Theoretical Studies of Modulated Protein-DNA Interactions and Their Impacts on Gene Regulation

Thesis submitted to the:

Indian Institute of Technology Ropar
In fulfillment of the requirements for the degree of:

Doctor of Philosophy

by

PANKAJ GAUTAM

Under the guidance of

Dr. Sudipta Kumar Sinha

Department of Chemistry
Indian Institute of Technology Ropar
March 2024

CERTIFICATE

This is to certify that the thesis entitled **"Theoretical Studies of Modulated Protein-DNA Interactions and Their Impacts on Gene Regulation"**, submitted by **Pankaj Gautam** to the Indian Institute of Technology Ropar, is a record of bonafide research work under my supervision and I consider it worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

Signature of the Supervisor

Dr. Sudipta Kumar Sinha

Department of Chemistry

IIT Ropar

DECLARATION

I certify that

a. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Gautam

Signature of the Student
Mr. Pankaj Gautam
2017CYZ0006
Date: 28-03-2024

Acknowledgments

List of Symbols

Symbols	Definition
μ	Chemical potential
λ	Protein's activity
Ξ	Grand canonical partition function
q	Site partition function
$\bar{o}$	Occupation index
ϵ_{PD}	Protein-DNA interaction energy
ϵ_{LP}	Free energy of DNA looping
k_l	Elastic constant of DNA
w_{PP}	Protein-protein interaction free energy
M	Number of DNA sites
$s_i,\ \sigma_i$	Binary numbers
i	Index of biomolecular species
k_p	Protein production rate
k_B	Boltzmann constant
k_d	Protein degradation rate
T	Temperature
τ	Correlation time

Notations and Abbreviations

TF	Transcription Factor
RNAP	RNA Polymerase II
GRN	Gene Regulatory Network
FBL	Feedback Loop
FFL	Feedforward Loop
FC	Fold Change
MCS	Monte Carlo Simulation
DR	Dose Response
SSPD	Steady-State Probability Distributions
SE	Super Enhancer
TE	Typical Enhancer
TAC	Transcription Activation Complex
CRR	Core Regulatory Region
PBC	Periodic Boundary Condition
UCNA	Unified Colored Noise Approximation

Abstract

Gene expression describes the process by which instructions encoded in an organism's DNA directs the synthesis of mRNA or protein. The relationship between an input signal and the amount of final gene products in a gene regulatory system is described by a response curve. However, the shape of these curves depends on the detailed interactions among the protein and DNA in gene regulatory networks. A comprehensive understanding of the shape of the response curve and its relationships with the underlying molecular mechanisms by which a gene is transcribed is still challenging from a theoretical and experimental point of view. Therefore, considerable attention has been paid to understanding the mechanisms that determine the shape of the response curve. With the advantage of available genomic data, one can develop predictive models that explore the relationship between the genotype and phenotype of an organism. Thus, theoretical models emerge as a suitable option that provides insight into the various routes of protein-DNA interactions with the response curve. This thesis investigates the relationship between the mechanisms of protein-DNA interactions and the shape of the response curve for gene regulatory networks.

First, we develop a statistical thermodynamic framework for response curves by considering the binding of a transcription factor with the promoter region of genomic DNA. The transcription factors follow various mechanisms during binding, such as cooperative interactions and DNA looping. In cooperative interaction, the transcription factors can spread from a nonspecific binding site to an adjacent specific binding site. DNA looping is another crucial alternative by which two bound transcription factors at large distances come close through protein-protein interaction. These two physical factors promote the self-assembly of transcription factors or the formation of higher-order oligomers on DNA. However, one can control their population by adding suitable input signals that perturb the protein-protein interactions. These input signals may be a selective binding of a small molecule to transcription factors or post-translational modifications such as phosphorylation or acetylation of an amino acid. Both modes alter the binding property of the transcription factors, controlling the population of a selective configuration of a protein-DNA complex. We develop a thermodynamic model in a grand canonical ensemble that corroborates the relationship between an input signal and the population of a selective protein-DNA complex at thermodynamic equilibrium. Precisely, this relationship is the response curve in our study. However, the calculations become difficult for a complex gene regulatory system. Therefore, we use grand canonical Monte Carlo simulation to calculate the response curves for those cases.

The equilibrium thermodynamic analysis of gene regulatory systems is a good starting point, but these systems often experience out-of-equilibrium events that result in alternative steady states. These alternative steady states of gene regulatory systems are critical factors for the functioning of an assembly network. We perform their stochastic dynamic analysis to explore their existence in a gene regulatory network. Here, the evo-

lution of the system is described by a Markov process as realized by a set of elementary reactions whose joint distribution is governed by a master equation. The gene regulatory systems often have correlated noise that alters their dynamics significantly. In this thesis, we explore the role of correlated noise in detail for a few gene regulatory systems.

We also show that our developed thermodynamic model can discern the fate of a cell. To explore this, we consider the p53 signaling network, where the binding of tetrameric phosphorylated p53 to the promoter regions of a few cell fate-determining genes. We use a minimum free energy model and the Ising-based network model to establish a connection between the network topology and cell fate. In particular, the minimum free energy model infers the existence of first-order phase transitions of a damaged cell upon binding of tetrameric phosphorylated p53. Further, we apply our network model to various cancer cell lines ranging from breast cancer (MCF-7), colon cancer (HCT116), and leukemia (K562) that exhibit different network topologies and determine the differential fate of a malignant cell. Together, this thesis investigates modulated protein-DNA interactions and their role in gene regulation in complex regulatory systems.

Keywords: Gene regulation, response curve, protein-DNA interactions, thermodynamic model, Monte Carlo simulation, master equation, cell fate

Contents

Contents

Contents

Chapter 1

Introduction

Gene expression is one of the main biological events by which the instructions encoded in a gene are turned into a function in all living systems. The cell interprets genetic information to guide the synthesis of proteins (Alberts, 2017). Various cells ranging from bacteria to humans, use gene expression for their development and differentiation (Bentovim et al., 2017; Zhou et al., 2019). The central dogma of molecular biology states that the instruction stored in a gene flows from DNA to protein is unidirectional for any organism, and that happens via two steps: the transcription (DNA $\xrightarrow{\text{TRANSCRIPTION}}$ mRNA) and the translation (mRNA $\xrightarrow{\text{TRANSLATION}}$ Protein) (Crick, 1970). During transcription, the instructions encoded on the DNA are transferred to a messenger RNA (mRNA) molecule upon binding an RNA polymerase (RNAP) to a promoter sequence near the beginning of a gene. In the following step, these mRNAs translate into proteins by ribosomal factories (Cech, 2000).

For gene expression, transcription is the critical step that controls the "on" and "off" of genes (Lee and Young, 2013). In this process, a set of regulatory proteins, known as Transcription Factors (TFs), activate (or, more rarely, inhibit) DNA transcription by binding to specific DNA sequences. These TFs bind to the promoter region of a gene. Overall, this process is a collective result of the interplay among TFs, stimuli, and various other biomolecules that work together to fine-tune the production of the number of copies of mRNA mediated via various mechanisms. Nevertheless, genes follow many control mechanisms that are known to determine gene expression (Kouno et al., 2013). These control mechanisms are crucial in molecular biology since they permit the cell/organism to counteract a wide range of inter- and intracellular stimuli and eventually decide a cell's responsive characteristics (Chen et al., 1998; Mazal et al., 2018). Both prokaryotes

and eukaryotes have different regulatory mechanisms for controlling the transcription, but they share a few common features, such as combinatorial control (Buchler et al., 2003). In combinatorial control, specific combinations of TFs can regulate the expression of multiple genes. Such combinations of TFs and their interactions with the promoter regions can unriddle the expression patterns of a complex regulatory system. Thus, transcription is a crucial biological step for gene expression in all living organisms.

Nevertheless, transcriptional regulation is a complex process that requires coordination among various biological events; their quantitative prediction is still challenging (Pennisi, 2020). This thesis explores the two important questions in gene regulation, calculating whether we can predict the shape of the gene expression response curve and its relationships with the underlying molecular mechanisms by which a gene is transcribed. The in-depth basis for exploring and understanding gene regulation comes from pioneering biophysical literature, which suggests the existence of various control layers that significantly affect and regulate the gene expression (Vilar and Leibler, 2003; van Dieck et al., 2009; Tolhuis et al., 2002; Stenger et al., 1994; Rhee et al., 1998; Ackers et al., 1982; Bintu et al., 2005b). Multiple factors that alter the expression include a) binding of particular types of proteins on DNA, b)strength of input signal, c) self-assembly of TFs on DNA, d) physical state of DNA: looped or unlooped form, etc. These biophysical events are crucial from the regulation perspective as they are known to affect the regulation of bacteria and eukaryotic cells. We discuss a few of these factors below.

1.1 Regulation of gene expression

Gene regulation is the process used to control the gene expression. Near the Transcription Start Site (TSS), many factors modulate the recruitment and activity of RNAP (Petrenko et al., 2019). Factors such as the binding of TFs can either enhance or suppress the recruitment of RNAP. A particular class of TFs acts as an activator that promotes the recruitment of RNAP upon their binding at TSS, whereas a repressor replaces the RNAP for the repression. Further, a stimulus can activate TFs that further modulate the recruitment of RNAP – a feature that can be controlled externally. These TFs often experience cooperative interactions and further influence the transcription process. We discuss their role at the transcription level in gene expression below.

1.1.1 TF controlled transcription

TFs control the transcription process by responding to an input signal sourced by diverse stimuli. The type of stimulus can be either the binding of small molecules to TFs, which we refer to as ligand binding, or the chemical modification mechanisms such as phosphorylation and acetylation of a TF (Wenta et al., 2008; Li et al., 2006). These stimuli often promote the formation of active homo- or hetero-oligomeric complexes that bind to the promoter regions of the genes for controlling the recruitment of RNAP (Burz and Ackers, 1996; van Dieck et al., 2009). For example, the phosphorylation of p53 (p53*) promotes the population of tetrameric p53 repressors, which further binds to the promoter regions of a few genes. These genes are expressed on the binding of p53* to promoters and determine if a cell is malignant (Wenta et al., 2008; Li et al., 2006). Thus, the binding of TFs to the promoter plays a significant role in various cellular processes, such as cell development, differentiation, and counteracting responses to cellular stresses. The stimuli-induced TF binding is often linked either with the onsetting or offsetting of diseases (Albert and Kruglyak, 2015; Chen et al., 2018; Lee and Young, 2013). Therefore, understanding the mechanisms behind the binding of modulated TFs to the DNA has become an active area of research for curing various diseases such as cancer, diabetes, and autoimmune disorders (Albert and Kruglyak, 2015; Chen et al., 2018; Lee and Young, 2013).

1.1.2 Strength of input signals

The free energy of interaction between stimuli and TF determines the strength of an input signal. Therefore, one can modify such interaction by altering the type of stimulus or by chemical modification. The stimulus-dependent gene expression is important for controlled gene regulation (Molina et al., 2013; Saravanan et al., 2020). Generally, we require a threshold amount of stimuli or a dose for controlled regulation. For gene regulatory systems, stimuli activate TFs in a dose-dependent manner. The binding of a ligand to a TF alters its functional activity. More often, ligand binding to TF does not affect the binding affinity of TF to DNA; instead, it alters the conformational state of the TF and promotes the population of active complexes (Rhee et al., 1998). The

activated TF-DNA complex further influences other biophysical processes such as DNA looping, altering the binding affinity of RNAP and DNA, and thereby modulating the gene expression. Thus, ligand-dependent modulation offers an avenue to control the intricate mechanism of gene regulation at the molecular level (Bashor et al., 2019).

1.1.3 Oligomerisation of TF

Moreover, TFs can self-assemble into several higher-order homo- or hetero-oligomeric species on DNA under the exposure of input signals (Chen et al., 1998). The self-assembly phenomenon plays a crucial role in modulating the control of gene expression (van Dieck et al., 2009; Vilar and Saiz, 2011). Examples such as tumor suppressor p53, the nuclear factor NF-κB, retinoid nuclear hormone receptor RXR, etc., are known for co-existing populations of dimers, trimers, tetramers, etc. (Kristjuhan et al., 1998; Michida et al., 2020; Vilar and Saiz, 2011). However, various environmental and intracellular signals can modulate their population. These oligomers are known to be the active species that promote differential gene expression (Vilar and Saiz, 2011). Since the population of only a few oligomers are active in gene regulation, they are also crucial for diseases. Therefore, one can relate these differential expressions with diseases ranging from cancer to autoimmune disorders (Wang et al., 2019b; Szymczak et al., 2021). However, we can tune these signals to control their population for developing molecular therapies against those diseases (Stenger et al., 1994; Li et al., 2006). Therefore, understanding the corresponding mechanism behind the modulation of various populations of complex assemblies and its correlation with gene regulation is necessary for developing molecular therapies against diverse diseases.

1.1.4 DNA loop driven transcription

Not only does an input signal promote the formation of higher-order oligomers, but the physical state of DNA also plays a crucial role in their formation. In this regard, DNA looping plays a vital role in forming higher-order oligomers. The transcriptional control for prokaryotes and eukaryotes involves the dependency on the activity of regulatory protein on DNA sites which are very far from direct contact between promoter-bound TFs (Yasmin et al., 2004; Vilar and Leibler, 2003; Tolhuis et al., 2002). These two

simultaneously bound protein or protein complexes interact at two distant DNA sites, forming DNA loops through the *in situ* formation of higher-order protein oligomers (Vilar and Saiz, 2011). In this process, the DNA can act as a tether, where a pair of DNA-bound proteins interacts even though they are a thousand nucleotide pairs away. For such instances, DNA looping is the crucial mechanism facilitating the communication between two distant bound regulatory proteins (Su et al., 1991). Thus, DNA looping is entrenched in various biological processes such as transcription and replication, which assists remote DNA sites' influence on one another (Schleif, 1992; Stenger et al., 1994). Two types of DNA loops exist in the literature: a) short DNA loops and b) Long DNA loops based on the physical forces that assist their formation (Bazett-Jones et al., 1994; Heenan et al., 2020). DNA elasticity is the critical factor for forming short DNA loops, whereas for long loops, uncertain motion in the two DNA parts of a cell until they discover one another is the plausible step (Saiz and Vilar, 2006a). Literature suggests that the main driving factor for long loops is entropy loss when two DNA sites come close from a far distance (Saiz and Vilar, 2006a). The DNA loop significantly affects the initiation of gene expression at the transcription level.

Classic examples of DNA looping have been experimentally demonstrated for the gene expressions in several operons in Escherichia coli (Cournac and Plumbridge, 2013). The *lac operon* is one example where a single tetrameric form of the Lac repressor can bind two operators simultaneously, looping out the intervening DNA. Simultaneous binding of two lac repressors to two operators strengthens the overall interaction with DNA and thereby leads to greater levels of repression in the cell (Garcia and Phillips, 2011). The reason behind such repression is the blocking of the activity of RNAP, which stops the gene expression in the lac operon. The repression can be either at the level of polymerase binding, e.g., due to competition between repressor and polymerase for overlapping binding sites, or at later stages, because RNAP can be trapped in the loop or not be able to recruit an activator (e.g., cyclic AMP receptor protein [CRP]). Not only in *lac operon*, the *gal operon* also share a similar feature, where the dimeric form of Gal repressor (GalR) binds to two operators, O_E and O_I, by forming a DNA loop (Dalma-Weiszhausz and Brenowitz, 1992). The formation of such a DNA loop blocks the accessibility of RNAP to promoters and hinders the formation of an open

complex, thereby promoting repression. External stimuli galactose interacts with GalR and changes its conformational state, which promotes unbinding of GalR and enhances activation. In Figure 1.1, we have presented a schematic view of DNA looping and the protein oligomers in *gal operon* in E. coli bacteria.
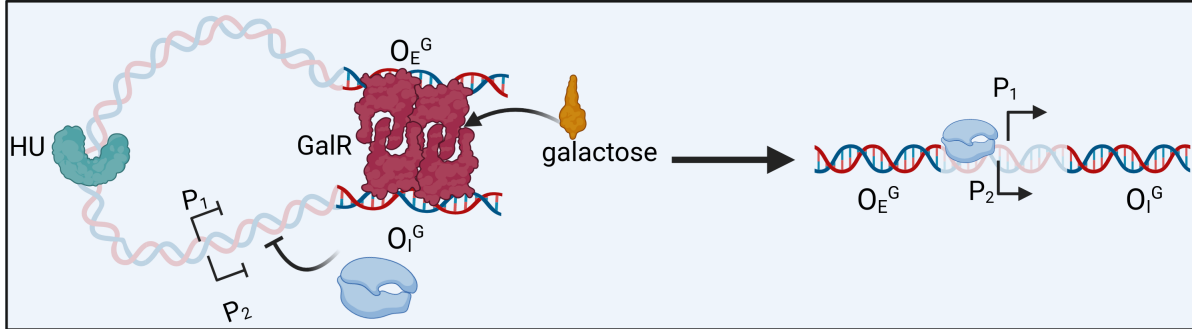
Figure 1.1: A schematic view of various biophysical events in *gal operon*, is presented. Gal repressosome is formed by DNA loop-assisted tetrameric GalR, which represses both promoters, consisting of GalR in the tetrameric form and bound to the two operators, and a histone-like protein HU responsible for bending the DNA.

1.1.5 DNA supercoiling driven transcription

DNA supercoiling is another physical factor that constrains the movement of RNAPs along the DNA during transcription (Kim et al., 2019). The DNA double helix follows an axis typically curved-creating a phenomenon called supercoiling (Hatfield and Benham, 2002). Supercoiling of closed DNA is ubiquitous in various biological processes. It is one of the ways by which DNA can store free energy. In this phenomenon, the amount of twist in a particular DNA strand determines the amount of strain on it. It can arise in two ways: a) when DNA winds around proteins and b) due to topological constraints known as under or overwinding. Understanding supercoiling is crucial for the understanding of the transcription process. At the transcription step, the individual RNAPs generate negative supercoiling upstream and positive supercoiling downstream in DNA (Kim et al., 2019; Sevier and Levine, 2017). It means that the transcription would result in over- or under-winding of a DNA strand from RNAP. Single-molecule experiments explore how molecular motors like RNAPs respond to mechanical interven-

tions, including DNA stretching and twisting (Tripathi et al., 2022). These experiments further show that the accumulation of negative DNA supercoils upstream of an RNAP inhibits the translocation of the polymerase due to torsional stress (Ma et al., 2013).

1.1.6 Modulation of expression by network motifs

In synthetic biology, design principles or finding the underlying gene regulatory network are an active area of research since they offer an understanding of the control mechanism behind the gene expression (English et al., 2021; Zhang et al., 2007; Shen-Orr et al., 2002). More often, these networks form a coordinated interaction pattern among protein and DNA, and they are known to be the network motifs in the gene regulatory systems (Ozbudak et al., 2004; Duddu et al., 2020; Milo et al., 2002). Such coordinated protein-protein and protein-DNA interactions form programmable complex assemblies whose apparent structures are very similar. However, their functional response strongly depends on the topology of the protein-DNA network motifs. The coordinated self-assembly creates gene regulatory network motifs that corroborate the existence of a precise functional response from them at the molecular level. Literature suggests various complex networks exist inside a cell that form decision-making feedback loops that further control the transcription process in a gene regulatory system (Buchler et al., 2003; Webster and Weixlbaumer, 2021; Milo et al., 2002; Mangan et al., 2003). Thus, the network topology is vital in producing phenotype diversity in regulatory circuits for controlling the transcription process. In Figure 1.2, we have depicted a few such intricate network motifs.

1.1.7 Modulation of expression by enhancer-promoter interaction

Enhancer-promoter interaction also contributes to regulating the transcription process (Zuin et al., 2022). An enhancer element is a small part of DNA (50-1500 bp), occupied by regulatory proteins, that increases the transcription of a particular gene by enhancer-promoter interaction (Nolis et al., 2009; Samee et al., 2015). These interactions can be regulated by forming long-range DNA loops, sometimes greater than one kilo of base pairs. For multicellular organisms, distant enhancers generally control promoters, which are megabase pair distance apart. Literature suggests (He et al., 2010; Nolis et al.,
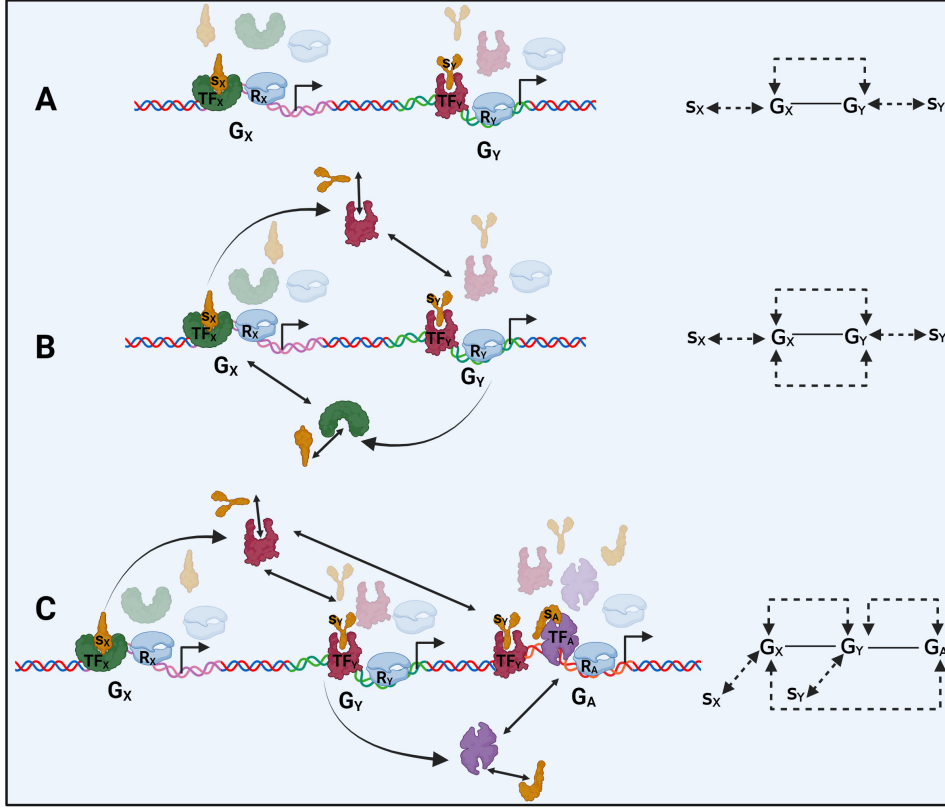
Figure 1.2: A schematic view of various intricate network motifs, (A) activation and repression, (B) feedback loop, and (B) feed-forward loop, is presented. Complex assemblies formed by the free energy of interactions among TFs, R, and s are shown on the promoter regions of G_X, G_Y, and G_A in the left-hand side. The double-headed arrows are shown to represent both the binding and unbinding events. The curly single-headed arrows are shown for the production of biomolecules as they form a specific complex on the promoter regions. The shaded shapes show unbound biomolecules in the figure. Wired diagrams for various network motifs are shown on the right-hand side. We represent the solid lines between two genes as hardware connections, and the dashed lines are for the software of the network motifs. The population of these complex assemblies is controlled by ligand molecules (s).

