i}{dt} = -\sum_j T_{ij}P_i + \sum_j T_{ji}P_j, \tag{3.4}$$

where T_{ij} represents the transition probability from state i to state j specified in details above. Here i and j are from 1 to $2^{11} = 2048$ states and $\sum_{i=1}^{i=2^{11}} P_i = 1$ We solved the $2^{11} = 2048$ master equations numerically of the yeast cell cycle (by using iterative method) to follow the evolution of the probability distribution of each state, with the initial condition of equal small probability of all the cell states ($P_i = 1/2048$). Both the time-dependent evolution and the steady-state probability distribution for each state are obtained.

3.1.3 Entropy Production and Dissipation

The network is an open system in non-equilibrium state. Although we can obtain the steady state probability and can define an equilibrium like quantity, the local flux from j to i ($F_{ji_{steady-state}} = -T_{ij}P_{i_{steady-state}} + T_{ji}P_{j_{steady-state}}$) is not necessarily equal to zero (no detailed balance). The flux defines a generalized force for the non-equilibrium state along with the associated generalized chemical potential (from j to i) $A_{ji} = ln(\frac{T_{ji}P_j}{T_{ij}P_i})$ [75, 78, 69]. There is a mapping between the cellular networks and electric circuits. The flux F_{ij} corresponds to current I and chemical potential A_{ij} corresponds to voltage V. The non-equilibrium cell network dissipates energy just as the electric circuits. In the steady state, the heat loss rate is related to the entropy production rate. The entropy production or dissipation characterizes "time irreversibility" and provides a lower bound for the actual heat loss in Boolean network. [75, 78, 74, 69]. The total entropy change is equal to the part from the system or source plus the part from the bath or sink (dissipation). Since in steady state the entropy change of the system is equal to zero, thus the total entropy change (source) is equal to the entropy change of the sink (dissipation). The total entropy change (source) = $\sum F_i j A_i j$ is the entropy production and the sink term is dissipation. Therefore in steady state, knowing the entropy production, we know the dissipation quantitatively. The entropy S from the system part is defined as $S = -\sum_i P_i ln P_i$ and entropy production rate (per unit time) $\frac{dS_{tot}}{dt}$ (system plus bath) is given by: $\frac{dS_{tot}}{dt} = \sum F_{ji}A_{ji} = \sum_{ij} T_{ji}P_j ln(\frac{T_{ji}P_j}{T_{ij}P_i})$. Entropy production rate is a

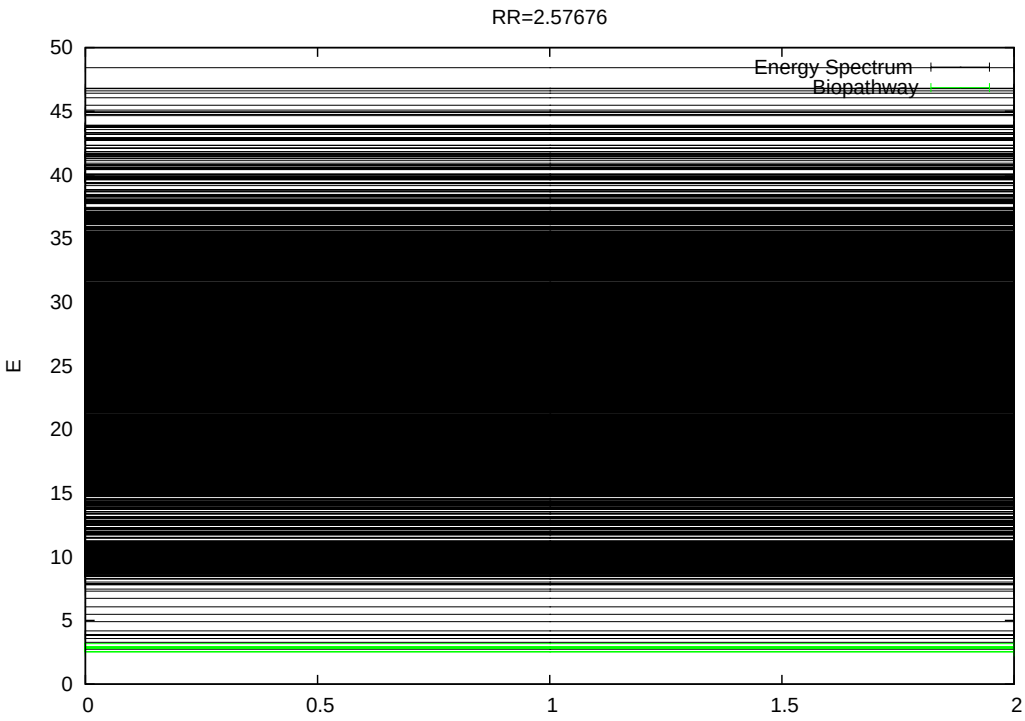


Figure 3.2: Potential landscape spectrum E, of the 2^{11} states, where $\mu = 5$, $c = 0 : 0001$ and $\gamma = 10$. The potential values of those 13 states of the biological pathway are in green lines. They are lower than the rest of the states.

characterization of the global properties of the network.

3.2 Results and Discussions

Since the dynamics can be decomposed into the detailed balance part quantified by the steady state probability and the non-detailed balanced part quantified by the fluxes, we start the exploration of the underlying potential and flux landscapes of the budding yeast cell cycle network.

3.2.1 Potential Landscape of Budding Yeast Cell Cycle

Fig. 3.2 shows the spectrum of the potential landscape U of the 2^{11} states as negative logarithm of steady state probability $U = -lnP_{ss}$, where $\mu = 5$, $c = 0.0001$ and $\gamma = 10$. 13 states are found to have distinct weights (steady state probability) [30, 32]. They are identified as biological paths for the system. The potentials of

the 13 states of the biological pathway are presented in green lines as shown in the spectrum. They are lower than the rest of the states. Since lower potential means higher probability, the dynamics tends to converge to the biological pathway. To quantify the topology of underlying potential landscape, we define the Robustness Ratio RR as the potential gap between the low potentials of the states (the 13 states on the biological path) and the average potentials of the rest of the states, versus the average variance of potentials ($RR = \delta U/\Delta U$ where $\delta U = |U_m - < U >|$ and $\Delta U = \sqrt{< U^2 > - < U >^2}$).

For our model of budding yeast cell cycle with excitation, we obtain $RR = 2.36757$. This shows that states along the biological path are separated from the rest. If the states of the biological paths are not separated from the rest of the other states, then the biological paths will be indistinguishable from the rest of the others. Yet, biological paths (going through stages of the cycle G1, SG2, M etc) are required to perform the biological function which is the cell cycle with high probability of appearance. Therefore, the separation of biological path from the rest of the other states guarantees the biological function of the cell cycle and stability of the system (with biological path of higher probability).

In order to illustrate the feature of the potential landscape, the U is shown in Fig. 3.3. 2048 states are projected into 2D square lattice, following the rule of minimizing the lengths of strongest connection. The vertical axis and color represent the potential level of each state in both the potential surface (3d) and the contour map (2d) laying on the bottom. We can see that the potential landscape shows a distinct topology with Mexican hat like close ring valley shape. Outside of the ring valley, the potential is high, and on the ring the potential is low. Therefore, the system has a tendency to be attracted down to the ring. As seen, the close ring valley has low potentials which corresponds to exactly the biological path. This can also be seen more clearly on the contour map. The biological path completes the circle or oscillation of the budding yeast cell cycle.

In our earlier study of this system without explicitly put in the excitations from G1 ground state to the Cln3 start signal [30, 32], we see quite different dynamics and landscape. There the potential landscape has a funnel shape. The system has one dominant basin of attraction pointing towards G1. The model can explain the dynamical process of budding yeast cell cycle once the start signal kicks off. Since the end state is always the G1 state (bottom of the funnel or basin of attraction), it

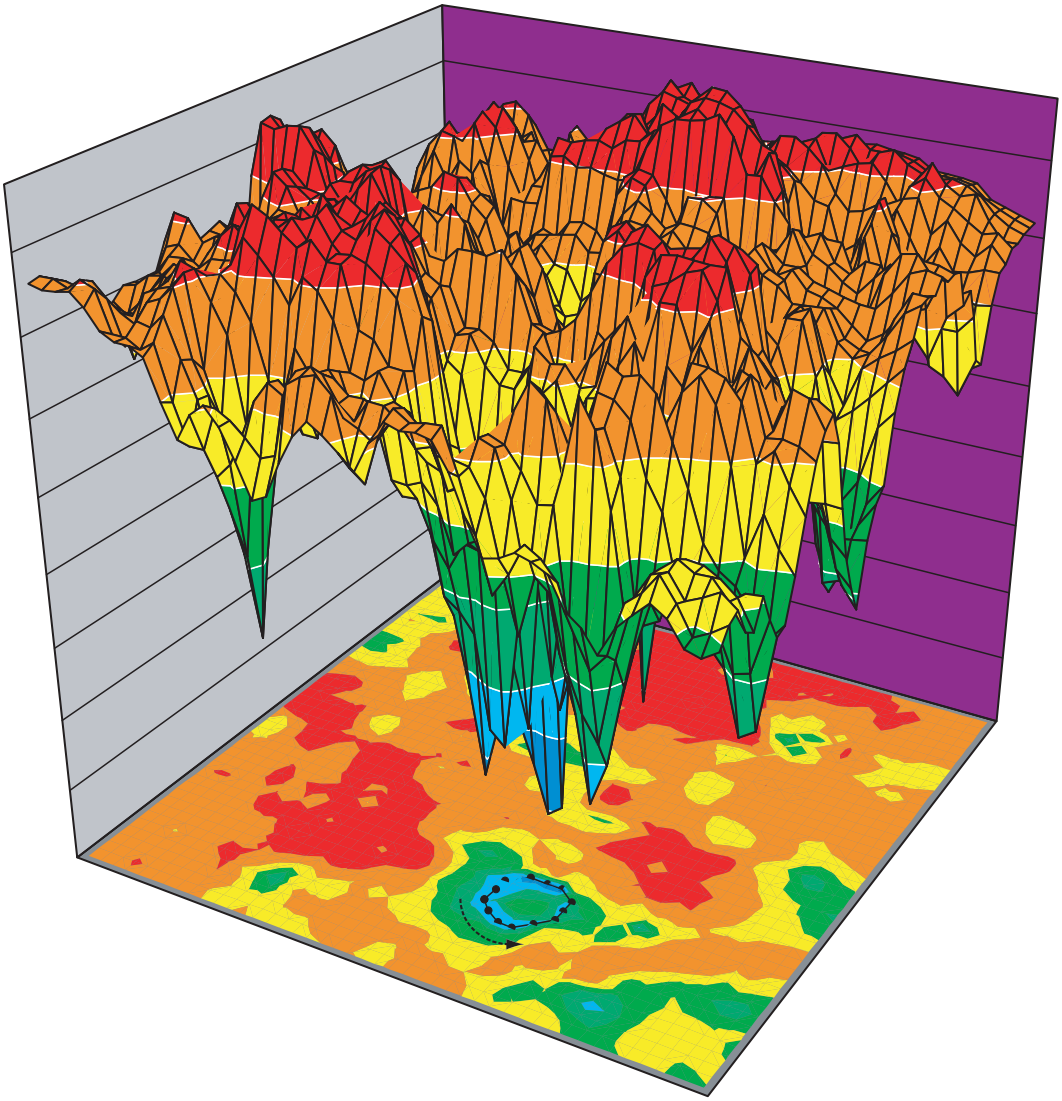


Figure 3.3: Three dimensional potential landscape and two dimensional contour in projected 2 dimensional state space. The vertical axis and color represent the potential level of each state in both the three dimension and the contour map laying on the bottom. The low potential valley of the potential is a circle or closed ring, which is exactly the biological cycle path with low potential level, and this can also be seen more clearly on the contour map.

does not explain the cycle part. With the excitation explicitly put in, we now see the potential landscape changes the shape from single attractor funnel which only explains the check point G1 behavior to Mexican hat shape which naturally leads to a oscillations dynamics. Clearly, the oscillation is maintained to be stable due to two driving forces: the potential landscape which tends to drive the system to the close ring valley and the flux flow part (which we will discuss in details below) originated of from excitation of G1 back to start signal to maintain the cycle on the close ring valley. Both are essential for the stability of the budding yeast cell cycle.

3.2.2 Robustness of potential landscape against parameter changes

To study how robust of the network is to the parameters, the *RR* is calculated with different parameter sets. By fixing two out of three parameters and changing the rest one, one can find out how the robustness ratio of the network or the potentials of the biological cycle path changes. Fig. 3.4(a) illustrated that large μ with more sharp switching rate of the transition probability or less fluctuation will lead to higher probability of the biological cycle path and make it more stable. Fig. 3.4(b) shows the robustness ratio of the underlying potential landscape versus different switching parameters μ under fixed $c = 0.001$ and $\gamma = 10$. As we see the robustness ratio increases with μ increases, which means a sharper transition or response from input to output gives more robust network compared with the smoother transition or response. The value μ can also be seen as a measure or characterization of the strengths of the noise from the intrinsic or extrinsic statistical fluctuations in the cellular environments [30]. The μ could then be related to the inverse of the temperature (temperature here is a measure of the strength of the noise level). The slight turnover of RR for large μ is caused by the existence of the other traps of the landscapes. These traps with lower probabilities are not biological and are the artifacts of the Bollean model. With slightly more fluctuations, the traps will be less stable and have lower probability, leading RR to increase (gap larger and variance smaller). Fig. 3.4(c) and Fig. 3.4(d) show the entropy production mimicking the heat dissipation of the system versus μ and RR. We can see that the sharper the switching is, and therefore the more stable the oscillation is, the more dissipation cost is. The stable oscillation needs dissipation costs or energy consumptions to maintain it.

Fig. 3.5(a) shows the probability of the biological circle versus different self-

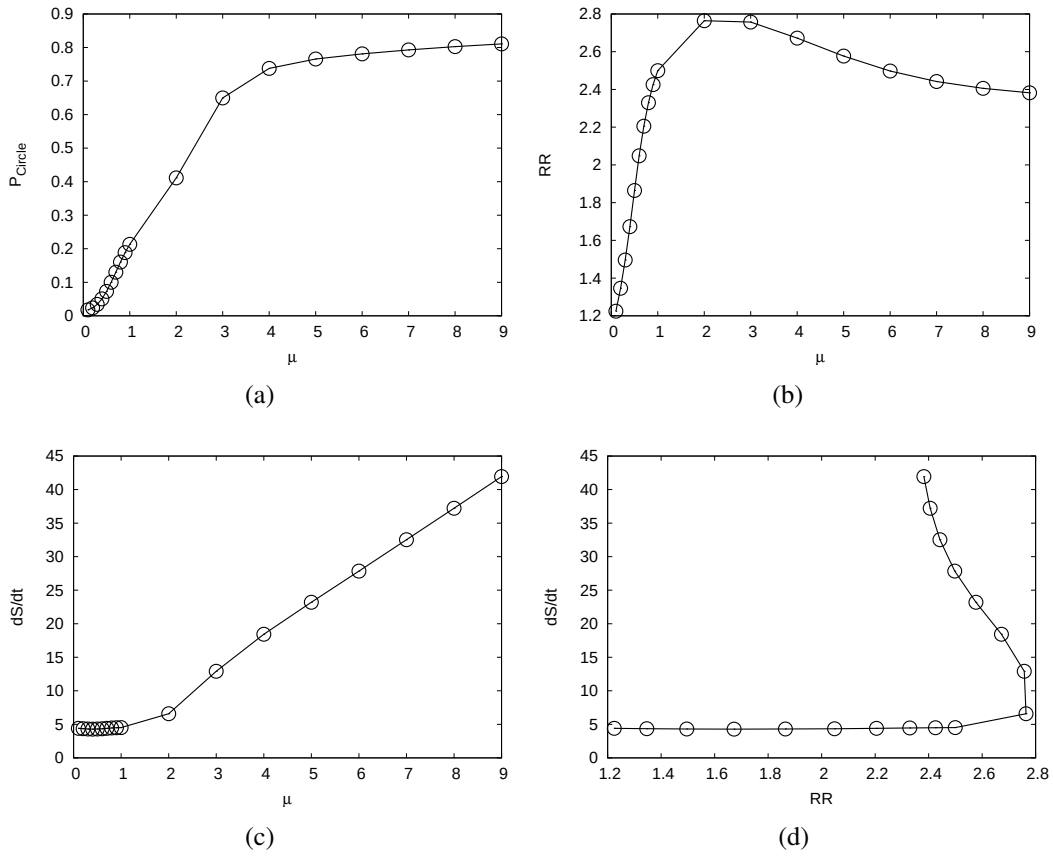


Figure 3.4: (a) Steady state probability of the biological cycle path versus the changes of switching or fluctuation parameter μ. (b) Robustness ratio of the potential landscape relative to the biological cycle path versus the changes of switching or fluctuation parameter μ. (c) Dissipation costs of the network versus the changes of switching or fluctuation parameter μ. (d) Dissipation costs of the network versus robustness ratio of the potential landscape relative to the biological cycle path when changing μ. ($c = 0.001$ and $\gamma = 10$)

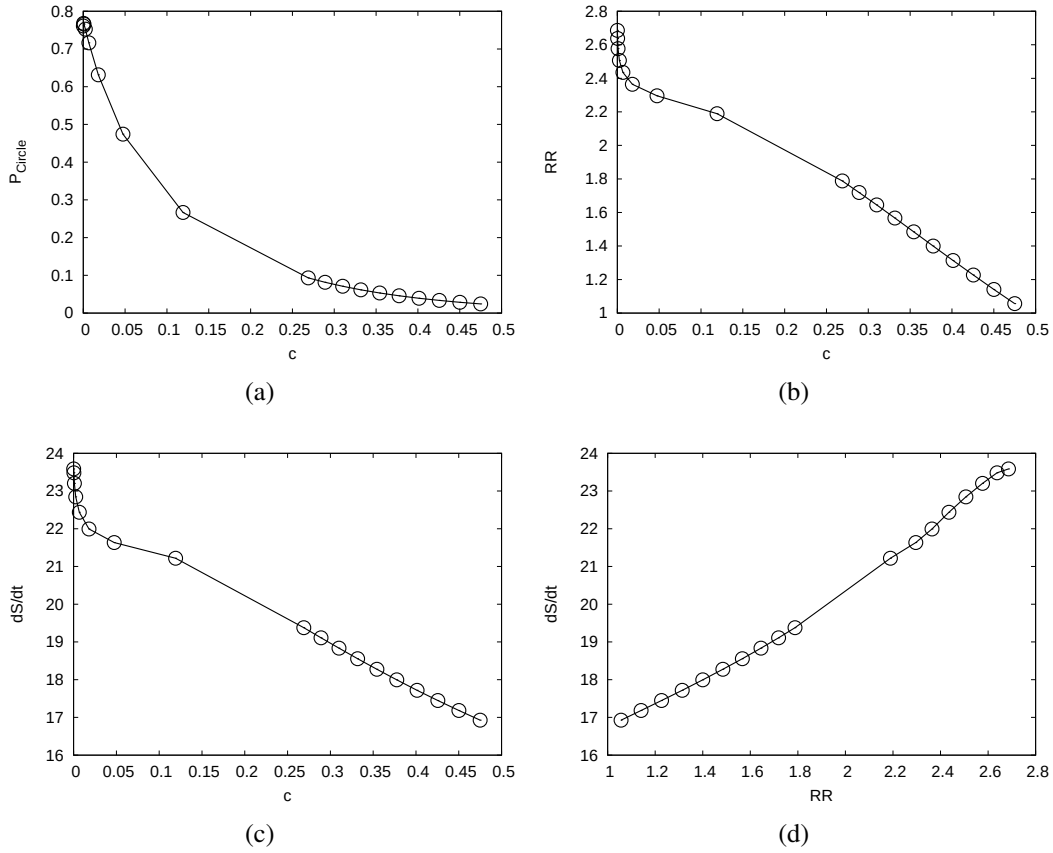


Figure 3.5: (a) Steady state probability of the biological cycle path versus the changes of degradation parameter c. (b) Robustness ratio of the potential landscape relative to the biological cycle path versus the changes of degradation parameter c. (c) Dissipation costs of the network versus the changes of degradation parameter c. (d) Dissipation costs of the network versus robustness ratio of the potential landscape relative to the biological cycle path when changing c. ($\mu = 5, \gamma = 10$)

degradation parameters c (at $\mu = 5$, $\gamma = 10$). We see that when c is large (small) indicating a large (small) self-degradation, the probability of the biological cycle path increases with decrease of c parameters. This means less degradation gives a more probable and stable biological cycle path of the budding yeast cell cycle phases (G1, S, G2, M), and therefore a more robust network. Fig. 3.5(b) gives the robustness ratio RR of the landscape versus the self-fluctuation parameters c (at $\mu = 5$, $\gamma = 10$). It is clear that the robustness ratio RR decreases with the increase of c parameters, which means that the landscape is less stable when the self-fluctuations become large. As we also see Fig. 3.5(c) and Fig. 3.5(d) the dissipation increases as degradation decreases and as the stability of the system increases. This again illustrates that more stable oscillations requires more dissipations or costs to maintain it.

From the definition, the value of γ is to mimic the excitation from the check point G1 back to the start signal. The larger (less) γ is, the faster (slower) or easier (more difficult) the cell cycle passes from the stationary G1 phase to the excitation of the next starting round of the budding yeast cell cycle.

Fig. 3.6(a) shows the probability of the biological cycle path versus different G1 check point nutrition pumping quantified by parameters γ (at $\mu = 5$, $c = 0.001$). We see that when γ is large (small) indicating a large (small) chance of nutrition pumping to the start signal node cln3 when the G1 stationary phase is reached, the probability of the biological cycle increases with increases of γ parameters. As we know, the pumping of the nutrition to cln3 in the G1 phase will excite the cell to be back to the starting phase and trigger a new cycle. This can keep the cell goes in cycles and therefore stabilize the biological cycle of four phases. Fig. 3.6(b) gives the robustness ratio RR of the landscape versus parameter γ (at $\mu = 5$, $c = 0.001$). The robustness ratio RR increases with the increase of γ parameters, which means that the landscape is more stable with more nutrition pumping in. Fig. 3.6(c) and Fig. 3.6(d) show that dissipations increase as the higher excitation or increase of the nutrition pumping. Again to maintain a stable oscillation cycle requires more dissipations or energy costs to maintain it.