2009) the role of enhancers in regulating the expression mediated by enhancer-promoter interactions in the three-dimensional nuclear domain. In Figure 1.3, we have presented a schematic diagram where NF-κB controls the gene expression by modulating enhancer-promoter interactions.
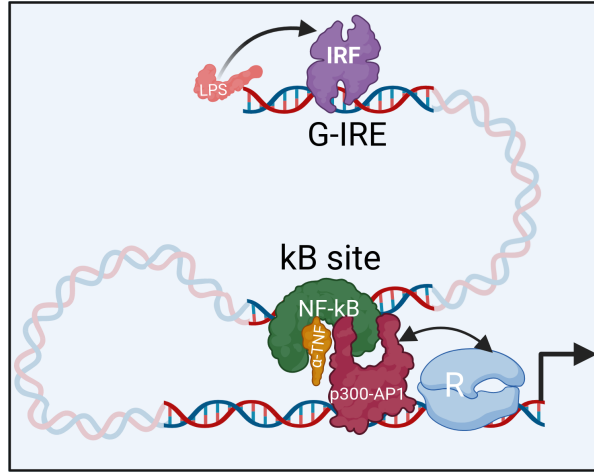
Figure 1.3: A schematic view of enhancer-promoter interactions mediated via DNA loop-ing and NF-κB-DNA interaction energy is presented. The population of these complex assemblies is controlled by ligand molecules such as $\alpha - TNF$ and LPS.

1.2 Biophysical modeling and quantification of gene regulation

Therefore, it is clear from the above discussion that a molecular-level organization hap-pens among TFs, chromatin remodellers, polymerases, and kinases on DNA (Coulon et al., 2013). These proteins and enzymes work concertedly inside the nucleus to regu-late the genes (Luo and Dean, 1999). Understanding the molecular mechanism behind the transcriptional outputs at the gene level is a central question in molecular biology. Recent advances in experimental techniques have enabled the analysis of transcriptional regulation *in vivo* at the scale of single-molecule level (Larson, 2011) to genome-wide (Stamatoyannopoulos, 2012) measurements. Due to an enormous increase in the avail-ability of single-molecule level and genomic data, it becomes necessary to have quantita-tive and predictive models for understanding gene regulation. There are two approaches, a) equilibrium and b) non-equilibrium, to model gene regulation. The equilibrium mod-els are built without considering the explicit dynamics. In contrast, the transcriptional dynamics in non-equilibrium models can exhibit a form of molecular memory so that the system's future behavior depends on its past (Pedraza and Paulsson, 2008; Ahsendorf

et al., 2014). These models are used to relate the concentration of regulators of a given gene to its transcriptional output. Overall, thermodynamic models are often applied to pre-initiation processes, such as binding of TFs and displacement of nucleosomes, and non-equilibrium models are often applied to transcription initiation and elongation (Boettiger et al., 2011; Dobrzyński and Bruggeman, 2009). In this thesis, we have constructed both models to understand gene regulation of various regulatory systems.

1.2.1 Equilibrium thermodynamic models of gene regulation

The biological systems are examples of systems that are out of equilibrium. However, there is a wealth of examples where equilibrium ideas are well justified based on the timescales of underlying process (Bintu et al., 2005b; Weinert et al., 2014; Phillips, 2015). To apply this argument to our gene regulatory systems, we assume (a) the average behavior of the network is invariant over time and (b) the binding kinetics is much faster than other cellular processes, such as cell growth. These two assumptions set the stage for the thermodynamic modeling of gene regulation, which uses simple ideas from statistical mechanics to infer and understand the distributions of regulatory proteins across genomic DNA. The thermodynamic modeling of gene regulation considers a few elementary protein-DNA and protein-protein interactions at equilibrium. These models allow us to compute the probability of each promoter configuration based on the association and dissociation of molecules and displacement of nucleosomes on DNA, without considering the kinetic details (Bintu et al., 2005b). In these models, the regulatory principle is based on the average occupancy of different complexes at promoter DNA, subsequently influencing the accessibility of specific sites such as activator sequences or the TATA box regions (Coulon et al., 2013). The key advantage of this method is that it provides steady-state properties of gene regulatory systems without explicit consideration of the dynamics.

Typically, one can describe the gene expression quantitatively by estimating two quantities: a) fold change and b) dose response. The fold change is defined as the relative population of the TF-RNAP complexes to free RNAP molecules on DNA (Vilar and Saiz, 2011; Saiz and Vilar, 2007a). Similarly, the dose-response is defined as the ratio between the activated TF-RNAP complexes as formed by the specific binding of a drug

molecule with the TFs and the free RNAP molecules on DNA (Chen et al., 1998; Mazal et al., 2018). Nevertheless, typical biological systems have many complexities, such as interacting sites mediated via the DNA loop. However, the statistical thermodynamic description of such complex regulatory systems provides a road map for Monte Carlo simulation using the Metropolis algorithm (Rabier et al., 2021; Gasic et al., 2021). The model, together with the simulation, predicts the occupation numbers of RNAP and active complexes on the DNA and infer the intricate behavior of a cell.

1.2.2 Nonequilibrium dynamics of gene regulation and stochastic models

Although it is worth demonstrating the molecules and macromolecular assemblies of the cell from an equilibrium perspective, their non-equilibrium version must be addressed when there is an energy expenditure (Coulon et al., 2013). The equilibrium models need to be revised when considering the temporal aspects of transcription. Inspired by single-cell level *in vivo* experiments, one can model these systems using a set of dynamical equations that bridge links with the experimental data and provide an insight into the mechanisms of gene regulation (Dar et al., 2012; Geva-Zatorsky et al., 2006). Moreover, these models are also helpful in estimating the timescales of gene regulatory activity (McNally et al., 2000; Dundr et al., 2002; Dion et al., 2007). In a non-equilibrium model, the transcription is directly proportional to the amount of time RNAP spent bound to the promoter (Buchler et al., 2003). Experimental biologists practice these descriptions to probe the mechanistic view of gene regulation since it considers the directionality in reactions and cycling molecular events. Moreover, the non-equilibrium models for transcriptional dynamics can exhibit a form of molecular memory so that the system's future behavior is intimately related to its past (Coulon et al., 2013).

To carry out gene expression dynamics, one can write the coupled differential equations based on mass action kinetics of the elementary biomolecular reactions of a biological system (Cao and Grima, 2018; Ahnert and Fink, 2016; Berg, 2008). These equations together retain the system's equilibrium character as a whole since it obeys the detailed balance principle. However, the detailed balance principle relaxes in the non-equilibrium models. These models take into account how the kinetic organization

of molecular events at gene promoters, which is vital in regulating the time course of transcription. More often, these models are deterministic and possess nonlinear features. Numerical techniques are adopted to solve them. However, the analysis based on the kinetic equations is limited in its ability to characterize probabilistic events, as they do not capture multi-modality in gene expression that arises from slow promoter binding (Ali Al-Radhawi et al., 2019; Thomas et al., 2014). Therefore, one can consider the birth and death processes governed by the elementary reactions in a gene regulatory network. Generally, the time evolution of a grand probability function of a system is governed by a chemical master equation (Vellela and Qian, 2009; Cao and Grima, 2018). However, their analytical tracking is limited as the complexity of the network increases. Thus, analytical methods cannot handle the solution of master equations for complex regulatory systems involving many nonlinear functioning promoters. Therefore, stochastic simulations become popular to infer transcriptional outcomes by considering elementary molecular events in a gene regulatory system (Gillespie, 1976; Saiz and Vilar, 2006b; Cao et al., 2020).

1.3 Literature Review

The complex bio-molecular interactions in a cell regulate and control fundamental cellular processes, including gene expression and signal transduction. Control over gene expression is the central theme of any living system, as it allows an organism to respond by changing external and internal signals. The genes are transcribed to mRNA molecules with perfect regulatory control, further translating to functional proteins (Alberts, 2017; Phillips et al., 2012). A particular class of protein called transcription factors (TFs) activates target genes (TGs) by binding with the upstream or promoter region of the gene that facilitates the recruitment of RNAP for the expression (Cao et al., 2020). A well-known fact in gene regulation is the modulation of interactions between TFs and the promoter region of TGs (Teif and Rippe, 2010; Alberts, 2017; Gerland et al., 2002). The strong interactions between proteins and DNA and preceding events control gene expression (Dangkulwanich et al., 2014). Several biophysical events such as the strength of input signals, ligand binding (Chen et al., 1998), acetylation (Li et al., 2006), phosphorylation (Wenta et al., 2008; Kang et al., 2009); self-assembly of TFs (Hanson et al.,

2005; van Dieck et al., 2009); and the physical state of DNA, etc. (Vilar and Saiz, 2005; Boedicker et al., 2013) control the final interaction between protein and DNA. In particular, these modulated protein-DNA interactions have been observed in a class of regulatory systems such as *g*al, *l*ac and *d*eo operons in *E.* coli (Nick and Gilbert, 1985; Schleif, 1992; Dalma-Weiszhausz and Brenowitz, 1992), the lysogenic to lytic switch in phage λ (Ptashne, 2004), the human β-globin locus (Tolhuis et al., 2002), the nuclear hormone receptor RXR (Yasmin et al., 2004) and tumor suppressor protein p53 (Stenger et al., 1994).

The response curve typically describes the relationship between input signals and the amount of final gene product. However, the shape of these curves depends on the detailed information of gene regulatory networks. A comprehensive understanding of complex genetic networks and their correlation with the shape of the response curves is challenging from an experimental and theoretical point of view. Characterizing and decoding the connectivity of gene regulatory networks is difficult since their regulatory activities are influenced by each other. Due to an increase in the availability of genomic data, it becomes necessary to have predictive models (Phillips, 2015; Wong and Gunawardena, 2020). Thus, theoretical models emerge as a suitable option that gives insight into the mechanism of molecular organization by considering the various possible bio-physical events linked with the gene regulation (Ackers et al., 1982; Bintu et al., 2005b; Teif, 2007). Such models put limits when considering a highly complex multi-component system with interacting genes and inter-particle interactions. In recent times, despite the vital role of the quantitative modeling of gene regulation in biological systems, the mechanisms that determine the shape of the response curve are still challenging.

Researchers have presented thermodynamic models wherein the probability of finding the system in a given regulatory state is a function of the free energy associated with each system state under quasi-equilibrium approximation. The activity of a gene is assumed to be proportional to the probability of a bound RNAP to the promoter sequence (Buchler et al., 2003; Alberts, 2017). In particular, regulation is quantified by the "response characteristics" indicating the gene expression level as a function of TF or ligand concentration (Estrada et al., 2016; Buchler et al., 2003; Bintu et al., 2005b). Therefore, calculating the occupancy index of RNAP on the promoter is crucial since

it quantitatively impacts the transcription process. However, kinetics-based models are also reported extensively for gene regulation, and they do not require as many assumptions but increase the number of parameters as complexity increases (Berg, 2008; Vilar and Saiz, 2014). Kinetic-based methods also have two types: 1) Deterministic and 2) Stochastic. The analysis based on the deterministic techniques is limited in characterizing probabilistic events. Moreover, these models do not capture multi-modality in gene expression that arises from slow promoter binding (Kapuy et al., 2009; Vellela and Qian, 2009; Sahoo et al., 2018). The stochastic models offer an understanding of the multiple steady states for the transcriptional outcomes (mRNA, protein) rendered by the solution of the master equation.

Another vital area of research in gene regulation is its link with various human diseases. The disruption of a cellular network is responsible for many human diseases, such as cancer, diabetes, and autoimmune disorders (Chen et al., 1998). Researchers have explored molecular mechanisms to understand how overexpression and underexpression of a gene are associated with the origin of various diseases. One such example includes the p53 repressor signaling system, which demonstrates how cells behave under environmental stress and determine their fate based on the overexpression or underexpression (Hanson et al., 2005; van Dieck et al., 2009; Navalkar et al., 2022). Thus, the detailed understanding of the interplay between modulated protein-DNA interactions and genetic responses in various systems is one of the central problems shaping the healthcare and pharmaceutical industry.

From the above literature review, there arise the following questions while investigating the impact of modulated protein-DNA interaction on gene regulation for a wide range of gene regulatory networks that exist in organisms ranging from bacteria to humans:

- How the expression of target genes can be quantitatively controlled by modulating TF-ligand interactions and DNA looping?

- How vital is the network topology in gene regulatory circuits for phenotype diversity?

- How do the long-distance protein-protein interactions influence the promoter-enhancer interaction?

- What are the impacts of fluctuating rate parameters in gene expression?

- Can we predict the fate of a malignant cell using thermodynamic analysis? In this regard, we chose the tumor suppressor, the p53 signaling network.

Thus, this thesis aims to provide detailed answers to the questions that are mentioned above by using theoretical and computational methods.

1.4 Objectives and Scope

This thesis aims to understand the role of modulated interactions between protein and DNA in regulating the gene expressions associated with a wide range of regulatory networks ranging from *lac operon* in bacteria to tumor suppressor p53 in humans. For this purpose, we have used statistical thermodynamics and stochastic modeling schemes.

The control over gene regulation is the central theme of any form of life, i.e., from bacteria to humans (Dalma-Weiszhausz and Brenowitz, 1992; Tolhuis et al., 2002). Despite the involvement of noisy and complex biomolecular processes, they execute an ordered genetic response in a biological cell. The origin of this order is related to the detailed protein-DNA interaction involved in various gene regulatory (GR) networks. However, the quantitative prediction of a response curve and its correlation with the mechanism needs scientific attention. The first research problem in this thesis is thus to explore the role of various biophysical processes, such as the binding of a TF to the promoter and their oligomerization on DNA, TF-ligand interactions, and DNA looping etc., in controlling gene expression.

Cooperative protein-protein and protein-DNA interactions form programmable complex assemblies (Bocci et al., 2023; Vilar and Saiz, 2014) by performing non-linear gene regulatory operations in signal transductions. The apparent structure of those complex assemblies is very similar, but their functional response strongly depends on the topology of the protein-DNA interaction networks (Huang et al., 2018; Buchler et al., 2003). Formulation of response function from a gene regulatory networks is crucial since it explores the mechanism behind the expression (Ackers et al., 1982; Bintu et al., 2005b; Teif, 2007). Further, the enhancer elements regulate gene expression patterns

mediated via promoter-enhancer interactions through DNA looping at a large distance (Levine et al., 2014; Vilar and Saiz, 2014; Yasmin et al., 2004). Considering this, we use theoretical analysis to examine how the internal connectivity of a gene regulatory network appears in cellular environments. We further explore their role in calculating the functional response. We validated our model by applying it to biological systems like *Saccharomyces* in yeast cells and NF-κB enhancer systems to investigate the impact of topology spatial regulation in modulating the gene expression pattern with systematic variations of the binding and DNA looping parameters in a thermodynamic model (Kumawat and Chakrabarty, 2019; Levine et al., 2014). This fact motivates us to pursue the second research problem of this thesis, where we have explored how a complex network of interactions can form a decision-making loop in a GR system.

It has been known that gene expression is inherently stochastic (Vellela and Qian, 2009; Golding et al., 2005). Moreover, many studies show that the higher-order networks exhibit multimodal distribution in the probability function for the transcriptional outcomes (Zhu et al., 2022; Faucon et al., 2014). A deterministic model does not offer any opportunity to understand the origin of such multimodal distribution or the existence of multistable features. However, the stochastic model considers various birth and death processes linked with protein formation or degradation. In many instances, one assumes that the rate parameters of those processes are constant, but this is not true in general, as the values of rate parameters may vary in the cellular environment. This argument motivates us to step into the third research problem, where we consider the role of colored noise in controlling the transcriptional outcomes, where the correlation time and noise strengths play a critical role in regulating gene expression.

The mutations and phosphorylation of p53 suppressor are integral to human cancer progression (Li et al., 2006). It controls many cellular processes, including cell cycle arrest, apoptosis, senescence, inhibition of angiogenesis, DNA repair, metastasis, etc (Stenger et al., 1994; van Dieck et al., 2009). This fact motivates us to study the role of tumor suppressor protein p53 in determining the cell fate of a malignant cell. The phosphorylated p53 binds with promoters of four genes, p21, Bax, p48, and PAI, for the gene expression (Bieging et al., 2014). However, the binding affinity of the phosphorylated p53 to them is different. Such heterogeneous binding produces differential gene

expression, which can be related to a malignant cell's fate (Reczek et al., 2003; Chen, 2016; Gomes and Espinosa, 2010). Various drug molecules, such as Nutlin and SIRT inhibitors, are known to control the stability of phosphorylated p53 repressor, which introduces an extra layer of control in gene expression (Vassilev et al., 2004; Peck et al., 2010). Therefore, we employ a thermodynamic model to calculate the expression of these genes upon binding phosphorylated p53 suppressors to their promoter regions. We also apply our modeling schemes to study the relation between the topology of gene regulatory networks and cellular fates for various cancer cell lines (Mirzayans et al., 2015; Mayo et al., 2005; Fan et al., 2020). We explore this because the topology of the gene regulatory networks for different cell lines is different. We apply our proposed model to explore the fate of Breast cancer (MCF-7), Colon cancer (HCT-116), and Blood cancer cell lines (K562).

1.5 Outline of the Thesis

After a brief introduction to the modulation of protein-DNA interactions for various gene regulatory networks, we move to particular research problems.

1.5.1 Formulation of response function for gene regulatory networks

In **Chapter 2**, we employ *in silico* binding studies of GR systems to show that the TF binds to multiple DNA sites with high cooperativity. In this regard, we develop a statistical thermodynamic formalism that considers the binding and unbinding events of TFs to DNA and calculates the probability of each promoter configuration. We apply this model to simple model systems and systems containing multiple interacting genes. Using these models, we explore the effect of TF oligomerization, TF-ligand interactions, and DNA looping on gene expression.

The thermodynamic models fall short when complex gene regulatory networks contain multiple interacting genes. In this regard, we employ grand canonical Monte Carlo simulations to study such a system and calculate the probability of promoter configurations. The predictions are validated against detailed grand canonical Monte Carlo simulations and published data for the *lac* operon system. Overall, our study reveals that

the expression of target genes can be quantitatively controlled by modulating TF-ligand interactions and the looping energy of DNA. In contrast with the earlier quantitative gene expression analysis studies, our modeling scheme based on the grand canonical ensembles renders flexibility and essential insights by linking genetic response with the detailed mechanism of protein-DNA interactions in various complex gene regulatory networks.

1.5.2 Functional responses of bio-molecular assembly networks

In **Chapter 3**, we demonstrate how the coordinated self-assembly of protein molecules on DNA produces gene regulatory network motifs. Our findings corroborate the existence of a precise functional response at the molecular level using thermodynamic analyses. We show that a complex network of interactions can form a decision-making loop, such as feedback and feed-forward circuits, only by a few molecular mechanisms. These networks or the self-assembled proteins on the DNA produce precise functional responses at thermodynamic equilibrium. We characterize each possible network of interactions by systematic variations of free energy parameters associated with protein binding to DNA and DNA looping. We further show that our proposed model can capture the boolean logic operations such as AND, NAND, NOR, and OR gates. In this regard, we consider the regulatory network of the NF-κB system, which takes into account multiple levels of protein organization in the DNA. Through the quantitative thermodynamic model of transcriptional regulation and systematic variation of promoter-enhancer interaction modes, we can account for the origin of various logic gates formed in this system. We further show that the interconversion or switching among various logic gates yields systematic binding and DNA looping parameter variations. Overall, our proposed model demonstrates that the coordinated self-assembly of protein and DNA interactions creates network motifs and logic operations at thermal equilibrium. These assemblies perform well-defined computations and amplify gene expression inside a biological cell.

1.5.3 Stochastic dynamics of gene regulatory networks driven by intrinsic molecular noise

In **Chapter 4**, we study the dynamics of gene regulatory networks. Describing the feedback loops for gene regulation in an equilibrium thermodynamic framework is challenging since the binding and unbinding of protein to DNA strictly follows the principle of detailed balance in this framework. Moreover, the detailed balance allows one to carry out the dynamics of the association and dissociation of protein on DNA using the well-known principle of mass action kinetics. However, the dynamics of the feedback loops for gene regulatory networks are interesting. The positive feedback loops of a network promote the system to stay away from equilibrium, whereas the negative feedback loops introduce multistability. Therefore, the stochastic models are best to describe the dynamics of these complex regulatory networks. These stochastic models are developed either by adding extra delta function correlated noise into the kinetic equation (extrinsic noise) or by considering the random molecular events based on the few elementary reactions (intrinsic noise) given for a regulatory system. The first approach gives us a set of coupled stochastic differential equations. However, the latter describes the time evolution of the grand probability function as described by master or Fokker-Planck equations. Using both approaches, We explore the Gal promoter dynamics in yeast cells and the dynamics in NF-κB signaling systems. We calculate the stochastic potentials from those stochastic trajectories. Our calculated stochastic potentials capture the signature of multistability in the feedback loops. However, the cellular noise sources have a finite correlation time measured experimentally (Wang et al., 2019a; Gupta and Khammash, 2022). In this chapter, we also discussed the effect of short-range correlation, introduced by the Ornstein-Uhlenbeck processes, to model noisy gene expression.

1.5.4 Role of network topology in controling the cellular fate under stressed condition: A tumor enigma

In **Chapter 5**, we develop thermodynamic models to determine the fate of a malignant cell as governed by the tumor suppressor p53 signaling network. The tumor suppressor p53 responds to stress by selectively triggering one of among many potential transcriptomes that influences many layers of input signal modification, ranging from

phosphorylation of p53, the wide range of binding affinity of p53 with the promoters of various genes, to internal connectivity among cell fate genes. We show that minimum free energy is a fundamental property of biological networks that establishes a connection to unriddle the enigma between the network topology and the state of the cell. Using this model, we show how the binding of the p53* to the promoter regions of the four cell fate-determining genes, Bax, p21, p48, and PAI, show phase transition characteristics. We apply our developed model to various cancer cell lines from breast cancer (MCF-7), colon cancer (HCT116), and leukemia (K562) that exhibit different network topologies. Our modeling scheme shows that the cell's fate is mainly related to the internal links among different cell fate genes and the free energy of binding among various biomolecules.

Finally, in **Chapter 6**, we summarize the key findings and discuss future research directions.

Chapter 2

Formulation of response function for gene regulatory networks

2.1 Introduction

The control over gene expression is the central theme of any form of life, as it allows an organism to respond by changing external and internal signals. With perfect regulatory control, the DNA of those genes are transcribed to mRNA molecules, which are further translated to functional proteins (Alberts, 2017; Phillips et al., 2012). A particular class of protein called transcription factors (TFs) activates target genes (TGs) by binding with the upstream or promoter region of the gene that facilitates the recruitment of RNA polymerase (RNAP) for the expression. A well-known fact in gene regulation is the modulation of interactions between TFs and the promoter region of TGs (Gerland et al., 2002; Alberts, 2017). The strong interactions between proteins and DNA and preceding events control gene regulation (Dangkulwanich et al., 2014). Several biophysical events such as the strength of input signals, ligand binding (Chen et al., 1998), acetylation (Li et al., 2006), phosphorylation (Wenta et al., 2008; Kang et al., 2009); self-assembly of TFs (Hanson et al., 2005; van Dieck et al., 2009); and physical state of DNA etc. (Vilar and Saiz, 2005; Saiz and Vilar, 2007b; Boedicker et al., 2013) control the final interaction between protein and DNA. In particular, forming a DNA loop due to the long-distance protein-protein interactions plays a vital role in the unusual burst of gene expression (Choi et al., 2008). This behavior has been observed in a class of regulatory systems such as *gal*, *lac*, and *deo* operons in E. Coli (Adhya, 1989; Schleif, 1992), the lysogenic to lytic switch in phage λ (Ptashne, 2004), the human β-globin locus (Tolhuis et al., 2002), the nuclear hormone receptor RXR (Yasmin et al., 2004) and tumor suppressor

protein p53 (Stenger et al., 1994; Cournac and Plumbridge, 2013). However, how an input signal is modulated by these biophysical events and its impact on transcriptional output remains unclear.

A gene regulatory network mediates the relationship between an input signal and the transcriptional output. Such networks inside the cell require a system-level understanding rather than a list of parts. Moreover, a cell is an open system, which introduces further complexity to study them by suitable experimental methods. The network of biomolecular interactions is commonly known as cellular networks, composed of molecules used by the cells to sense and respond to the environment (Saiz and Vilar, 2008, 2006b; Saiz, 2012; Marbach et al., 2010). Such networks regulate and control the fundamental cellular processes, including gene expression and signal transduction, in all types of organisms, from bacteria to humans. The disruption of a cellular network is responsible for many human diseases, such as cancer, diabetes, and autoimmune disorders (Chen et al., 2018). With this view, an idea has emerged in gene regulation that can be used as a guiding principle in the ongoing development of molecular therapies against diverse diseases.

Many studies have been performed to understand the intricate behavior of gene regulatory networks (Djordjevic et al., 2003; Ong et al., 2010; Wang et al., 2009; Marbach et al., 2010; Haldane et al., 2014; Landman et al., 2017; Wong and Gunawardena, 2020). Single-molecule level experiments such as atomic force microscopy (AFM) (Harada et al., 1999; Friedman et al., 2013; Wang et al., 2013; Lee, 2019), fluorescence-based methods (Guthold et al., 1999; Bustamante et al., 1999; Suzuki et al., 2012), magnetic tweezers assays (Revyakin et al., 2006; Kapanidis et al., 2006), cryo-electron microscopy (Liu et al., 2017) etc. have emerged as a guiding tool to explore them in detail. Researchers have shown that RNAP first binds with TFs that aid the enzyme to recognize and bind to the promoter region of the DNA to initiate transcription using these experiments (Dangkulwanich et al., 2014). These studies also show that transcription initiation promotes DNA unwinding and influences the early phase of transcription elongation. Further, these studies explore how short- and long-range interactions among TFs affect the DNA binding properties of TFs. Moreover, the mechanism behind the local and distant interactions between DNA and TF has also been revealed from those studies

(Dangkulwanich et al., 2014; Alberts, 2017; Phillips et al., 2012).

Statistical thermodynamics-based models serve a quantitative framework to describe transcriptional regulation for a variety of regulatory motifs (Phillips et al., 2012; Ackers et al., 1982; Bintu et al., 2005b; Segal and Widom, 2009; He et al., 2010; Vilar and Saiz, 2011; Marzen et al., 2013; Weinert et al., 2014; Samee et al., 2015; González et al., 2019). In particular, calculations under canonical and grand canonical ensembles are helpful to predict the equilibrium or static behavior of response function for gene regulation (Ackers et al., 1982; Vilar and Saiz, 2011; Weinert et al., 2014). However, most of these models are developed under the non-interacting regime and are often considered an isolated genetic unit (Saiz and Vilar, 2007a; Vilar and Saiz, 2011; Bintu et al., 2005b; Landman et al., 2017). TFs act on multiple genes, often accumulating locally on DNA sites by various mechanisms.

In the static model, the probability of finding the promoter configuration in a given regulatory state is a function of the free energy of interactions among the components present in the system under quasi-equilibrium approximation. The gene's activity is assumed to be proportional to the probability of an RNAP being bound to the promoter sequence (Phillips et al., 2012). However, there are many kinetics-based models for gene regulation, and they do not require as many assumptions but increase the number of parameters (Ko, 1991; Berg, 2008; Sánchez and Kondev, 2008; Vilar and Saiz, 2014).

Quantitative modeling of gene regulations requires understanding the mechanisms that determine the shape of response function. In this regard, we develop a novel computational method by employing statistical mechanics to analyze a cellular network. In particular, we employ the grand canonical ensemble approach to obtain an explicit expression for a fraction of RNAP bound to DNA promoter sites. We consider short-range interactions as the local interactions between DNA sites and TF, the nearest neighbor interactions among TFs on DNA, and the formation of a DNA loop due to the interactions between two distant TFs bound to DNA sites (Dangkulwanich et al., 2014; Alberts, 2017; Phillips et al., 2012; Rydenfelt et al., 2014; Liu et al., 2016). Then, we obtain the expressions for Fold-Change (FC) and Dose-Response (DR) in the presence of TFs and small molecules or ligand. We also perform a Grand Canonical Monte Carlo (GCMC) simulation for the GR systems. The GCMC simulation considers the binding

and unbinding events under quasi-equilibrium conditions using the Metropolis-Hastings algorithm. Finally, we validate our models and computational methods by revisiting the published results for the cellular networks formed in the *lac* operon system of E. Coli.

2.2 Model system

The phenomenon of gene regulation is complex and not limited to only protein-DNA interactions. Figure 2.1 presents a general schematic picture of protein-DNA interaction found in nucleosomes/nucleoids when viewed as a lattice of possible binding sites. The simplified lattice model is our starting point for the quantitative gene expression calculations. The long-distance and short-range interactions between protein and DNA are shown explicitly in the figure. With these varieties of biomolecular interactions, the system forms a cellular network, and we assume that a) the average behavior of the network is invariant over time and b) the binding kinetics is much faster than other cellular processes, such as cell growth. We first develop a theoretical model for systems without inter-particle interactions and then introduce it by nearest-neighbor interactions. We employ grand canonical ensemble formalism to obtain an explicit expression for the average protein and its complexes bound to the promoter region of a gene. This modeling scheme can include more complex regulatory features, opening the door to consider gene regulation in natural cellular systems. Further, the method has the advantage over its canonical ensemble counterpart since the former considers an open system and provides simple analytical solutions for different competition scenarios. Effects of DNA looping and protein-ligand interactions on gene regulation are analyzed using our in-house developed GCMC code.

2.2.1 Non-Interacting Sites

We map the DNA segment to a one-dimensional lattice of M equivalent, distinguishable, and independent binding sites. We consider there are N protein molecules in the system that bind with the DNA sites. Each DNA site can accommodate up to m protein molecules. These proteins may oligomerize locally at a DNA site. We define a new variable a_o, the total number of sites with the protein occupancy, o. If the configuration
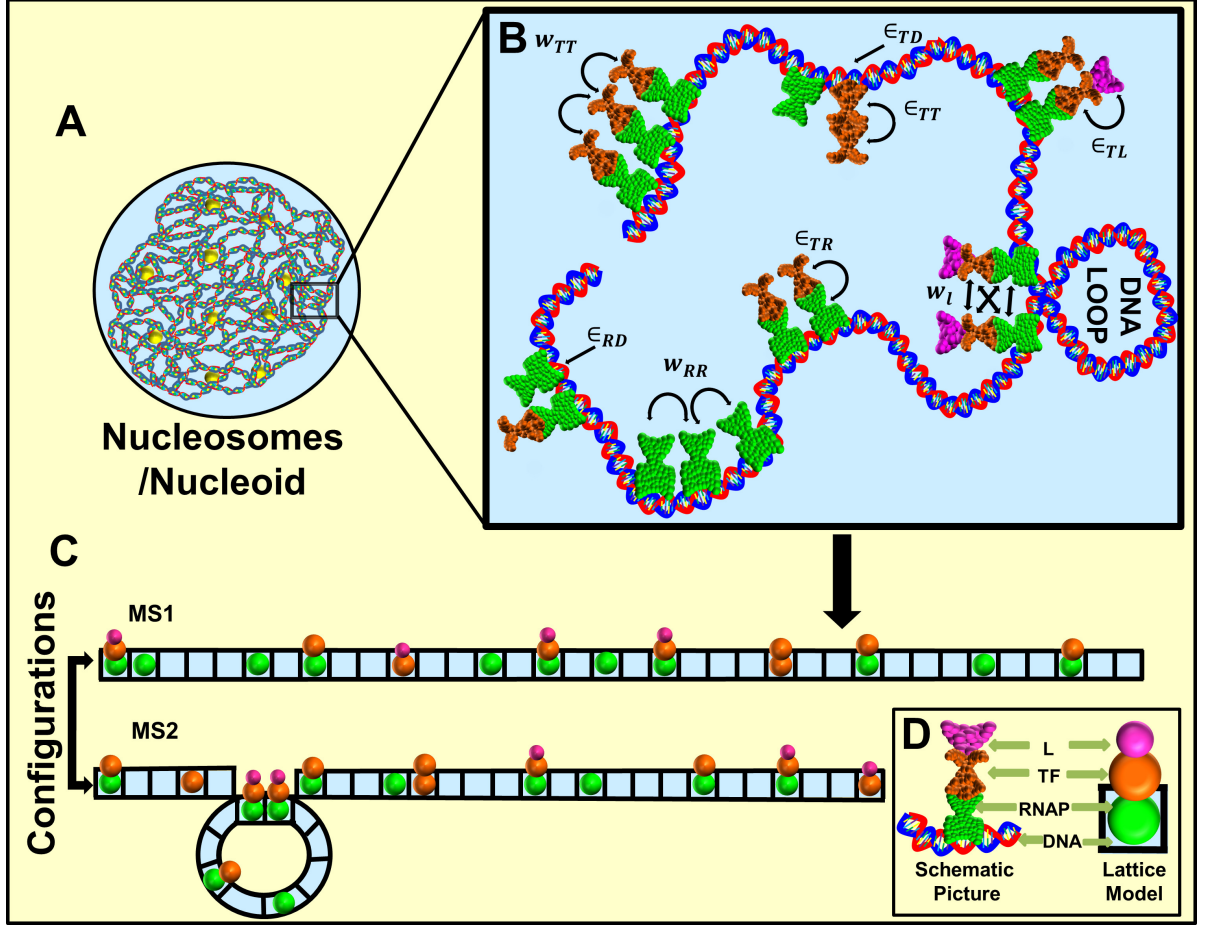
Figure 2.1: A) Schematic view of the complex structure of nucleosomes/nucleoid B) The zoomed view of a section of nucleosomes/nucleoid. Different types of protein-DNA interactions control the populations of a specific configuration. We denote RNAP-DNA, TF-DNA, TF-RNAP, and TF-ligand by ϵ_{RD}, ϵ_{TD}, ϵ_{RT}, ϵ_{TL}. The nearest neighbor interactions for TF and RNAP on the DNA lattice are w_{TT} and w_{RR}. w_l represents the long-distance interactions among RNAP and TF molecules. The color-coding schemes for RNAP, TF, and ligand are green, orange, and purple. The formation of the DNA loop is also shown here. An elastic constant k_l defines the DNA elasticity for looping. C) Protein-DNA interactions are modeled as a lattice of possible binding sites that TFs can occupy. Two possible configurations out of an enormous number of microstates are shown for representation. D) A correspondence between the bead and 3-D structures of protein is shown.

of the system is given by the set of numbers $\{a_o\}$, then the total energy of interactions is $E = \sum_{o=0}^{m} a_o \epsilon_j(o)$; where $\epsilon_j(o)$ is the jth level site with occupancy o. The grand canonical partition function for this non-interacting system is $\Xi(\lambda, M, T) = \xi(\lambda, T)^M$. The symbol,

ξ, which is the promoter partition function, takes the following form $\sum\limits_{o=0}^{m} q(o)\lambda^{o}$; where $q(o) = \sum_{j} e^{-\frac{\epsilon_j(o)}{k_B T}}$ is the partition function for a site and $\lambda = e^{\frac{\mu}{k_B T}}$ is the absolute activity. The symbols, $k_B T$, μ and T are the thermal energy and chemical potential of the protein and the absolute temperature of the system. The average occupation number of bound protein per DNA site, $\bar{o} = \frac{\bar{N}}{M}$ is given by,
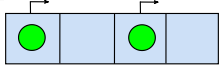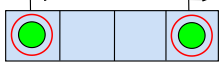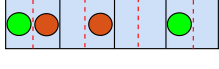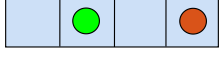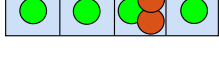
$$\bar{o} = \frac{\lambda}{M}\left(\frac{\partial \ln \Xi}{\partial \lambda}\right)_{M,T} = \frac{\sum\limits_{o=0}^{m} o\, q(o)\lambda^{o}}{\sum\limits_{o=0}^{m} q(o)\lambda^{o}} \tag{2.1}$$

Calculation of the quantity $\bar{o}$ is central since it quantitatively impacts the transcription process.

Different types of sites in DNA are known in gene regulatory systems as various TFs bind to the promoter regions of DNA either specifically or nonspecifically. For example, the TFs and RNAP bind specifically to the different binding sites of the gene's promoter region for the initiation of transcription. The binding of a TF to the different operator regions (O_R) of DNA is standard in bacterial gene regulation (Shea and Ackers, 1985; Einav et al., 2018). In a classic work, Shea and Ackers have studied the efficient O_R control system of bacteriophage lambda, where they consider the interactions of cI repressor and cro proteins at three different sites of the right O_R and the binding of RNAP at different promoters (Shea and Ackers, 1985). With this view, we extend the above method for the distinguishable and independent binding sites. In Table 1, we present the obtained expressions for different GR systems. In particular, we derived the grand canonical partition function and the fraction of bound protein molecules per site for those cases. We consider cases such as a) simple regulation, when RNAP binds with DNA, and b) controlled regulation, when TF and ligand control binding of RNAP (Hawley and McClure, 1980). The detailed derivation of each of these equations is presented in Appendix 1.

Table 2.1: We consider seven different GR Systems (GRS). The system has a total of M equivalent, distinguishable, independent promoters. M_S and M_O are the total number of specific and non-specific or other binding sites. The grand canonical partition function (Ξ); the promoter partition function (ξ); and the fraction of bound RNAP, RNAP-TF (C), and RNAP-TF-L (L) complexes ($\bar{o}$, $\bar{o}_C$, $\bar{o}_L$) are calculated, when 1) when RNAP binds with DNA, 2) when RNAP binds specifically (S) and non specifically (O) to DNA, 3) when TF-RNAP complex binds with DNA as a pair, 4) recruitment of RNAP due to the binding of TF, 5) stimulation of RNAP by TF, 6) cooperative stimulation by the dimeric TFs and 7) activation due to the binding of a ligand to TFs. The symbols λ_R, λ_T, λ_L are the absolute activities for RNAP, TF, and ligand. q_R, q_T, q_L are the site partition functions for RNAP, TFs, ligand, and q_S and q_O are the site partition functions for specific and non-specific sites.

GRS	Ξ	ξ	$\bar{o}/\ \bar{o}_C/\ \bar{o}_L$
(diagram)	ξ^M	$1 + q_R\lambda_R$	$\dfrac{q_R\lambda_R}{\xi}$
(diagram)	$\xi_S^{M_S}\xi_O^{M_O}$	$\xi_S = 1 + q_S\lambda_R$ $\xi_O = 1 + q_O\lambda_R$	$\dfrac{1}{M}\left(\dfrac{M_S q_S \lambda_R}{\xi_S} + \dfrac{M_O q_O \lambda_R}{\xi_O}\right)$
(diagram)	ξ^M	$1 + q_R\lambda_R + q_T\lambda_T + q_R q_T e^{-\frac{\epsilon_{TR}}{k_B T}}\lambda_R\lambda_T$	$\dfrac{q_R\lambda_R\left(1+q_T\lambda_T e^{-\frac{\epsilon_{TR}}{k_B T}}\right)}{\xi}$
(diagram)	ξ^M	$1 + q_R\lambda_R(1 + q_T\lambda_T)$	$\dfrac{q_R\lambda_R(1+q_T\lambda_T)}{\xi}$
(diagram)	ξ^M	$1 + q_R\lambda_R\left(1+q_T\lambda_T e^{-\frac{\epsilon_{TR}}{k_B T}}\right)$	$\dfrac{q_R\lambda_R q_T\lambda_T e^{-\frac{\epsilon_{TR}}{k_B T}}}{\xi}$
(diagram)	ξ^M	$1 + 2q_T\lambda_T + q_T^2 e^{-\frac{\epsilon_{TT}}{k_B T}}\lambda_T^2$	$\dfrac{q_T\lambda_T + q_T^2 e^{-\frac{\epsilon_{TT}}{k_B T}}\lambda_T^2}{\xi}$
(diagram)	ξ^M	$1 + q_R\lambda_R\big(1 + q_T\lambda_T e^{-\frac{\epsilon_{TR}}{k_B T}}(1 + q_L\lambda_L e^{-\frac{\epsilon_{TL}}{k_B T}})\big)$	$\dfrac{q_R q_T q_L e^{-\frac{(\epsilon_{TR}+\epsilon_{TL})}{k_B T}}\lambda_R\lambda_T\lambda_L}{\xi}$

2.2.2 Interacting Sites

2.2.2.1 Ising-based Model:

We extend our calculations for interacting sites. Here, we consider a DNA consisting of M equivalent, distinguishable sites, but they interact if protein molecules occupy both sites. We introduce nearest neighbor interaction, w for such interaction. In this model, the interaction energy between protein and DNA at i^{th} site is ϵ. Under this framework, the model becomes the Ising model, known for studying various systems such as gas adsorption, gas-liquid phase transition, binary alloy, etc. (Hill, 1986). This model can be solved by various analytical techniques such as combinatorial, transfer matrix, etc., methods (Huang, 2009). We adopt the combinatorial approach to calculate the fraction of bound RNAP for these systems, which is presented in Appendix 2. We have chosen the combinatorial path over other methods since the former is consistent with our previous analysis.

For the sake of simplicity, we consider each of the DNA sites can be either occupied or unoccupied, then the total energy of interaction for the interacting system is:

$$E = \epsilon \sum_{i=1}^{M} s_i + w \sum_{i=1}^{M} s_i s_{i+1}, \tag{2.2}$$

where s_i is the binary variable, which takes a value of either 0 or 1 depending on the occupancy of ith binding site. With this description of total energy, we obtain the fraction of bound protein, $\bar{o}$ as:

$$\bar{o} = \frac{1}{2} - \frac{e^{w'} \sinh(\mu'/k_B T)}{2[e^{-2w'} + e^{2w'} \sinh^2(\mu'/k_B T)]^{\frac{1}{2}}}, \tag{2.3}$$

where, the effective chemical potential, $\mu' = (\epsilon + w - \mu)/2$ and $w' = -w/4k_B T$. The above analytical approach is straightforward for simple regulation, such as binding between the DNA site and RNAP. However, the method is challenging for controlled regulation, where RNAP, TF, and ligand do not bind independently with DNA and TF. Further, introducing oligomerization of TFs and forming a loop in DNA enhances the problem's complexity. Therefore, the computational approach, such as the GCMC technique, can

be an alternative option for calculating $\bar{o}$ for such systems.