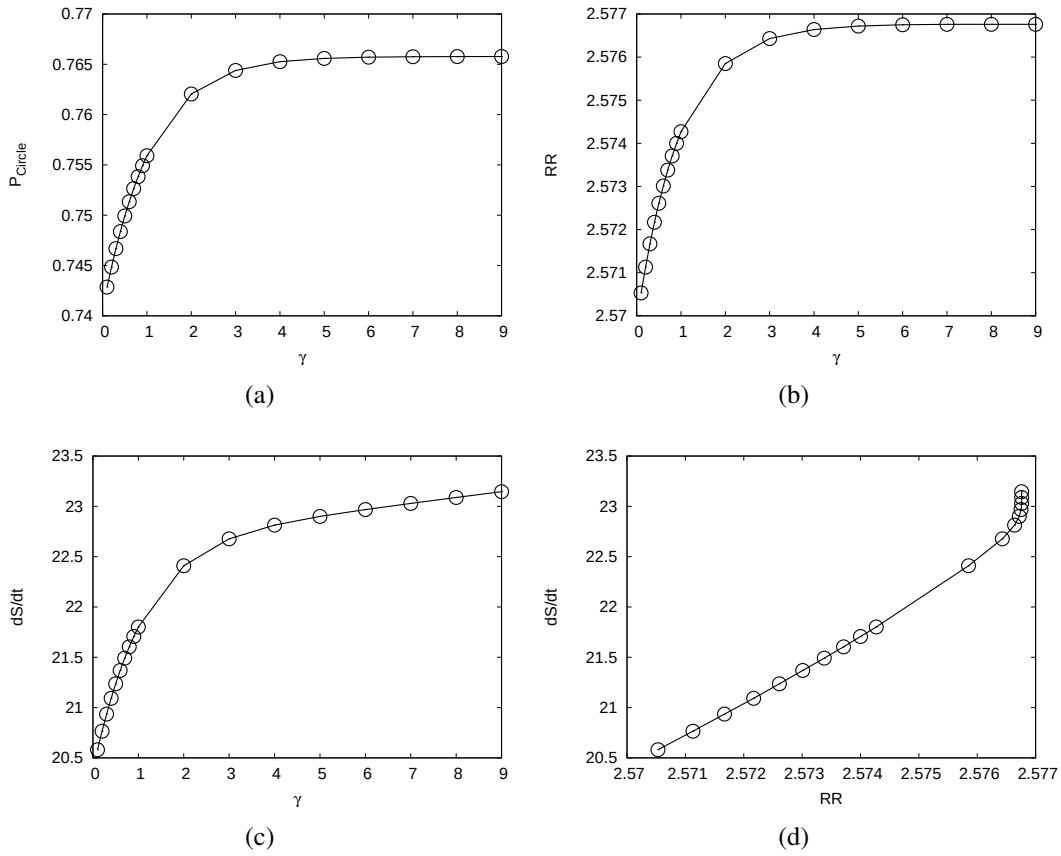


Figure 3.6: (a) Steady state probability of the biological cycle path versus the changes of the excitation nutrition pumping parameter γ. (b) Robustness ratio of the potential landscape relative to the biological cycle path versus the changes of the excitation nutrition pumping parameter γ. (c) Dissipation costs of the network versus the changes of the excitation nutrition pumping parameter γ. (d) Dissipation costs of the network versus robustness ratio of the potential landscape relative to the biological cycle path when changing γ. ($\mu = 5, c = 0.001$)

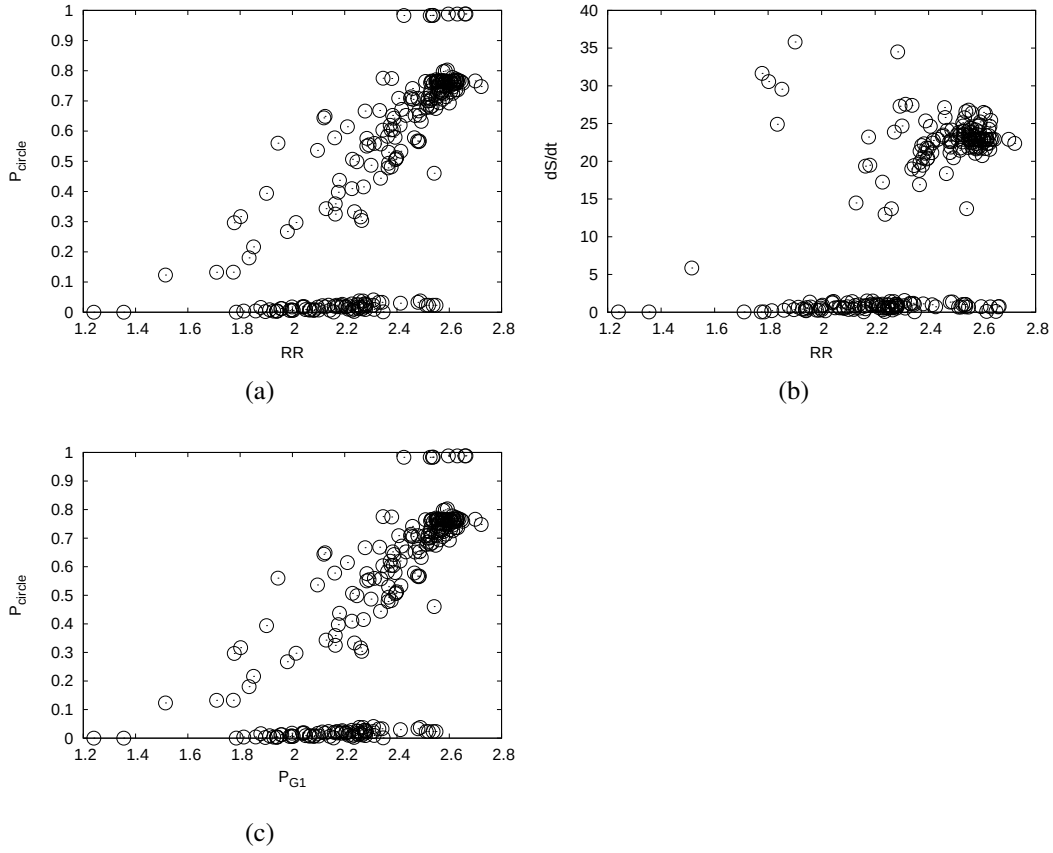


Figure 3.7: (a) The steady state probability of biological path versus robustness ratio of the potential landscape relative to the biological path against perturbations of network topology of wirings. (b) Dissipation cost versus robustness ratio of the potential landscape relative to the biological path against perturbations of network topology of wirings. (c) The steady state probability of biological path versus steady state probability of G1 state against perturbations of network topology of wirings. The perturbations are the ones from deleting an interaction arrow, adding an activating or repressing arrow between the nodes that are not yet connected, switching an activating arrow to a repressing arrow or vice versa, and deleting an individual node. ($\mu = 5$, $c = 0.001$ and $\gamma = 10$).

3.2.3 The key wirings or structure motifs of the network for global stability and robustness from perturbation landscape analysis

Based on the original network wiring diagram in Fig. 3.1, we can do perturbations such as deleting an interaction arrow, adding an activating or repressing arrow between the nodes that are not yet connected, switching an activating arrow to a repressing arrow or vice versa, and deleting an individual node. Take the perturbed network wiring diagram as the new input, we calculate the robustness ratio, *RR*. Fig. 3.7(a) shows robustness ratio, *RR* of the cell cycle network versus the steady-state probability of the biological cycle path (with $\mu = 5$, $c = 0.001$ and $\gamma = 10$) against various perturbations. There is a monotonic relationship between the steady-state probability of biological cycle path and robustness ratio *RR*. When the biological cycle path is more(less) stable with higher probability, the *RR* is larger (smaller), the landscape is more (less) robust. For those points with extremely low probability of the biological paths, no dominant states are preferred, therefore the values of *RR* are not reliable due to the variances of the potential values of all the states. Fig. **??** shows again that more stable oscillations requires more energy dissipations (ignoring the points of low dissipations with large RR values where the actual probability of the biological cycle path is low). Fig. 3.7(c) shows the relationship between the probability of the stationary G1 state and the probability of biological cycle path. Since stationary G1 state plays an important roll in the biological circle, the stable G1 state will result in a stable biological circle. However, the biological circle can remain stable when P_{G1} is not significantly large.

With landscape topology quantified by the probability of biological cycle path, the robustness ratio and the probability of G1 state, we can characterize the budding yeast cell cycle biological function globally. We can see how those features are changed with the network topology of wirings. The network topologies can be changed from the environmental influences, evolution and epigenetics etc. Fig. 8 identifies the most important links, mutating which will make the network unstable and loose its functions (low probability of biological cycle path, low robustness ratio, low probability of G1 state etc). Fig. 3.8(b) indicate the most important links to be added to the original wiring diagram, while Fig. 3.8(a) indicate the most important links which cannot be deleted, and Fig. 3.8(c) indicate the most important

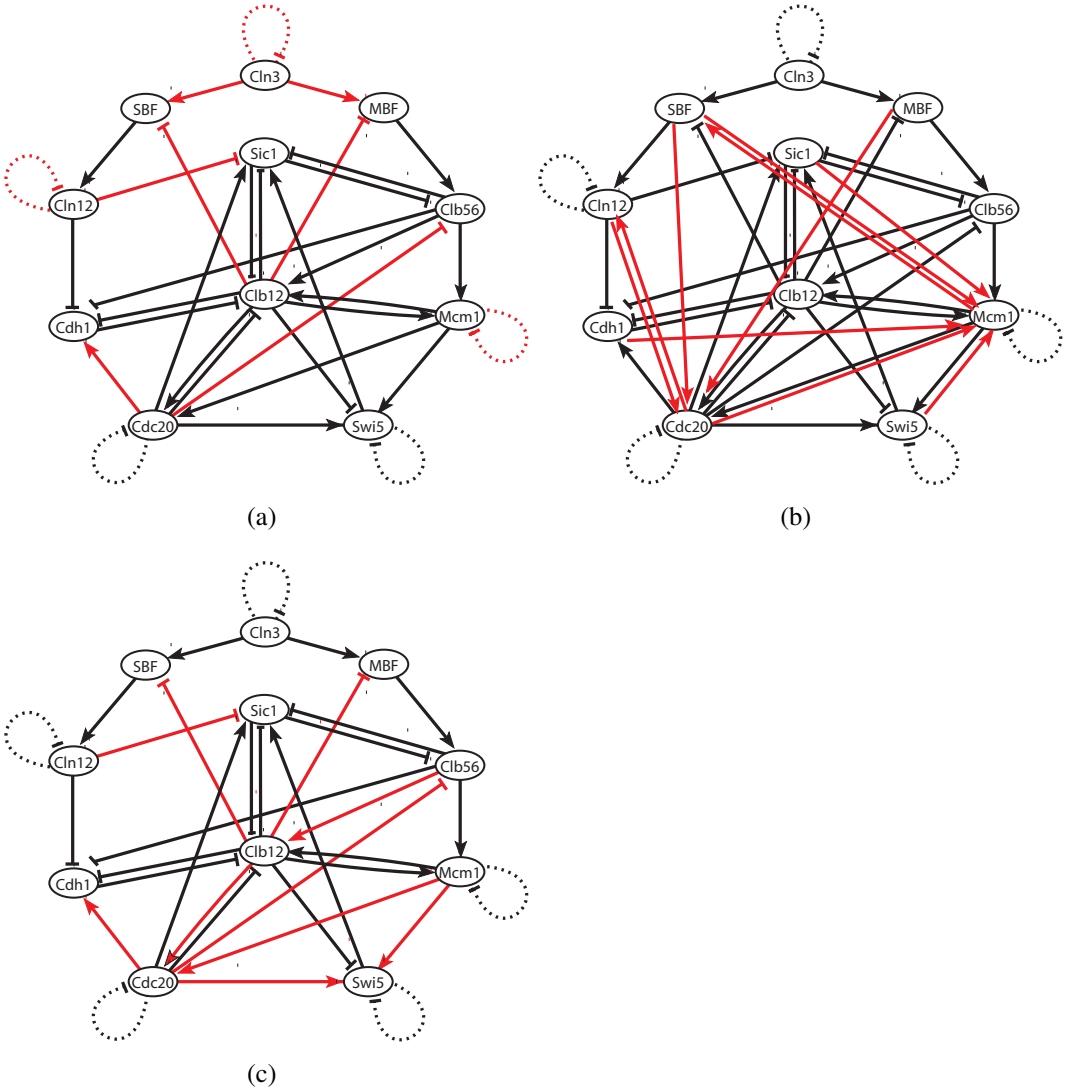


Figure 3.8: The most important links leading the network unstable and loosing its functions. (a) indicates the most important links which cannot be deleted. (b) indicates the most important links to be added to the original wiring diagram, and (c) indicates the most important links which cannot be switched into the opposite in activation and repression. ($\mu = 5$, $c = 0.001$ and $\gamma = 10$).

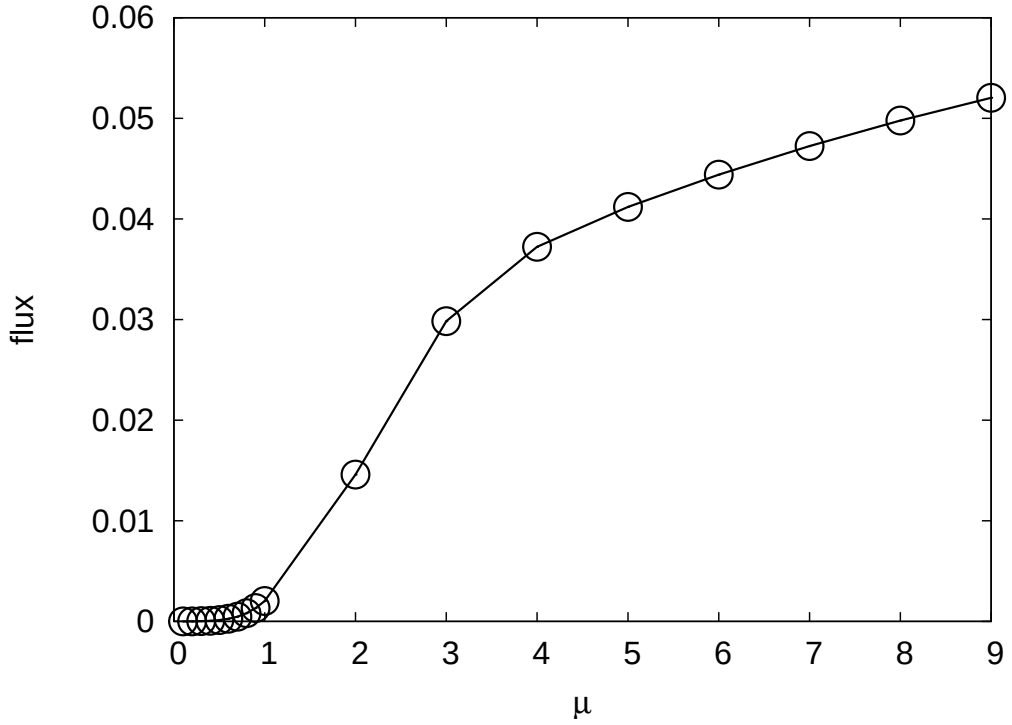


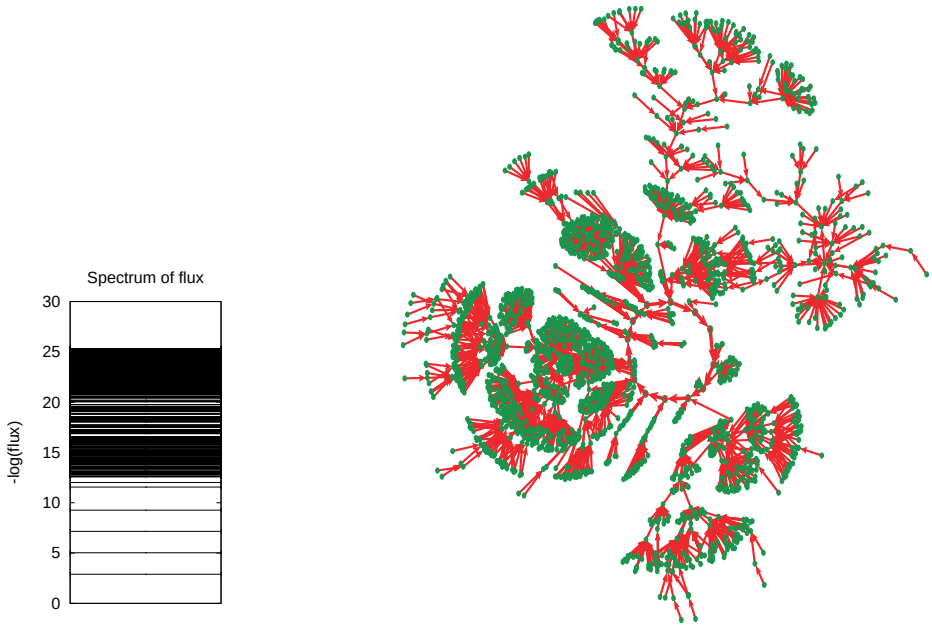
Figure 3.9: Flux value of the biological cycle path versus changes of the switching of the transition probability or fluctuation parameter μ. ($c = 0.001$ and $\gamma = 10$)

links which cannot be switched into the opposite in activation and repression. So through the global analysis of the underlying landscape of the networks, one can identify the key network structure elements or motifs (hot spots) responsible for biological function. This is potentially useful for synthetic network design and network drug design.

3.2.4 Funneled flux landscape leads to stable and robust oscillations

Here we illustrate how the flux landscape emerge, how that changes with system parameters, and how the flux landscape is connected to the stability and robustness of the network.

Fig. 3.9 shows the flux value along the biological cycle path after the decomposition of fluxes into loops versus μ. The monotonic relationship between them shows that when the fluctuation is weak or sharper switching rate of the transition probability, the cell will mostly stay on the biological cycle path when the steady-



(a) (b)

Figure 3.10: The flux landscape ($U = -ln(Flux)$) spectrum: statistical distribution of probability flux values of flux loops. The bottom line above 0 of the flux spectrum is the biological cycle path with high flux values. ($\mu = 5$, $c = 0.001$ and $\gamma = 10$)

state is reached. On the other hand, high fluctuation or shallower switching rate of the transition probability will destroy the network from moving along the biological cycle path.

We can explore the flux landscape through the flux spectrum (statistical collections of fluxes in individual loops) resulting from our decomposition of the non-detailed balance part of the dynamics (the transition probability or rate matrix) into individual flux loops. In Fig. 3.10(a), we see clear separation between the dominant probability flux of the biological cycle and the rest of the fluxes from other loops. Notice that the flux spectrum is in log scale. This defines a funneled flux landscape. The flux landscape can be defined as $U_{flux} = -lnFlux_{loop}$ where $Flux$ is the probability flux value of the individual flux loop. We can also define robustness ratio for the flux landscape as flux landscape gap between the biological cycle path and the averaged fluxes from the other loops versus the variations of the flux values $RR_{flux} = (U_{flux}^{biopath} - < U_{flux} >)/\sqrt{< U_{flux}^2 > - < U_{flux} >^2}$. Large robustness ratio quantifies the large separation of the biological path flux with the rest of the other fluxes. This means there are many flux loops and among them only biological path

loop stands out to have distinctly larger values of flux (or low value of the flux landscape values in U) compared with the flux values from the rest of the other loops. This implies a funnel flux landscape which guarantees the stability of the biological path and related biological function. In Fig. 3.10(b), we see the flux landscapes with dominant loop of the biological cycle path explicitly shown (we did not show other loops for the purpose of clearness on the figure) and dominant flux flow direction between states. The thickness of the arrows represents the magnitude of flux values. As we can see the other states converge to the stable biological cycle path.

The Robustness Ratio of the flux landscapes, RR_{flux} is calculated to illustrate how separate the flux of the biological circle from the rest of the cycle loops. The results are shown in Fig. 3.11.

Fig. 3.11(a) shows the relationship between the robustness ratio of the flux landscape versus probability of the biological cycle path as the switching rate of the transition probability or fluctuation strength μ changes. We see an essential correlation between the two. From unstable to stable oscillation, we can see that larger weights of the biological cycle path often gives more funneled flux landscape with clear separation between the biological path flux and the rest of the other flux loops. The turnover behavior can again be explained by the presence of artificial traps of the Bollean networks, which become unstable with non-zero fluctuations. This leads to higher RR (or peak) value for flux landscape (larger gap and smaller variations) for certain fluctuations (μ).

Fig. 3.11(b) shows that the robustness ratio quantifying the degree of the funnel of flux landscape (or the separation of probability flux between biological path and the rest of the other flux loops), correlates with the robustness ratio of the potential landscape quantifying the separation of potentials between the biological path and the rest of the other states, with respect to the changes of μ. To guarantee the stable biological function of cell cycle, the flux landscape must be funneled. The corresponding potential landscape relative to the biological path (not just a single G1 state) should also be funneled. Therefore the landscape from flux loop perspective or from potential perspective (treating the oscillation as a continuous attractor: limit cycle) are all consistent to lead to a funnel for stability.

Fig. 3.11(c) shows the robustness ratio of flux landscape versus the switching rate of the transition probability or fluctuation strengths. We see lower fluctuations and sharper switch rate of transition provides a more funneled flux landscape,

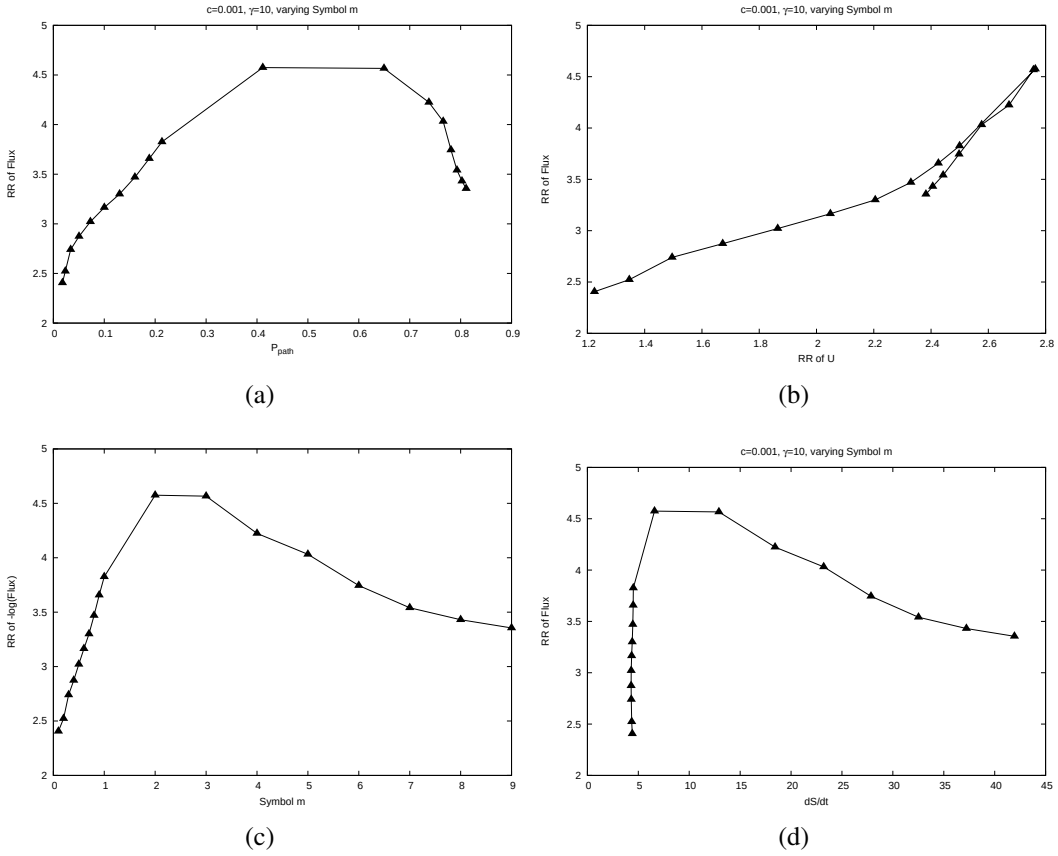


Figure 3.11: (a) Robustness ratio of flux landscape versus steady state probability of biological cycle path against changes of switching or fluctuation parameter μ. (b) Robustness ratio of flux landscape versus robustness ratio of the potential landscape relative to biological cycle path against changes of switching or fluctuation parameter μ. (c) Robustness ratio of flux landscape versus the changes of switching or fluctuation parameter μ. (d) Robustness ratio of flux landscape versus dissipation costs against changes of switching or fluctuation parameter μ. ($c = 0.001$ and $\gamma = 10$)

leading to more stable biological oscillation (ignoring the effects of turn over from artificial traps of the Bollean model). Since the switching rate represents the probability of switching upon inputs from the other nodes, or response function. It reflects the input-output relationship. As μ is large the tanh[μI] is highly nonlinear, giving sharp response. On the other hand, as μ is small, the input-output relationship becomes linear (tanh[μI] μI). So as we can see a stable and robust oscillation requires the needs the nonlinear input-output relationships between the network links or connections. Therefore, nonlinearity significantly increase the chance of the dominant flux loop emerging from the rest of others.

Fig. 3.11(d) shows the robustness ratio of flux landscape versus the dissipation of the system. It implies that funneled flux landscape giving more stability (larger probability of biological cycle path, see Fig. 11a) needs dissipation costs and energy nutrition supply to maintain. (Again ignoring the effects of turn over from artificial traps of the Bollean model).

3.3 Conclusions

We explore the global natures of the networks. We found the network dynamics and global properties are determined by two essential features: the potential landscape and the flux landscape. While potential landscape quantifies the probabilities of different states forming hills and valleys, the flux landscape quantifies the probability fluxes of different loops through states. These two landscapes can be quantitatively constructed through the decomposition of the dynamics into the detailed balance part and non-detailed balance part. While funneled landscape is crucial for the stability of the single attractor networks, the funneled flux landscape is crucial for stable oscillation networks.

This may provide a new interpretation of the origin of the limit cycle oscillations: There are always many cycles and loops forming the flux landscapes, each with a probability flux value going through the loop. The oscillation only emerges when one specific loop stands out and carries much more flux than the rest of the others. This happens when the nonlinearity of the inherent dynamical systems increases.

We studied the budding yeast cell cycle as an example to illustrate the idea. We found the flux landscape of the budding yeast cell cycle oscillations is funneled ,

which guarantees the global stability. The stability and robustness of the oscillations are quantified through a dimensionless ratio of the steepness versus the averaged variations or roughness of the landscape (measuring funnelness as we termed as robustness ratio RR). We explore how RR changes with respect to the changes in topology of the network (wirings) and stochastic fluctuations. This allows us to identify the key factors and structure elements of the networks in determining the stability and robustness of the budding yeast cell cycle oscillations.

Chapter 4

Adiabatic and Non-Adiabatic Non-Equilibrium Stochastic Dynamics of Single Regulating Genes

4.1 Introduction

Uncovering the global principle and underlying mechanisms of gene regulatory networks is crucial for understanding the associated biological functions. In living cells, gene expressions are regulated by a complex and diverse genetic networks,which involve an intricate set of biochemical reactions[67]. The large networks are often made of smaller modules or motifs [34, 2]. Therefore understanding the motifs as the building blocks of the whole network is very important.