2.2.2.2 Monte Carlo Simulation (MCS):

MCS offers an intriguing opportunity to calculate the same quantity for the more complex interacting system. We assume the protein's bound phase is in thermodynamic equilibrium with the free unbound phase at fixed chemical potential, μ, the total number of DNA sites M, and temperature, T. Therefore, we perform the grand canonical Monte Carlo Simulations for protein-DNA interactions. In short, the MCS realizes the binding and unbinding of proteins to DNA as the Markov process and generates many configurations under a few thermodynamic constraints. The MCS is advantageous over the analytical approaches because we can implement it effortlessly for the complex interacting protein-DNA system. For example, (a) the inclusion of long-distance interactions between two distant DNA sites, or (b) the binding and formation of different oligomeric species of TFs on the DNA site, or (c) the long-range correlation effects through nearest-neighbor interactions are quite challenging using analytical techniques. However, the MCS offers to include such types of complex interactions present in protein-DNA systems.

To perform MCS, we model DNA as a coarse-grained one-dimensional (1-D) lattice of size 180 sites. Each of the sites of the lattice corresponds with 40 DNA base pairs. We consider TFs, RNAP, and ligands to perform our simulation for the gene expression. The nearest neighbor interactions, oligomerization of TFs, and the DNA loop formation are incorporated into the simulation. The periodic boundary condition (PBC) is applied to avoid the end effects in our simulation. The simulations are performed using the Metropolis update rule. Our MC simulation code is written in MATLAB, and the outline of the program is written in Appendix 3.

2.2.3 Fold change and dose response

The TFs can be an activator or a repressor protein. Since the activator enhances mRNA production, the number of mRNAs, F can be directly proportional to the probability of bound activator-RNAP complex to DNA, i.e., $\bar{o}$. Thus, we can write the number of mRNA production in the presence of an activator, $F^A = \Gamma^A \bar{o}$, where Γ^A is the

proportionality constant. In the case of GR network systems, multiple active species are present in the system. Moreover, the population and activity of each of the species are heterogeneous. Therefore, the net production of mRNA for an activator may be written as $F^A = \Gamma_0^A \bar{o}_0 + \sum_{i=1} \Gamma_i^A \bar{o}_i$. The first term is the basal expression, which appears due to the binding of the RNAP to the DNA, and other terms in the summation are for activated transcription. Similarly, the binding of a repressor to DNA suppresses the production of mRNA. Therefore, the number of mRNAs, F for the repressor-RNAP complex bound to DNA is $F^R = \Gamma^R(1 - \bar{o})$, where again the parameter, Γ^R is proportionality constant. Thus, the net production of mRNA for a repressor is $F^R = \Gamma_0^R \bar{o}_0 + \sum_{i=1} \Gamma_i^R(1 - \bar{o}_i)$. We define the fold change for an activator or a repressor as the relative change in the population of mRNA in the presence of TFs to the population in the absence of TFs, i.e., $FC^{A/R} = F^{A/R}/\Gamma_0^{A/R}\bar{o}_0$. The normalized fold change (NFC) for both activator and repressor is defined as $NFC = (FC - FC_{min})/(FC_{max} - FC_{min})$, where FC_{max} and FC_{min} are maximum and minimum values of the FC function.

We can modify the molecular interaction in GR systems by adding other small molecules, such as ligands, to control such gene regulation. The ligand can specifically bind with the specific site of TFs and further modulate the interactions. Therefore, we define Dose-Response (DR) as $DR^{A/R} = F^{A/R}/\Gamma_0^{A/R}\bar{o}_0$ for an activator (or a repressor) is the relative change in the population of active complexes that produce mRNA due to the binding of a ligand to an activator (or a repressor) and the population of active complexes in absence of ligand. The ratio is given by $DR^R = F^{A/R}/\Gamma_0^{A/R}\bar{o}_0$. The gene regulation can be enhanced or suppressed depending on the amount of ligand and strength of interactions between TF and ligand. Thus, we can calculate the DR of a particular ligand based on our developed model. In the subsequent sections, we present the results for various GR systems.

2.3 Results

2.3.1 Average Occupation Number

Using the above theoretical and computational methodologies, we first compute the average occupation number, $\bar{o}$ for different GR systems as a function of the activity of

RNAP, λ_R for both interacting and non-interacting systems. The results obtained from our derived analytical function and MCS are presented in Figure 2.2. The range of λ that we have chosen for our calculation is based on the average concentration of proteins found in a cylindrical cell of major and minor axes ≈3 and 1 μM (Cossart and Gicquel-Sanzey, 1985). We show the results for the interacting case in Panel D. It is apparent from that figure that our theoretical and computational results are matched well for both the non-interacting and interacting cases. The values of $\bar{o}$ quickly reached saturation as the interaction between RNAP and DNA strengthened. In this situation, DNA binding sites continue to recruit RNAP molecules until they are saturated. However, in the case of binding of RNAP with specific and non-specific DNA sites in terms of their binding affinity to RNAP, we find two saturation points for $\bar{o}$. The first corresponds with the specific sites, and the other corresponds with the non-specific sites. Two different binding interactions between the RNAP molecule and DNA binding site can explain the origin behind the two saturation points.

In Figure 2.2C, we show the results obtained for the simultaneous binding of TF with the RNAP at the promoter region of the gene. Here, we choose a fixed value activity (Einav et al., 2018), $\lambda_T = 1.5 \times 10^{-6}$ for TF to calculate the $\bar{o}$ (Buchler et al., 2003). The $\bar{o}$ quickly reached saturation as the interaction between RNAP and TF strengthened. In other words, the TF facilitates the recruitment of more RNAP molecules as the interaction between RNAP and TF enhances. Thus, the binding of RNAPs to the DNA can be controlled by modulating RNAP and TF interaction. Further, it is clear from Panel D that there is a sharp rise in the function, $\bar{o}$, as the nearest neighbor interactions among RNAP become strong. In other words, DNA sites enhance the recruitment of RNAP molecules as the attractive interactions among RNAP are strengthened. It is a clear signature of cooperative effect among RNAP, and often, it describes the presence of long-range correlation among RNAP on the DNA (Ackers et al., 1982). As the nearest-neighbor interaction is enhanced, the function $\bar{o}$ becomes steeper, a clear sign of the TG activation due to the cooperative effect.

Since the binding of TF with DNA controls the gene expression, we calculate the $\bar{o}$ as a function of λ_T at fixed value of RNAP activity, $\lambda_R = 10^{-4}$ (Landman et al., 2017). We consider that the TFs recruit RNAP to the gene promoter and stimulate the
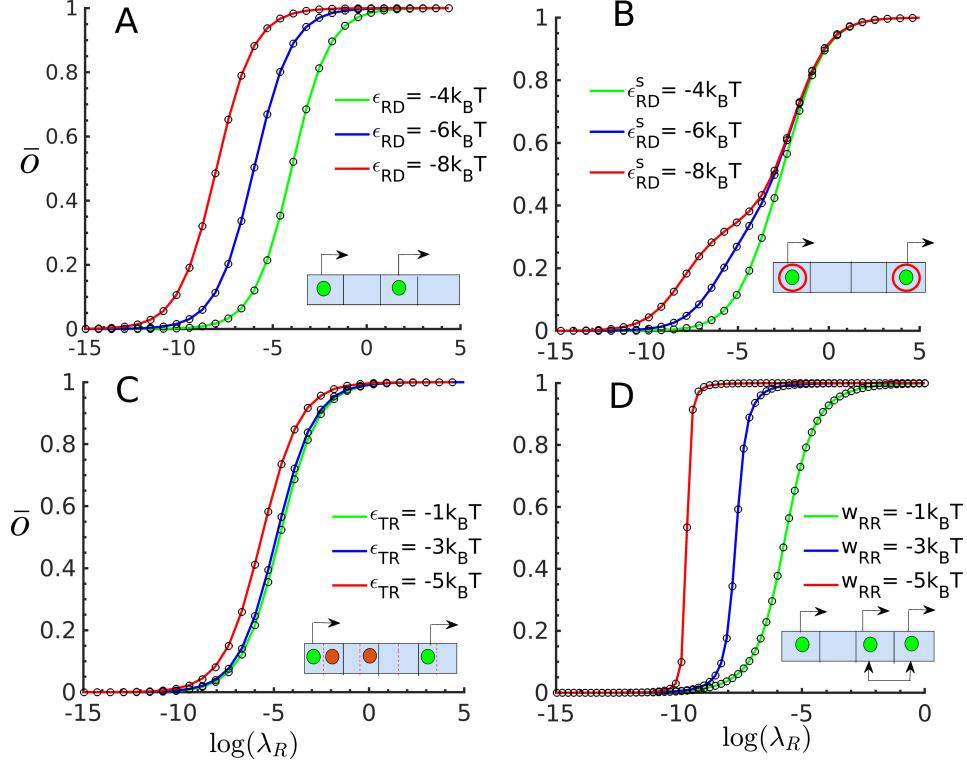
Figure 2.2: Average occupation number of RNAP and RNAP-TF complex, $\bar{o}$ for non-interacting and interacting systems as a function of RNAP activity. Solid lines and the symbol circles represent theoretical and simulation results. The calculations are done for three different values of interaction energies as shown by three line colors: red, blue, and green. Schematic lattice configurations of short segments of protein-DNA complexes are shown in the respective panels. Panel A) simple binding between RNAP and DNA. Panel B), when RNAP binds specifically and non-specifically to DNA. The value of binding interaction for RNAP with nonspecific DNA binding sites, $\epsilon_{RD}^{O} = -2k_BT$ is considered. Three values of ϵ_{RD} are considered for both analyses. Panel C), simultaneous binding of RNAP and TFs on DNA. The binding interactions of RNAPs and TFs with DNA, $\epsilon_{RD} = -4.68k_BT, \epsilon_{TD} = -8.88k_BT$ and $\lambda_T = 1.5 \times 10^{-6}$ are considered. Three values of ϵ_{TR} are considered. D) The nearest neighbor interactions (w_{RR}) are introduced when RNAP occupies both sites. Three values of w_{RR} are considered for the calculation.

transition from a closed to an open complex (Ye et al., 2022). We also consider that the TF and RNAP molecules are in close contact for such a stimulated transition. We present the results in Figure 2.3. Once again, we find an active correspondence between theoretical and simulation results. In specific, we show the results for the recruitment

of RNAP due to the binding of TF (panel A), stimulation of RNAP by TF (panel B), cooperative stimulation by the dimeric TFs (Panel C), and activation due to the binding of a ligand to TFs (Panel D). We find that as the interaction between TF and DNA is enhanced, TF recruits more RNAP and saturates promoter sites quickly as we increase the activity of TF, λ_T. We also find that stimuli in the system enhance the interaction between TF and DNA, further increasing the RNAP-TF complex population at equilibrium. In general, such stimulation can be activation or repression, depending on the nature of TFs. We also find that as the dimeric interaction between two TFs is enhanced, there is a sharp rise in the function, $\bar{o}$. Such a change in slope in that function is a clear signature of cooperative effect. As shown in panel D, the ligand activates the TFs, facilitating the recruitment of more RNAP on the promoter region of a gene. As a result, the promoters become saturated earlier with the increase in ligand activity, λ_L.

The above discussion demonstrates that our theoretical and computational results correspond closely. Therefore, any of these methods can calculate $\bar{o}$. Since the theoretical approaches have mathematical difficulty for complex interacting GR systems, we adopt MCS for further calculations.

2.3.2 Self Assembly of TF and DNA Looping

We discuss the results obtained from MCS for complex GR systems in this section. Here, we mainly introduce a few complexities into the GR systems: a) formation of oligomeric species, b) DNA looping, c) ligand-TF interactions, etc. Such complexities are common in the prokaryotic and eukaryotic cells (Matthews, 1992; Cournac and Plumbridge, 2013; Burz and Ackers, 1996). In Figure 2.4, we show the possible complexes that formed on the DNA in our simulation. As mentioned in the figure, both free and ligand-bound TFs participate in the oligomerization of DNA. Moreover, forming a loop in DNA promotes the population of higher-order oligomers in the system. For example, in the bacteriophage lambda model, the cI genetic switch includes extra cooperativity through octamerization of the cI repressor protein, mediated by long-range DNA looping (Dodd et al., 2005). We also consider a few other bio-inspired facts for the gene regulation, such as a) specific binding of ligand to the TFs, b) simultaneous binding of TFs with both RNAP and the DNA binding sites, and c) only RNAP unable to participate for
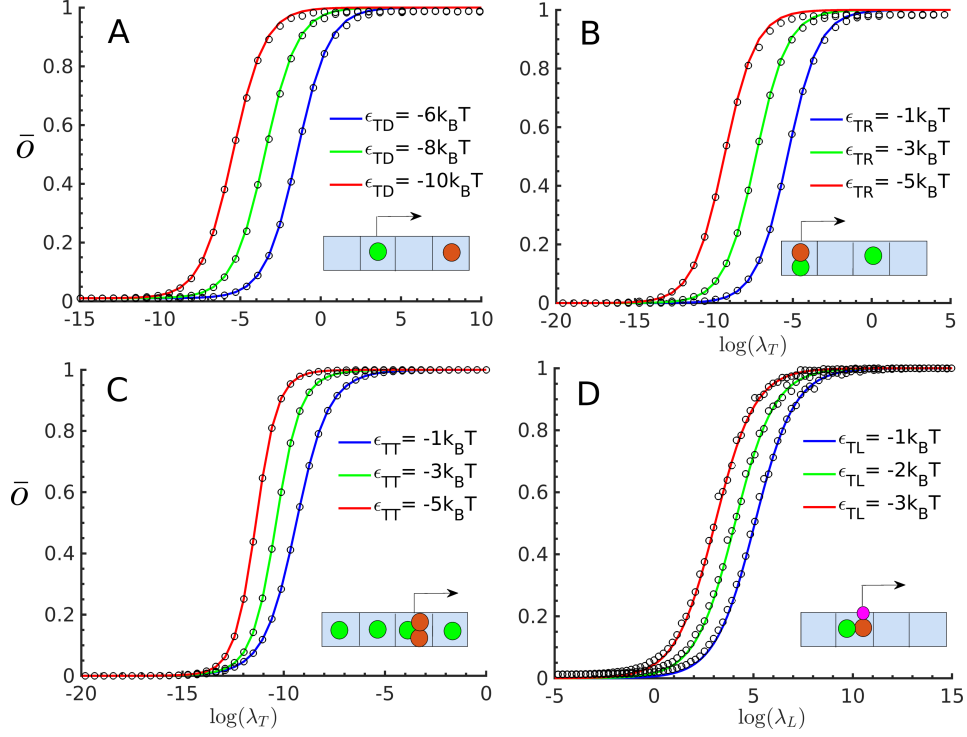
Figure 2.3: Average occupation number ($\bar{o}$) of RNAP, TF-RNAP, and TF-RNAP-L complexes for non-interacting systems as a function of λ_T and λ_L. Solid lines and the symbol circles represent theoretical and simulation results. Schematic lattice configurations of short segments of protein-DNA complexes are shown in the respective panels. A) recruitment of RNAP by TF, B) stimulation of RNAP due to the binding of TF, C) cooperative stimulation by the dimeric TFs, and D) activation due to the binding of a ligand to TFs. Following binding interaction parameter are used: $\epsilon_{RD} = -4.68 k_B T$, $\epsilon_{TD} = -8.88 k_B T$, $\lambda_R = 10^{-4}$, $\lambda_T = 1.5 \times 10^{-6}$.

the loop formation, in our calculations. Nevertheless, these binding events generate a new set of microstates, directly impacting the populations of different complexes. We further consider that 1) DNA sites can accommodate up to two TFs on a linear lattice, and 2) two distant occupied sites of DNA can participate in the loop formation in our calculations. However, one can relax such restrictions in a simulation and perform MCS for any arbitrary situation.
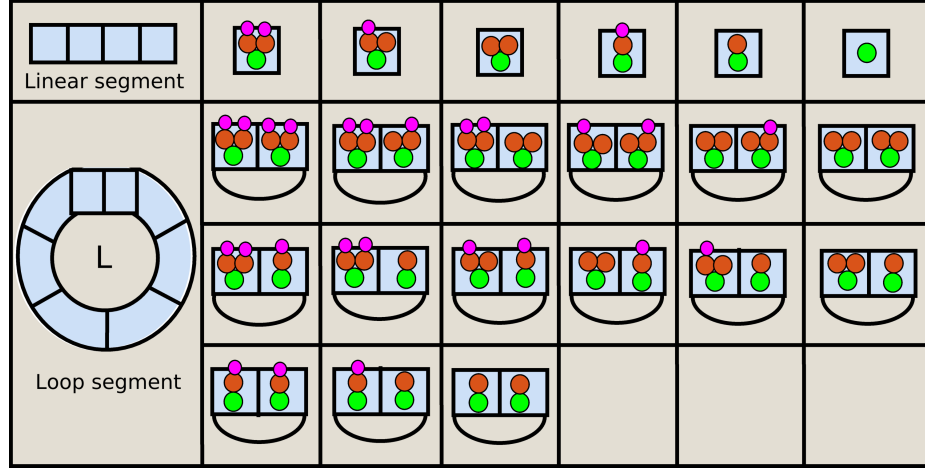
Figure 2.4: Various complexes formed in our MCS are shown in this figure. Green, orange, and magenta beads are for the RNAP, TF, and ligand. The arc shown in the figure is for the DNA loop. Complexes formed on a linear DNA segment are shown in the first row. In the subsequent rows, we show the possible complexes formed due to DNA loop formation.

2.3.2.1 Binding of TFs on a Linear DNA Segment:

We present the $\bar{o}$ for six possible types of complexes formed among TFs, ligands, and DNA as a function of ligand activity, λ_L in Figure 2.5. To obtain these results, we first consider that the DNA is sufficiently stiff and does not undergo loop formation at the expense of interactions among TFs. For comparison, we further include the results obtained from linear segments when some parts of a flexible DNA undergo loop formation. It is clear from the figure that each of the complexes' relative populations varies significantly with the increase in ligand activity. We see that the population of the highest-order oligomer is enhanced, reaching a saturation point with increased ligand activity. It happens because plenty of ligand molecules are available to bind with TF, further enhancing the attraction among TFs as ligand activity increases. We also found that the lower-order oligomer population reaches a maximum and then depletes as ligand activity increases. We also note that their magnitude is significantly smaller than the higher-order oligomer. These non-uniform populations of different oligomers are the origin of differential gene regulation as it is sensitive to a specific oligomeric species (Vilar and Saiz, 2011). The figure also shows that the favorable interactions between

TF and ligand enhance the overall population of ligand-bound TF complexes, promoting higher-order oligomerization and DNA looping (Bintu et al., 2005b,a). Therefore, it is clear from our analysis that the population of various oligomeric complexes on the DNA can be controlled quantitatively by ligand activity.
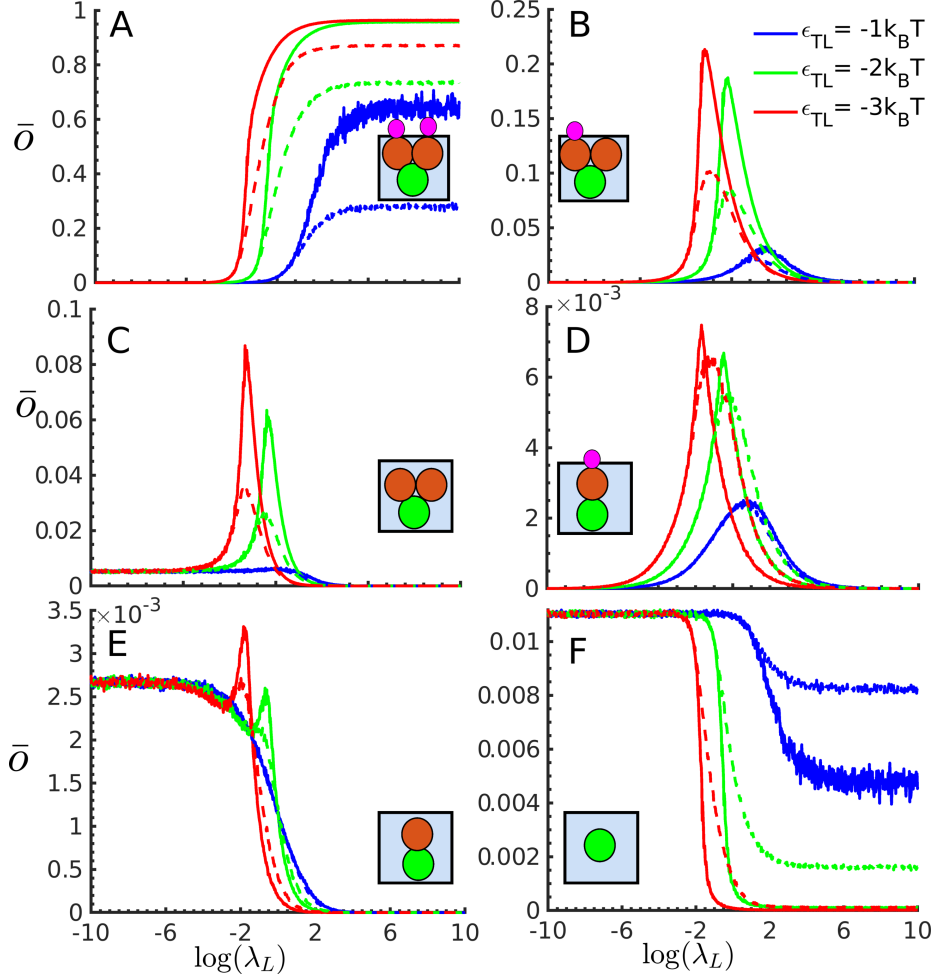
Figure 2.5: Fraction of different species formed on linear DNA segments for both the stiff (solid lines) and flexible (dashed lines) DNAs are shown as a function of ligand activity, λ_L. The following values of energy of interactions for various complex formations are considered, $\epsilon_{RD} = -4.68k_BT$, $\epsilon_{TD} = -8.88k_BT$, $\epsilon_{TR} = -3k_BT$, $\epsilon_{TT} = -2k_BT$, $w_l = -0.2k_BT$, $w_{RR} = -5k_BT$ $w_{TT} = -1k_BT$, $k_l = 3.8\frac{k_BT}{d^2}$. The calculations are done at fixed values of activities for RNAP and TF, those are given by $\lambda_R = 10^{-4}$, $\lambda_T = 1.5 \times 10^{-6}$ respectively.

2.3.2.2 Effect of DNA looping:

The formation of a DNA loop in these systems is crucial since it modifies the relative population of higher-order oligomeric species such as trimer, tetramer, etc. Proteins are the key element for the DNA loop formation (Bintu et al., 2005b,a; Rydenfelt et al., 2014; Liu et al., 2016). The origin behind DNA loop formation is the strong interactions between two long distant TFs bound to the DNA sites. These interactions must overcome the elastic free energy of DNA to form a loop. An increase in ligand activity and the enhancement of TF-ligand interaction promotes the binding ability of TFs to DNA. If they are far apart, then the DNA loop forms at the expense of TF-TF interactions.

We present the $\bar{o}$ as a function of ligand activity, λ_L in Figure 2.6. We see that the populations of all those species are small, but they are very crucial in gene regulation. The gene regulation depends on the population of a specific complex formed in a cellular system. The overall population of oligomeric species enhances as the interaction between TF and ligand strengthens. However, we find that both ligand activity and the TF-ligand interaction modulate the relative population of the complex species formed on DNA due to forming a loop. For example, we find an enhancement of the population of ligand-bound tetrameric species as the interaction between TF and ligand is strong. Enhancement of the population of tetrameric species and the expression through them are well known for the RXR receptor, p53 repressor signaling systems (Vilar and Saiz, 2011). We present the average number of loops per configuration in the last panel of the figure. We see an increase in the number of loops for strong TF-ligand interactions. Our calculation demonstrates how a ligand controls the population of active complexes in a gene regulatory system.

2.3.3 Quantification of Activation and Repression

This section discusses the GR system's functional response from the population of active species present at thermal equilibrium. We calculate FC as defined in the materials and methods section to quantify the activation and repression. We take stoichiometric values of oligomerization for the parameters, Γ_i^A and Γ_i^R in our calculation. For example, we consider Γ_i values for monomer, dimer, trimer, and tetramer as 1, 2, 3, and 4,
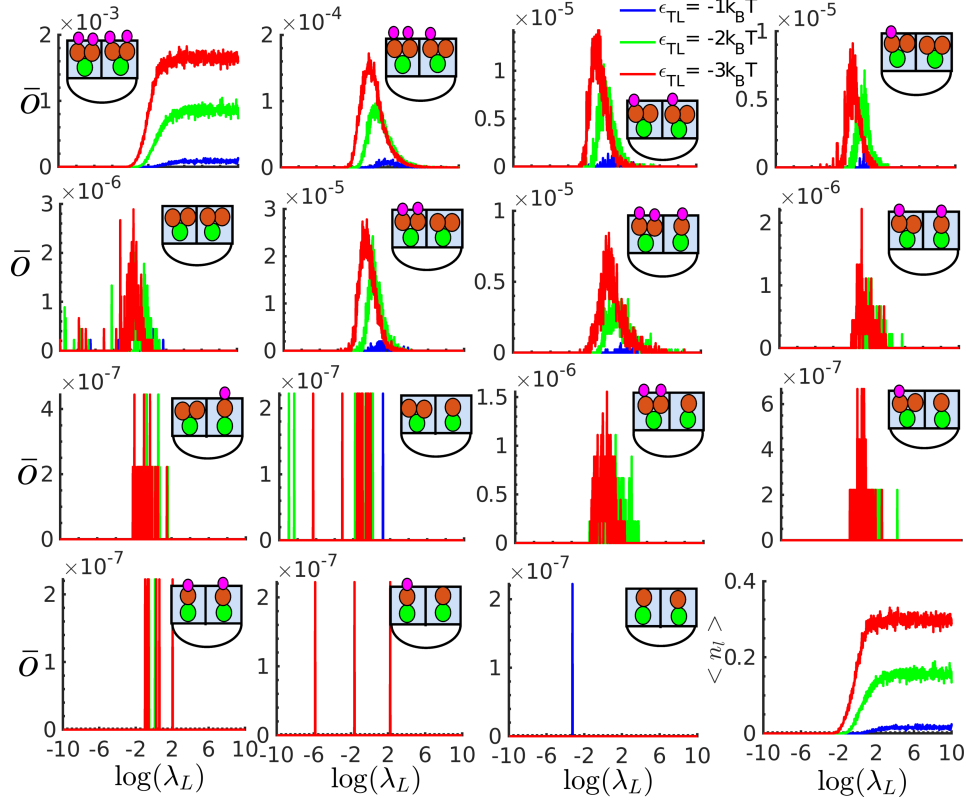
Figure 2.6: The population of higher-order oligomeric species formed on the flexible DNA due to the looping. The following values of energy of interactions for various complex formations are considered, $\epsilon_{RD} = -4.68k_BT, \epsilon_{TD} = -8.88k_BT$, $\epsilon_{TR} = -3k_BT$, $\epsilon_{TT} = -2k_BT$, $w_l = -0.2k_BT$, $w_{RR} = -5k_BT$, $w_{TT} = -1k_BT$, $k_l = 3.8\frac{k_BT}{d^2}$. The calculations are done at fixed values of activities for RNAP and TF, those are $\lambda_R = 10^{-4}$, $\lambda_T = 1.5 \times 10^{-6}$ respectively. The average number of loops per configuration is shown at the extreme bottom right panel of the figure.

respectively. However, those parameters are system-specific and can be obtained from the experiment. In Figure 2.7, we present the FC for the activation and repression as a function of TF activity for the system where TFs form at maximum dimer on linear DNA segment and tetramer through DNA looping. The results for DR are shown in the same figure as a function of the ligand activity.

In the top two panels, we show the results for NFC as a function of TF activity. If the TF acts as an activator, then NFC rises to 1. We further notice that with strong favorable TF-TF nearest-neighbor interactions for an activator, the function is steeper than the unfavorable TF-TF interactions. An opposite effect is also true for a

repressor. The NFC for the repressor quickly reaches 0 as we increase the TF-TF nearest-neighbor interactions. Therefore, our theoretical models clearly show the activation and repression of GR systems. In the two panels on the bottom, we show the results for DR as a function of ligand activity. We show that the DR rises to a saturation point when the ligand activates the activator. Our quantitative calculation shows that the strength of TF-ligand interaction modulates the population of active species, which further influences the DR. We see that the favorable interaction between protein and ligand produces strong DR. However, we find an opposite effect when the ligand activates a repressor. Therefore, it is clear from our analysis that we can control these functional responses quantitatively by including ligands in the system. Measurement of DR in a quantitative way has substantial implications in the pharmacological industry since it provides valuable information on a drug molecule's potency.
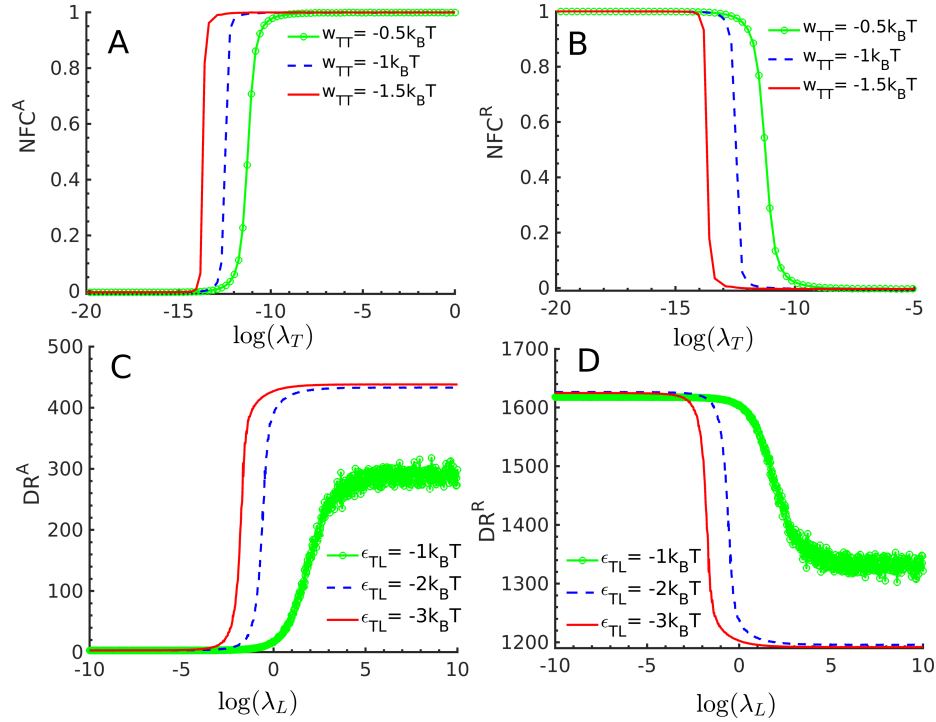
Figure 2.7: Quantification of functional responses in GR systems. A) NFC for the activation as a λ_T function. B) NFC for repression as a λ_T function. C) Control activation with a ligand, i.e., DR^A as a function of λ_L. D) Control of repression with a ligand, i.e., DR^R as a function of λ_L.

2.3.4 Transcriptional Control Of Lac Operon

We finally use the *lac* operon in E. coli bacteria as a prototype system to illustrate the applicability and limitations of our modeling scheme. An isolated *lac* operon gene's promoter region consists of an activator binding site, an RNAP binding site adjacent to the activator site, and three operator sites (Liu et al., 2017). Different binding regions of the *lac* operon promoter and the interaction network are shown schematically in figure 2.8. The *lac* operon gene expression does not happen only due to the binding and unbinding of RNAP with the promoter site of DNA, but there are also many layers of control mechanism. In particular, two proteins, namely *lac* repressor (LacR) and cAMP activator proteins (CRP), are involved in the controlled repression. In short, the binding of dimeric CRP protein to the activator site of the promoter forms a complex with RNAP. An allosteric effector, cAMP binds with the CRP protein of the complex that activates transcription. The active form of tetrameric LacR binds to one of the three operator sites of the *lac* operon gene that represses transcription. Repression by LacR becomes activation upon binding of lactose or synthetic gratuitous inducer isopropyl β-D-thiogalactopyranoside (IPTG). It is evident from experimental studies that the fold change is substantial due to the involvement of LacR-mediated DNA looping (Choi et al., 2008). We can capture this complex mechanism's final response function quantitatively from our analysis.

We show the results in Figure 2.9. In Figure 2.9A, we show the FC as a function of LacR activity. The results are also obtained for three different values of the activity of IPTG. It is evident from the figure that the LacR down-regulates the *lac* operon gene. It happens because the binding of LacR inhibits the binding of RNAP to the promoter that exhibits a downregulation of the *lac* operon gene. We also see that the gene is up-regulated in the presence of IPTG.

We find that as we increase the IPTG activity, the LacR molecules quickly unbind from the promoter region of the *lac* operon gene, which in turn up-regulates the expression. We observed a delay in repression as we increased the magnitude of λ_I from 10^{-3} to 10^{-1}. As the activity, λ_I is increased, LacR molecules are more susceptible to IPTG, which impedes LacR-IPTG complex binding and promotes binding between RNAP and DNA. Our simulation studies also find that the nearest-neighbor interactions among
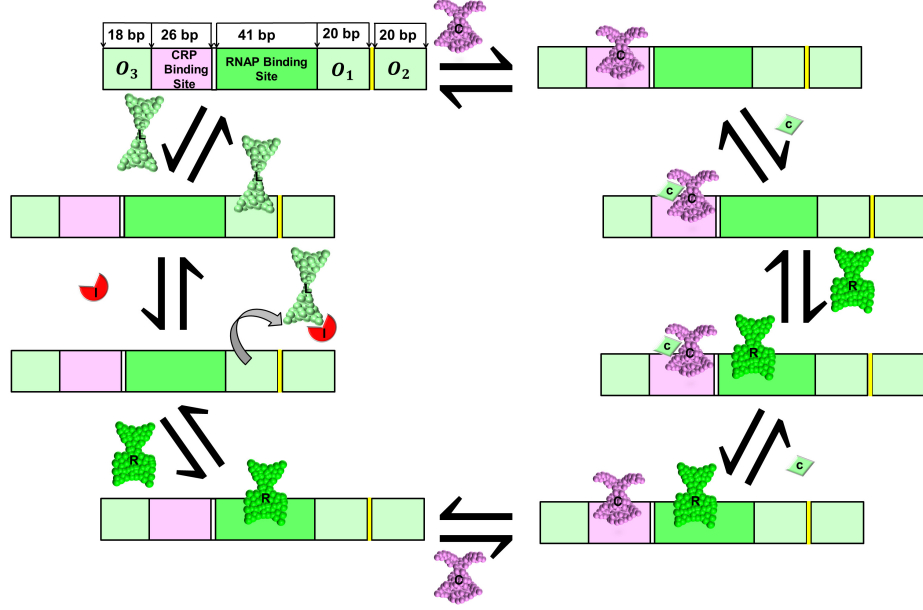
Figure 2.8: The Protein-DNA interaction network of *lac* operon system at thermodynamic equilibrium is presented. Different binding regions of the promoter of the *lac* operon gene are depicted by different shades of color in the lattice. O_1, O_2, O_3 are three operators where LacR can bind. CRP and RNAP bind to their respective binding regions. Different shapes are used to represent the different biomolecules in the figure. The letters inside the shapes, such as R, L, C, I, and c, correspond with the RNAP, LacR, CRP, IPTG, and cAMP molecules in the figure.

RNAP do not impact the repression process. However, the binding of CRP up-regulates the gene expression, as evident from Figure 2.9B. It happens because the CRP protein enhances the recruitment of RNAPs on DNA. The analysis also shows that the increased activity of cAMP further up-regulates expression. It occurs because the binding between cAMP and CRP strengthens the interaction between CRP and RNAP. Such binding facilitates the recruitment of RNAP and stimulates the bound RNAP on the promoter region of the *lac* operon gene. We found a strong activation by introducing the nearest-neighbor interaction among RNAP molecules. The long-range correlation among RNAP molecules promotes the recruitment of more RNAP on DNA. Such recruitment of RNAP molecules on DNA shows a strong activation effect in our results.

Our calculations also quantitatively predict the up and down regulations mechanisms of *lac* operon system as a function of IPTG and cAMP. We show the results in panels C and D of Figure 2.9 under two conditions: a) in the presence and b) in the absence
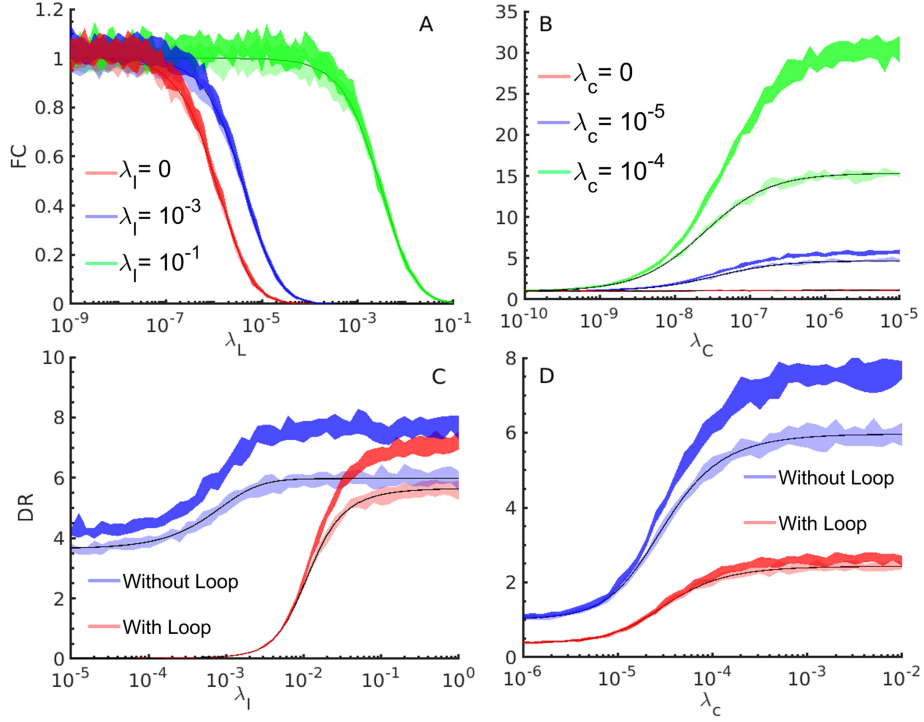
Figure 2.9: Fold Change (FC) and Dose-Response (DR) as a function of biomolecular activities, λ_L, λ_C, λ_I and λ_c. Solid lines represent theoretical results, and the results obtained from 20 independent GCMC simulation runs are shown by shaded colors. A) Repression as a function of λ_L. The calculations are done at three λ_I values. B) Activation as a function of λ_C. The calculation is also done at three constant values of λ_c. C) Activation as a λ_I function. The calculations are done for both stiff and flexible DNAs. D) Activation as a function of λ_c. Here, we also consider two cases: with and without loop cases. The results obtained from GCMC simulations for the interacting system, where we consider nearest-neighbor interactions among RNAP molecules ($w_{RR} = -5k_BT$), are shown by solid green and blue colors in each panel.

of DNA loops. Our analysis shows that the binding of IPTG to LacR facilitates the unbinding of LacR from DNA, which up-regulates the gene expression. However, removing LacR molecules from DNA using IPTG is not energetically favorable if they are involved in DNA loop formation. This fact appears in our results as the down-regulation of LacR-mediated gene expression in the presence of DNA loops. However, we observed a strong leaky effect in DR for a stiff DNA at a low value of λ_I as evident from our results. Such leaky expression disappears for flexible DNA. It happens because the LacR promotes

the loop formation of flexible DNA that tightens the binding between LacR and DNA. As a result, the low value of λ_I is insufficient to unbind LacR from the DNA binding sites and is observed as down-regulation for the flexible DNA. Since the DNA loops only form under the binding of LacR, cAMP binding to CRP up-regulates the gene expression. However, the effect is not linear since the binding of cAMP indirectly controls the binding of LacR (i.e., it decreases), further up-regulating the expression. Once again, we observe a weak leaky effect for stiffer DNA compared to the flexible one. The cAMP activates only the CRP and does not affect the unbinding of LacR. Since both ligands activate the gene expression, we find a strong long-range correlation for both cases, as shown in panels C and D. Both ligands ultimately promote the recruitment of RNAP, and the recruited RNAP molecules undergo nearest neighbor interactions on the DNA. This is a clear signature of the cooperative effect, which we observe a strong activation in the response function.

2.4 Discussions

Recent experimental and theoretical studies show that gene regulation strongly depends on the network of protein and DNA interactions. A series of unprecedented physico-chemical events control the information (energy) flow from extracellular to intracellular environments before binding a protein with DNA. Such interactions are responsible for the over-expression and under-expression of a gene. Since most diseases are intimately related to the over-expression and under-expression of a gene, it is crucial to explore those molecular mechanisms in detail. In this study, we develop novel theoretical and computational approaches based on statistical mechanics to investigate the gene expression of several regulatory motifs, as shown in Table 2.2.1. The probability of promoter occupancy is quantified by simple statistical thermodynamic relations that contain only a few experimentally measurable parameters. Our theoretical model includes a few biophysical effects, such as oligomerization of TFs, DNA looping, and TF-ligand interactions. We show that the population of different oligomers on DNA is controlled by modifying the protein-protein, protein-ligand interactions, and DNA elasticity at thermal equilibrium.

The central theme of our study is calculating the average population of RNAP and RNAP-TF complexes formed on DNA and finding the functional response. Our model's

key feature is that it is based on equilibrium thermodynamics. We apply this modeling scheme to an isolated gene and a wide variety of GR networks under different competition scenarios. The model requires only a few molecular parameters that can affect the shape of the response curves in gene regulation. Calculating the populations of various complex species on the promoter (or the promoter configuration) is crucial for determining the genotypic to phenotypic changes in the cellular environment. Extracting that information from experiments is challenging, but finding the relationships between the molecular mechanism behind gene regulation and the shape of the functional response is central in molecular biology. Our model successfully captures such facts without ignoring the complexity of the problem.

We calculate the $\bar{o}$ of RNAP on the model DNA lattice as a function of the activity of RNAP for the basal expression. We show that the $\bar{o}$ quickly reaches the saturation point for the favorable interactions between RNAP and DNA. However, we find that binding RNAP with specific and non-specific sites of DNA modifies the shape of $\bar{o}$ function. It happens due to the competition of binding of RNAP with specific and non-specific sites of DNA. The origin of such hidden heterogeneity in the function, $\bar{o}$ arises from the different molecular interactions and the possible number of available states of the system (Vilar and Saiz, 2011; Saiz, 2012; Phillips, 2015; Landman et al., 2017). We also show that the cooperative transcriptional activation results from cooperative interaction among RNAP molecules. However, the free energy of interactions between RNAP and DNA is particular and varies within a narrow range for a natural genetic system. The free energy of interactions between RNAP and DNA is modified by the interactions of TFs with DNA. We show that the strong interactions between TFs and DNA significantly alter the function, $\bar{o}$ of RNAP-TF complexes. Such modifications in the population of complexes influence the functional response, i.e., mRNA production. Thus, it is clear from our discussion that a TF controls the production of mRNA.

Moreover, TF in a genetic system controls gene regulation by external interventions, such as introducing a small molecule. These small molecules can be drug molecules that only bind to the ligand-binding domain of a TF and alter network configurations. Including a drug molecule in the system opens up an ample opportunity to discover a new drug molecule that controls the over-expression and under-expression of a gene

using this method. The ligand that alters the RNAP-TF complex population affects the functional response, i.e., mRNA production. Our analysis shows that the proposed modeling scheme successfully captures such ligand-controlled gene regulation mechanisms at thermal equilibrium.