The simplest motifs are the self regulators: activator and repressor. "Activator" is a regulatory motif protein that increases the level of transcription, and a "repressor" is another motif protein which decreases the level of transcription. The transcriptional process is suppressed when the promoter site of the DNA is occupied by a repressor (gene is off) and enhanced when the repressor is dissociated from DNA (gene is on). Therefore the states where gene is switched on and off are important for the transcription process. The gene regulation process involves at least two kinds of biochemical reactions: binding/unbinding reactions of regulatory proteins to the genes and synthesis/degradation reactions of the RNAs and proteins (see Figure. 5.2a and Figure. 5.2b). Conventionally, it was often assumed that the binding/unbinding is significantly faster than the synthesis and degrada-

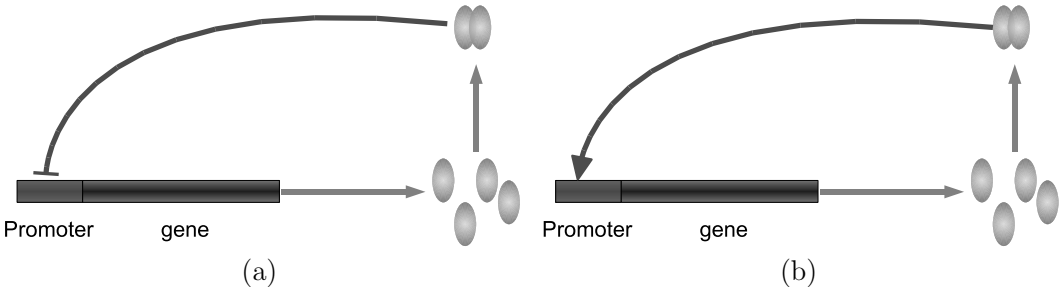


Figure 4.1: A schematic representation of self-regulating gene circuit. The dimer of the product protein binds in the promoter and controls the protein synthesis rate, forming a negative feedback loop for self-repressors (a) and a positive feedback loop for self-activators (b).

tion (adiabatic limit) [1]. This leads to the expected single stable state for a self repressor which could be measured in experiments [7]. While this condition may hold in some of the prokaryotic cells (for example E. Coli) at certain conditions, in general there is no guarantee it is true. In fact, one expects in eukaryotic cells and some prokaryotic cells, binding/unbinding can be comparable or even slower than the corresponding synthesis and degradation. This can lead nontrivial stable states appearing as a result of new time scales introduced, [37, 79, 45, 15] which is confirmed by recent experiments [80, 11, 110].

Since most cells have a relatively small number of molecules (typically range from 1 to 10^3), the intrinsic statistical fluctuations can make large contributions to gene expression and must be included in the studies of gene networks [17, 38, 47]. The underlying dynamics can not be described by the conventional deterministic bulk chemical rate equations[77, 60, 66, 84, 86]. Instead, one should follow the evolution of the probabilities which obey the probabilistic master equations. In bulk kinetics, the deterministic equations usually give only information about local stability and function. Global stability issue is hard to address. A probabilistic description is essential for understanding the global biological function.

In this study, we explore the stochastic dynamics of self regulative genes due to intrinsic fluctuations with finite number of molecules in the cell and the fluctuations from switching of on and off gene states due to regulatory proteins binding/unbinding to the genes. This is quantified through steady state probability distribution and Fano factors. We link this with recent experiments [80, 110]. We further study the inhibition and promotion curves in adiabatic and non-adiabatic regime

and compare with previous studies. We derive the non-equilibrium phase diagrams of mono-stability and bi-stability in adiabatic and non-adiabatic regimes. We study the dynamical trajectories of the self regulating genes on the underlying landscapes from non-adiabatic to adiabatic limit [79] and show the analogy to the problem of electron transfer between two energy wells [37, 79]. We study the stability and robustness of the systems through mean first passage time from one peak (basin of attraction) to another. We also explore global dissipation by entropy production and correlate with the stability through landscape topography and kinetics.

4.2 Model

Here is a brief discussion of the assumptions and methods used in this work for self-regulators. The DNA can be bound by a dimer of regulatory protein with the rate $\frac{1}{2}h_{\alpha\beta}n(n-1)$ and the dimer can be dissociated from DNA with the rate f:

$$\mathcal{O}_\alpha^1 + q_{\alpha\beta}\mathcal{M}_\beta \quad \xrightarrow{h^R_{\alpha\beta}} \quad \mathcal{O}_\alpha^0 \tag{4.1}$$

$$\mathcal{O}_\alpha^0 \quad \xrightarrow{f^R_{\alpha\beta}} \quad \mathcal{O}_\alpha^1 + q_{\alpha\beta}\mathcal{M}_\beta \tag{4.2}$$

where $O_\alpha^{1(0)}$ stands for the unbound(bound) state of gene α which will synthesize protein α, $\mathcal{M}_\beta$ means the monomer regulatory protein of gene β, and $q_{\alpha\beta}$ is for the multimer-type of proteins [77]. For example, if $q_{AB} = 2(4)$, dimer(tetramer) proteins of gene β regulate the expression of gene α. Notice that the superscript "1(0)" in $\mathcal{O}_\alpha^{1(0)}$ means the unbound(bound) state of the operator and does not mean the bound state of regulatory protein. We will also say "gene is on(off)" when the operator of the gene is active(inactive). Gene will be on when it is occupied by activators or when repressors are dissociated from it.

Also, we combine transcription and translation into one step and ignore the roles of mRNAs for simplicity. Therefore, the transcription step is:

$$\emptyset \quad \xrightarrow{g_{\alpha 1}} \quad \mathcal{M}_\alpha \quad \text{for} \quad \mathcal{O}_\alpha^1 \tag{4.3}$$

$$\emptyset \quad \xrightarrow{g_{\alpha 0}} \quad \mathcal{M}_\alpha \quad \text{for} \quad \mathcal{O}_\alpha^0 \tag{4.4}$$

$$\mathcal{M}_\alpha \quad \xrightarrow{k_\alpha} \quad \emptyset \tag{4.5}$$

where $\emptyset$ is used to denote a protein sink or source. $g_{\alpha 1(0)}$ is α protein synthesis rate

when the gene is unoccupied(occupied) and k_α is a degradation rate of protein α.

So for a dimer self-regulator, the master equations will be as follows:

$$\frac{dP_1(n)}{dt} = -\frac{h}{2}[n(n-1)]P_1(n) + fP_0(n-2)$$
$$+k[(n+1)P_1(n+1) - nP_1(n)] + g_1[P_1(n-1) - P_1(n)] \tag{4.6}$$

$$\frac{dP_0(n)}{dt} = \frac{h}{2}[(n+2)(n+1)]P_1(n+2) - fP_0(n)$$
$$k[(n+1)P_0(n+1) - nP_0(n)] + g_0[P_0(n-1) - P_0(n)] \tag{4.7}$$

where we ignore the decay of proteins bound on the gene. Also, for probability $P_\alpha(n)$, $(\alpha = 1, 0)$, the dimer (2 protein) bound on the gene is not counted for protein number n. Details of the solution of master equation is given in supporting information.

4.3 Results

4.3.1 The steady-state distributions, phase diagrams and underlying landscape of self regulating gene

As discussed above, the network under stochastic fluctuations is best characterized by the probability distribution which gives a global quantitative measure of the biological function (phenotypes). The fluctuations not only come from protein concentration changes but also the changes of gene states. The steady state distribution has two contributions: one is from when gene is on and the other is from when gene is off. $P(n) = P_{on}(n) + P_{off}(n)$ [94].

In order to take into account both gene state and protein states, we need to define four critical parameters and three crucial ratios governing the dynamics: binding rate constant h, unbinding rate constant f, protein synthesis rate g_{on} when gene state is on and protein synthesis rate g_{off} when gene state is off, and finally the protein degradation rate k; the ratio $\omega = f/k$ quantifies the unbinding relative to the speed of degradation; the equilibrium constant $X_{eq} = f/h$ quantifies the relative balance between unbinding and binding; the ratio $X_{ad} = (g_1 + g_0)/2k$ quantifies the protein production rate relative to the degradation. In all our calculations, we set $g_0 = \mathbf{8}$ for

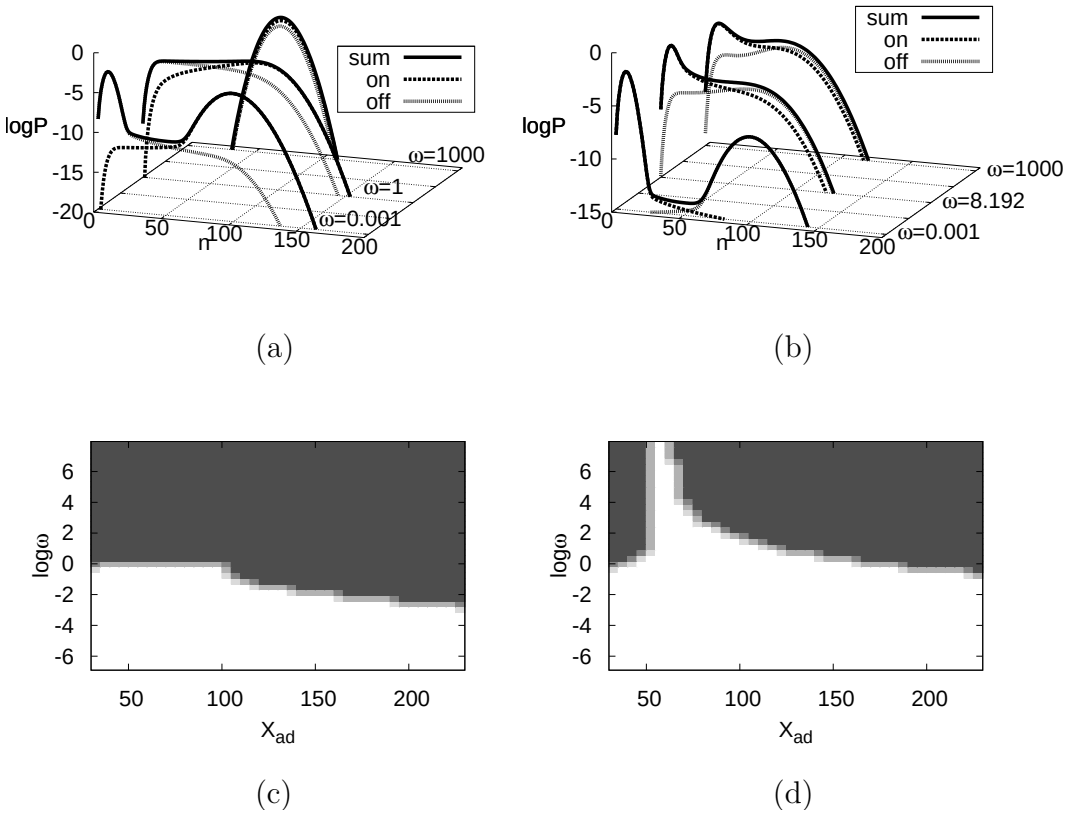


Figure 4.2: Steady state distributions for the on state, off state and the sum as a function of the number of proteins of, for self-repressors with $X_{eq} = 5000, X_{ad} = 54$ (a) and self-activators with $X_{eq} = 1400, X_{ad} = 54$ (b). Phase diagram respect to different ω and X_{ad} for for self-repressors with $X_{eq} = 5000$ (c) and self-activators with $X_{eq} = 1400$ (d). Purple region represents monostability (single peak in $P(n)$) and yellow region represents bistability (double peak in $P(n)$).

self-repressors and $g_1 = \mathbf{8}$ for self-activators.

We solve the master equation and show the results. We see in Figure. 4.2a and 4.2b, in the nonadiabatic limit where binding/unbinding is significantly slower compared with the corresponding synthesis and degradation, i.e., $\omega = f/k = 0.001$, the steady-state distributions $P(n)$ in protein concentration of self-repressors and self-activators all have two peaks at around g_1/k and g_0/k. These two peaks are separately from the contributions of protein distributions when gene state is on with high protein synthesis and when gene state is off with low protein synthesis. We can think of the two gene states are distinct and there are almost no"interactions" between them in the non-adiabatic limit. Two peaks are therefore independent with one another.

Although both self repressor and self activator can have two peaks in protein concentrations from on and off gene state in non-adiabatic limit. The origin and trend are different. For self repressor (activator), at high protein concentration peak when gene is on, more proteins will tend to suppress (promote) more, produce less (more) and make the state less (more) stable and move towards less (more, bounded by degradation) protein side. At low protein concentration peak when gene is off, less proteins will tend to suppress (activate) less, produce more (less) and make the state less (more) stable and move towards more (less, bounded by zero protein number) protein side. So there is a tendency of merging two peaks into one for self repressor and self stablization of two peaks for self activator.

When the binding/unbinding is fast compared to synthesis/degradation, the gene state changes rapidly. Interactions and mixing become stronger between the two gene states and therefore the two protein concentration peaks. For self repressor, at a fixed equilibrium constant of unbinding versus binding of regulatory protein to the gene, increasing the unbinding promotes production of proteins which in turn represses more, leading the high concentration peak to move downward. Increasing the binding represses production of proteins which in turn promotes more leading the low concentration peak to move outward. As a result, the two peaks meet in the adiabatic limit and merge to one for self repressor. At any moment, the protein production has a single peak at an average of the on and off gene states (average production rate of the on and off gene states). Adiabatic limit with one peak in protein concentration has been studied before but the underlying mechanism is not so clear as stated here. The non-adiabatic case with two peaks can also appear and deserve further explorations [37, 79, 94, 45, 15].

For self activator, adiabatic fast binding leads to the shifts of the balance or preference between the on and off peaks. As a result, we can have either one peak in protein concentration at on state when we have large protein productions or one peak at off state when we have small protein productions, reaching mono-stability. For self activators, there is no tendency as in the self repressor case of merging the two peaks together in adiabatic limit. Here either on or off peaks is preferred (not merged).

For intermediate protein productions relative to the degradation at fixed equilibrium constant, there is no preference for the system to be in either on or off states. In fact the range of intermediate protein production scales with the equilib-

rium constant. This guarantees that even for small or large unbinding rate there is still a chance for co-existence of on and off peaks due to the balance of the corresponding protein productions. Therefore bi-stability can emerge for self activators in adiabatic limit.

Figure. 4.2c and 4.2d show the phase diagrams of self-repressors and self-activators. It is easy to notice the similarity between self-repressors and self-activators, which is due to the symmetry of their master equations: in the non-adiabatic regime with bistability (two peaks in protein concentrations) and mono-stability (one peak in protein concentration) in some parts of the adiabatic regime although the origin of bi-stability to mono-stability (merging versus one peak dominant the other) and position of single peak (average versus on or off peak) are different. The self activators can also have a regime with bistability in the adiabatic and non-adiabatic limit.

From the steady-state distribution $P(n)$, we can quantify the generalized potential function $U(n)$ of the non-equilibrium network analogous to equilibrium Boltzmann relationship between potential and probability: $U(n) = -\ln(P(n))$. This will map to the potential landscape [77, 4, 73, 98, 97, 51, 31, 52, 102, 106, 103, 104].

This is analogous to electron transfer problem [57, 77], where two distinct potential wells with each well labeled by the corresponding electronic state (here potential well is labeled by corresponding gene state) emerge in non-adiabatic limit. The adiabatic case gives the limit of mixing of two potential wells and leads to an effective potential.

The adiabatic to non-adiabatic transformations under different conditions can be tested from experiments [110, 80] through the single or multiple peaks in fluorescence intensity, which correspond to the phenotypes. By varying inducer concentrations and mutations, one can change either the binding/unbinding or degradation and therefore effectively ω to induce new phenotypes (additional peak compared to the conventional fast equilibrium one peak case). Our model provides a theoretical explanation of the recent experiments on self activators [110, 80].

To understand the global stability, we need to quantify the degrees of fluctuations [53, 49]. This can be measured by Fano factor, defined as the variance/mean of a specific distribution: $F = \sigma^2/\mu$, where σ and μ are the standard deviations and the mean of the probability distribution, $F = 1$ if the distributions were exactly Poisson, a purely independent random one. The σ and μ can be measured quantita-

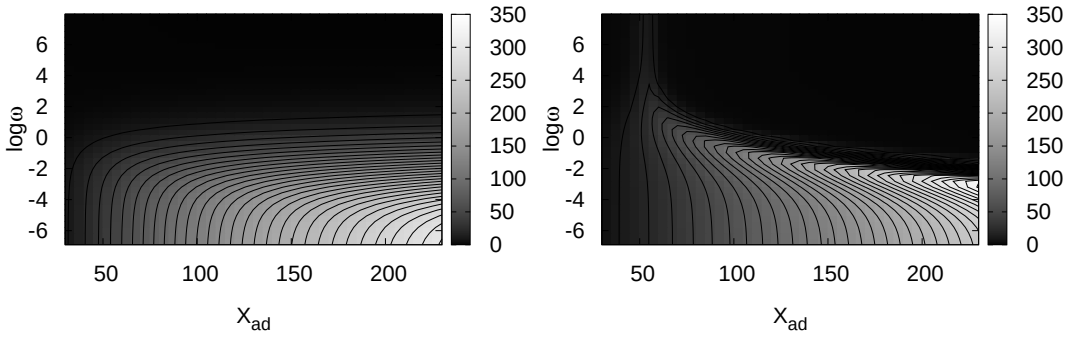


Figure 4.3: Total Fano factor $F = \sigma^2/\mu$ for self-repressors with $X_{eq} = 5000$ (a) and self-activators with $X_{eq} = 1400$ (b) respect to different ω and X_{ad}. $F = 1$ for Poisson like distributions. Large Fano factors implies large statistical fluctuations deviating from Poisson distributions.

tively in the experiments [80, 11, 110]. Large Fano factors implies large statistical fluctuations deviating from Poisson. In Figure. 4.3a, 4.3b, the Fano factor F is plotted as a function of adiabaticity parameter ω and relative production parameter X_{ad}. Qualitatively, the Poisson distribution should be a good approximation for on and off states separately in the nonadiabatic limit: $\omega \ll 1$, because each gene state can produce proteins almost independently of gene switching. But the overall Fano factor for the combined probability distribution is much larger than 1. Because the system is close to two Poisson distributions with different means added together, producing large statistical fluctuations deviating from the single Poisson distribution.

As ω increases, effects from mixing between different gene states should produce larger fluctuations to the Poisson distributions leading to large Fano factors in the intermediate region for ω. The Poisson is poorly suited in these two gene states and the distribution is spread out with "fatty" tails implying large statistical fluctuations. In this regime, it is necessary to include the fluctuations from gene switching.

While in the adiabatic limit $\omega \gg 1$, gene state is fluctuating so rapidly and the probability distribution for "on" or "off" or the combination are all Poisson like with the two peaks in the same position for self repressor. The Fano factor approaches 1 and the the system becomes Poisson-like. For self activators, the Fano factors are much bigger than 1 in the bistable region even in the adiabatic limit with large

statistical fluctuations.

In the intermediate adiabaticity regime, for self-repressors and self activators, the Fano factor increases with protein production X_{ad}. For self-repressors, the distance between "on" peak and "off" peak increases with X_{ad} and the tail for each peak is caused by the mixing effect from the other peak. So larger peak distance will lead to bigger tail for each peak and larger Fano factor. For self-activators, the larger X_{ad} will suppress the "off" state more. Thus the mixing effect from "off" to "on" state will be suppressed and the mixing effect from "on" to "off" state will be promoted, which causes smaller fluctuation in "on" state and larger fluctuation in "off" state. Fano factor including both on and off states increases as X_{ad} increases for self activators. This is consistent with the phase diagrams.

Above all, we noticed that, for self regulators, fluctuations are not only from intrinsic noise of protein numbers but also from the switching of gene states. The intrinsic fluctuations are $\sqrt{< n^2 > - < n >^2}/ < n >= 1/\sqrt{n}$ for Poisson. When protein number n (X_{ad}) is large (small), intrinsic fluctuations are small (significant). The control for fluctuations of switching gene states or adiabaticity is the relative binding/unbinding to synthesis/unbinding ω. If ω is larger(smaller), the fluctuations caused by gene state switching are in general smaller (larger). These adiabatic/non-adiabatic fluctuations can be directly probed from the ongoing and future single molecule single gene experiments [110, 80].

Another measure of the change of gene fluctuations with ω can be made through the efficiency of repressor and activator for gene regulations through the probability of inhibition for self-repressors and the probability of promotion for self-activators: $P_\beta = c_0$ which can be quantitatively measured in the experiments [1]. In the approximation of Ackers *et al*[1], large ω limit is taken (equilibrium): $P_\beta = \langle n \rangle^2/(\langle n \rangle^2 + 2X_{eq})$. While as we see in the moment equations , for self-repressors, $P_\beta = c_0 = (\langle n^2 \rangle_1 - \langle n \rangle_1)/(\langle n^2 \rangle_1 - \langle n \rangle_1 + 2X_{eq})$ where $\langle n \rangle_1$ is the number of proteins in unbound gene state. Only in the adiabatic limit, $\langle n \rangle = \langle n \rangle_1$ and with Poisson assumption, these two expression will be the same. Figure. 4.4a shows that for self-repressors, the inhibition curves converge to equilibrium approximation in large ω limit. While for self-activators, as shown in Figure. 4.4b, even in the adiabatic region, if it is bistable, $\langle n \rangle \neq \langle n \rangle_1$ and the promotion curves won't converge to Ackers' approximation.

In addition, for different ω, the order of inhibition curves for self-repressors and

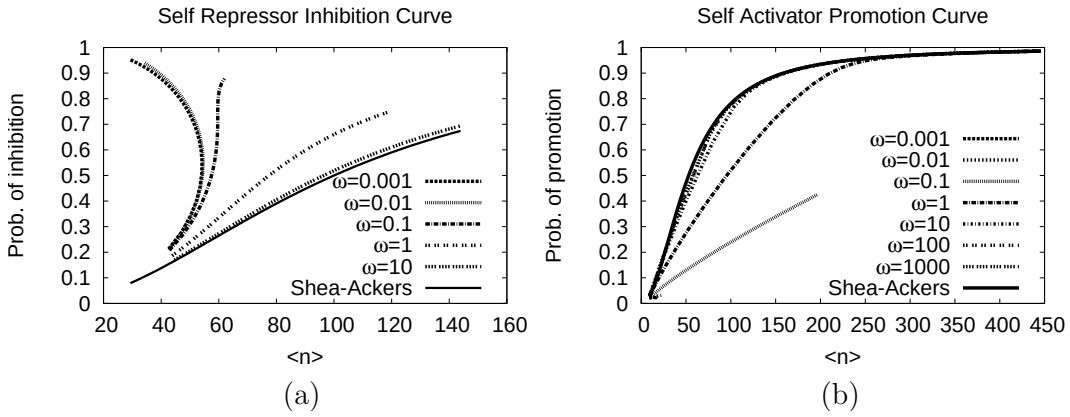


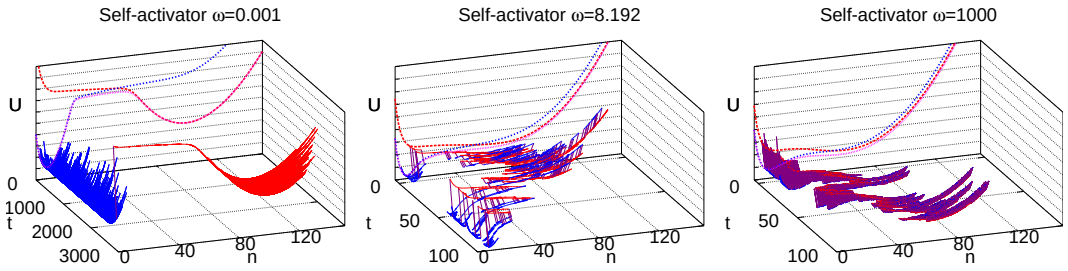
Figure 4.4: Total probability P_β of the DNA in the bound state as a function of the average protein number $\langle n \rangle$: inhibition curves for self-repressors with $X_{eq} = 5000, X_{ad} = 54$ (a) and activation curves for self-activators with $X_{eq} = 1400, X_{ad} = 54$ (b). As ω increasing, P_β for both self-repressors and self-activators approach Shea-Ackers approximation: $P_\beta = \langle n \rangle^2/(\langle n \rangle^2 + 2X_{eq})$. For small ω, we can observe new turn over behavior in inhibition curves for self-repressors.

promotion curves for self-activators are opposite. It is because unbound state is "on" state in self-repressors while "off" state in self-activators. For self-repressors, the "on" state peak will decrease with ω. This decreases the probability of inhibition as adibaticity increases. While for self-activators, the "off" state peak will increase with ω. This increases the probability of promotion as adibaticity increases. The inhibition and promotion curves will have opposite order in ω.