The TF-ligand interactions modulate the population of RNAP complexes and modify the population of oligomeric species formed by TFs. Often, a specific oligomeric complex is an active species for gene expression (Vilar and Saiz, 2011; Landman et al., 2017). We have also shown that the oligomerization of TFs on DNA further modifies the population of the complexes. The formation of such species on DNA is crucial since one can correlate the functional response with a specific oligomeric species. For example, dimeric and tetrameric species are found active for the RXR receptor system; tetrameric species are found active for the p53 receptor system; all the monomeric, dimeric, and tetrameric species are found active for octamer binding proteins (Vilar and Saiz, 2011). We further show that long-distance interactions among TFs can happen through the formation of the DNA loop. DNA looping is crucial in gene regulation since it modifies the population of oligomeric species and generates a functionally active looped configuration of DNA. Many detailed studies have been performed on *lac* represor system to explore the DNA looping and its consequences (Vilar and Saiz, 2011; Landman et al., 2017). The origin behind looping is associated with DNA elasticity that alters the population of oligomeric species at thermal equilibrium (Smith et al., 1992).

Softening or hardening of DNA can happen due to the change in cellular environments, such as alteration of pH, ionic strength, etc., or chemical modification of DNA such as DNA methylation (Choy et al., 2010; Dangkulwanich et al., 2014; Severin et al., 2011; Baumann et al., 1997; Ngo et al., 2016). Our analysis shows that the formation of the loop and the population of oligomeric species are correlated. They significantly influence FC and DR. Further, the sensitivity of the biophysical parameters to the response curve shows that the complexes' population formed within a narrow range of them. In other words, they are less sensitive to the population of the complexes. We consider only three biologically relevant parameters such as k_l, ϵ_{PL}, and λ_L to show their effect on the population of complexes. Our analysis shows competition among the parameters, but they are less sensitive to the response curve.

We finish our analysis by considering the *lac* operon promoter sequence in which multiple cellular organizations are observed. We have chosen this system because many detailed studies have been done in this direction. In particular, Vilar *et. al.* (Vilar and Saiz, 2005, 2014, 2011), Saiz *et. al.* (Saiz and Vilar, 2008, 2007b) and Philips *et. al.* (Phillips, 2015; Boedicker et al., 2013; Cournac and Plumbridge, 2013) have developed similar thermodynamic models for the regulation of the *lac* operon system. We have performed detailed, comprehensive work for this system using our developed model and validated it against GCMC simulation in this work. Our analysis shows that the cAMP activates the CRP activator protein, enhancing the promoter-RNAP complex population and finally upregulating the expression. Similarly, the IPTG facilitates the LacR unbinding, further upregulating the expression. We also show that LacR-mediated DNA looping down-regulates a few orders of magnitude in the gene expression compared to unlooped configuration. On a final note, the *lac* operon system analysis is not new, but our calculations provide new insights into gene regulation at the system level in a quantitative manner.

Appendix: 1) Average occupation number for non-interacting sites

The central quantity is to find the population of various complexes formed at thermodynamic equilibrium. We define the total energy function for protein-DNA interaction for a j_{th} energy state as below :

$$E_j = \sum_{i=1}^{M} \sum_{o=0}^{m} s_i^o \epsilon_j(o) \tag{2.4}$$

Here, we introduce a binary variable s_i that takes care of a biomolecule occupancy on the i_{th} site of the DNA lattice. The index and superscript o stands for the occupation of biomolecules at that site. The variable, $s_i^o = 1$, when i_{th} site of DNA is occupied with o protein molecules and $s_i^o = 0$ otherwise. The symbol $\epsilon_j(o)$ is the j_{th} level site with occupancy o. With these descriptions, we define the canonical partition function for our system that takes the following form

$$Q(N, M, T) = M! \sum_j e^{-\frac{E_j}{k_B T}} = M! \sum_j e^{-\sum_{i=1}^{M} \sum_{o=0}^{m} s_i^o \frac{\epsilon_j(o)}{k_B T}} \tag{2.5}$$

At this point, we can group the number of DNA sites with o protein molecules. We define a new variable a_o, the total number of sites with occupancy, o. The variables a_o and o are related by the following relation.

$$\sum_{o=0}^{m} a_o = M, \quad \sum_{o=0}^{m} o a_o = N \tag{2.6}$$

If the system's configuration is given by those set of numbers $\{a_o\}$, then the above partition function takes the following form.

$$
\begin{aligned}
Q(N, M, T) &= M! \frac{\sum_j e^{-\sum_{o=0}^{m} a_o \frac{\epsilon_j(o)}{k_B T}}}{\prod_{o=0}^{m} a_o!} \\
&= M! \prod_{o=0}^{m} \frac{(\sum_j e^{-\frac{\epsilon_j(o)}{k_B T}})^{a_o}}{a_o!} \\
&= M! \prod_{o=0}^{m} \frac{q(o)^{a_o}}{a_o!}
\end{aligned}
\tag{2.7}
$$

where $q(o) = \sum_j e^{-\frac{\epsilon_j(o)}{k_B T}}$ is the site partition function, where the site is occupied with o protein molecules. The above mathematically awkward restriction (2.6) can be avoided by introducing the grand canonical partition function instead.

$$
\begin{aligned}
\Xi(\mu, M, T) &= \sum_{N=0}^{mM} Q(N, M, T) \lambda^N \\
&= \sum_{N=0}^{mM} \frac{M! q(0)^{a_0} q(1)^{a_1} q(2)^{a_2} \cdots q(m)^{a_m}}{a_0! a_1! a_2! \cdots a_m!} \lambda^N \\
&= \sum_{\{a_o\}} \frac{M! q(0)^{a_0} [q(1)\lambda]^{a_1} [q(2)\lambda^2]^{a_2} \cdots [q(m)\lambda^m]^{a_m}}{a_0! a_1! a_2! \cdots a_m!}
\end{aligned}
\tag{2.8}
$$

where $\lambda = e^{\mu/k_B T}$ is the absolute activity, which converts a many-body problem to a one-body problem as a power series of λ. We have used the second restriction in (2.6) in rewriting λ^N in the last step, and the only first restriction (2.6) on the sets $\{a_o\}$ is.

Use of the multinomial theorem, we can simplify the function Ξ

$$\Xi(\mu, M, T) = \xi(\lambda, T)^M \tag{2.9}$$

where $\xi = \sum_{o=0}^{m} q(o)\lambda^o$, The average number of bound protein molecule per site, $\bar{o}$ is given as

$$\bar{o} = \frac{\bar{N}}{M} = \frac{\lambda}{M}\left(\frac{\partial ln\Xi}{\partial \lambda}\right)_{M,T} = \frac{\sum_{o=0}^{m} oq(o)\lambda^o}{\sum_{o=0}^{m} q(o)\lambda^o} \tag{2.10}$$

The quantity $\bar{o}$ is of primary interest since it is directly related to the DNA molecule transcription. We can extend the above derivation for different, distinguishable, and independent sites. If all the sites are distinguishable but still independent of each other, then the grand partition, $\Xi(\mu, M, T)$ function will be $\prod_{i=1}^{M} \xi_i$, where ξ_i is a sum, $\sum_{o=0}^{m} q_i(o)\lambda^o$, for the i_{th} site only. Different types of DNA binding sites are common as the protein binds to DNA, either specifically or nonspecifically. For example, the RNAP binds specifically to the gene's promoter region to initiate transcription. The TF can also bind either specifically or nonspecifically. In this case, the specific binding of RNAP to a gene can only happen due to forming a particular complex between TF and RNAP. Therefore, if there are S types of equivalent sites present and the number of DNA binding sites for that type is M_S, then the grand canonical partition function takes the following form $\Xi(\mu, M, T) = \prod_{t=1}^{S} \xi_t^{M_t}$, where $M = \sum_{t=1}^{S} M_t$ and $\xi_t = \sum_{o=0}^{m} q_t(o)\lambda^o$.

Further, the above derivation can be extended for the multi-component system if the DNA sites are still independent and equivalent. In cellular systems, different TFs and enzymes can bind with DNA sites. If we consider there are P types of protein molecules present in the system, each of the types protein possesses N_P molecules, and they bind up to m_P protein molecules with a DNA site, then we can calculate all the binding properties of the system. We further consider system is open with respect to all the components, then the grand canonical partition function, $\Xi(\mu_1, \mu_2....\mu_P, M, T) = \xi(\lambda_1\lambda_2...\lambda_P, T)^M$, where $\xi = \sum_{o_1=0}^{m_1} \sum_{o_2=0}^{m_2} ... \sum_{o_P=0}^{m_P} q(o_1, o_2...o_P)\lambda_1^{o_1}\lambda_2^{o_2}...\lambda_P^{o_P}$.

In a similar fashion, we can also calculate the fluctuations of occupation (σ_o). We start with the following relation and then differentiate both sides with respect to λ,

dividing by Ξ and multiplying with λ.

$$\bar{N} \sum_N Q(N,M,T)\lambda^N = \sum_N NQ(N,M,T)\lambda^N \tag{2.11}$$

$$\lambda \left(\frac{\partial \bar{N}}{\partial \lambda}\right)_{M,T} + \frac{\bar{N}}{\Xi} \sum_N NQ(N,M,T)\lambda^N = \frac{1}{\Xi} \sum_N N^2 Q(N,M,T)\lambda^N \tag{2.12}$$

or,

$$\bar{N^2} - (\bar{N})^2 = {\sigma_N}^2 = \lambda \left(\frac{\partial \bar{N}}{\partial \lambda}\right)_{M,T} \tag{2.13}$$

Here, ${\sigma_N}^2$ indicates the variance. Now, we put $\bar{N}$ from Eq. (1.10) and obtain

$$\bar{N^2} - (\bar{N})^2 = {\sigma_N}^2 = \lambda \left(\frac{\partial \left(\lambda \left(\frac{\partial ln\Xi}{\partial \lambda}\right)_{M,T}\right)}{\partial \lambda}\right)_{M,T} \tag{2.14}$$

Now, we divide Eq. 1.11 by M^2 and obtain fluctuations of occupation numbers per site, which takes the following form

$$\bar{o^2} - (\bar{o})^2 = {\sigma_o}^2 = \frac{\lambda}{M^2} \left(\frac{\partial \left(\lambda \left(\frac{\partial ln\Xi}{\partial \lambda}\right)_{M,T}\right)}{\partial \lambda}\right)_{M,T} \tag{2.15}$$

Let us compute the fraction of bound molecules for a few gene regulatory systems with the above formulation. First, we consider a straightforward case, i.e., the binding of RNAP with DNA. It is a case of uncontrolled gene expression. We consider RNAP binds with the DNA site in a binary fashion, i.e., the occupancy, o is either 0 or 1. Since mRNA synthesis is proportional to the number of RNAP molecules bound to the promoter region, we primarily want to calculate the fraction of bound RNAP per DNA site. The site partition function $q(o)$ for this case is either $q(0) = 1$ for unoccupied site or $q(1) = q_R$ for the occupied site. The quantity ξ will be $1 + q_R\lambda_R$ and the grand canonical partition, $\Xi = \xi^M$. Thus, the fraction of bound RNAP is,

$$\bar{o} = \frac{\lambda_R}{M} \left(\frac{\partial ln\Xi}{\partial \lambda_R}\right)_{M,T} = \frac{q_R\lambda_R}{1 + q_R\lambda_R} \tag{2.16}$$

Appendix: 2) Average occupation number for interacting sites

We consider a DNA segment as a one-dimensional lattice of which M equivalent, distinguishable sites are present. However, there is the potential energy of interaction, w between two neighboring sites when protein molecules occupy both. The energy of interaction between protein and DNA at i_{th} site is ϵ_i. The introduction of such a nearest neighboring interaction effect into the model converts our previous model to a more mathematically challenging model. However, one can map this model with the 1-d lattice gas model or more formally known as the Ising model for magnetic systems. This model is solved by a variety of analytical techniques such as combinatorial, transfer matrix, etc. This prototypical model is the starting point for a wide variety of real systems, such as gas adsorption, gas-liquid phase transition, binary alloy, protein-DNA interaction, etc. Here, we adopt this model for the calculation of occupation numbers for a weakly interacting system, where the long-range correlation is important. According to our previous notation, the total energy for this interacting system for a jth energy state is given by

$$E_j = \sum_{i=1}^{M} \sum_{o=0}^{m} s_i^o \epsilon_j^N(o) + \sideset{}{'}\sum_{ik} \sum_{o=0}^{m} \sum_{o'=0}^{m} s_i^o s_k^{o'} \epsilon_j^{NN}(o) \tag{2.17}$$

The index i and k are related by the nearest neighbor, and the periodic boundary condition is applied to the lattice (i.e., $M+1$ site is the first site). The variable, s_i^o, takes value 1 when the site is occupied by o molecules and 0 otherwise, and the primed sum extends over nearest-neighbor interactions. For the sake of simplicity, we consider each of the DNA sites can be either occupied or unoccupied (i.e., the value for $m = 1$), and the energy of interaction between protein and DNA at i_{th} site is $\epsilon_j^N(o) = \epsilon$, and the nearest neighbor interaction between two occupied sites is $\epsilon_j^{NN}(o) = w$. Thus, the total energy expression for a j_{th} energy state, takes the form

$$E_j = \epsilon \sum_{i=1}^{M} s_i + w \sum_{i=1}^{M} s_i s_{i+1} \tag{2.18}$$

where s_i is the binary variable, which takes a value of either 0 or 1. At this point, we can group two types of fundamental interactions, such as (*a*) N protein-DNA interactions, (*b*) N_{11} nearest neighbor interactions. The system is characterized thermodynamically by M sites, of which N are occupied at temperature T. There will be N_{11} nearest neighbor pairs of occupied sites; the interaction potential is $N_{11}w$. However, it is more convenient to use N_{01}–the number of nearest-neighbor pairs of sites with one site empty and the other filled with protein. Both variables are related, and the following argument can establish it. If a line is drawn from each occupied site to its two neighboring sites, we will have drawn $2N$ lines. Also, in this process, two lines are placed between each nearest neighbor, 11 pairs, and one line between 01 pairs. Therefore, these arguments constitute two relations (*a*) $2N = 2N_{11} + N_{01}$ for occupied sites and (*b*) $2(M - N) = 2N_{00} + N_{01}$ for empty sites, which further suggest that only one of N_{11}, N_{01}, N_{00} is independent; we chose N_{01}. We assume M is very large, and therefore end effects from relations (*a*) and (*b*) can be neglected. The potential energy of nearest neighbor interaction is $N_{11}w = (N - N_{01}/2)w$. Since the system is open, we consider a grand canonical ensemble for the calculation of the average number of bound proteins. The total energy of interaction for this interacting system, which can be written in an explicit way as

$$E_j = N\epsilon + N_{11}w = N\epsilon + (N - N_{01}/2)w \tag{2.19}$$

The canonical partition function is

$$
\begin{aligned}
Q(N, M, T) &= \sum_j g_j e^{-\frac{E_j}{k_B T}} \\
&= \sum_E \sum_{N_{01}} g(N, M, N_{01}) e^{-\frac{\{N\epsilon + (N - N_{01}/2)w\}}{k_B T}} \\
&= \sum_E \left[e^{-\frac{(N\epsilon - \beta w N)}{k_B T}} \sum_{N_{01}} g(N, M, N_{01}) \left(e^{\frac{w}{2k_B T}} \right)^{N_{01}} \right] \\
&= \left(q e^{-\frac{w}{k_B T}} \right)^N \sum_{N_{01}} g(N, M, N_{01}) \left(e^{\frac{w}{2k_B T}} \right)^{N_{01}}
\end{aligned}
\tag{2.20}
$$

where $q = \sum_E e^{-\frac{\epsilon}{k_B T}}$, is the site partition function. The function $g(N, M, N_{01})$ is the

degeneracy for the configurational arrangements N molecules in M sites giving N_{01} pairs of type 01. The function $g(N, M, N_{01})$ takes the following form

$$g(N, M, N_{01}) = \frac{N!(M-N)!}{(N-N_{01}/2)!(M-N-N_{01}/2)![(N_{01}/2)!]^2} \tag{2.21}$$

The details of the construction of this function can be found elsewhere (Hill, 1986). The grand canonical partition function is

$$\begin{aligned}
\Xi &= \sum_N Q(N, M, T)\lambda^N \\
&= \sum_N \left(q\lambda e^{-\frac{w}{k_B T}}\right)^N \sum_{N_{01}} g(N, M, N_{01}) \left(e^{\frac{w}{2k_B T}}\right)^{N_{01}} \\
&= \sum_N \sum_{N_{01}} x^N g(N, M, N_{01}) y^{N_{01}} \\
&= \sum_N \sum_{N_{01}} t(N, N_{01})
\end{aligned} \tag{2.22}$$

where, $x = e^{w/2k_B T}$, and $y = q\lambda e^{-w/k_B T}$. Use of maximum term method, one can approximate $\ln \Xi$, as the dominant term of the double summation over N and N_{01} to determine the Ξ

$$\begin{aligned}
\ln \Xi &= \ln \sum_{N, N_{01}} t(N, N_{01}) \\
&= \ln t_{max}
\end{aligned} \tag{2.23}$$

where $\ln t = \ln g + N \ln y + N_{01} \ln x$. To find $\ln t_{max}$ we set $\frac{\partial \ln t}{\partial N} = 0$, and $\frac{\partial \ln t}{\partial N_{01}} = 0$ and obtain

$$\begin{aligned}
\frac{\bar{o}}{1-\bar{o}}\left(\frac{1-\bar{o}-\alpha}{\bar{o}-\alpha}\right) &= \frac{1}{y} \\
\frac{(\bar{o}-\alpha)(1-\bar{o}-\alpha)}{\alpha^2} &= \frac{1}{x^2}
\end{aligned} \tag{2.24}$$

where $\bar{o} = \frac{N^*}{M}$, $\alpha = \frac{N_{01}^*}{2M}$. Upon rearrangement of the first equation of (2.24), we first obtain α and then get $\bar{o}$ after substituting it into the second expression.

$$\begin{aligned}
\alpha &= \frac{\bar{o}(1-\bar{o})(y-1)}{\bar{o}y - (1-\bar{o})} \\
\bar{o} &= \frac{4y + x^{-2}(y-1)^2 + (y-1)[4yx^{-2} + x^{-4}(y-1)^2]^{\frac{1}{2}}}{2[4y + x^{-2}(y-1)^2]} \\
&= \frac{1}{2} - \frac{e^{w'}\sinh(\mu'/k_B T)}{2[e^{-2w'} + e^{2w'}\sinh^2(\mu'/k_B T)]^{\frac{1}{2}}}
\end{aligned} \tag{2.25}$$

where, the effective chemical potential, $\mu' = (\epsilon + w - \mu)/2$ and $w' = -w/4k_BT$.

Appendix: 3) Grand canonical Monte Carlo simulation

In the previous subsections, we discussed the calculation of the fraction of bound protein, which is the key quantity to study gene regulation using analytical approaches. However, we noticed that introducing simple nearest neighboring interactions into the model introduces serious mathematical complications and limits its applicability to more complex interacting systems. In this regard, Monte Carlo simulation offers an intriguing opportunity to calculate the same quantity for a more complex interacting system. Therefore, we continue our calculation with the framework of the grand canonical Monte Carlo simulation (MCs) for protein-DNA interactions. The protein's bound phase is in thermodynamic equilibrium with the free unbound phase at fixed chemical potential, μ, the total number of DNA sites M, and temperature, T. We validated our exact solution with this simulation. For example, (a) the inclusion of long-distance interactions among TFs occupied at two distant DNA sites, or (b) the binding and formation of different oligomeric species on the DNA site, or (c) the long-range correlation effects mediated through DNA loop are quite challenging to study through analytical techniques. However, the MCs offer to include any type of complex interactions present in protein–DNA systems. To perform MCs, we first define the total energy of interaction for a microstate in the protein-DNA model system. We still consider the DNA as a 1-D lattice in which different types of molecular interactions are present. Specifically, we consider three types of biomolecules, namely, RNAP(responsible for the gene expression), TFs, and ligands, that control the expression. All sorts of possible complex interactions are considered to prove the richness and flexibility of the model. In specific, the total energy of such an interacting system is given by

$$
\begin{aligned}
E_j = \sum_{i=1}^{M} \sum_{o=0}^{m} s_i^o \epsilon_j^N(o) + \sum_{i=1}^{M} \sum_{o=0}^{m} \sum_{o'=0}^{m} s_i^o s_{i+1}^{o'} \epsilon_j^{NN}(o) + \sum_{i=1}^{M} \sum_{i'=1}^{M} \\
\sum_{o=0}^{m} \sum_{o'=0}^{m} \{ s_i^o s_{i'}^{o'} \epsilon_j^{DIS}(o) + \frac{1}{2} k_l (|\Delta M_{ii'}| - L_0)^2 \}
\end{aligned}
\tag{2.26}
$$

where, M is the total number of lattice points for DNA; m is the maximum occupancy per site; s_i^o is the binary variable, which takes value 1 when the i_{th} site is occupied with o molecules, otherwise it takes value 0. ϵ_j^N is the host-guest interaction between DNA and other molecules, such as TFs, RNAP, and ligand. ϵ_j^{NN} is the different types of nearest-neighbor interactions among different bound molecules present on DNA; ϵ_j^{DIS} is the interaction between two molecules at two distal bound DNA sites, i and i'; k_l is the force constant for bending of DNA molecule. $|\Delta M_{ii'}|$ is the absolute value of the difference between two sites, i and i'; and L_0 is the reference value of site difference for which the loop energy is minimum. The last term constitutes the competition between the intermolecular interactions and the elastic energy of DNA, and these two energies are opposite in nature.

Once we defined the system's total energy, we performed the simulation in a grand-canonical ensemble using the Metropolis-Hastings algorithm. We calculate the change in energy, ΔE for the transition from old to new microstates. We consider a promoter site containing an RNAP binding site and a TF binding site that accommodates RNAP and TF molecules in our simulation. We further consider the promoter site can accommodate up to 2 TFs and one RNAP. A single ligand molecule binds specifically with the TF. We follow the following Markov chain proposal transitions.

1) We chose a random DNA site and proposed to insert TF or, RNAP or ligand. If the sites are saturated, we do not accept the proposal. If unsaturated, accept the proposal with probability

$$P(N \rightarrow N+1) = min\left\{1, \left(\frac{M}{\Lambda^3}\right)\frac{\lambda}{(N+1)}e^{-\frac{\Delta E}{k_B T}}\right\} \qquad (2.27)$$

2) We chose a random bound DNA site to remove it. We accept the proposal with the probability

$$P(N \rightarrow N-1) = min\left\{1, \left(\frac{\Lambda^3}{M}\right)\frac{N}{\lambda}e^{\frac{\Delta E}{k_B T}}\right\} \qquad (2.28)$$

If there are no bound molecules in the system, we proceed with the next Markov

chain transition proposal.

3) We translate the molecules on the lattice, and we accept the proposal with the following probability

$$P(o \rightarrow n) = min\left\{1, e^{-\frac{\Delta E}{k_B T}}\right\} \tag{2.29}$$

The periodic boundary condition is applied to avoid the end effects in our simulation. Our MC simulation code is written in MATLAB, and the relevant **pseudo code containing algorithm for carrying out Grand Canonical Monte Carlo (GCMC) simulation for 1-D DNA lattice** is written below :

Parameters

$(k_B T, V, activity, MCSteps, AddRemSteps) \rightarrow (0.53, 180, 10^{-9}, 25000, 100)$

$\epsilon \rightarrow$ All sort of interaction energies

Initialisation

X=zeros(V,1);   **Array containing coordinate of proteins**

N=sum(X);   **Number of proteins**

U=0;   **Total energy of the system**

$\bar{o} = 0;$ **Average number of proteins per microstate per site**

Monte carlo code

for $step = 1 : MCSteps$

for $substep = 1 : AddRemSteps$

if $rand < 0.5$ **For addition of protein**

$i = ceil(rand * V); , j = 1;$

if $X_{i,j} == 0$

$\delta U = \epsilon;$ **Interaction energies by applying PBC**

if $rand < \frac{exp(-\frac{\delta U}{k_B T}) * activity * V}{(N+1)}$ **Metropolis algorithm**

$X_{i,j} = 1; U = U + \delta U; N = N + 1;$

end

end

else   **For deletion of protein**

if $N > 0$

$i1 = ceil(rand * V); , j1 = 1;$

if $X_{i1,j1} > 0$

$\delta U = -\epsilon;$ **Interaction energies by applying PBC**

if $rand < \frac{exp(-\frac{\delta U}{k_B T}) * N}{activity * V}$ **Metropolis algorithm**

$X_{i1,j1} = 0; U = U + \delta U; N = N - 1;$

end

end

end

end

end

if $N > 0$ **Translation of protein on lattice**

for $substep = 1 : V$

$i2 = ceil(rand * V); , j2 = 1;$

if $X_{i2,j2} > 0$

$i3 = ceil(rand * V); , j3 = 1;$

if $X_{i3,j3} == 0$

$\delta U1 = -\epsilon;$ **Interaction energies by applying PBC**

$X_{i2,j2} = 0$

$\delta U2 = \epsilon;$ **Interaction energies by applying PBC**

$\delta U = \delta U1 + \delta U2;$

$X_{i2,j2} = X_{i2,j2} + 1;$

if $rand < exp(-\frac{\delta U}{k_B T})$ **Metropolis algorithm**

$X_{i2,j2} = X_{i2,j2} - 1; X_{i3,j3} = X_{i3,j3} + 1; U = U + \delta U;$

end

end

end

end

end

$N_P = sum(X)$

$\bar{o} = \bar{o} + \frac{N_P}{V}$;

end

$\bar{o} = \frac{\bar{o}}{MCSteps}$;   **Final average no. of proteins per microstate per site**

Functional responses of bio-molecular assembly networks

3.1 Introduction

In the previous chapter, we discussed how protein-DNA interaction networks modulate the shape of the response function in gene regulation. In this chapter, we characterized them and infer that these networks form extremely reliable functional units in a cell (Hartwell et al., 1999). These networks of interaction form a diverse range of biomolecular assemblies that perform nonlinear regulatory operations involved in cellular decision-making and signal processing (Bashor et al., 2019; English et al., 2021; Gyorgy et al., 2023; Kalir and Alon, 2004; Mangan and Alon, 2003). Moreover, they bind combinatorially to produce functional responses in converting the TF inputs into a switching-like transcriptional output (Buchler et al., 2003). Many examples are reported in literature where the self-assembly mechanism is critical for forming signaling complexes with tens of different molecular species in a biological cell (Liu et al., 2017; Bashor et al., 2019).

The network motif in the gene regulatory system is defined as the statistically significant smaller protein-DNA interaction network or pattern in a large biomolecular network. The self-assembly of protein and DNA is the critical mechanism for forming such network motifs in a cell (Vilar and Leibler, 2003; Teif, 2005; Widom, 2005; Wong and Gunawardena, 2020). The most common network motifs are the feedback and feedforward loops, known in both prokaryotic and eukaryotic cells (Kaplan et al., 2008; Avendaño et al., 2013; Siegele and Hu, 1997; Jenkins and Macauley, 2017).

For example, the *lac* operon system exhibits a feedback mechanism upon binding *lac* repressors to the *lac* operators (Mitrophanov and Groisman, 2008; Yildirim and

Mackey, 2003). The circuit constitutes positive feedback upon binding an inducer allolactose with the LacI repressor. The inducer promotes unbinding the repressor, further repressing transcription of the *lacZYA* operon. As a result, the presence of more allolactose in the system elevates the expression of the transport protein LacY, which in turn enhances the rate of lactose intake and its conversion into allolactose. The system also exhibits negative feedback because the LacZ metabolizes allolactose, thus decreasing its availability to LacI, eventually leading to the lacZYA operon's repression. The feedback loops are also known to exist in eukaryotic cells. In plant cells, the hormone Auxin induces the production of a family of proteins, AUX and IAA, which act as repressors for the expression of its gene. They also interact with the Auxin Response Factor (ARF) that promotes repression (Teale et al., 2006). In yeast cells, the galactose activates the TetR repressor, which represses the yellow GFP production. However, a signaling molecule, anhydrotetracycline (ATC), controls the DNA binding activity of Tet repressor protein that leads to a negative feedback loop in this system (Nevozhay et al., 2009).

Feedforward loops (FFLs) are also common in the GRN of *E.Coli* bacteria, yeast and human cells (Milo et al., 2002; Shen-Orr et al., 2002). An incoherent FFL is formed in the *gal* operon system of *E. coli* bacteria, where an activator, cAMP, induces CRP recruitment. Such recruitment activates both *galS* and *galETK* operons. In addition to that, the galactose induces the *galS* operon that represses *galETK* operon and its promoter (Mangan et al., 2003). Higher-order organisms such as human stem cells, OCT4, SOX2, and NANOG also form a regulatory circuit consisting of autoregulatory and feedforward loops (Boyer et al., 2005).

Long-distanced enhancer elements participate in gene expression by forming specific assemblies in various systems (Levine and Davidson, 2005; Stathopoulos and Levine, 2005). NF-κB is one such system, where the enhancer elements, κB, and IRE are far from their promoter. Two transcriptional activators, NF-κB and Interferon regulatory factors (IRF), form heterodimers and bind to their respective enhancer sites. These activators get activated in macrophages after exposure to pathogens. Formation of such protein assembly at the regulatory region of NF-κB system determines the origin of the Boolean logic in the NF-κB gene regulatory circuits (Cheng et al., 2011).

Wang *et. al.* have reported that the IRF combinatorially control NF-κB target genes through their computational modeling and predicted AND and OR logic gates for this system (Wang et al., 2021b). Further, both of the activators fluctuate between active and inactive states upon exposure to a stimulus, lipopolysaccharide (LPS), (Saravanan et al., 2020; Puc et al., 2015) and that promote binding to their respective enhancer elements of target genes (Michida et al., 2020). Typically, the source of these stimuli are pathogens, and they are powerful pro-inflammatory agents and potent activators in monocytes/macrophages (Tucureanu et al., 2018; Idriss and Naismith, 2000). However, the biological function of NF-κB is complex, producing diverse cellular variability in response to stimuli. The mechanisms behind the selective participation of NF-κB at the enhancer regions are still unclear (Lee et al., 2009; Tay et al., 2010).

Various experimental techniques based on single-molecule experiments, such as Atomic Force Microscopy (AFM) (Harada et al., 1999; Lee, 2019), magnetic tweezers assays (Revyakin et al., 2006), cryo-electron microscopy (Liu et al., 2017), and DNA sequence analysis help to find the parameter information for such networks. Single-cell level experiments based on Green Fluorescence Protein (GFP) provide quantitative details of the amount of mRNA produced or expression (Soboleski et al., 2005). Further, the electromobility shift assays (Hellman and Fried, 2007) (EMSA) or "footprinting" assays (Ragnhildstveit et al., 1997) are used to obtain the detailed protein-DNA interaction energies. Recent advances in cryo-electron microscopy allow one to identify the Transcription Active Complexes (TAC) and establish a direct relationship between their structure and gene expression (Liu et al., 2017; Bashor et al., 2019). However, most of the experimental techniques follow a reductionist approach, where the studies are performed on the part of a biological system. Therefore, it introduces a challenge in modern biology to integrate molecular-level data and the properties of biological systems.

The structural characterization of TAC using microscopy is sometimes sufficient to relate with the desired functional responses (Buchler et al., 2003; Bintu et al., 2005b; Teif, 2007; Bashor et al., 2019). In molecular biology, the structurally identical TACs can produce different functional responses (Buchler et al., 2003; Bashor et al., 2019). The origin of such variations can be linked with the topology of the biomolecular networks. Therefore, theoretical and computational approaches emerge to understand the

network properties of biomolecular interactions at the system level. Typically, equilibrium thermodynamic models identify the most likely formed biomolecular assembly networks under a few constraints (Buchler et al., 2003; Wong and Gunawardena, 2020; Biddle et al., 2019). The method first calculates the probability of each complex as a function of free energy parameters by considering all possible microstates for a given network topology. We express the functional responses as a function of the average value of the most probable configuration of the topology in a dose-dependent manner.

This Chapter explores how the coordinated interaction among biomolecules forms functional assemblies for the feedback (FL) and feedforward (FFL) network motifs. We further explore the relationship between the network topology of a biomolecular assembly and the Boolean logic responses in detail. In this regard, we developed a statistical thermodynamical model to characterize those configurations. We further show how DNA flexibility, or its chemical modification, can cause a switch in the logic behavior, such as transitioning from AND into OR behavior. We show that the assembly of biomolecules in the gene regulatory network forms simple, functional units in the parameter space. These assemblies perform well-defined computations that finally appear in phenotypic diversity in a biological cell. We discuss them below.

3.2 Models and Methods

3.2.1 Theoretical Background and Simulation Details:

We model the transcriptional regulation of complex gene regulatory networks. We propose a generic thermodynamic model, which considers explicit networks of protein-DNA and protein-protein and protein-ligand interactions. To provide a realistic view of the theoretical models, we validated them against GCMC. As mentioned above, the production of a protein from an active assembly is a tightly regulated process; therefore, the population of the final product is proportional to the relative abundance of the complex assembly. The thermodynamic approach captures such information by considering all possible network configurations for a set of free energy parameters.

3.2.1.1 Thermodynamic Model for Gene Transcription:

The transcription rate may be defined in many instances by the amount of time spent bound to the RNAP bound to the promoter. Here, we quantify the degree of gene transcription by the equilibrium binding probability ($\bar{o}$) of RNAP-promoter binding due to the interactions with bound TFs. For a single promoter, the probability under a grand canonical ensemble can be written as

$$\bar{o} = \frac{\xi_{ON}}{\xi_{ON} + \xi_{OFF}}, \tag{3.1}$$

where the ξ_{ON} and ξ_{OFF} are the partition functions for the TF (P) bound and unbound RNAP (R) microstates. The explicit form of ξ_{ON} and ξ_{OFF} for this simplest case are $\xi_{OFF} = 1 + q_P\lambda_P$ and $\xi_{ON} = q_R\lambda_R(1 + q_P\lambda_T e^{-\epsilon_{RP}/k_BT})$. The subscripts P, R, and RP are for the TF, RNAP and RNAP-TF complex molecules. The strength of interactions between RNAP and TF is denoted by ϵ_{RP}. We define the exponential term ($e^{-\epsilon_{RP}/k_BT}$) as the cooperativity factor, ω_{RP}, that can be tunable to a certain degree by the relative placement of these sites in the regulatory region. In general, the variable ω_{RP} takes values 0, 1, $\omega > 1$ and $\Omega \geq 5\omega$ for the mutual exclusion, independent binding, pairwise cooperativity if they are bound to adjacent sites and cooperativity due to DNA looping. Here, the symbol $\lambda = e^{\frac{\mu}{k_BT}}$ is the absolute activity and $q = \sum_j e^{-\frac{\epsilon_j}{k_BT}}$ is the site partition function, when a TF is bound to the DNA site. Here, each site of the DNA is associated with many energy states, and we denoted them using the j index. The symbols k_BT, T, and μ are the systems' thermal energy, absolute temperature, and chemical potential. The above formalism for the single promoter case can be extended for more complex GRNs containing multiple promoters interacting through TF binding and DNA looping. If there are a total of L TF binding sites in a Core Regulatory Region (CRR), the weight for each microstate of site occupation is still a simple product of $q_{P_i}\lambda_{P_i}$ and $\omega_{P_iP_j}$ values, under the assumption that the TF-TF interaction is glue-like (Buchler et al., 2003). The subscripts, i and j are the labels for various promoter sites. We introduce a binary variable $\sigma \in [0, 1]$ to denote the occupation of each site i. Therefore, we can write the weight for each microstate:

$$W[\sigma_1, \sigma_2....\sigma_i] = \prod_{i=1}^{L} q_{P_i}^{\sigma_i} \lambda_{P_i}^{\sigma_i} \prod_{i<j}^{L} \omega_{P_i P_j}^{\sigma_i, \sigma_j} \tag{3.2}$$

Immediately, we can write the ξ_{ON} and ξ_{OFF} as the sum

$$\xi_{OFF} = \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} .... \sum_{\sigma_i=0,1} W[\sigma_1, \sigma_2....\sigma_i], \tag{3.3}$$

and

$$\xi_{ON} = \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} .... \sum_{\sigma_i=0,1} Q[\sigma_1, \sigma_2....\sigma_i] W[\sigma_1, \sigma_2....\sigma_i] \tag{3.4}$$

The function, Q is the additional weight factor due to the interaction of the RNAP with the bound TFs. Note that an RNAP can bind either simultaneously with multiple TFs or with a single TFs. Taking these facts into consideration, we define Q as follows.

$$Q = q_R \lambda_R \prod_{i=1}^{L} [1 - \sigma_i \delta(\omega_{p,i}, 0)][1 + \omega \sum_{j=1}^{L} \sigma_j \delta(\omega_{p,i}, \omega)] \tag{3.5}$$

The first factor ensures that the RNAP can not bind to a site i if it is bound with a repressor (i.e., with $\omega_{p,i} = 0$). The second factor is the additional weight gained by the cooperative interaction of TFs with RNAP (i.e., with $\omega_{p,i} = \omega$). We apply the above prescription for a wide range of gene regulatory network motifs and obtain the explicit form of the grand canonical partition functions. The grand partition functions are used to obtain the $\bar{o}$, which we convert to fold change (FC). The FC is defined as the ratio between the occupation numbers of the specific complex assembly when RNAP is bound with TFs in the network ($\bar{o}_c$), and the R-DNA complex ($\bar{o}_b$) respectively (Gautam and Kumar Sinha, 2021), i.e., $FC = \frac{\bar{o}_c}{\bar{o}_b}$.

3.2.1.2 GCMC Simulations:

The partition function-based analysis of a protein-DNA interactions network is increasingly complex as the components increase. For example, considering all possible pairwise interactions and a system-level analysis of such biomolecular networks using the above analytical method is a real challenge. Therefore, we perform GCMC simulation that offers a numerical way to analyze such complex networks without an explicit analytical solution. The MC simulation realizes all possible binding and unbinding events as the

Markov process and generates many configurations of the Metropolis-Hastings algorithm (Gautam and Kumar Sinha, 2021). First, we map the Core Regulatory Region (CRR) consisting of L binding sites to a one-dimensional (1-D) lattice. The TFs or RNAPs walk randomly and bind reversibly to their respective binding sites by following the topology of the protein-DNA interaction network. We further consider M identical units for averaging in our simulation. Other technicalities, such as consideration of pairwise interactions among all the biomolecules and periodic boundary conditions (PBC), are also applied in our simulation.

3.2.2 Complex Assembly as Network Motifs:

We apply the above formalism for activation and repression, feedback, and feed-forward loop assemblies. The biomolecular network of interactions of each of the units is presented in Figures 3.1, 3.2, 3.3, and 3.4. To demonstrate those processes, we consider a unit consisting of the promoter regions of regulatory and target genes as denoted by G_X and G_Y. These two genes produce two TFs, namely, TF_X and TF_Y. In the following step, we deduce the grand canonical partition function for M identical units using our proposed scheme, $\Xi(\lambda, M, T) = (\xi_{OFF} + \xi_{ON})^M$ by considering all possible configurations of a biomolecular network for each of the cases. We then calculate the fraction, $[\bar{o_R}]_Y = \frac{\xi_{ON}}{\xi_{OFF} + \xi_{ON}}$ of bound complexes on G_Y from the $\Xi(\lambda, M, T)$ under different competition scenario.

3.2.2.1 Activation and Repression:

In Figure 3.1, we present the biomolecular network of interactions for the simple and induced mechanisms. The network consists of two genes, G_X, a regulatory gene, and G_Y, a target gene. Here, the word activation means the G_X activates G_Y through a network of interactions, as mentioned in the figure. We calculate the conditional population, $(o[(R, \bar{C})_Y | R_X])$ i.e. the population of RNAP (R) or R-mediated complexes (C) on G_Y when another R is bound on G_X at thermal equilibrium. In short, the mechanism for activation can read as the R binds with G_X and produces the TF_Y that again binds with the G_Y along with the binding of R. As the interaction between TF_Y and R is enhanced at the promoter region of G_Y, the G_Y is activated. In contrast to the activation, the word

repression means the G_X represses G_Y through a biomolecular network of interactions, as also presented in the figure. Again, we calculate the conditional population, $(o[R_Y^-|R_X])$ to quantify repression. Here, the binding of R to G_X produces the TF_Y that binds with the G_Y that hinders the binding of R. As a result, the expression of G_Y is suppressed. As shown in the table, we control the whole process by a signaling molecule s_Y which binds with the TF_Y and further controls the gene expression. Binding of s_Y to the TF_Y strengthen the interaction of TF_Y with the DNA and R for activation. Similarly, forming a complex between TF_Y and s_Y enhances the excluded volume interaction that further represses the expression. We again calculate the conditional populations of those specific assemblies for induced activation and repression to quantify the induced responses.
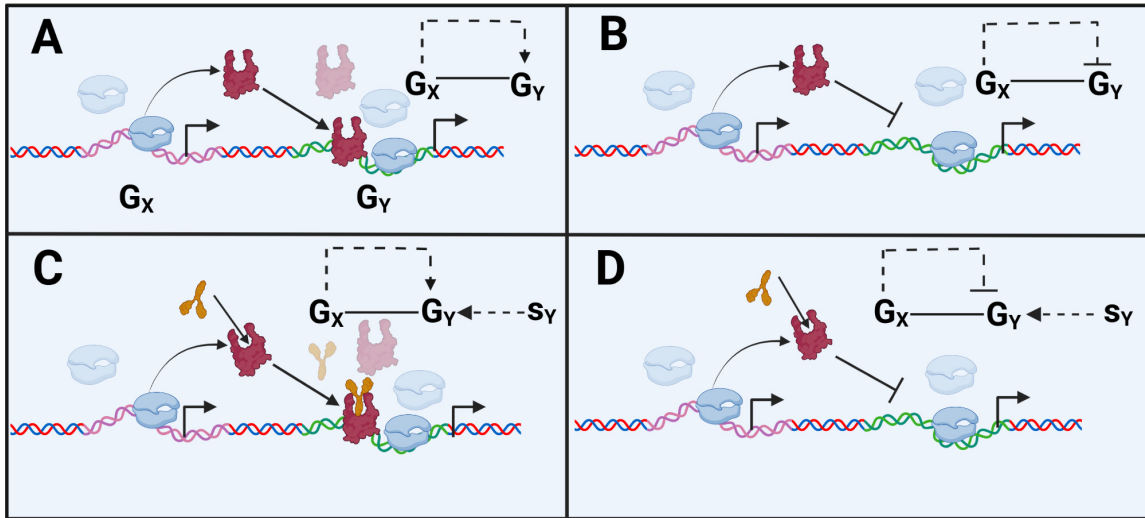
Figure 3.1: Panel A and B represent the complex assemblies for simple activation and repression for two gene systems. Complex assemblies for induced activation and repression are presented in panels C and D. We adopt the same symbolic scheme as shown in Figure 1.2 to represent each biomolecule. The arrow $\rightarrow$ and $\dashv$ are used for activation and repression. The wired network version for each complex assembly is shown in the inset.