Another interesting observation is that inhibition curves turn over for self-repressors in non-adiabatic region giving two different trend of inhibition curve in the same regime of average protein number. This is caused by the dimer formation for gene regulations. In non-adiabatic region, $\langle n_1 \rangle = g_1/k$ and $\langle n_0 \rangle = g_0/k$. $\langle n \rangle = c_1\langle n_1 \rangle + c_0\langle n_0 \rangle = \frac{2X_{eq}}{\langle n_1 \rangle^2+2X_{eq}}\langle n_1 \rangle + \frac{\langle n_1 \rangle^2}{\langle n_1 \rangle^2+2X_{eq}}\langle n_0 \rangle = \langle n_0 \rangle + \frac{2X_{eq}(\langle n_1 \rangle - \langle n_0 \rangle)}{\langle n_1 \rangle^2+2X_{eq}}$. For fixed $\langle n_0 \rangle = g_0/k$, the $\langle n \rangle$ vs $\langle n_1 \rangle$ curve is not monotonic while the inhibition rate is monotonic with $\langle n_1 \rangle$: $c_0 = \frac{\langle n_1 \rangle^2}{\langle n_1 \rangle^2+2X_{eq}}$ which make the inhibition rate non-monotonic with $\langle n \rangle$ with dimer binding/unbinding. But for monomer binding/unbinding, $\langle n \rangle = \frac{2X_{eq}+\langle n_0 \rangle}{\langle n_1 \rangle+2X_{eq}}\langle n_1 \rangle$ which leads to the monotonic inhibition curve.

We point out here that the adiabatic and non-adiabatic dynamics of gene networks can also be probed by the measurement of promotion and inhibition curves in the experiments in addition to Fano factors.

From the trajectory of monte carlo simulations, we can see how the dynamics



(a) (b) (c)

Figure 4.5: Underlying Landscapes $U = -lnP(n)$ and typical trajectories of protein concentrations in transition from "off" basin(the peak on $P_{off}(n)$) to "on" basin(the peak on $P_{on}(n)$) for self-activators of $X_{eq} = 1400, X_{ad} = 54$ and (a) $\omega = 0.001$, only one gene switch happenes; (b) $\omega = 8.192$, some gene switches form "eddy-induced" diffusion; (c) $\omega = 1000$, many binding-growth-unbinding-decay cycles and it can be approximated by diffusions in protein concentrations. Blue (red) represents off (on) state of gene.

occurs and switches between two gene states from non-adiabatic to adiabatic regime for self-repressors (bi-stability to mono-stability) and self-activators (bi-stability to bi-stability and bi-stability to mono-stability). Some details are shown in Supporting Information.

In the nonadiabatic limit case with small ω, as shown in Figure. 4.5a, only a few binding/unbinding processes happened. The system stays mostly in a protein number basin (specific gene state) fluctuating around until occasionally the gene state is changed through binding/unbinding event of regulatory proteins to the genes, the system will then be switched from one basin to the other potential surface quickly and moves on the switched potential surface towards the new basin of attraction (due to the slow binding/unbinding process) before turning back.

In the intermediate ω region, as in Figure. 4.5b, there are more binding/unbinding processes happened. The system stays first in a protein number basin fluctuating around until the gene state is changed through binding/unbinding event, and will then be switched from one basin to the other potential surface quickly and stays on the switched potential surface for a while. Then it will turn back to the original potential surface through unbinding/binding event. This forms a cycle. The processes continues, with each cycle of binding-growth-unbinding-decay, and eventually reaches the other basin of attraction, resembling of the eddy current in the "eddy-induced" diffusion[93].

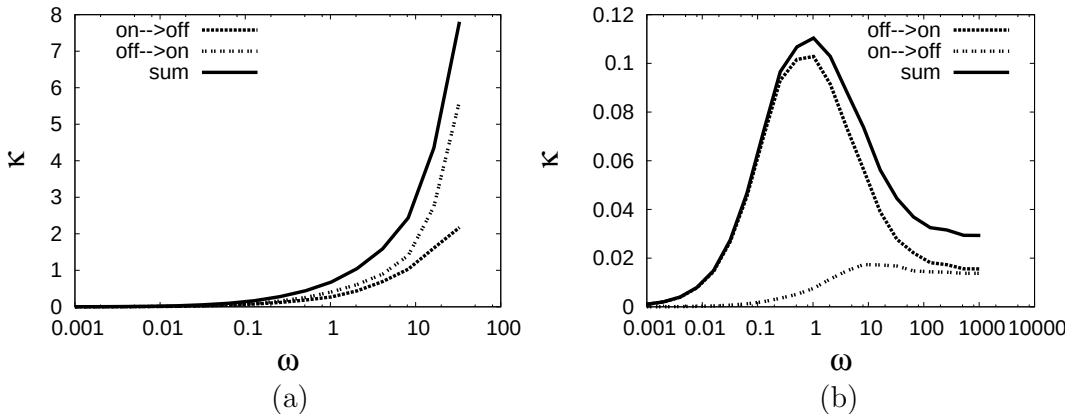


Figure 4.6: Mean transition rate $\kappa \sim 1/MFPT$, from "off" basin to "on" basin, from "on" basin to "off" basin and sum of them, for self-repressors of $X_{eq} = 5000, X_{ad} = 54$ (a) and self-activators of $X_{eq} = 1400, X_{ad} = 54$ (b). It is monotonic for self-repressors but non-monotonic for self-activators.

In typical adiabatic limit as in Figure. 4.5c, there are many binding/unbinding events at very small intervals of protein concentration changes. The system switches the gene state rapidly through binding/unbinding event from one potential surface to the other, moves on the switched potential surface slightly and quickly turns back to the original potential surface, moves on that surface slightly before the next binding/unbinding event to bring the system back to the other potential surface to start another cycle. There are many such cycles of binding-growth-unbinding-decay with small incremental changes of protein concentrations. The system can be approximated as an diffusion in protein concentration space with an effective barrier by averaging over many binding/unbinding events. This analysis is analogous to electron transfer problem in complex systems [57].

4.3.2 Mean Transition Rate

In an attractor landscape, the lifetime of an attractor reflects its stability, which can be measured by the Mean First Passage Time (MFPT). MFPT is the average transition time induced by noise between attractors on a landscape, since the traversing time represents how easy to communicate from one place to another. Here we calculated mean transition rate $\kappa \sim 1/MFPT$ from the attractors of gene on (off) to gene off (on) state as well as the combined time for both self-repressors and self-activators.

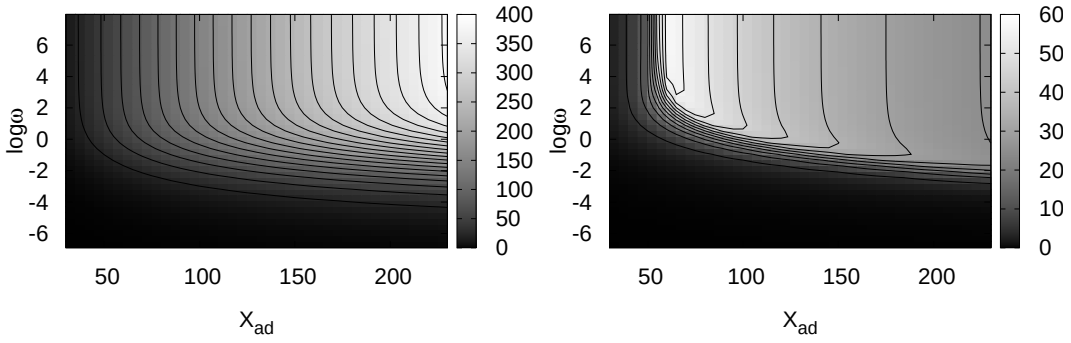
In Figure. 4.6a we found that for self repressor the mean transition rate $\kappa \sim 1/MFPT$ increases from non-adiabatic to adiabatic case. This can be understood as follows: In the non-adiabatic regime, the rate limiting step is the change of the gene state, increasing the binding/unbinding versus synthesis/degradation increases the chance of changing gene state and therefore will boost the kinetic rate. This can also be seen from the fact that the distance between the two peaks becomes smaller as ω increases and off peaks decrease the amplitude significantly while on peaks increases the amplitude steadily. For adiabatic regime, the position of the on and off peak tends to be the same, so transition is fast. Similar behavior is seen for the transition of self activator from mono-stable to mono-stable regime (more details see Supporting Information).

While for self-activators with bi-stability in both non-adiabatic and adiabatic case (Figure. 4.6b), there are turn overs for $\kappa \sim 1/MFPT$ both from "on" state to "off" state and "off" state to "on" state. In non-adiabatic limit, the rate limiting step is again the binding/unbinding events. Therefore, increasing ω will accelerate the kinetics. On the other hand, as ω increases, faster binding/unbinding events and relatively slower protein number fluctuations will lead to the rate limiting step determined more by the effective barrier in protein numbers separating the two basins of attraction (slow time step). As a result, the transitions will be more difficult with the increasing effects from the protein number barrier and κ will be smaller. Combining these two mechanisms together , the mean transition rate κ will first increase, then decrease. Once ω reaches the adiabatic limit, the transition will totally depend on the effective barrier in protein number, which will not change with ω any more, and the κ will reach a plateau, which agree with [93].

We point out here that by measuring the MFPT in the experiment in different conditions, one can also probe the the stability and adiabatic/non-adiabatic dynamics of the gene network.

4.3.3 Entropy Production Rate

In Figure 4.7a, 4.7b, we show the Entropy Production Rate(EPR) for self-repressors and self-activators with different parameters. In the non-adiabatic regime, for both self-repressors and self-activators [68]. EPR increases monotonically with the increase of ω. It is easy to understand here because the distribution is bistable and larger ω leads to more binding/unbinding reactions between two gene states (two



(a) (b)

Figure 4.7: Entropy Product Rate(EPR) respecting to different ω and X_{ad}, for self-repressors with $X_{eq} = 5000$ (a) self-activators with $X_{eq} = 1400$ (b). EPR measures the energy consuming rate for keeping the steady state.

peaks), which will consume more energy. As ω reaches the adiabatic limit, EPR also reaches a limit and won't increase with ω any more. It is because when ω is big enough, the binding/unbinding processes reach equilibrium and even larger ω won't consume more energy. In the intermediate region of ω from bistable in non-adiabatic case to monostable in adiabatic case, EPR of the self-activator increases to a maximum value then decreases, because increasing ω will also suppress the peak either in the unbinding state(off state) or in the binding state(on state) depending on the value of equilibrium constant so that larger ω won't consume more energy in binding/unbinding reactions. But for self-activators with bi-stability in both adiabatic and non-adiabatic conditions, such suppression is much weaker, which leads to a monotonically increases EPR with ω.

On the other hand, in the nonadiabatic limit, EPR for both self-repressors and self-activators increase monotonically as X_{ad} increases. This is because in the nonadiabatic limit there are two peaks and the distance between two peaks(or wells in potential) increases with X_{ad}, which makes system consume more energy for communications between two peaks and thus have higher EPR. In the adiabatic limit, the two peaks of self-repressors in binding and unbinding status overlaps. As X_{ad} increases, the location of both peaks increases, which leads to higher binding rate $\frac{h}{2}n(n-1)$, more binding/unbinging processes and eventually consume more energy. While for self-activators, EPR reaches the maxim value for same ω when it is bistable, because the communication between two peaks will consume more energy. Away from this bistable region, the unbinding states are more and more suppressed

with the increase of X_{ad} while the binding states are more and more suppressed with the decrease of X_{ad} , which means the binding/unbinding reactions are less and less likely to happen and EPR gets lower and lower.

4.4 Discussion

We explore the stochastic dynamics of self regulative genes. We found when the binding/unbinding are relatively fast (slow) compared with the synthesis/degradation of proteins in adiabatic (non-adiabatic) case, the self regulators can exhibit one or two peak (two peak) distributions or basins of attractions in protein concentrations. It shows even with the same architecture (topology of wiring), networks can have quite different functions (phenotypes), consistent with recent single molecule single gene experiments [80, 110]. We further found the inhibition and promotion curves consistent with previous results in adiabatic regime but show significantly different behaviors in non-adiabatic regime due to gene fluctuations. We study the dynamical trajectories of the self regulating genes on the underlying landscapes from non-adiabatic to adiabatic limit, provide a global picture of understanding and show the analogy to the problem of electron transfer between two energy wells. We study the stability and robustness of the systems through mean first passage time from one peak (basin of attraction) to another and found both monotonic and non-monotonic turnover behavior from adiabatic to non-adiabatic regimes. We also explore global dissipation by entropy production and correlate with the stability through landscape topography and kinetics. This study suggests many theoretical predictions for experiments to verify.

4.5 materials

For a dimer self-regulator, the master equations will be as follows:

$$\begin{aligned}\frac{dP_1(n)}{dt} &= k[(n+1)P_1(n+1) - nP_1(n)] + g_1[P_1(n-1) - P_1(n)] \\ &\quad -\frac{h}{2}[n(n-1)]P_1(n) + fP_0(n-2) \end{aligned} \tag{4.8}$$

$$\frac{dP_0(n)}{dt} = k[(n+1)P_0(n+1) - nP_0(n)] + g_0[P_0(n-1) - P_0(n)] + \frac{h}{2}[(n+2)(n+1)]P_1(n+2) - fP_0(n) \quad (4.9)$$

where we ignore the decay of proteins bound on the gene. Also, for probability $P_\alpha(n)$, $(\alpha = 1, 0)$, the dimer (2 protein) bound on the gene is not counted.

The m^{th} moment is defined as[96]:

$$\langle n_\alpha^m \rangle = \sum_n n^m P_\alpha(n) \quad (4.10)$$

To get m^{th} order moment equations, multiply n^m and then sum over n on both sides of equ. (4.8) and (4.9). In principle, moment equations are equivalent to original master equations if we can include all moment equations to infinite order. For moment equations up to second order, which is accurate enough for most conditions, we assume the Gaussian distribution for protein numbers. The moment equations up to the second moment is

$$\frac{dc_1}{dt} = -\frac{1}{2}hc_1\langle n_1^2\rangle + \frac{1}{2}hc_1\langle n_1\rangle + fc_0 \quad (4.11)$$

$$\frac{dc_0}{dt} = \frac{1}{2}hc_1\langle n_1^2\rangle - \frac{1}{2}hc_1\langle n_1\rangle - fc_0 \quad (4.12)$$

$$\frac{d(c_1\langle n_1\rangle)}{dt} = \frac{1}{2}hc_1(\langle n_1^2\rangle - 3\langle n_1\rangle\langle n_1^2\rangle + 2\langle n_1\rangle^3) + fc_0(\langle n_0\rangle + 2) + g_1c_1 - k_pc_1\langle n_1\rangle \quad (4.13)$$

$$\frac{d(c_0\langle n_0\rangle)}{dt} = \frac{1}{2}hc_1(-3\langle n_1^2\rangle + 2\langle n_1\rangle + 3\langle n_1\rangle\langle n_1^2\rangle - 2\langle n_1\rangle^3) - fc_0\langle n_0\rangle + g_0c_0 - k_pc_0\langle n_0\rangle \quad (4.14)$$

$$\frac{d(c_1\langle n_1^2\rangle)}{dt} = -\frac{1}{2}hc_1(3\langle n_1^2\rangle^2 - 2\langle n_1\rangle^4 - 3\langle n_1\rangle\langle n_1^2\rangle + 2\langle n_1\rangle^3) + fc_0(4 + 4\langle n_0\rangle + \langle n_0^2\rangle) + g_1c_1(2\langle n_1\rangle + 1) + k_pc_1(\langle n_1\rangle - 2\langle n_1^2\rangle) \quad (4.15)$$

$$\frac{d(c_0\langle n_0^2\rangle)}{dt} = \frac{1}{2}hc_1(3\langle n_1^2\rangle^2 - 2\langle n_1\rangle^4 - 15\langle n_1\rangle\langle n_1^2\rangle + 10\langle n_1\rangle^3 + 8\langle n_1^2\rangle - 4\langle n_1\rangle) - fc_0\langle n_0^2\rangle + g_0c_0(2\langle n_0\rangle + 1) + k_pc_0(\langle n_0\rangle - 2\langle n_0^2\rangle) \quad (4.16)$$

In above equations, the zero moments, c_1 and c_0 measure the possibility that the gene is unbound and bound; $\langle n_1\rangle$ and $\langle n_0\rangle$ measure the average protein number

when the gene is unbound and bound; $\langle n_1^2\rangle - \langle n_1\rangle^2$ and $\langle n_0^2\rangle - \langle n_0\rangle^2$ measure the lowest order of protein number flustration when the gene is unbound and bound.

The master equations and moment equations are basically the same for self-repressors and self-activators, except $g_1 > g_0$ for self repressors while $g_1 < g_0$ for self activators. The bound state is a "off" state for the self-repressor while an "on" state for the self-activator. So this observation tells us the existing symmetry between them. But there is a big difference between self-repressors and self-activators when the synthesis rate of the off state (g_0 for self-repressors, g_1 for self-activators) goes to zero. Consider the steady-state solution for them by setting $\frac{dP_\alpha(n)}{dt} = 0$. For (4.8), when $n = 0$ and $n = 1$, there are two equations:

$$\frac{dP_1(0)}{dt} = k_P P_1(1) - g_1 P_1(0) = 0 \rightarrow P_1(1) = \frac{g_1}{k_p} P_1(0) \tag{4.17}$$

$$\frac{dP_1(1)}{dt} = k_P[2P_1(2) - P_1(1)] + g_1[P_1(0) - P_1(1)] = 0$$
$$\rightarrow P_1(2) = \frac{1}{2}(\frac{g_1}{k_p})^2 P_1(0) \tag{4.18}$$

When $g_1 \rightarrow 0$ for self-activators, $P_1(1) \rightarrow 0$ and $P_1(2) \rightarrow 0$. Then, for $n = 2$,

$$\frac{dP_1(2)}{dt} = 3k_P P_1(3) - 2k_P P_1(2) + g_1(P_1(1) - P_1(2)) - hP_1(2) + fP_0(0) = 0$$
$$\rightarrow 3k_P P_1(3) + fP_0(0) = 0 \tag{4.19}$$

because $P_1(1) \rightarrow 0$ and $P_1(2) \rightarrow 0$. For both $P_1(3) \geq 0$ and $P_0(0) \geq 0$, $P_1(3) \rightarrow 0$ and $P_0(0) \rightarrow 0$. Checking equations by equations, we find all $P_1(n) \rightarrow 0$ and $P_0(n) \rightarrow 0$ except $P_1(0)$. While for self-repressors, the steady state is not such a trivial solution when $g_0 \rightarrow 0$. Because of the term $-fP_0(0)$ in

$$\frac{dP_0(0)}{dt} = k_P P_0(1) - g_0 P_0(0) + hP_1(2) - fP_0(0) = 0 \tag{4.20}$$

even $g_0 \rightarrow 0$, we can't have any similar conclusion as the self-activator case. So, for self-activators, the off-state is unbound state, once the cell enter state $P_0(0)$, it can't leave anymore because neither binding/unbinding process can happen nor synthesis/degradation can happen if $g_1 = 0$; but for self-repressors, off-state is bound state, the cell can leave the state $P_1(0)$ through unbinding process.

4.6 Exact solution of Steady State

In general, the stochastic model been used to describe biochemical systems with considering the intrinsic noise can be described by master equations:

$$\frac{d}{dt}P(t) = P(t)Q(t) \tag{4.21}$$

where $P_i(t)$ is the probability of the state i at time t and $Q(t)$ is the rate matrix.

For most of the stochastic systems, the distribution will eventually become stationary when they reach steady-state. If a steady-state distribution $\pi(t)$ exists, then it necessarily follows that the rate of change of $\pi(t)$ at steady-state is zero, i.e., $d\pi(t)/dt = 0$, which leads to

$$\pi Q = 0 \tag{4.22}$$

For a general Markov chain, the Q matrix can be decomposed into the sum of three matrixes such as an upper triangular matrix U, a lower triangular matrix L and a diagonal matrix D, i.e. $Q = L + U + D$. There are a variety ways to find the numerical exact solution for (4.22). One is the Jacobi method, expressed as:

$$\pi^{(k+1)} = -D^{-1}(L + U)\pi^{(k)} \tag{4.23}$$

where k is the iteration count. The other one is GaussCSeidel iteration:

$$\pi^{(k+1)} = (D + L)^{-1}(-U)\pi^{(k)} \tag{4.24}$$

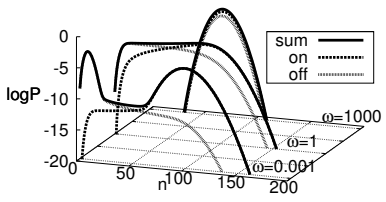
And a more general and efficient method is called Successive over-relaxation(SOR).

$$\pi^{(k+1)} = (D + \omega L)^{-1}[-\omega U + (1 - \omega)D]\pi^{(k)} \tag{4.25}$$

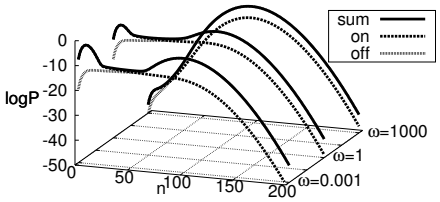
Here we will show the steady state distributions from the exact solutions both for self-repressors and self-activators. Though theoretically the protein number can reach any large number due to the fluctuation in the system, in reality this number is finite and usually within the range under the common parameters. So, one can always choose a large n as a boundary($n = 400$ here), beyond which the probabilities of the protein number states are as small as possible, because the states are barely reached. On the boundary, the zero flux boundary condition is provided here in case there is an overflow. This means no synthesis/decay or binding/unbinding happens

at the state across the boundary.

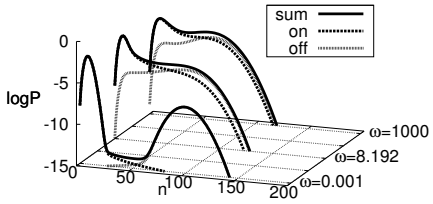
The steady state distributions for self-repressors and self-activators are shown in Fig. 4.8(a), 4.8(b), 4.8(c) and 4.8(d). For self-repressors in Fig. 4.8(a), interactions increase as ω increasing and peaks for the "on" state and "off" state approaches to each other and eventually merge together. For self-activators with $X_{eq} = 50$ in Fig. 4.8(b), there are two strong peaks in both "on" state and "off" state in the non-adiabatic limit when ω is small; the "off" peak will be suppressed as as ω increasing and finally the system will reach mono-stability("on" state side) in the adiabatic(large ω) limit. For self-activators with $X_{eq} = 1400$ in Fig. 4.8(c), for small ω, there are also two peaks in both "on" state(week) and "off" state(strong); the "off" peak will be suppressed as as ω increasing and finally the system will reach bi-stability in the adiabatic(large ω) limit. For self-activators with $X_{eq} = 1400$ in Fig. 4.8(d), for small ω, there are also two peaks in both "on" state(week) and "off" state(strong); the "off" peak is hard to be suppressed as as ω increasing and the system will reach mono-stability("off" state side) in the adiabatic(large ω) limit.



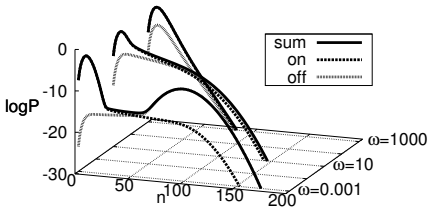
(a) Self-repressors for $X_{eq} = 5000$.



(b) Self-activators for $X_{eq} = 50$.



(c) Self-activators for $X_{eq} = 1400$.

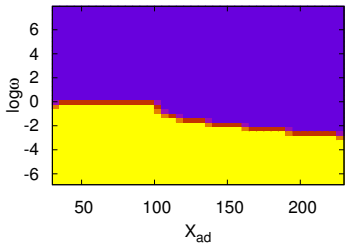


(d) Self-activators for $X_{eq} = 3000$.

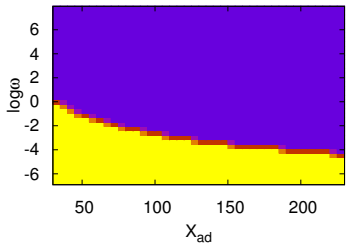
Figure 4.8: The steady state distributions for self-regulators.

Phase diagrams for monostable and bistable are shown in Fig. 4.9(a), 4.9(b),

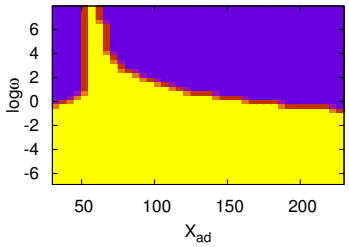
4.9(c) and 4.9(d). For self-repressors in Fig. 4.9(a), the bi-stability appears in small ω region and the mono-stability appears in large ω region. For self-activators in Fig. 4.9(b), 4.9(c) and 4.9(d), the bi-stabilities show not only in in small ω region but also in large ω region for proper X_{ad}(there is no bi-stability region for large ω in Fig. 4.9(b) because the corresponding X_{ad} is too small).



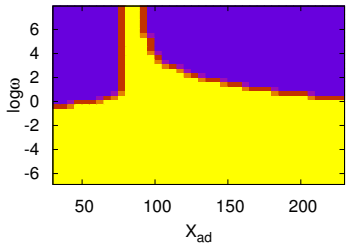
(a) Self-repressors for $X_{eq} = 5000$.



(b) Self-activators for $X_{eq} = 50$.



(c) Self-activators for $X_{eq} = 1400$.



(d) Self-activators for $X_{eq} = 3000$.

Figure 4.9: Phase diagrams for self-regulators. Blue represents monostable and yellow represents bistable.

4.6.1 Fano Factor

Fano factor is defined as the variance/mean of a specific distribution: $F = \sigma^2/\mu$, where σ and μ are the standard deviations and the mean of the probability distribution, and would be equal to 1 for Poisson distribution.

In Fig. 4.10(a), 4.10(b), 4.10(c), 4.10(d), 4.10(e), 4.10(f), 4.10(g), 4.10(h), 4.10(i), 4.10(j), 4.10(k) and 4.10(l), for self-repressors and self-activators with different X_{eq}, the Fano factor F vs ω and X_{ad} are plotted. Fig. 4.10(a), 4.10(d), 4.10(g) and 4.10(j) are the Fano factors of "on" states, Fig. 4.10(b), 4.10(e), 4.10(h) and

4.10(k) are the Fano factors of "off" states, Fig. Fig. 4.10(c), 4.10(f), 4.10(i) and 4.10(l) are the total Fano factors. The Poisson distribution is a good approximation for the on and off states separately in the nonadiabatic limit: $\omega \ll 1$, where the interaction of different gene state is weak and the birth-death term is dominant. But the total Fano factor for the combined probability distribution is much larger than 1, since the summation of two Poisson distributions are not a Poisson distribution. As ω increases, effects from mixing between different gene states produce larger fluctuations into the Poisson distributions and large Fano factors in the intermediate region for ω. It means that the Poisson distribution very poorly suited in these two individual states and the distribution is spread out much, likely with "fatty" tails implying large statistical fluctuations. In this region, stochastic treatment is necessary for taking care of the fluctuations caused by the gene switch. In the adiabatic limit $\omega \gg 1$, for self-repressors and self-activators in the mono-stable region, the Fano factor approaches 1 and the the system becomes Poisson-like again for "on" states, "off" states and total distributions. In this limit, gene state fluctuates rapidly and the system can be considered as the single birth-death one again with an effective synthesis rate from the averaging "on" and "off" states. So the probability distribution for "on" or "off" or the combination are all Poisson like with the two peaks in the same position. For self-activators in the bi-stable region, Fano factors for "on" states, "off" states and total distributions are still much larger than 1 because there are two peaks with "fatty" tails because of strong interactions between "on" and "off" states. It is easy to notice that for self-activators with different X_{eq}, Fano factor diagrams are very similar except some shift in X_{ad}.

On the other hand, in the intermediate adiabatic region, for self-repressors, the Fano factor increases with X_{ad} for "on" state (Fig. 4.10(a)) and "off" state (Fig. 4.10(b)), while for self-activators, the Fano factor increases with X_{ad} for "off" state (Fig. 4.10(e), Fig. 4.10(h), Fig. 4.10(k)) but decreases with X_{ad} for "on" state (Fig. 4.10(d), Fig. 4.10(g), Fig. 4.10(j)). The reason of this happening is as follows: For self-repressors, the distance between "on" peak and "off" peak increases with X_{ad} and the tail for each peak is caused by the mixing effect from the other peak. So larger peak distance will lead to bigger tail for each peak and larger Fano factor. For self-activators, the larger X_{ad} will suppress the "off" state more. Thus the mixing effect from "off" to "on" state will be suppressed and the mixing effect from "on" to "off" state will be promoted, which causes smaller fluctuation in "on" state and

larger fluctuation in "off" state.

4.6.2 Inhibition Curve

The change of gene fluctuations can also be made through the efficiency of repressor and activator: the probability of inhibition for self-repressors and the probability of promotion for self-activators: $P_\beta = c_0$. In the approximation of Ackers *et al*[1] for large ω: $P_\beta = \langle n\rangle^2/(\langle n\rangle^2 + 2X_{eq})$. While as we see in the moment equations (4.11) ~ (4.16), for self-repressors, $P_\beta = c_0 = (\langle n^2\rangle_1 - \langle n\rangle_1)/(\langle n^2\rangle_1 - \langle n\rangle_1 + 2X_{eq})$ where $\langle n\rangle_1$ is the number of proteins in unbound gene state. Only in the limit of $\langle n\rangle = \langle n\rangle_1$ and with Poisson assumption, these two expression will be the same. Fig. 4.11(a) and 4.11(b) show that for self-repressors, the inhibition curves converge to the equilibrium approximation in the limit of large ω. Also, inhibition curves turn over for self-repressors in non-adiabatic region. For the same number of proteins, one can have a solution with unbound protein increase and bound protein decrease as well as a solution with bound protein increase and unbound protein decrease, giving two different trend of inhibition curve in the same regime of average protein number. This is caused by the dimer formation and binding/unbinding for gene regulations. In non-adiabatic region, $\langle n_1\rangle = g_1/k$ and $\langle n_0\rangle = g_0/k$

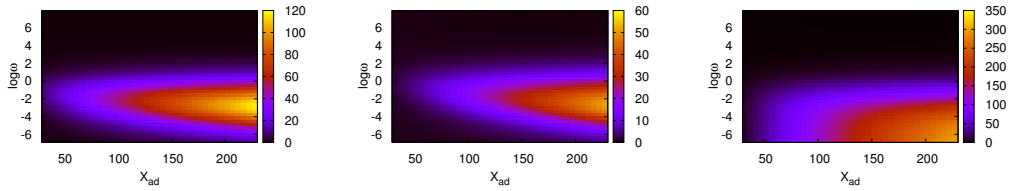
$$\begin{aligned}\langle n\rangle &= c_1\langle n_1\rangle + c_0\langle n_0\rangle \\ &= \frac{2X_{eq}}{\langle n_1\rangle^2 + 2X_{eq}}\langle n_1\rangle + \frac{\langle n_1\rangle^2}{\langle n_1\rangle^2 + 2X_{eq}}\langle n_0\rangle \\ &= \langle n_0\rangle + \frac{2X_{eq}(\langle n_1\rangle - \langle n_0\rangle)}{\langle n_1\rangle^2 + 2X_{eq}} \end{aligned} \tag{4.26}$$

For fixed $\langle n_0\rangle = g_0/k$, the $\langle n\rangle$ vs $\langle n_1\rangle$ curve is not monotonic while the inhibition rate is monotonic with $\langle n_1\rangle$:

$$c_0 = \frac{\langle n_1\rangle^2}{\langle n_1\rangle^2 + 2X_{eq}} \tag{4.27}$$

which make the inhibition rate non-monotonic with $\langle n\rangle$ with dimer binding/unbinding. But for monomer binding/unbinding,

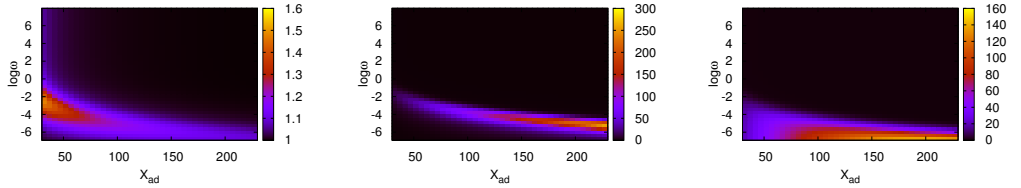
$$\langle n\rangle = \frac{2X_{eq} + \langle n_0\rangle}{\langle n_1\rangle + 2X_{eq}}\langle n_1\rangle \tag{4.28}$$



(a) Fano factors of self-repressors for on state ($X_{eq} = 5000$).

(b) Fano factors of self-repressors for off state ($X_{eq} = 5000$).

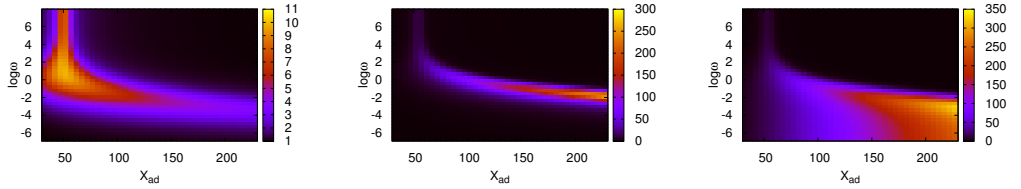
(c) Total Fano factors of self-repressors ($X_{eq} = 5000$).



(d) Fano factors of self-activators for on state ($X_{eq} = 50$).

(e) Fano factors of self-activators for off state ($X_{eq} = 50$).

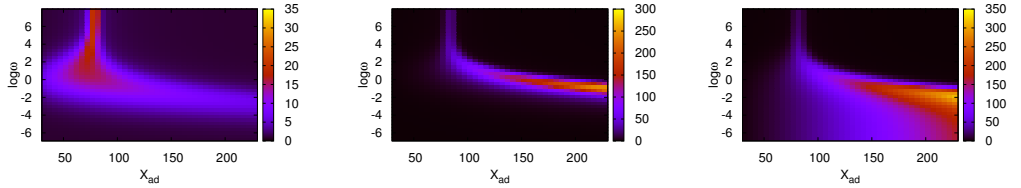
(f) Total Fano factors of self-activators ($X_{eq} = 50$).



(g) Fano factors of self-activators for on state ($X_{eq} = 1400$).

(h) Fano factors of self-activators for off state ($X_{eq} = 1400$).

(i) Total Fano factors of self-activators ($X_{eq} = 1400$).



(j) Fano factors of self-activators for on state ($X_{eq} = 3000$).

(k) Fano factors of self-activators for off state ($X_{eq} = 3000$).

(l) Total Fano factors of self-activators ($X_{eq} = 3000$).

Figure 4.10: Fano factors for self-regulators.

which leads to the monotonic inhibition curve.

While for self-activators, as shown in Fig. 4.11(c)), 4.11(d), 4.11(e), 4.11(f), 4.11(g) and 4.11(h), even in the adiabatic region, if it is bistable, $\langle n \rangle \neq \langle n \rangle_1$ and the promotion curves won't converge to Ackers' approximation. Only in the monostable region of adiabatic limit, $\langle n \rangle = \langle n \rangle_1$ and the promotion curves converge to Ackers' approximation. Not liking self-repressors, the promotion curves don't have turn over behavior in non-adiabatic ω. For self-activators, $\langle n_1 \rangle = g_1/k = 8$ is a constant. According to equ. (4.26), the $\langle n \rangle$ vs $\langle n_0 \rangle$ curve is monotonic and equ. (4.27), the promotion rate is also monotonic with $\langle n_0 \rangle$, which make the promotion rate monotonic with $\langle n \rangle$.

For different ω, the order of inhibition curves for self-repressors and promotion curves for self-activators are opposite. It is because unbound state is "on" state in self-repressors while "off" state in self-activators. When ω increase, the interaction between "on" state and "off" state becomes stronger, which will push two peaks of "on" state and "off" state towards each other. For self-repressors, the "on" state peak will decrease with ω, while for self-activators, the "off" state peak will increase with ω. For self repressor, the unbound state corresponds to gene on state, therefore decreases as binding/unbinding increases. This decreases the probability of inhibition as adibaticity increases. For self activator, the unbound state corresponds to gene off state, therefore increases as binding/unbinding increases. This increases the probability of promotion as adibaticity increases. According to equ. (4.27), their inhibition(promotion) curves will have opposite order in ω.

4.6.3 Mean First Passage Time

We plot the transition rate κ (inverse of mean first passage time: MFPT) from "on" peak to "off" peak (κ_{on}) and vice versa (κ_{off}) as well as the combined ($\kappa_{sum} = \kappa_{on} + \kappa_{off}$)in order to explore the kinetics of traversing from one state to another on the landscape. This is essential in characterizing the stability of the system. We found that for self repressor Fig. 4.12(a) κ increases from non-adiabatic to adiabatic case. This can be understood as follows: in the non-adiabatic regime, the rate limiting step is the change of the gene state, increasing the binding/unbinding versus synthesis/degradation increases the chance of changing gene state and therefore will boost the kinetic rate. Also, as the binding/unbinding increases, the distance between the two peaks becomes smaller, "off" peaks decrease significantly while

"on" peaks increases steadily, which also make the transition easier. For adiabatic regime, the position of the on and off peak tends to be the same, so transition is fast.

For self-activators: Fig. 4.12(b), 4.12(c) and 4.12(d), we have several different cases: In Fig. 4.12(c), we observe non-monotonic turn over behavior of κ_{on}, κ_{off} and κ_{sum} with respect to ω from bistable non-adiabatic case to bistable adiabatic case for $X_{eq} = 1400$. The understand is as following: When ω is small and system is in the non-adiabatic limit, the transition between attractors is more dependent on the rare gene binding/unbinding processes since the binding/unbinding event is the rate limiting step of the whole kinetics. Once the transition is made from one gene state to another (from one potential surface to another), the rest is a fast downhill process towards the potential basin of the other gene state. So increasing ω will lead to faster binding transition and smaller MFPT. On the other hand, as ω increases, faster binding/unbinding events and relatively slower protein number fluctuations will lead to the rate limiting step determined more by the effective barrier in protein numbers separating the two basins of attraction (slow time step). As a result, the transitions over the increasing effects from the protein number barrier will be more difficult and MFPT will be longer. Combining these two mechanisms together , the MFPT will first decrease, then increase. Once ω reaches the adiabatic limit, the transition will totally depend on the effective barrier in protein number, which will not change with ω any more, and the MFPT will reach a plateau.

Fig. 4.12(d) shows monotonic increase of κ_{on}, κ_{off} and κ_{sum} versus ω, from bistable non-adiabatic case to mono-stable adiabatic case for $Xeq = 3000$. While Fig. 4.12(b) shows monotonic increase of κ_{on} and κ_{sum} versus ω but there is a week turn over for κ_{off}, from bistable non-adiabatic case to mono-stable adiabatic case for $Xeq = 50$. Though the basic mechanisms are same for self-activators, here the disappearance of turn overs for κ_{on}, κ_{off} and κ_{sum} in Fig. 4.12(d) and κ_{on} and κ_{sum} in Fig. 4.12(b) is cause by the disappearance of bistabilities in the adiabatic region, i.e., the landscapes become mono-stable before the $kappa$s reach the turn over point.

4.6.4 Entropy Production Rate

Entropy Production Rate(EPR) measures the energy consumed and converted into heat in a unit time by the system to keep the system in a steady state. It measures

the nonequilibrium level for the whole system and is generally defined as:

$$EPR = \sum_{i,j} (M_{ij}P_j - M_{ji}P_i) \ln \frac{M_{ij}P_j}{M_{ji}P_i} \tag{4.29}$$

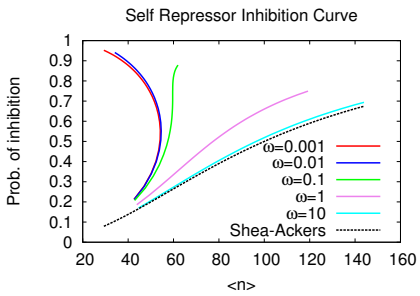
where M_{ij} is the transition rate from state j to state i and P_i is the possibility distribution of state i. EPR of self-regulators are shown in Fig. 4.13(a), 4.13(b), 4.13(c) and 4.13(d).

In the non-adiabatic region, for both self-repressors and self-activators. EPR increase monotonically with the increase of ω. It is because the distribution is bistable and larger ω leads to more binding/unbinding reactions between two gene states, which will consume more energy and product larger EPR. As ω reaches the adiabatic limit, EPR also reaches a limit and won't increase with ω any more because the binding/unbinding processes reach equilibrium and even larger ω won't consume more energy.

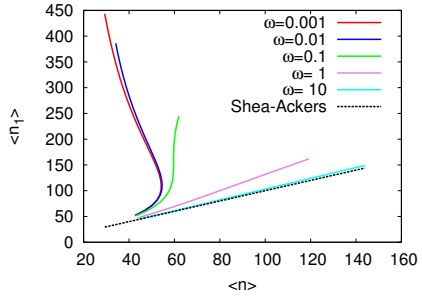
For self-activators, in the intermediate region of ω from bistable in non-adiabatic case to monostable in adiabatic case, EPR of the self-activator increases to a maximum value then decreases, as in Fig. 4.13(b), 4.13(c) and 4.13(d), because increasing ω will also suppress the peak either in the unbinding state(off state) or in the binding state(on state) so that larger ω won't consume more energy in binding/unbinding reactions. But for self-repressors and self-activators with bi-stability in both adiabatic and non-adiabatic conditions, such suppression is much weaker and EPR monotonically increases with ω.

From Fig. 4.13(a), 4.13(b), 4.13(c) and 4.13(d), in the nonadiabatic limit, EPR for both self-repressors and self-activators increase monotonically as X_{ad} increasing because the distance between two peaks increase with X_{ad}, which makes system consume more energy for two-peak communications and thus have higher EPR. While in the adiabatic limit, the two peaks of self-repressors in binding and unbinding status overlaps. As X_{ad} increases, the location of both peaks increase, which leads to higher binding rate $\frac{h}{2}n(n-1)$, more binding/unbinging processes and eventually consume more energy (larger EPR), as shown in Fig. 4.13(a). For self-activators in adiabatic limit, Fig. 4.13(b), 4.13(c) and 4.13(d), EPR reach the maxim value for same ω when it is bistable, because the communication between two peaks will consume more energy. Away from this bistable region, the unbinding states are more and more suppressed with the increase of X_{ad} while the binding

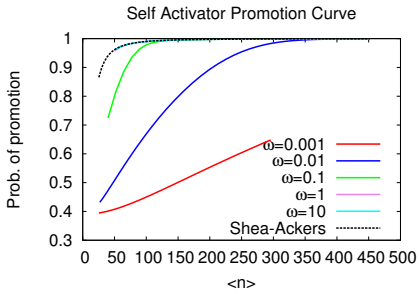
states are more and more suppressed with the decrease of X_{ad} , which means the binding/unbinding reactions are less and less to happen and EPR gets lower and lower. Again, we observed the similarity for EPR diagrams of Fig. 4.13(b), 4.13(c) and 4.13(d) for self-activators with different X_{eq}s.



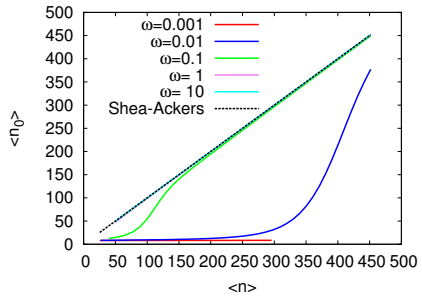
(a) Self-repressors for $X_{eq} = 5000$.



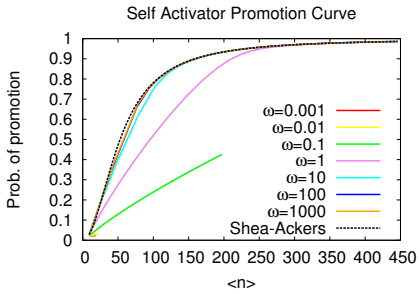
(b) Self-repressors for $X_{eq} = 5000$.



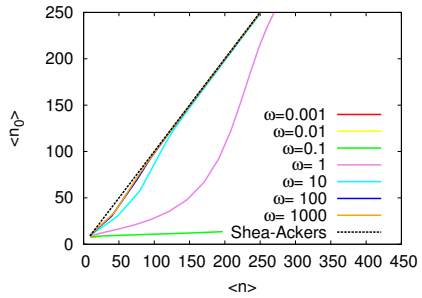
(c) Self-activators for $X_{eq} = 50$.



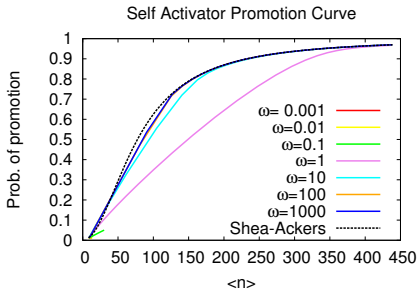
(d) Self-activators for $X_{eq} = 50$.



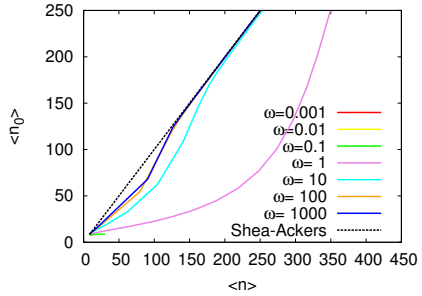
(e) Self-activators for $X_{eq} = 1400$.



(f) Self-activators for $X_{eq} = 1400$.

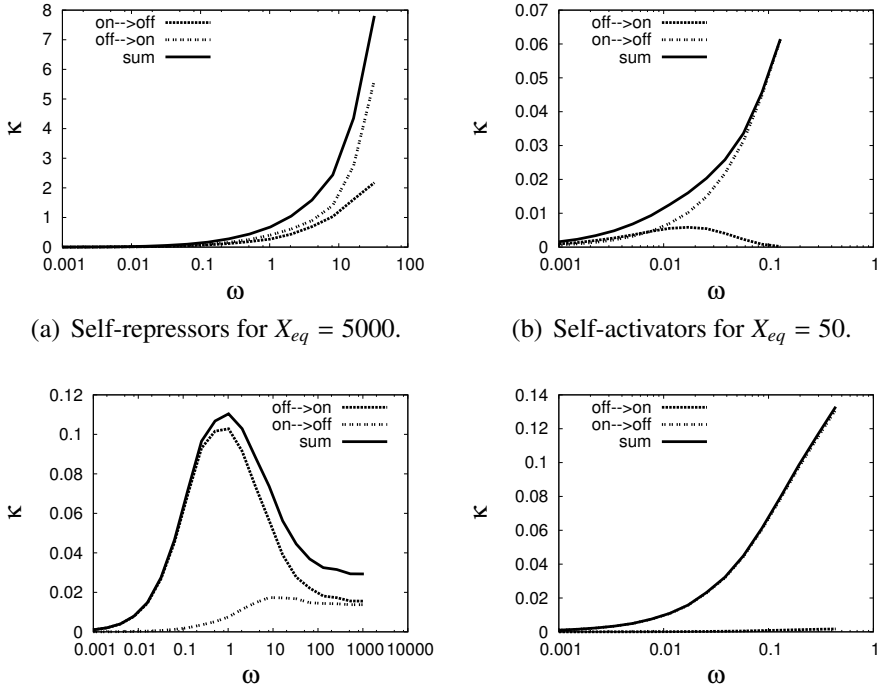


(g) Self-activators for $X_{eq} = 3000$.



(h) Self-activators for $X_{eq} = 3000$.

Figure 4.11: Inhibition(promotion) Curve for self-regulators.



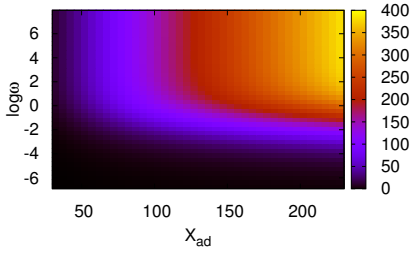
(a) Self-repressors for $X_{eq} = 5000$.

(b) Self-activators for $X_{eq} = 50$.

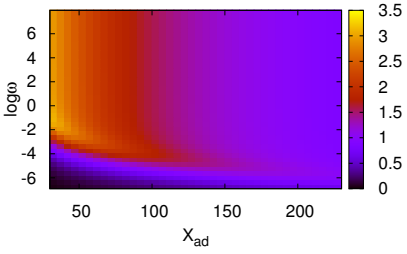
(c) Self-activators for $X_{eq} = 1400$.

(d) Self-activators for $X_{eq} = 3000$.

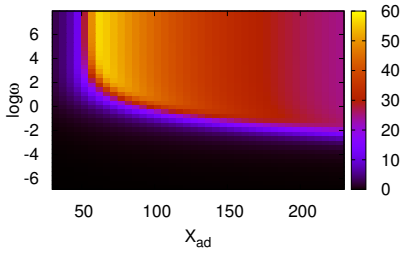
Figure 4.12: Transition rate κ for self-regulators.



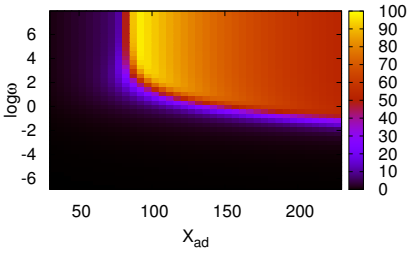
(a) Self-repressors for $X_{eq} = 5000$.



(b) Self-activators for $X_{eq} = 50$.



(c) Self-activators for $X_{eq} = 1400$.



(d) Self-activators for $X_{eq} = 3000$.

Figure 4.13: Entropy Production Rate (EPR) for self-regulators.

Chapter 5

Dominant Kinetic Paths of Complex Systems: Gene Networks

5.1 Introduction

Identifying the dominant paths from one state to another is crucial in understanding the underlying kinetic mechanisms for complex systems [22]. In living cells, biological functions are regulated by gene networks formed through the interactions among genes and associated proteins [17]. Due to the relatively small number of molecules (typically less than 1000) in the cell, the stochastic nature of biochemical reactions can be very important in determining the behaviors in gene expression and gene network patterns[17]. Such stochastic processes can be described mathematically by Markov chains with *master equations*[25]. Epigenetic states as the inherited gene expression patterns are very stable because spontaneous switching events in most genetic systems are rare[77, 33]. Several approaches have been proposed to identify the most dominant paths for stochastic processes between arbitrary states and between stable states to uncover the underlying kinetic mechanisms. One formalism applied to complex systems, protein dynamics and folding is to convert master equations into *Fokker-Planck* diffusion equations in the continues and adiabatic approximations and quantify the dominant paths through the functional variations of the path integral actions in continuous space [64, 100, 56, 6, 105, 91, 90, 19]. Another way is to study the dominant paths for master equations in discrete space after Laplace transformation[100, 82]. Unfortunately, the short microscopic single jumping time scale in gene networks

($10^{-3} \sim 10^{-2}s$) makes computations extremely challenging for macroscopic transition time scales from basin to basin ($10 \sim 1000s$). In this chapter, we present a new formalism to calculate the weight of a path with a relatively large time scale of a single step and a new way to find the dominant paths for general Markov chains. In particular, when the time scale of each step is not so small, the method will give coarse grained dominant paths which can be compared with the Monte Carlo Gillespie simulations [27] or experimental observations. As an example, here we study the dominant paths for gene network motif: self-activators from the basin of "off" state when gene is turned off to the "on" state when gene is switched on. We will also compare dominant paths with the transition trajectories from Gillespie simulations. Finally, we apply our method to calculate the transition rates and transition time scales between basins.

5.2 Method and Material

Consider the master equations in general:

$$\frac{dP(i,t)}{dt} = \sum_j M_{ji} P(j,t) \tag{5.1}$$

where $P(i,t)$ is the possibility of state i at time t and the M_{ji} is from state j to i. The off diagonal term M_{ji} is the probability rate jumping from state j to state i, while the diagonal term M_{ii} is the escape rate from state i. The steady state distribution $P_{SS}(n)$ satisfies 5.1 as $\sum_j M_{ji} P_{SS}(j) = 0$, which can give the underlying potential landscape $U(n) = -\ln(P_{SS}(n))$. With an initial distribution $P(i, t = 0)$, we separate time 0 to T into N intervals, $\delta t = T/N$, the solution of master equations (5.1) can be written in the form of $P(i,T) = \sum_j [\prod_{n=1}^{N} e^{\mathbf{M}_n \delta t}]_{ji} P(j,0), N \to \infty$ with $\mathbf{M}_n$ represents the transition matrix at the time $t_n = n\delta t$. When matrix $\mathbf{M}$ is independent on t, $P(i,T) = \sum_j [e^{\mathbf{M}T}]_{ji} P(j,0)$.

Particularly, choosing initial condition $P(i',0) = \delta_{i',i}$, 5.1 gives the transition probability from state i at $t = 0$ to j at $t = T$, with the formal solution: $P(i,0|j,T) = \sum_k [\prod_{n=1}^{N} e^{M_n \delta t}]_{kj} \delta_{k,i}, \quad N \to \infty$. So to the first order of small δt, we can define $U(i,j,\delta t) \equiv P(i,0|j,\delta t) = \sum_k [e^{M\delta t}]_{kj} \delta_{k,i} = (I + M\delta t)_{ij}$, with $U(i,j,\delta t) = M_{ij}\delta t$ for $i \neq j$ and $U(i,i,\delta t) = 1 + M_{ii}\delta t = e^{M_{ii}\delta t}$. It is equivalent to say: with a N-step specific path from $t = 0$ to $t = T$: $(i_0, 0; ...; i_n, n\delta t; ...; i_N, T)$, the weight or the

transition probability is

$$P(i_0, 0; ...; i_n, n\delta t; ...; i_N, T) = \prod_{n=0}^{N-1} U(i_n, i_{n+1}, \delta t), \quad N \to \infty \tag{5.2}$$

Here, if $i_n \neq i_{n+1}$, $U(i_n, i_{n+1}, \delta t) = M_{i_n, i_{n+1}}\delta t$ and if $i_n = i_{n+1}$, $U(i_n, i_{n+1}, \delta t) = e^{-K_{i_n}\delta t}$ where K_i is the total escape rate from state i: $-M_{ii}$.

However, the formal solution in 5.2 is valid only when the time interval $\delta t \to 0$ or equivalently $N \to \infty$, which means that the number of steps will be huge if we want calculate the transition probability as in 5.2. The computation cost for N time intervals will be m^N where m is the number of nonzero transitions between two states. For typical gene networks, the average time scale for single jump is $10^{-3} \sim 10^{-2}s$ and the systems need at least thousands of steps to finish the transition. To find the most dominant path among $\sim m^{1000}$ paths can be a tremendous computational challenge. Another disadvantage of using 5.2 directly for the dominant path identification is that the resulting dominant path can be unrepresentative because the number of total possible paths will be very big and the weight of each single path will be very small. As $N \to \infty$, even the weight of the most dominant path will be 0.

Here, instead of 5.2, we provide another way to calculate the weight of paths: coarse graining. Mathematically, 5.2 is correct only when $\delta t \to 0$. So when the time intervals δt are finite, $U(i, j, \delta t)$ should be replaced by the the transition possibility $\pi(i, j, \delta t) \equiv P(i, 0|j, \delta t)$ from state i to state j between two neighbor times, which can be solved from 5.1. For each i, set the initial distribution $P(i', t = 0) = \delta_{i',i}$, then the numerical solution $P(j, \delta t)$ of 5.1 represents the transition possibility from state i at $t = 0$ to state j at $t = \delta t$, i.e., $P(i, 0|j, \delta t)$. Thus, the transition probability from state i at $t = 0$ to state j at $t = T$ can be written as

$$P(i, 0|j, T) = \sum_{paths} [\prod_{n=0}^{N-1} \pi(i_n, i_{n+1}, \delta t)] \tag{5.3}$$

Then, the weight (or transition possibility) for a single path $(i_0, i_1, ..., i_N)$ is the product part

$$W(i_0, 0|i_N, T) = \prod_{n=0}^{N-1} \pi(i_n, i_{n+1}, \delta t) \tag{5.4}$$

The dominant path is a path which maximizes the product in 5.4, which can be realized through simulations. We have to point that 5.4 gives the absolute weight for a particular path and the relative weight can also be computed from: $w(i_0, 0|i_N, T) = W(i_0, 0|i_N, T)/P(i, 0|j, T)$.

The advantage of 5.3 is that now we can use a few finite steps, 50 steps for instance, to explore the dominant paths. The result can be called "coarse grained dominant paths". When $N \to \infty$, the expression in 5.2 is equivalent with 5.3. From 5.4, we can find not only the most dominant path among all possible paths, but also the relative weight for that path. One interesting result is that the weight for a single dominant path will become smaller and smaller as number of intermediate steps N get bigger and bigger. It is because as N increases, one step will split into multiple steps, the single path for that step will split into multiple paths and only one of them is more dominant than others.

Then, our new algorithm for quantifying the dominant paths between initial state and final state for Markov chains can be summarize as following:

- **For all is, solve $\pi(i, j, \delta t) \equiv P(i, 0|j, \delta t)$ from 5.1 with initial condition $P(i', t = 0) = \delta_{i',j}$.**

- Starting with an existing path with equal time interval $\delta t = T/N$ for each step, $(i_0, 0; ...; i_n, n\delta t; ...; i_N, T)$, randomly select two states along the path i_{n_α} at $n_\alpha \delta t$ and i_{n_β} at $n_\beta \delta t$, calculate the weight of this path $W(i_{n_\alpha}, n_\alpha \delta t|i_{n_\beta}, n_\beta \delta t)$ following 5.4.

- For each simulation, randomly choose a new path between i_{n_α} at $n_\alpha \delta t$ and i_{n_β} at $n_\beta \delta t$, calculate the weight for the new path $W'(i_{n_\alpha}, n_\alpha \delta t|i_{n_\beta}, n_\beta \delta t)$ following 5.4.

- If the weight W' of the new path bigger than the old path W, replace the old path by the new path. Return to the second step.

As a example. we computed the dominant path from the basin of "off" state when gene is turned off to the basin of "on" state when gene is switched on for self-activators. The circuit is shown as in 5.1. On the DNA, the promotor region can be bound by a dimer of regulatory protein with the rate $\frac{1}{2}h_{\alpha\beta}n(n-1)$ and the dimer can be dissociated from there with the rate f. We use subindex "1" for the DNA bound state and "0" for DNA unbound state. The synthesis rate for the bound

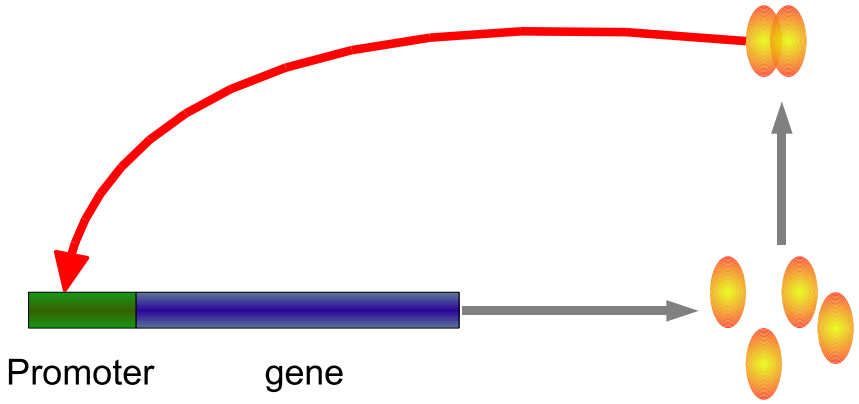


Figure 5.1: Model of self-activators. The dimer of the product protein promotes the protein synthesis rate, forming a positive feedback loop.

state g_1 is bigger than the synthesis rate for the unbound state g_0: $g_1 > g_0$ (activator, gene is activated when regulatory proteins are bound), and the protein degradation rate is k. Then the master equations can be written as

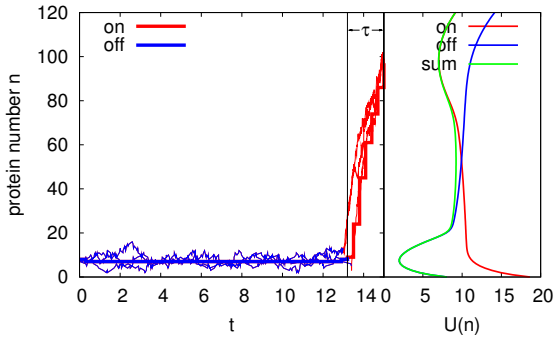
$$\begin{aligned}\frac{dP_1(n)}{dt} &= k[(n+1)P_1(n+1) - nP_1(n)] - \frac{h}{2}[n(n-1)]P_1(n) \\ &\quad + fP_0(n-2) + g_1[P_1(n-1) - P_1(n)] \qquad (5.5)\\ \frac{dP_0(n)}{dt} &= k[(n+1)P_0(n+1) - nP_0(n)] + \frac{h}{2}[(n+2)(n+1)]P_1(n+2) \\ &\quad - fP_0(n) + g_0[P_0(n-1) - P_0(n)] \qquad (5.6)\end{aligned}$$

where $P_\alpha(n)$ is the possibility for n proteins with the promotor region occupied when $\alpha = 1$ and unoccupied when $\alpha = 0$. Here we ignore the decay of proteins bound on the gene. Also, for probability $P_\alpha(n)$, $(\alpha = 1, 0)$, the dimer (2 protein) bound on the gene is not counted. In this chapter, we won't specify the unit of time t and set $k = 1$, $g_1 = 100$, $g_0 = 8$, scaled parameters $\omega = f/k$, $X_{eq} = f/h = 1400$.

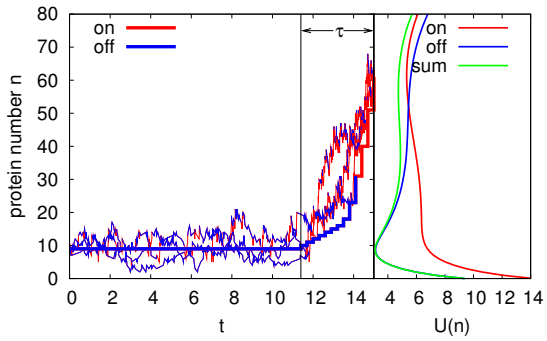
In principle, the dimension of transition matrix of self-activators is almost infinity because the protein number can reach any large number due to the statistical fluctuations. In reality, this number is finite and usually within the range under normal parameters. So, we can always choose a large n as a boundary. Beyond this boundary, the probabilities are zero, because these states rarely can be reached. On the boundary, the zero flux boundary condition is provided here in case there is an overflow. This means no synthesis, decays, binding and unbinding happens passing the state on the boundary. Then, with a finite transition matrix, we can numerically

solve 5.1, and eventually find the dominant path.

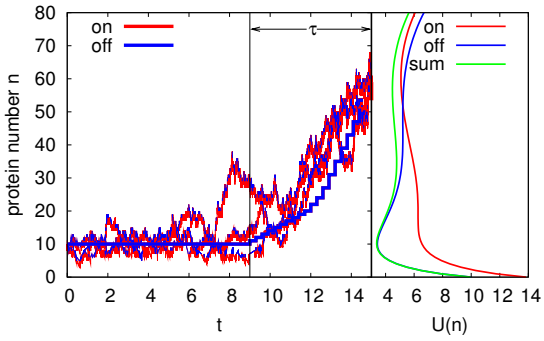
The simulation results of the dominant paths for $T = 15$ and $N = 50$ steps performed for different ω are shown in 5.2 (a), 5.2 (b) and 5.2 (c) as well as typical trajectories from Gillespie simulations [27]. It is easy to notice that dominant paths from 5.3 are quite similar with the typical paths from Gillespie algorithm. When $\omega = 0.1$, the system is in the non-adiabatic region where the binding/unbinding rate is much slower than synthesis/degradation rate. For the dominant path in 5.2 (a), we can observe that the system will stay on the initial state for a long time with gene state unchanged. Once the gene state is changed, the protein number will increase sharply from the "off" state basin to the "on" state basin. It is easy to be understood from the mechanism of self-activators. When the binding/unbinding rate is much slower than synthesis/degradation rate, the chance of gene state switch is really rare and protein number will wait in the "off" state until the gene state is changed. Once the gene state is changed, the protein number will increase from 8 to 97 very fast and jump to the other basin, during which the gene state rarely has chances to switch back. While for the adiabatic region $\omega = 1000$, where the binding/unbinding rate is much faster than synthesis/degradation rate, the protein number increases smoothly from "off" basin to "on" basin. In this case, the binding/unbinding processes happen so frequently that the the average occupancy of DNA determines the average local rate of protein synthesis and degradation. Then, the system acts as if it is diffusing along an effective potential $V(n) = \frac{fg_0+\frac{1}{2}hn(n-1)g_1}{f+\frac{1}{2}hn(n-1)} - kn$ and the DNA occupied probability will be determined locally as $c_{on} = \frac{\frac{1}{2}hn(n-1)}{f+\frac{1}{2}hn(n-1)}$. From the dominant paths in 5.2 (c), we can see that the protein number slowly goes through the one dimensional barrier of $V(n)$ by diffusions, then reaches the bottom relatively fast. Also, the gene state switches from "off" to "on" at $n_{eq} = 53.4$ which satisfy $f \approx \frac{1}{2}hn_{eq}(n_{eq} - 1)$. It doesn't mean that gene state won't change before this point. On the contrary, the gene state switch on and off all the times. The "off" state is more dominant (frequently appearing) when $n < n_{eq}$ and the "on" state is more dominant (frequently appearing) when $n > n_{eq}$. For the intermediate $\omega = 50$ in 5.2 (b), the trend of dominant path is between the adiabatic and non-adiabatic path, as expected. It agrees with the previous classification of dynamics for self-activators as non-adiabatic, adiabatic, and eddy cycles [95]. For all three paths, the system will stay around the most stable state on the path (initial state here) all the time except the necessary transition time. It is very reasonable because otherwise the stable state can't be



(a) $\omega = 0.1$



(b) $\omega = 50$



(c) $\omega = 1000$

Figure 5.2: The paths from "off" basin to "on" basin and potential landscape $U(n)$ for self-activators of three different ω: thick step lines for the dominant path; thin fuzzy lines for typical trajectories. Gene state is represented by color: red for "on" state and green for "off" state.

stable.

In Monte Carlo simulations, we can see many fluctuations in protein number for non-adiabatic ω , binding/unbinding processes for adiabatic ω and eddy cycles for intermediate ω, which do not show up in dominant paths. Because the time interval we use here is $\delta t = 0.3s$ which is much longer than the average time for a single binding/unbinding or synthesis/degradation jump, the typical paths fluctuate around the dominant path. So the dominant paths represent the most important group or tube of transition paths. The dominant paths also provide an rough approximate standard to relative weights for paths appeared in simulations or real experiments. Trajectories close to the dominant paths often have higher weights than trajectories far away from the dominant paths.

Total transition time T play an important role for the transition paths. Different T gives different dominant paths with the same initial and end positions. We can look at the rate of transition probability from $T - \Delta T$ to T, defined as $R(i, j, T, \Delta T) = \frac{1}{\Delta T}[P(i, 0|j, T) - P(i, 0|j, T - \Delta T)]$. $R(i, j, T, \Delta T)$ can be found from the solution $P(i, 0|j, T)$ of 5.1. For the self-activators, the transition probability rate from "off" basin to "on" basin with transition time T is shown in 5.3 for different ω. When T is shorter than the necessary transition time τ as in 5.2 (a), 5.2 (b) and 5.2 (c), the system should go straightly from initial state to final state and the transition probability rate is pretty low. As T increases, the system will have more and more chance reaching final state and $R(T)$ increases with T first. When T is longer than τ, the system has to spend a long period of time by repeatedly visiting certain set of states (the initial state here as in 5.2 (a), 5.2 (b) and 5.2 (c) to arrive the final state on time T. So the longer T, the lower the $R(T)$ is, which leads to the exponential decrease $\propto \exp(-\kappa T)$ as observed in 5.3. The time scale $1/\kappa$ measures relative stability of the attractor and in 5.4, the rate κ vs ω is plotted. In non-adiabatic regime, the rate limiting step for jumping from one basin to another is the binding/unbinding event. κ increases with ω due to the increase of the binding/unbinding rate. Further increases of ω creates an effective barrier in the adiabatic limit which slows the transitions down. So we see the turnover behavior in kinetics of basin to basin transitions from non-adiabatic to non-adiabatic regimes which is important for future experiments [95].

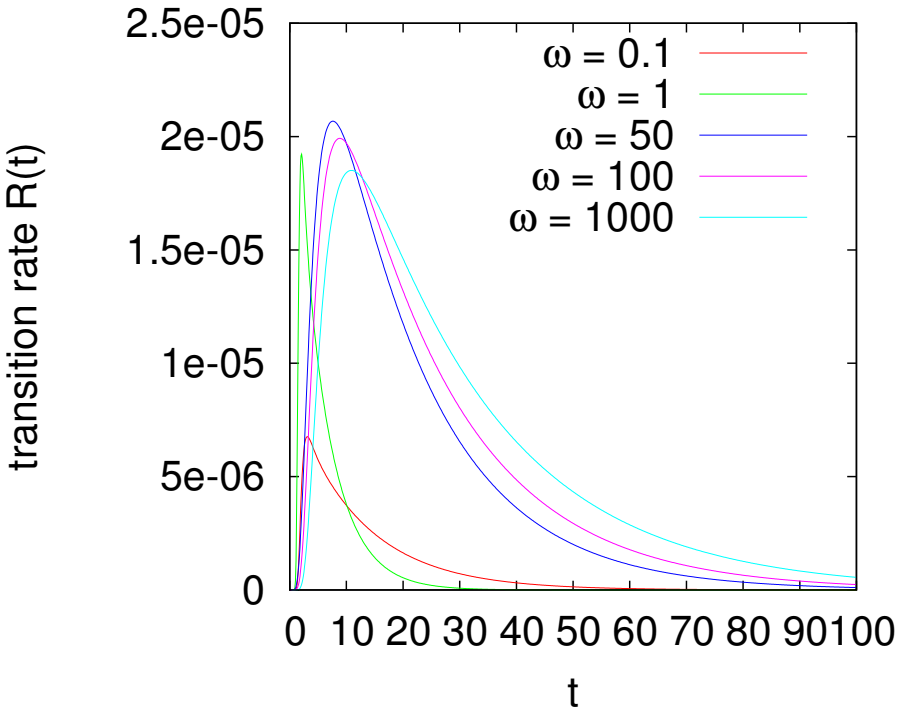


Figure 5.3: The transition probability rate curve $R(T)$ from "on" basin to "off" basin of self-activators for different ω.

5.3 Discussion

As a summary, in this chapter we calculated the dominant paths in discrete space for gene networks. The definition of the weight of a single path provides the possibility of finding other subdominant paths whose weights are of certain percentage, for instance 50%, of the most dominant path. Furthermore, we can split any intermediate step δt into multiple shorter steps to study detail dominant paths within this time interval δt. All formalisms in this chapter can be universally extended to stochastic processes of equilibrium (with detailed balance) and non-equilibrium (without detailed balance) complex systems which can be described by Markov chains.

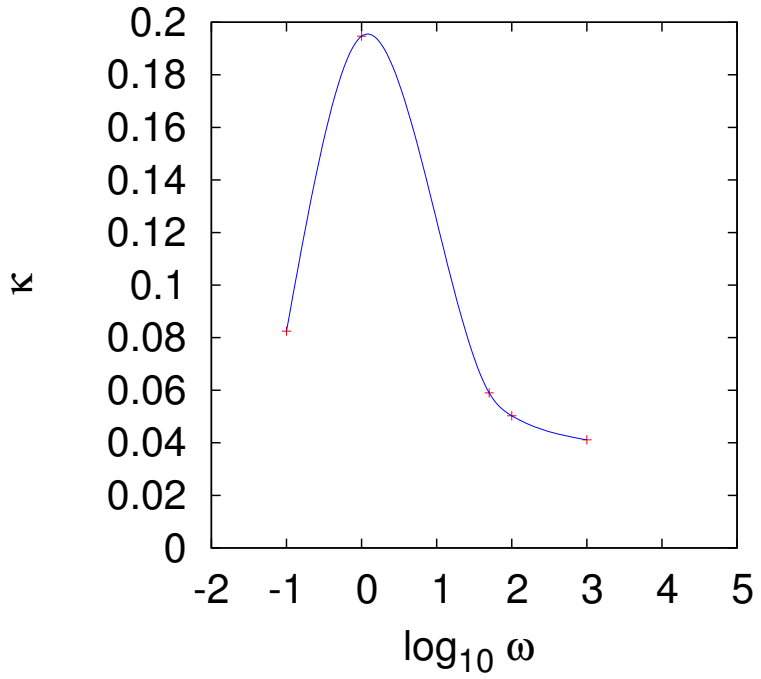


Figure 5.4: Rate κ from "on" basin to "off" basin of self-activators for different ω.

Chapter 6

Landscape and global stability of Non-adiabatic and adiabatic oscillations in gene networks

6.1 Introduction

The dynamical systems and networks are everywhere, ranging from biological and non-biological systems, such as cells, organisms, living creatures, evolution, ecological, social, and economical, atmosphere weather, chemical reactions, energy and information transport, stars and planets, galaxies and the universe. The global stability of these systems is essential for the function. However, it is still a great challenge nowadays to explore the global natures for the complex dynamical systems.

The hint comes from the fact that in equilibrium systems. The dynamics of the system is determined by the gradient of the interaction potential energy. The global stability of complex systems can be quantitatively studied once the underlying interaction energy is known or in other words, the potential energy landscape is known. However, most of the dynamical systems are not integrable non-equilibrium systems. It implies that the dynamics of the system can not be written as a gradient of the potential energy as the equilibrium system. Since the potential landscape is not known, the global stability can not be easily studied. But, analogous to the equilibrium case where the interaction potential energy is related to the equilibrium probability distribution, we can define the underlying potential landscape for dy-

namical non-equilibrium system as logarithmically related to the steady state probability distribution of the system. The non-equilibrium potential landscape defined this way cover the entire state space of the system and can be used to probe the global stability, especially for the oscillatory dynamics as in this study.

In a living cell, dynamic biological behaviors are regulated by complex and diverse genetic networks. The oscillatory behavior as the "biological clocks" is one of the most interesting enigmatic phenomena. Such rhythms exist at many levels in living organisms, from the cell proliferation cycle to the circadian sleep-wake cycle of higher organisms [9, 111, 21, 54, 14]. Recently, many timing mechanisms accompanied with periodic behaviors are intensely studied experimentally and theoretically, including three-gene repressilators [55], self-repressors with explicit time delays [81, 12], circadian clock networks [96] and engineered two component motif from synthetic biology with the interplay of positive feedback and negative feedback(activated repression) [81].

On the other hand, intrinsic fluctuations of gene networks, arise from the number of proteins available in the cell. Furthermore, another type of fluctuations arise from the the biochemical reactions of protein binding/unbinding to the genes, can be significant for oscillatory dynamics. It will be crucial to study of oscillatory behavior in an integrated and coherent way. Conventionally, it was often assumed that the binding/unbinding is significantly faster than the synthesis and degradation (adiabatic limit) [1]. This leads to the expected single stable state for a self repressor which could be measured in experiments [7]. While this condition may hold in some of the prokaryotic cells at certain conditions, in general there is no guarantee it is true. In fact, one expects in eukaryotic cells and some prokaryotic cells, binding/unbinding can be comparable or even slower than the corresponding synthesis and degradation. This can lead nontrivial stable states appearing as a result of new time scales introduced, [37, 79, 45, 15] which is confirmed by recent experiments [80, 11, 110].

In this study, we will establish a spatial landscape framework to explore the global stability and robustness of the dynamical systems and networks. We explore in particular a gene network motif appeared in the experimental synthetic biology studies of two genes mutually repress and activate each other with self activation and repression (activated repression). This network has been engineering in the experiment from synthetic biology and generated robust oscillations in *Escherichia*

coli [81]. In this designed system, as shown in Fig. 6.1 (a), the hybrid promoter ($P_{lac/ara-1}$) is composed of the activation operator site from the araBAD promoter placed in its normal location relative to the transcription start site, and repression operator sites from the lacZYA promoter placed both upstream and immediately downstream of the transcription start site. It is activated by the AraC protein (A) and repressed by the LacI protein (R). The araC, LacI genes are under the control of $P_{lac/ara-1}$ to form co-regulated transcription modules. It was found that if two identical promoters $P_{lac/ara-1}$ control the transcription of AraC and LacI proteins, the network can generate robust oscillations when the binding/unbinding speed is fast [81]. However, we found with the circuit wiring, if the two hybrid promoters controlling AraC and LacI proteins are not identical, robust oscillations can be generated even when the binding/unbinding speed is very slow (non-adiabatic). When the binding/unbinding speed is very slow (non-adiabatic), the gene network provides another level of complexity for the dynamical process. Instead of looking only different proteins, we need also to monitor simultaneously the states of genes. The transcriptional process is suppressed (activated) when the promoter site of the DNA is occupied by a repressor (activator) and enhanced (repressed) when the repressor (activator) is dissociated from DNA. Therefore, the gene state of the promoter is switched on(activated) or off(repressed) is important for the transcription process and the production of functional proteins. So, gene regulation processes involve at least two kinds of biochemical reactions: binding/unbinding reactions of regulatory proteins to the promoters and synthesis/degradation reactions of proteins. Stochastic noise generated by protein binding/unbinding processes was studies in single gene regulation circuits [13, 28], toggle switches [13] and competence cycles in Bacillus subtilis [14]. Here, we found that it can be a possible mechanism for robust oscillation or limit cycles.

In addition to the conventional deterministic chemical rate equations, we will explore the underlying master equations describing intrinsic fluctuations. The solution of the master equations will result in a global probability landscape which can address the stability or robustness of oscillations. We found that coherent limit cycle oscillations emerge in two regimes, adiabatic and non-adiabatic regimes. In both regimes, the spatial landscape has a topological shape of Mexican hat in protein concentrations. The shape of the Mexican hat changes from one spatial location to the other. The shape of the Mexican hat provides the quantitative description of the

capability of the system to communicate with each other. Therefore, the topology of the landscape quantitatively determines the global stability and robustness of the dynamical systems and networks. The coherence of the oscillations are shown to be correlated with the shape of Mexican hat characterize by the height from the top to the ring of the hat.

In the adiabatic regime, the binding/unbinding of regulatory proteins to the promoters are fast compared with the synthesis and degradation rate of the proteins. Mexican hat shape topology of the landscape is mostly determined by the averaged (over the fast binding and unbinding) nature of protein synthesis and decay. The oscillations are more stable as the binding is faster compared with the synthesis. In the non-adiabatic regime, the binding and unbinding of regulatory proteins to the promoter are slow compared with the synthesis and decay rate of the proteins. Mexican hat shape topology of the landscape is mostly determined by the binding-unbinding of regulatory proteins to the genes. The oscillations are more stable as the binding is slower compared with the synthesis. The two regimes give the two mechanisms of producing the spatial temple oscillations: from the adiabatic regime with nonlinear cooperative interactions and from the non-adiabatic regime with time delays due to the slow binding to the gene. Such oscillations are robust in the large range of parameter. With change the binding/unbinding rate, the oscillation period can be easily tuned without change the amplitude much. Such design is suitable for biological rhythms like heartbeats and cell cycles which require a near constant output over a range of frequencies [54]. We also generated robust oscillation both in deterministic and stochastic sense with single 2-step negative feedback loop with suitable time delay due to the slow binding/unbinding process. It means positive feedback is not necessary for oscillations but can make oscillations more robust.

Our landscape framework is general and can be applied to the other dynamical systems and networks to explore the global stability and function.

6.2 Methods and Materials

In a gene regulating network, the hybrid promoter α can be bound by the regulatory protein β with the binding rate $h_{\alpha\beta}$ and dissociation rate $f_{\alpha\beta}$ (both $h_{\alpha\beta}$ and $f_{\alpha\beta}$ can be depend on protein number n_β). The synthesis of protein α is controlled by the gene state of promoter α. For a system with activated repression, there is two types

of genes A and R to be translated into activators A and repressors R respectively. The activators A can bind to the promoter of gene $A(R)$ to activate the synthesis rate of $A(R)$; repressors R can be can bind to gene $A(R)$ to repress the synthesis rate of $A(R)$. Then each gene has 4 state and total system has 16 gene states. Then, with the degradation of proteins and taking transcription and translation as an one-step process (ignore the roles of mRNAs), this model can be expressed by the following chemical reactions:

$$\mathcal{O}_\alpha^{11} + 2\mathcal{M}_\mathcal{A} \quad \underset{f_{\alpha A}}{\overset{h_{\alpha A}}{\rightleftharpoons}} \quad \mathcal{O}_\alpha^{01} \tag{6.1}$$

$$\mathcal{O}_\alpha^{10} + 2\mathcal{M}_\mathcal{A} \quad \underset{f_{\alpha A}}{\overset{h_{\alpha A}}{\rightleftharpoons}} \quad \mathcal{O}_\alpha^{00} \tag{6.2}$$

$$\mathcal{O}_\alpha^{11} + 4\mathcal{M}_\mathcal{R} \quad \underset{f_{\alpha R}}{\overset{h_{\alpha R}}{\rightleftharpoons}} \quad \mathcal{O}_\alpha^{10} \tag{6.3}$$

$$\mathcal{O}_\alpha^{01} + 4\mathcal{M}_\mathcal{R} \quad \underset{f_{\alpha R}}{\overset{h_{\alpha R}}{\rightleftharpoons}} \quad \mathcal{O}_\alpha^{00} \tag{6.4}$$

$$\mathcal{O}_\alpha^{\chi} \quad \xrightarrow{g_\chi^\alpha} \quad \mathcal{M}_\alpha \tag{6.5}$$

$$\mathcal{M}_\alpha \quad \xrightarrow{k_\alpha} \quad \emptyset \tag{6.6}$$

where $\alpha = A$ for activators and $\alpha = R$ for repressors. For the gene state O_α^{ij}, the first index $i = 1(0)$ stands for the activator protein A unbound(bound) on promoter α; the second index $j = 1(0)$ stands for the repressor protein R unbound(bound) on promoter α. $g_{\alpha\chi}$ is the synthesis rate of protein α when the gene α in state χ. k_α is a degradation rate of proteins α. $\mathcal{M}_\beta$ indicates the monomer regulatory protein β. The distribution of the microstate is indicated as $P_{a_a a_r r_a r_r}(n_A, n_R)$ where index $a_a(a_r)$ present gene G_A occupation state by protein $A(R)$ and index $r_a(r_r)$ present gene G_R occupation state by protein $A(R)$. Here activators A bind on gene A and R as a dimer with binding rate $\frac{1}{2}h_{AA}n_A(n_A-1)$ and $\frac{1}{2}h_{RA}n_A(n_A-1)$ respectively; activators R bind on gene A and R as a tetrimor with binding rate $\frac{1}{4!}h_{AR}n_R(n_R-1)(n_R-2)(n_R-3)$ and $\frac{1}{4!}h_{RR}n_R(n_R-1)(n_R-2)(n_R-3)$.

Then, there will be 16 Master equations but we will list them in the supporting information. In the adiabatic limit, the binding/unbinding processes are much faster than the synthesis/degradation. Then the gene states can be averaged out and the system can be simplified into a 2 dimensional birth/death precess with effective

synthesis rate:

$$g^A_{eff} = \kappa_A \frac{[f_a^{-1} + \frac{1}{2!}\frac{n_A^2}{X^A_{eq}}][f_r^{-1} + \frac{1}{4!}\frac{n_R^4}{X^R_{eq}}]}{[1 + \frac{1}{2!}\frac{n_A}{X^A_{eq}}][1 + \frac{1}{4!}\frac{n_R^4}{X^R_{eq}}]} \tag{6.7}$$

$$g^R_{eff} = \kappa_R \frac{[f_a^{-1} + \frac{1}{2!}\frac{n_A^2}{X^A_{eq}}][f_r^{-1} + \frac{1}{4!}\frac{n_R^4}{X^R_{eq}}]}{[1 + \frac{1}{2!}\frac{n_A}{X^A_{eq}}][1 + \frac{1}{4!}\frac{n_R^4}{X^R_{eq}}]} \tag{6.8}$$

The steady state distribution satisfies

$$\frac{dP^{(ss)}_{ijlk}(n_A, n_R)}{dt} = 0 \tag{6.9}$$

for all i, j, l, k. Then the total probability distribution is $P^{(SS)} = \sum_{ijlk} P^{(ss)}_{ijlk}$. One direct way to find the steady state P^{SS} is through Gillespie kinetic simulations.

Also, it is helpful to study deterministic moment equations [12].The m^{th} moment is defined as:

$$\langle n^m \rangle_\gamma = \sum_n n^m P_\gamma(n) \tag{6.10}$$

where γ indicate the general gene state for the whole cell, 1111, 1100, 0001, etc. In principle, moment equations are equivalent to original master equations if we can include all moment equations to infinite order. However, the Hartree-type approximation, an approximations for electron wave functions in multielectron atoms, will be useful. It considers the probability distribution for each type of protein separate from that of the other and only has a mean-field type of effect on the other. With the simplest Poisson assumption, moment equations only involve 16 equations.

From the steady-state distribution $P^{(SS)}$, we can quantify the generalized potential function U of the non-equilibrium network analogous to equilibrium Boltzmann relationship between potential and probability: $U(n_A, n_B) = -\ln(P^{(SS)}(n))$. This will map to the potential landscape.

6.3 Deterministic trajectories and Stochastic Trajectories

The parameters are set as following. The protein degradation rate $k_A = 0.2$ for the activator A and $k_R = 0.005$ for the repressor R. Both genes have maximum protein synthesis rate when they are occupied by the activator A and unoccupied by the repressor R: $g^A_{01} = \kappa_A = 4g^R_{01} = 4\kappa_R$. Here, in $g_{\alpha\beta}$, the first index indicates the state of the promoter site for A and the second index for R. Also, we set the activation factor $f_a = 100$ and repression factor $f_r = 100000$: $g^A_{01} = f_a g^A_{11} = f_r g^A_{00} = f_a f_r g^A_{10}$, $g^R_{01} = f_a g^R_{11} = f_r g^R_{00} = f_a f_r g^R_{10}$. The binding/unbinding precesses are asymmetric between gene A and gene R:

$$h_{RA} \neq h_{AA} = h_A, \quad h_{AR} = h_{RR} = h_R \tag{6.11}$$

$$f_{RA} \neq f_{AA} = f_A, \quad f_{AR} = f_{RR} = f_R \tag{6.12}$$

We fix $\omega_A = f_A/k_A = \omega_R = f_R/k_R = 1000$, same derivative parameter: equilibrium constant $X^A_{eq} = f_{RA}/h_{RA} = f_{AA}/h_{AA} = 450$, $X^R_{eq} = f_{AR}/h_{AR} = f_{RR}/h_{RR} = 33750$ in all binding/unbinding processes. The parameter $\omega_{RA} = f_{RA}/k_A$ indicate the binding/unbinding speed of activator A to the gene R. Solutions of deterministic moment equations for different binding/unbinding rate ω_{RA} of the activator A on gene R are shown in Fig. 6.2.

From the deterministic solutions, we found the system can keep robust oscillations and limit cycles in a large range of parameters, as shown in phase diagram of Fig. 6.1 (c), which are also demonstrated by experiments [81]. However, it is noticed that the oscillation mechanism changes from the adiabatic region (fast binding/unbinding rate ω_{RA}) to the non-adiabatic region (slow binding/unbinding rate ω_{RA}). In adiabatic region, oscillations come from with nonlinear cooperative interactions of negative feedbacks in gene circuits. The trajectories of the activators A are relatively smooth, as shown in Fig. 6.2 (e). While in the non-adiabatic region, activators(A) behaviors like sparking in periods, as shown in Fig. 6.2 (a). Such oscillations are due to the non-adiabatic binding/unbinding processes, whose kinetic mechanism is very different with oscillations in adiabatic region.

In addition, stochastic trajectories from Gillespie simulations for different ω_{RA}s are given in Fig. 6.2 (c), (f) and (i) for different ω_{RA}. In the small ω_{RA} non-adiabatic

region, since the binding rate of A on gene R is small, system will waiting long time in states with small n_A until A is dissociated from gene R. Then the oscillation is initiated by the dissociation and the concentration of the repressor R will decrease without the activation from A. The low concentration of repressor R will have less repressions on activator A and a spark happens which make n_A rocket to a large value. With a large concentration of A, the hybrid promoter on gene R will be higher chance to be bound by the activator A because the binding rate is $\sim \frac{h_{RA}}{2} n_A^2$. Once gene R is occupied by the activator A, n_R will increase and repress n_A to a small value for a long time (because unbinding rate f_{RA} is small in non-adiabatiic limit) until next dissociation of activator A from gene R happens, which starts the the next spark and another round of oscillation (limit cycle). In this process, fluctuations arising from the the biochemical reactions of protein binding/unbinding to the promoters are significant for oscillatory dynamics. While in the large ω_{RA} adiabatic region, the state of gene R switches all the time and the effective protein synthesis rates $g_{eff}(n_A, n_R)$ are determined by the average weight of each gene state which depend on the molecular concentration of A and R. The oscillation process in nonadiabatic region is controlled by the time delay of the negative feedback because of the slow binding/unbinding of activator A. It means the period of the oscillation or limit cycle is more controlled by the binding/unbinding rate of the activator A on gene R. In Fig. 6.3, we demonstrated the changing of oscillation period and amplitude with the changing of ω_{RA}. We noticed that the oscillation period decrease monotonically with the increasing of ω_{RA}. So it can be tuned in a large range by adjusting the binding/unbinding rate ω_{RA} without changing the oscillation amplitudes very much, which is verified in experiments [81]. Such gene expression design is important for biological rhythms like heartbeats and cell cycles which require a near constant output (amplitude) over a range of frequencies [54]. However, in the adiabatic region, because of the different oscillation mechanism (nonlinear cooperative interactions), the oscillation period won' t change any more with the change of the binding/unbinding rate.

6.4 Distributions and Landscapes

The robustness of oscillations can also be shown by probability distributions or potential landscapes. For a stable oscillation, the landscape in A-R plane will be a

clear Mexican hat shape, because the stochastic trajectories fluctuates around the limit cycle and lead a higher probability on the ring than the center. Mexican hat shape landscape is a typical landscape for robust limit cycles (closed loop 2 dimensional oscillation) [96].

By Gillespie simulations, we observed sharp Mexican hat landscapes in both the adiabatic region (large ω_{RA}) and nonadiabatic region (small ω_{RA}), but blurred in the intermediate region between them, as shown in Fig. 6.4. In the adiabatic region, the oscillation mechanism is the nonlinear cooperative interactions of negative feedbacks which works only in large ω_{RA} limit. In the non-adiabatic region, the oscillation mechanism is the time delay of the negative feedbacks which need the condition of small ω_{RA} rate. It is noticed that in the adiabatic region (large ω_{RA}), the ring of the limit cycle in the landscape is relative smooth. However, in the non-adiabatic region, there is disconnected gaps on the right side of the ring and left side of the ring, which is due to the gene state jumping.

Barrier height is a good quantity to measure how sharp the landscape is a Mexican-hat like. It is defined as the potential($U = -lnP^{SS}$) height different between the peak inside the limit cycle and the peak on the limit cycle loop. Barrier heights for different ω_{RA} are shown in Fig. 6.5 (a). With the increase of ω_{RA}, the barrier height first decrease then increase. Such turn over behavior confirmed our observation that there are robust oscillation(sharp Mexican hat) in both the adiabatic region (large ω_{RA}) and nonadiabatic region (small ω_{RA}) but only week oscillation (blurred Mexican hat) in the between. It also confirms our previous conclusions from deterministic analysis.

In addition, we calculated the degree of coherence in oscillation measured by "phase coherence". "phase coherence" is defined as [55]:

$$\xi = \frac{2\sum_i \theta(\phi(t))\phi(t)}{\sum_i |\phi(t)|} - 1 \tag{6.13}$$

where θ is a step function and phase angle $\phi(t)$ is the angle between $\mathbf{N}(t)$ and $\mathbf{N}(t+\tau)$. $\xi \approx 0$ when the trajectory move randomly without any coherence. ξ increases when the regularity of the oscillation increases. ξ approaches 1 when the oscillation is coherent. The phase coherence for different binding/unbinding rate with $\tau = 0.01/k_R$ is shown in Fig. 6.5 (b). We noticed that strong coherent oscillation happened in adiabatic and non-adiabatic region. It correlates with the shape of Mexican hat

characterize by the barrier heights: the height from the top to the ring of the hat.

Both the barrier heights and phase coherence indicate two mechanisms oscillations. In the adiabatic regime (fast binding and unbinding), Mexican hat shape topology of the landscape is mostly determined by the averaged (over the fast binding and unbinding) nature of protein synthesis and decay. The oscillations emerge from nonlinear cooperative interactions and are more stable as the binding is faster compared with the synthesis. In the non-adiabatic regime (slow binding and unbinding), Mexican hat shape topology of the landscape is mostly determined by the binding-unbinding of regulatory proteins to the genes (averaged over the fast synthesis and degradation of proteins). The oscillations emerge from time delays due to the slow binding to the gene and are more stable as the binding is slower compared with the synthesis.

6.5 Oscillation with only Negative Feedback Loops

In addition, we discovered another region in parameter space that would support oscillatory behavior controlled by a time delayed negative feedback loop, with the network circuit shown as in Fig. 6.1 (b) and chemical reactions:

$$\mathcal{O}_A^1 + 4\mathcal{M}_\mathcal{R} \quad \underset{f_R}{\overset{h_R}{\rightleftharpoons}} \quad \mathcal{O}_A^0 \tag{6.14}$$

$$\mathcal{O}_R^1 + 2\mathcal{M}_\mathcal{A} \quad \underset{f_A}{\overset{h_A}{\rightleftharpoons}} \quad \mathcal{O}_R^0 \tag{6.15}$$

$$\mathcal{O}_A^\chi \quad \xrightarrow{g_\chi^A} \quad \mathcal{M}_A \quad , \quad \mathcal{O}_R^\chi \quad \xrightarrow{g_\chi^R} \quad \mathcal{M}_R \tag{6.16}$$

$$\mathcal{M}_\alpha \quad \xrightarrow{k_\alpha} \quad \emptyset \tag{6.17}$$

The time delay effect is from the intermedia step that R can bind on gene A and repress the synthesis of A, A can bind on gene R and enhance the synthesis of R, which forms a two-step negative feedback loop. Thus, the time delay effect strongly depends on the binding/unbinding rate. With suitable time delay and parameters: $k_A = 0.2$, $k_R = 0.05$, $g_1^A = 4g_0^R = \kappa = 80$, $f_A = 100, f_R = 100000$, $\omega_A = f_A/k_A, \omega_R = f_R/k_R = 1000$, $X_{eq}^A = 450$, $X_{eq}^R = 33750$, robust oscillations can be generated from this negative only feedback loop, as shown in Fig. 6.6. Because of the relatively slow binding/unbinding speed ω_A of the activator A, the dynamics is quite similar to the nonadiabatic oscillation case in the dual-feedback loops. According to the deterministic trajectories in Fig. 6.6 (a), (c) and stochastic trajec-

tories in Fig. 6.6 (b), the activator A still oscillates like sharp sparking, just as the case with positive feedback. The oscillation is initiated by the dissociation of the activator A from promoter site of gene R. Then the concentration of R decreases to a level that a sparking of the concentration A is triggered and the activator A rebinds to gene R which promotes the concentration of R and a limit cycle finishes. It also gives a sharp Mexican hat shape landscape as shown in Fig. 6.6 (d). It was discovered that with multiple intermedia steps such as transcription, translation, monomers to dimers, dimers to tetrimors, tetrimors binding on promoters, oscillations can be generated [12, 81]. Here we found that with the non-adiabatic binding/unbinding, robust oscillations can be generated by a 2-step negative only feedback loop. So the positive feedback is not necessary for a stable oscillation, but it will make the oscillation more robust. However, the positive feedback can make the oscillation more robust as shown in phase diagrams Fig. 6.1 (c) and (d). With positive feedback, the oscillations robustly exist from small ω_{RA} to large ω_{RA}, while without positive feedback, oscillations are not robust and exist only in a narrow range of relatively small ω_{RA}. When the binding/unbinding rate is too slow or too fast, the oscillation and limit cycle will perish and the system will return to the monostable steady state. The barrier heights for different ω_{RA} as in Fig. 6.5 (a) show that in the robust oscillation region there are sharp Mexican hat landscapes and the central barrier of the limit cycle is high, which means that the trajectories are kept on the path on the ring and has little chance of crossing the centre barrier. The phase coherence results (for $\tau = 0.01/k_R$) as in Fig. 6.5 (b) also agree with the barrier heights which characterize the shape of Mexican hat: the higher central barrier height, the more robust and coherent oscillations.

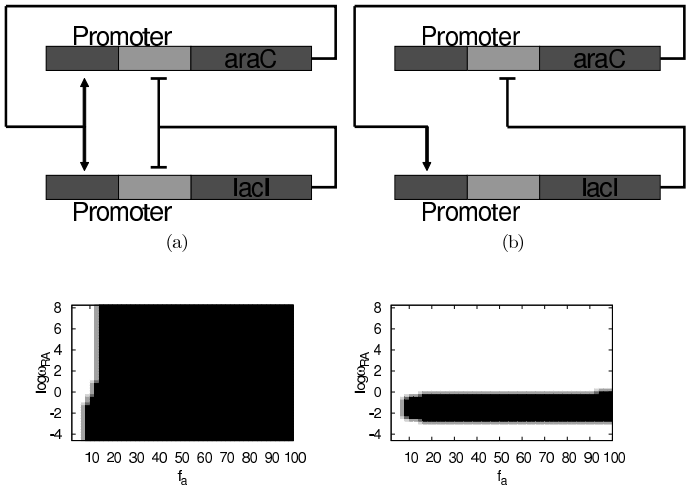
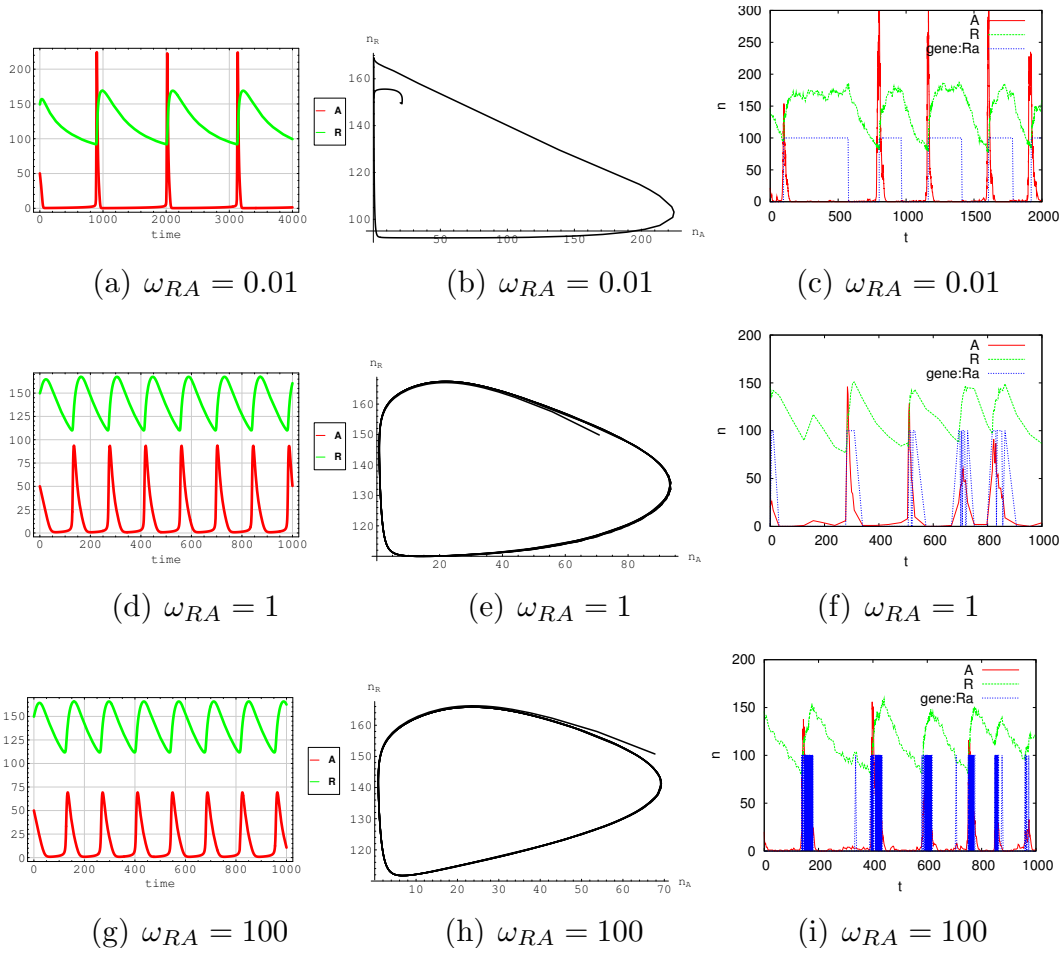


Figure 6.1: (a) Network diagram of the dual-feedback network: two genes mutually repress and activate each other with self activation and repression. (b) Network diagram of single loop negative feedback with one intermedia steps. (c) Phase diagram of dual-feedback network (black: oscillation region; white: monostable region). (d) Phase diagram of single loop negative feedback network (black: oscillation region; white: monostable region).



(a) $\omega_{RA} = 0.01$ (b) $\omega_{RA} = 0.01$ (c) $\omega_{RA} = 0.01$

(d) $\omega_{RA} = 1$ (e) $\omega_{RA} = 1$ (f) $\omega_{RA} = 1$

(g) $\omega_{RA} = 100$ (h) $\omega_{RA} = 100$ (i) $\omega_{RA} = 100$

Figure 6.2: The deterministic trajectory (a), (b), (d), (e), (g), (h) and stochastic trajectories (c), (f), (i) of activator (red) and repressor(green).

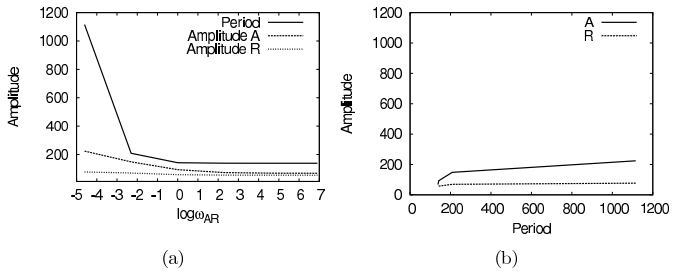


Figure 6.3: Oscillation Period and amplitude from deterministic trajectories.

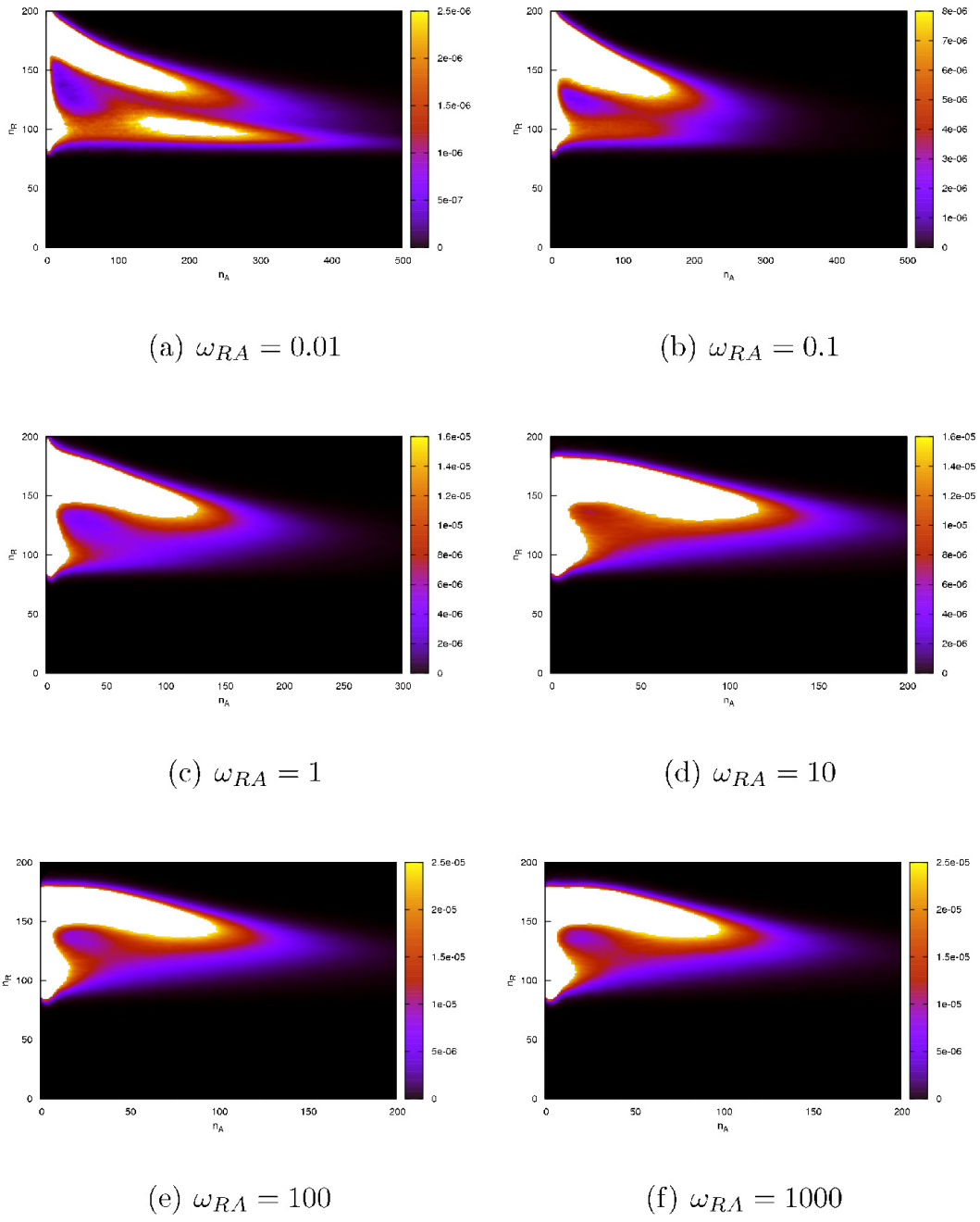
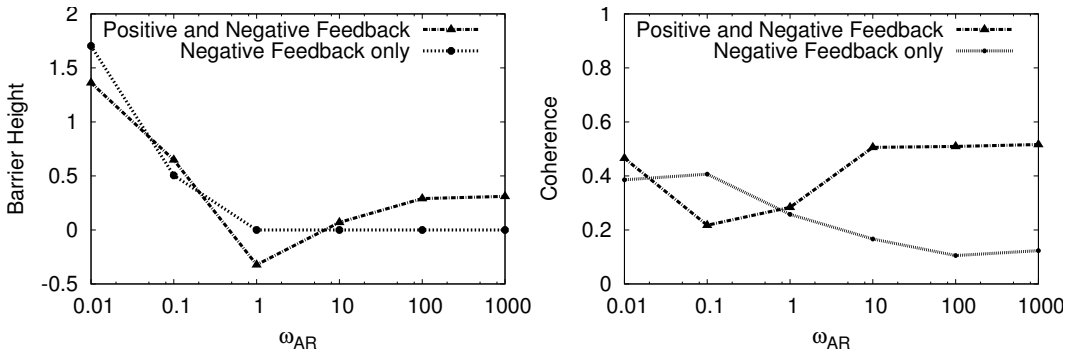


Figure 6.4: Probability distribution landscapes for different binding/unbinding rate ω_{RA} have Mexican hat shape. Most robust oscillations in nonadiabatic (a) and adiabatic region (f) companied by sharpest Mexican hat.



(a) (b)

Figure 6.5: (a) Barrier height of dual-feedback loops and single negative feedback loop with time delay for different binding/ubinding rate ω_{RA}. (b) Phase coherence of dual-feedback loops and single negative feedback loop with time delay for different binding/unbinding rate ω_{RA}.

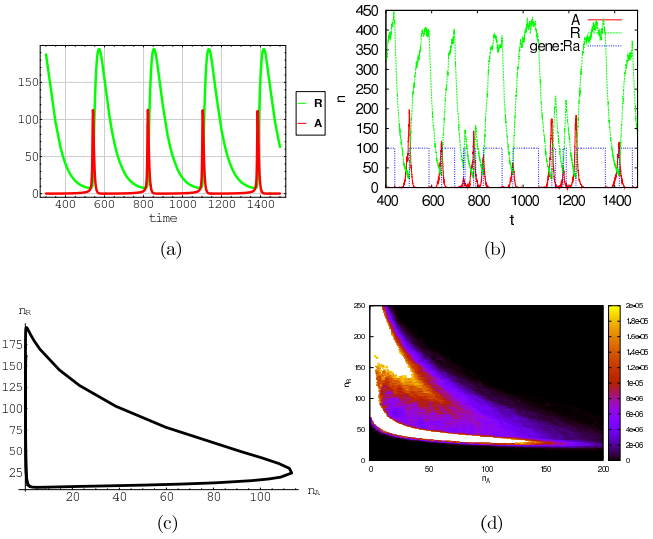


Figure 6.6: Oscillations with negative only feedback with time delay. (a) Deterministic oscillation trajectory. (b) Stochastic trajectory. (c) Deterministic limit cycle. (d) Mexican hat shape probability distribution landscapes.

Bibliography

[1] G. K. Ackers, A. D. Johnson, and M. A. Shea. Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci.*, 79:1129, 1982.

[2] U. Alon, editor. *An Introduction to Systems Biology:Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, FL, 2006.

[3] P. Ao. Potential in stochastic differential equations: novel construction. *Journal of Physics a-Mathematical and General*, 37(3):L25–L30, 2004. 770EQ Times Cited:15 Cited References Count:20.

[4] P. Ao. Potential in stochastic differential equations: Novel construction. *J. Phys. A Math. Gen.*, 37:L25–L30, 2004.

[5] P. Ao. Emerging of stochastic dynamical equalities and steady state thermodynamics from darwinian dynamics. *Communications in Theoretical Physics*, 49:1073, 2008.

[6] E. Aurell and K. Sneppen. Escape Problem for Irreversible Systems. *Phys. Rev. Lett.*, 88:048101, 2002.

[7] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature*, 439:608, 2006.

[8] A. L. Barabasi and E. Bonabeau. Scale-free networks. *Scientific American*, 288(5):60–69, 2003. 667YW Times Cited:51 Cited References Count:4.

[9] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell*, 15:3841, 1998.

[10] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Molecular Biology of the Cell*, 15(8):3841–3862, 2004. 841SK Times Cited:47 Cited References Count:107.

[11] P.J. Choi, L. Cai, K. Frieda, and S. Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322:442–446, 2008.

[12] H. Feng D. Lepzeltera and J. Wang. Oscillation, cooperativity, and intermediates in the self-repressing gene. *Chem. Phys. Lett.*, 490:216–220, 2010.

[13] J. N. Onuchic D. Schultz and P. G. Wolynes. Understanding stochastic simulations of the smallest genetic networks. *J. Chem. Phys.*, 126:245102, 2007.

[14] J. N. Onuchic D. Schultz, E. B. Jacob and P. G. Wolynes. Molecular level stochastic model for competence cycles in bacillus subtilis. *Proc. Natl. Acad. Sci.*, 104:17582, 2007.

[15] J. Das, M. Kardar, and A.K. Chakraborty. Purely stochastic binary decisions in cell signaling models without underlying deterministic instabilities. *Proc. Natl. Acad. Sci.*, 104:18598–18963, 2007.

[16] E. H. et.al. Davidson. A genomic regulatory network for development. *Science*, 295:1669–1673, 2002.

[17] E.H. Davidson and D.H. Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311:796–800, 2006.

[18] M. H. A. Davis. *Markov Models.* Chapman and Hall Publishers, New York., 1993.

[19] C. Dellago, P. Bolhuis, F. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108:1964–1977, 1998.

[20] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[21] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[22] R. P. Feynman and A. R. Hibbs, editors. *Quantum Mechanics and Path Integrals*. McGraw-Hill, New York, 1 edition, 1965.

[23] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

[24] D. Freedman. *Markov Chains.* Springer-Verlag Publishers, Heidelberg, Germany., 1983.

[25] C.W. Gardiner, editor. *Handbook of Stochastic Methods for Physics & Chemistry*. Springer-Verlag, Berlin, 2004.

[26] H. Ge, H. Qian, and M. Qian. Synchronized dynamics and non-equilibrium steady states in a stochastic yeast cell-cycle network. *Mathematical biosciences*, 211(1):132–152, 2008.

[27] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. of Comp. Phys.*, 22:403–434, 1976.

[28] B. Han H. Feng and J. Wang. Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *J. Phys. Chem. B*, 115:1254, 2010.

[29] H. Haken. *Advanced synergetics*. Springer-Verl Berlin etc., 1987.

[30] B. Han and J. Wang. Quantifying robustness and dissipation cost of yeast cell cycle network: The funneled energy landscape perspectives. *Biophysical Journal*, 92(11):3755–3763, 2007. 166TB Times Cited:0 Cited References Count:40.

[31] B. Han and J. Wang. Quantifying robustness and dissipation cost of yeast cell cycle network: The funneled energy landscape perspectives. *Biophys. J.*, 92:3755, 2007.

[32] B. Han and J. Wang. Least dissipation cost as a design principle for robustness and function of cellular networks. *Physical Review E*, 77(3):031922, 2008.

[33] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: Fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, 1990.

[34] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From Molecular to Modular Cell Biology. *Nature*, 402:C47–C52, 1999.

[35] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 79(8):2554–2558, 1982.

[36] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes. Self-regulating gene: An exact solution. *Physical Review E*, 72(5):–, 2005. Part 1 988LQ Times Cited:3 Cited References Count:28.

[37] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes. Self-regulating gene: An exact solution. *Phys. Rev. E.*, 72:051907, 2005.

[38] C. Y. Huang and J. E. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci.*, 93:10078, 1996.

[39] C. Y. & J.E. Jr. Ferrell Huang. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci*, 93:10078–10082, 1996.

[40] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–934, 2001. 429KB Times Cited:674 Cited References Count:29.

[41] T. et al. Ideker. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–933, 2001.

[42] E.A. Jackson and R.W. Rollins. Perspectives of nonlinear dynamics, vol. 1. *American Journal of Physics*, 59:381, 1991.

[43] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000. 362HP Times Cited:871 Cited References Count:28.

[44] S. Kauffman. Homeostasis and differentiation in random genetic control networks. *Nature*, 224:177–178, 1969.

[45] T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81:3116–3136, 2001.

[46] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *European Journal of Biochemistry*, 267(6):1583–1588, 2000. 299TB Times Cited:91 Cited References Count:41.

[47] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267:1583–1588, 2000.

[48] B.N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267:1583–1593, 2000.

[49] K. Kim, D. Lepzelter, and J. Wang. Single molecule dynamics and statistical fluctuations of gene regulatory networks: A repressilator. *J. Chem. Phys.*, 126:034702, 2007.

[50] K. Y. Kim and J. Wang. Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *Plos Computational Biology*, 3(3):565–577, 2007.

[51] K. Y. Kim and J. Wang. Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Comput. Biol.*, 3:e60, 2007.

[52] S. Lapidus, B. Han, and J. Wang. Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of mapk signal transduction. *Proc. Natl. Acad. Sci.*, 105:6039, 2008.

[53] D. Lepzelter, K. Kim, and J. Wang. Dynamics and intrinsic statistical fluctuations of a gene switch. *J. Phys. Chem. B*, 111:10239–10247, 2007.

[54] F. T. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786, 2004. 811GU Times Cited:65 Cited References Count:28.

[55] W. Inoue M. Yoda, T. Ushikubo and M. Sasai. Roles of noise in single and coupled multiple genetic oscillators. *J. Chem. Phys.*, 126:115101, 2007.

[56] R.S. Maier and D.L. Stein. Escape Problem for Irreversible Systems. *Phys. Rev. E*, 48:931–938, 1993.

[57] R. A. Marcus. Chemical and electrochemical electron-transfer theory. *Annu. Rev. Phys. Chem.*, 15:155–196, 1964.

[58] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002. 549KF Times Cited:299 Cited References Count:14.

[59] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 94(3):814–819, 1997.

[60] R. Metzler and P. G. Wolynes. Number fluctuations and the threshold model of kinetic switches. *Chem. Phys.*, 284:469–479, 2002.

[61] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. 607KR Times Cited:468 Cited References Count:24.

[62] F. Moss and P.V.E. McClintock. *Noise in nonlinear dynamical systems*, volume 1. Cambridge Univ Pr, 2007.

[63] B. Novak and J. J. Tyson. Modeling the control of dna replication in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 94(17):9147–9152, 1997. Xr765 Times Cited:54 Cited References Count:36.

[64] R. Olender and R. Elber. Calculation of Classical Trajectories with a Very Large Time Step: Formalism and Numerical Examples. *J. Chem. Phys.*, 105:9299–9315, 1996.

[65] J. Paulson. Summing up the noise in gene networks. *Nature*, 427(6973):415–418, 2004.

[66] J. Paulsson. Number fluctuations and the threshold model of kinetic switches. *Nature*, 427:415, 2004.

[67] M. Ptashne, editor. *A Genetic Switch*. Cell Press and Blackwell Science, Cambridge, 2 edition, 1992.

[68] H. Qian. Entropy production and excess entropy in nonequilibrium steady-state of single macromolecules. *Physical Review E.*, 65:021111, 2002.

[69] H. Qian. Open-system nonequilibrium steady state: Statistical thermodynamics, fluctuations, and chemical oscillations. *The Journal of Physical Chemistry B*, 110(31):15063–15074, 2006.

[70] H. Qian. Entropy demystified the" thermo"-dynamics of stochastically fluctuating systems. *Methods in enzymology*, 467:111, 2009.

[71] H. Qian and D. A. Beard. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chemistry*, 114(2-3):213–220, 2005. 919IZ Times Cited:26 Cited References Count:45.

[72] H. Qian and T. C. Reluga. Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Physical Review Letters*, 94(2):–, 2005. 889YN Times Cited:8 Cited References Count:36.

[73] H. Qian and T. C. Reluga. Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Phys. Rev. Lett.*, 94:028101, 2005.

[74] H. Qian, P. Shi, and J.H. Xing. Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity. *Phys. Chem. Chem Phys.*, 11:4861, 2009.

[75] & P. Mazur. S. R. de Groot. *Non-Equilibrium Thermodynamics. Dover Publications*. Inc. New York, 1984.

[76] M. Sasai and P. G. Wolynes. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2374–2379, 2003. 652DR Times Cited:31 Cited References Count:31.

[77] M Sasai and P.W. Wolynes. Gene regulatory networks and the evolution of animal body plans. *Proc. Natl. Acad. Sci.*, 100:2374–2379, 2003.

[78] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern physics*, 48(4):571, 1976.

[79] D. Schultz, Jos N. Onuchic, and P. G. Wolynes. Understanding stochastic simulations of the smallest genetic networks. *J. Chem. Phys.*, 126:245102, 2007.

[80] A. Singh and L. S. Weinberger. Stochastic gene expression as a molecular switch for viral latency. *Current Opinion in Microbiology*, 12:460–466, 2009.

[81] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456:516–520, 2008.

[82] S. X. Sun. Path Summation Formulation of the Master Equation. *Phys. Rev. Lett.*, 96:210602, 2006.

[83] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12795–12800, 2002.

[84] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci.*, 99:12795, 2002.

[85] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8614–8619, 2001.

[86] M. Thattai and A. Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci.*, 98:8614, 2001.

[87] N. V. Torres. Modeling approach to control of carbohydrate-metabolism during citric-acid accumulation by aspergillus-niger .1. model definition and stability of the steady-state. *Biotechnology and Bioengineering*, 44(1):104–111, 1994.

[88] J. J. Tyson. Modeling the cell-division cycle - cdc2 and cyclin interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16):7328–7332, 1991. Gb701 Times Cited:92 Cited References Count:47.

[89] J. J. Tyson, K. C. Chen, and B. Novak. Network dynamics and cell physiology. *Molecular Biology of the Cell*, 15:236a–237a, 2004. Suppl. S 864QU 1305 Times Cited:0 Cited References Count:2.

[90] C. Valeriani, R.J. Allen, M.J. Morelli, and Pieter Rein ten Wolde. Computing stationary distributions in equilibrium and non-equilibrium systems with forward flux sampling. *J. Chem. Phys.*, 127:114109, 2007.

[91] C. Valeriani, R.J. Allen, and M.J. and Morelli. Computing stationary distributions in equilibrium and non-equilibrium systems with forward flux sampling. *J. Chem. Phys.*, 127:114109, 2007.

[92] J. M. G. Vilar, C. C. Guet, and S. Leibler. Modeling network dynamics: the lac operon, a case study. *Journal of Cell Biology*, 161(3):471–476, 2003.

[93] A. M. Walczak, J. Onuchic, and P.G. Wolynes. Absolute Rate Theories of Epigenetic Stability. *Proc. Natl. Acad. Sci.*, 102:18926–18931, 2005.

[94] A. M. Walczak, M. Sasai, and P. G. Wolynes. Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J.*, 88:828, 2005.

[95] A.M Walczak, J. Onuchic, and P.G. Wolynes. Absolute Rate Theories of Epigenetic Stability. *Proc. Natl. Acad. Sci.*, 102:18926–18931, 2005.

[96] G. M. Verkhivker. Wang, J. Energy landscape theory, funnels, specificity and optimal criterion of biomolecular binding. *Phys. Rev. Lett.*, 90:188101–4, 2003.

[97] J. Wang, B. Huang, X. F. Xia, and Z. R. Sun. Funneled landscape leads to robustness of cell networks: Yeast cell cycle. *PLOS Comp. Biol. 2*, e147:1385, 2006.

[98] J. Wang, B. Huang, X. F. Xia, and Z. R. Sun. Funneled landscape leads to robustness of cellular networks: Mapk signal transduction. *Biophysical Journal*, 91(5):L54–L56, 2006.

[99] J. Wang, B. Huang, X. F. Xia, and Z. R. Sun. Funneled landscape leads to robustness of cellular networks: Mapk signal transduction. *Biophys. J. Lett.*, 91:L54, 2006.

[100] J Wang, J. Onuchic, and P.G. Wolynes. Statistics and Kinetic Pathways of Protein Folding on Complex Energy Landscapes. *Phys. Rev. Lett.*, 76:4861–4864, 1996.

[101] J. Wang, L. Xu, and E. Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271, 2008.

[102] J. Wang, L. Xu, and E. K. Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci.*, 105:12271, 2008.

[103] J. Wang, L. Xu, and E. K. Wang. Robustness, dissipations and coherence of the oscillation of circadian clock: Potential landscape and flux perspectives. *PMC Biophysics*, 1:7, 2008.

[104] J. Wang, L. Xu, and E. K. Wang. Robustness and coherence of a three-protein circadian oscillator: Landscape and flux perspectives. *Biophys. J.*, 97:3038, 2009.

[105] J. Wang, K Zhang, H Lu, and E.K. Wang. Quantifying Kinetic Paths of Protein Folding. *Biophys. J.*, 89:1612–1620, 2005.

[106] J. Wang, K. Zhang, and E. K. Wang. Robustness and dissipation of mitogen-activated protein kinases signal transduction network: Underlying funneled landscape against stochastic fluctuations. *J. Chem. Phys.*, 129:135101, 2008.

[107] S. Y. Wang, Y. P. Zhang, and Q. Ouyang. Stochastic model of coliphage lambda regulatory network. *Physical Review E*, 73(4):–, 2006.

[108] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267(5204):1619–1620, 1995.

[109] P.G. Wolynes, J.N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267(5204):1619, 1995.

[110] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311:1600–1603, 2006.

[111] W. Yu and P. E. Hardin. Circadian oscillators of drosophila and mammals. *Journal of Cell Science*, 119:4793, 2006.

[112] Y. P. Zhang, M. P. Qian, Q. Ouyang, M. H. Deng, F. T. Li, and C. Tang. Stochastic model of yeast cell-cycle network. *Physica D-Nonlinear Phenomena*, 219(1):35–39, 2006. 066ZQ Times Cited:1 Cited References Count:19.

[113] X.M. Zhu, L. Yin, L. Hood, and P. Ao. Calculating biological behaviors of epigenetic states in the phage? life cycle. *Functional & integrative genomics*, 4(3):188–195, 2004.

[114] RKP Zia and B. Schmittmann. Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states. *Journal of Statistical Mechanics: Theory and Experiment*, 2007:P07012, 2007.