3.2.2.2 Feedback Loops:

We further decode the feedback loops (FLs) through our thermodynamic formalism, shown in Figure 3.2. The fundamental units for the biomolecular network of interactions

for various feedback loops are shown in the same figure. Again, we consider two genes, G_X, a regulatory gene, and a target gene, G_Y. Here, both of these genes activate or repress each other simultaneously. Explicitly, we define the feedback loop as the production of the gene products of G_X and G_Y are affected through their products, namely TF_Y and TF_X. In the case of positive feedback loops, the gene G_X is activated by binding TF_X and the gene G_Y is activated by binding TF_Y along with the binding of R. However, the production of TF_X and TF_Y hinders the binding of R to the G_X and G_Y for the negative feedback loop cases. Once again, we calculate conditional population of R and its complexes on G_X and G_Y, $o[(R,C)_Y^-|(R,C)_X]$ and $o[(R,C)_X^-|(R,C)_Y]$, to find the response functions.

Figure 3.2: Panel A, B, C, and D are for the complex assemblies of PFL, NFL1, NFL2, and FNFL feedback loops. We adopt the same symbols as shown in Figure 1.2 to represent each biomolecule. The arrow $\rightarrow$ and $\dashv$ are used for activation and repression. The wired network version for each complex assembly is shown in the inset.

3.2.2.3 Feedforward Loops:

The feed-forward loop (FFL) is another common network motif found in various organisms. The fundamental unit consists of two regulatory genes, G_X and G_Y, and a structural or target gene G_A. The genetic output of G_X is the TF_Y that regulates the genetic output of G_Y, TF_A. Furthermore, the TF_Y and TF_A control the expression of G_A through an interaction with the R at the promoter region of G_A. These TFs can either be an activator or a repressor. Depending on the binding of TF_Y and TF_A to their respective promoter regions, there are eight possible configurations, as shown in Figures 3.3 and 3.4 (Mangan and Alon, 2003; Milo et al., 2002). The regulation of G_A by TF_Y is called the direct path, but the same regulation through TF_A is called the indirect path. Among them, four configurations are coherent, and the other four are incoherent feed-forward loops (Mazal et al., 2018; Milo et al., 2002; Kaplan et al., 2008). Here, the word coherent means that the overall sign of the gene regulation G_A along the direct and the indirect paths is the same. In the case of incoherent feed-forward loops, the overall sign of the regulation of gene G_A along the direct and indirect paths is the opposite. A classic example of incoherent FFL is the *gal operon* in E.Coli bacteria, where the CRP and gal repressor protein, cAMP, and galactose control the expression of the gal gene, i.e., the production of gal_{ETK} enzyme by forming an incoherent FFL among themselves.

3.2.3 Complex Assembly as Logic Gate Operation :

Figure 3.5 shows various configurations for AND, OR, NAND, and NOR logic gates for the Nuclear factor κB system. The origin behind various configurations of complex assemblies in the parameter space relies on the free energy of interactions and the activities of biomolecules. In this model, we aim to control the population of different configurations by varying stimuli activities and DNA looping energies. We define the active and inactive states of TFs depending on whether the stimuli randomly activate TF. We further consider that the binding of active TFs is more potent to bind with DNA than its inactive form. The TFs can access the promoter region by various modes and play a critical role in forming active configurations, thereby modifying gene regulation. An active configuration produces a unit of mRNA, demonstrating that the AND and NOR gates produce 25 % of mRNA, and OR and NAND produce 75 % of mRNA on
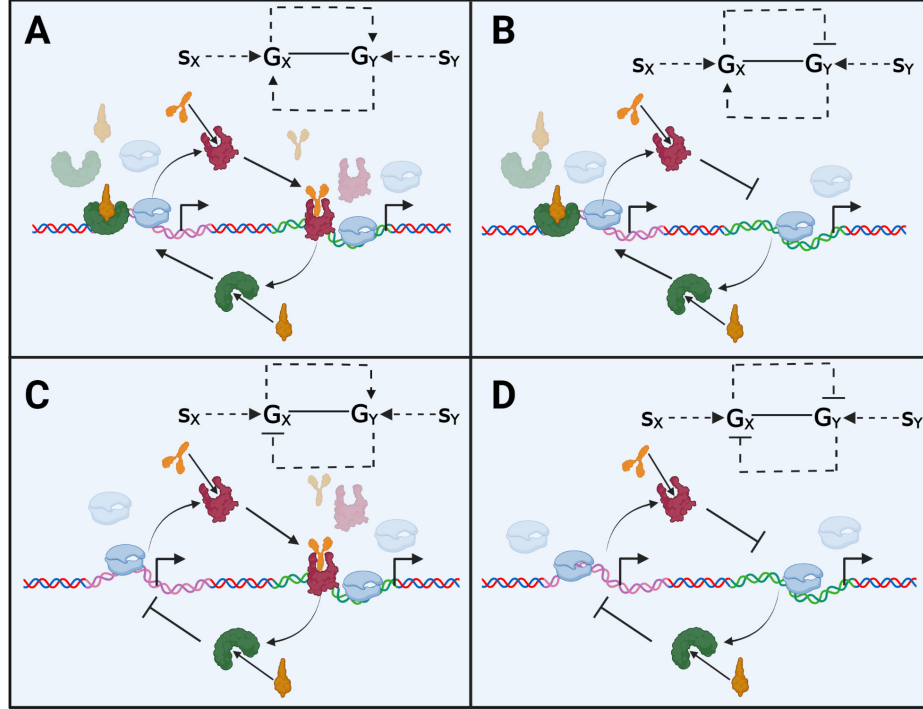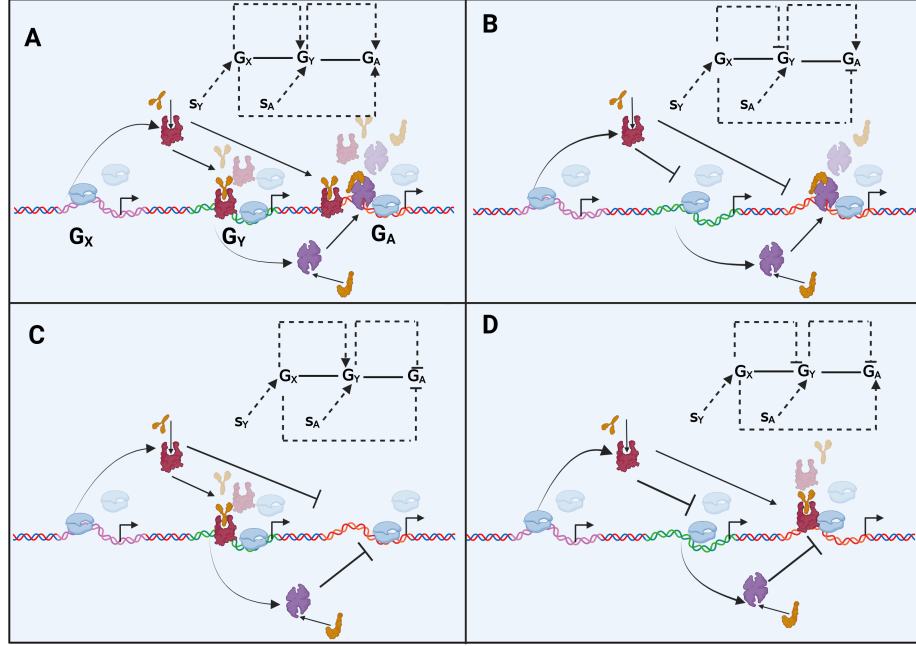
Figure 3.3: Panels A, B, C, and D are for the complex assemblies of types 1, 2, 3, and 4 coherent FFLs. We adopt the same symbols as shown in Figure 1.2 to represent each biomolecule. The arrow → and ⊣ are used for activation and repression. The wired network version for each complex assembly is shown in the inset.

average. In the case of AND and OR gates, the TFs access the promoter by forming loops, whereas the TFs access the promoter region by diffusion for NAND and NOR gates.

The complex assembly forms an active configuration in the presence of external stimuli. The presence of stimuli is controllable; one can achieve various gates by tuning them. The TFs can bind to the enhancer elements and participate in mRNA production when they access the promoter region of the genes. The DNA forms two loops between enhancer and promoter regions for an AND operation through the interaction between TFs and TATA-binding protein. We refer to it as an active configuration since it can produce a unit of mRNA (configurations shown in Figure 3.5). This particular configuration is possible when the DNA is flexible ($\epsilon_{LP} = 0k_BT$) and a saturated level of stimuli (LPS and TNF-α) weakly induce TFs ($\epsilon_{L-TF} = -1k_BT$). One can control the population of active TFs by increasing stimuli activities. The presence of stimuli promotes enhancer-TF interaction, further facilitating DNA loop formation between enhancer and promoter
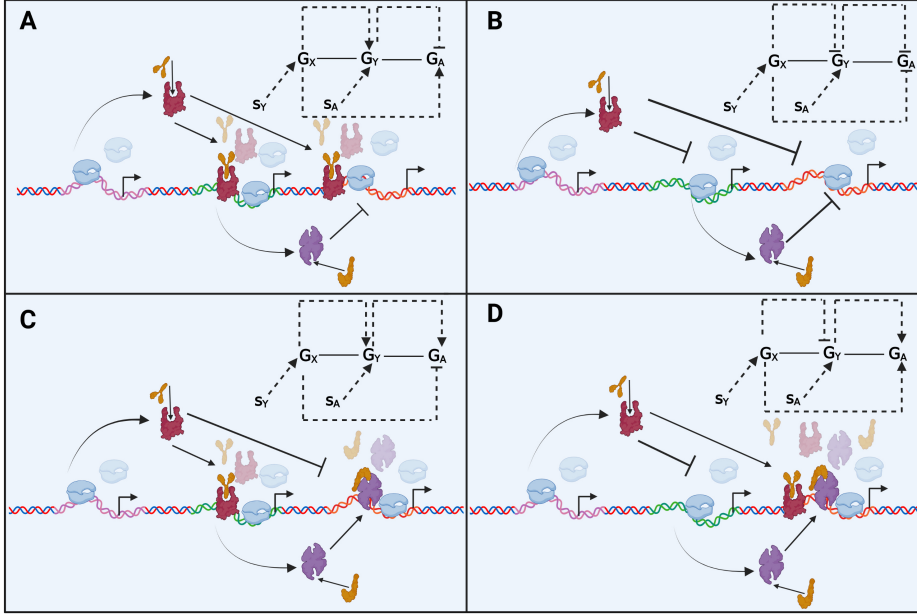
Figure 3.4: Panels A, B, C, and D are for the complex assemblies of types 1, 2, 3, and 4 incoherent FFLs. We adopt the same symbols as shown in Figure 1.2 to represent each biomolecule. The arrow → and ⊣ are used for activation and repression. The wired network version for each complex assembly is shown in the inset.

regions through protein-protein interactions. Therefore, one can realize a unique configuration for AND assembly containing two DNA loops at high values of stimuli activities. Note that the protein-protein cooperative interactions among TFs and TATA-binding proteins become very strong in this case, which rules out the formation of a single DNA loop configuration. Therefore, it promotes only one active configuration with two DNA loops to produce a single mRNA molecule. However, a strong induction of stimuli to TFs ($\epsilon_{L-TF} = -8k_BT$), which have bound to the enhancer elements of flexible DNA, allows to form three unique configurations of assemblies containing a single loop or double loops in the parameter space. Production of such configurations corresponds with the OR-like gates. In the case of the OR gate, we notice three active configurations produce three mRNA molecules.

As we decrease the flexibility of DNA (ϵ_{LP}=12 k_BT), the long-distanced TF-p300 interactions through DNA looping are stopped. Under this situation, only the facilitated tracking or diffusion-like mode allows inactive TFs to reach proximity to the promoter region and control the gene expression. Depending on the strong ($\epsilon_{L-TF} = -11k_BT$)

and weak ($\epsilon_{L-TF} = -3.5k_BT$) stimuli induced activation of TF, we obtain another set of unique configurations in the parameter space upon varying stimuli activities. With a small to moderate stimuli activity, it weakly induces TF that promotes the accessibility of inactive TFs towards the core promoter region and forms three unique configurations of complex assemblies that produce three mRNA molecules. We found this signature for the NAND gate. As stimuli strongly induce TF, the active TFs rarely visit the promoter region. Therefore, the inactive TFs visit the promoter region only at low stimuli activities and form a unique configuration of complex assembly for the NOR gate. However, at the saturated level of stimuli activities, the movement of TFs is entirely restricted for both NAND and NOR gates, and the gene stays almost at the off state that corresponds with a basal mRNA level.

3.3 Results

3.3.1 Thermodynamic Analysis of GRN:

We calculate the population of various complex assemblies on the target genes, as shown in Figures 3.6, 3.7, 3.8, and 3.9. We express them using normalized FC in our analysis. The activity for R molecule is considered fixed, $\lambda_R = 2.9 \times 10^{-4}$ throughout our analysis (Landman et al., 2017; Gautam and Kumar Sinha, 2021).

3.3.1.1 Activation and Repression

The FC obtained from our theory and GCMC simulation for simple and induced processes are presented in Figure 3.6. We find a good agreement between theory and simulation results. As evident from our results (panel A), the FC rises to the saturation point as the free energy of interaction between TF_Y and R_Y, i.e., ϵ_{R_Y-Y} is enhanced. In other words, the TF_Y activates the transcription process by forming a complex consisting of TF_Y, R_Y, and the promoter region of the G_Y. However, the strong interaction between TF_Y and R_Y facilitates the recruitment of R molecules, enhancing such complexes' population. An increase in the population of such complexes enhances the activation, which we observed in our FC analysis. However, the interactions present in a GRN are challenging to control because interactions among biomolecules are specific in a GRN.
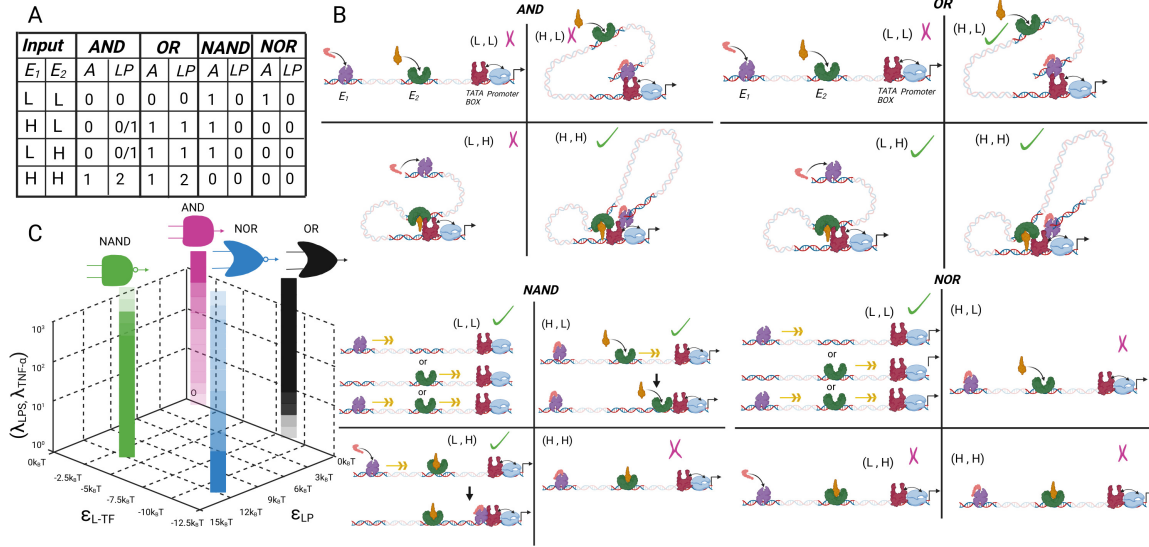
Figure 3.5: Schematic for forming various possible logic gates and their configuration for the NF-κB system. (A) H and L indicate the high and low stimuli activities in the table. The symbols E1 and E2 are the two enhancer elements of the NF-κB system, namely IRE, $\kappa\beta$ sites. The symbol A is the Boolean expression of the corresponding gate operations, and LP is the number of enhancer-promoter loops. (B) Various configurations for AND, OR, NAND, and NOR logical operations. The active configurations are marked with the tick symbols in the figure. Here, we use the purple and green colors cartoon for the protein, IRF, and NF-κB. These two proteins are stimulated by the LPS and $TNF-\alpha$, as shown by the orange and brown color cartoons. The red and light blue color cartoons are used to indicate the heterodimeric complex of p300-AP-1 and RNAP molecules. (C) A schematic view of various logic gates in parameter space. We show the logic gates as a function of a few controllable parameters such as protein-stimuli interaction strengths, $\epsilon_{LPS/TNF-\alpha}$, the strength of DNA loops ϵ_{LP} and the activities of stimuli, λ_{LPS} and $\lambda_{TNF-\alpha}$.

A change in free energy of interactions among these biomolecules may result from their chemical modifications inside a cell (Kim et al., 2008). Nevertheless, we can control that interaction by introducing a signaling molecule, s_Y, that binds specifically with the TF_Y. Furthermore, the s_Y can be chemically modified in the laboratory that binds with altered free energy (ϵ_{s-Y}), changing the FC quantitatively. We show those results in panel B. The activity of TF_Y, $\lambda_{TF_Y} = 1.5 \times 10^{-6}$ kept fixed for this analysis.

In contrast to activation, we find a sharp decline in FC for the repression (panel C of Figure 3.6). Once again, the agreement between theory and simulation is evident from

our analysis. We also find that the strong interactions between TF_Y and G_Y hinders the binding of R with G_Y. Such depletion of R from G_Y is a clear signature of binding competition between R and TF_Y with G_Y, as observed in repression analysis. Since we consider an excluded volume interaction throughout our calculation for repression, the presence of both TF_Y and R on G_Y in a configuration cannot be realized in the counting of microstates. Therefore, the FC decreases sharply and drops to 0 as the magnitude of λ_Y increases. We also find the FC drops down to the 0 value beforehand as the interactions between TF_Y and G_Y are enhanced. Nevertheless, the population of the desired complex during repression is reduced, and such reduction of the population of the complexes is directly proportional to the transcription. We show them in terms of FC in our analysis. Like activation, the s_Y molecule also enhances the repression dose-dependent, as shown in panel D.

The above prototypical model systems for the activation and repression are common in bacteria (Milo et al., 2002). For example, the *lacI* gene acts as the regulatory gene that produces the *lac*-repressor for the *lac* target gene in the *lac*-operon system (Mangan and Alon, 2003; Liu et al., 2017). Similarly, the cyclic adenosine 3',5'-monophosphate receptor protein (CAP) activates the transcription of *E. Coli.* bacteria by forming a complex assembly containing CAP dimer, RNAP, and the specific protein promoter region and de novo synthesized RNA oligonucleotide. Recently, the structure of such an active complex has been identified experimentally and is considered the TAC (Liu et al., 2017). An increase in the population of such TAC in the system enhances the transcription process. In another bacterial system, the *trp* operon, a repressor protein trpR binds to the regulatory region and blocks the binding of RNAP to that region, which finally inhibits the transcription (Kulasiri et al., 2008). However, a small molecule, tryptophan, binds with the repressor, enhances the binding probability of it to the regulatory region, and represses the gene in a dose-dependent manner.

3.3.1.2 Feedback Loops (FL):

FLs are common network motifs in various organisms. For example, the *lac operon* in bacteria have a FL. In this network, the lactose indirectly activates lactose permease expression, forming a positive FL. Also, there exist ara-FGH and xyl-FGH operons in
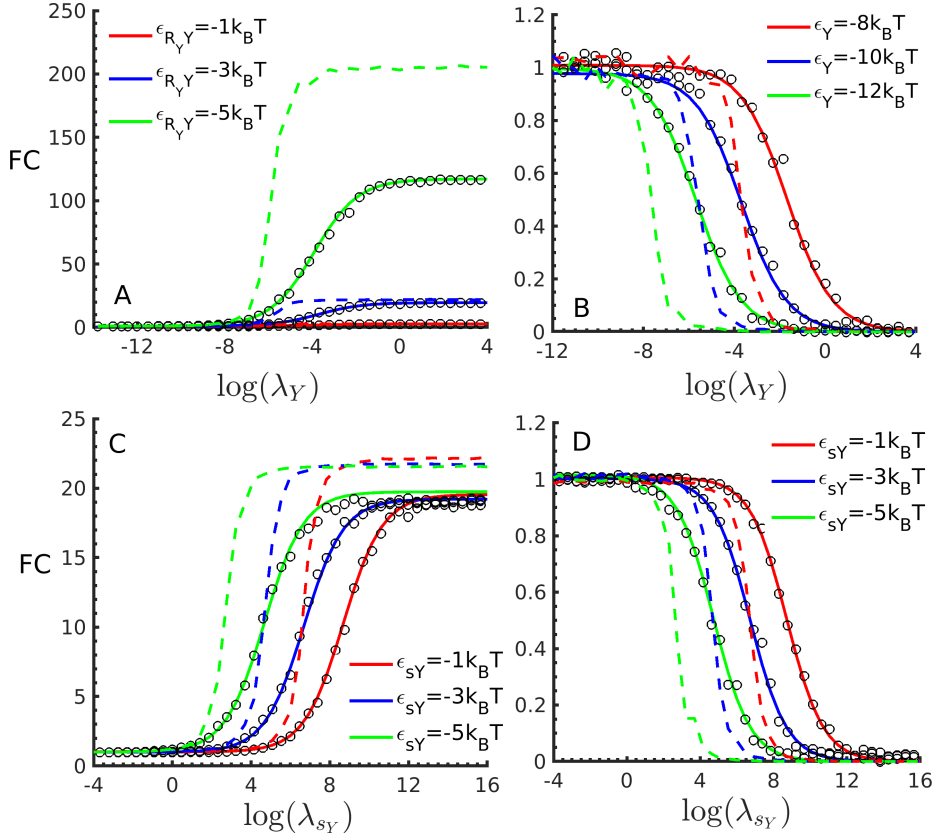
Figure 3.6: Results for the simple activation and repression for two gene systems in the absence and presence of signaling species are shown. For panels A and B, the x-axis indicates the activities of respective TFs. For panels C and D, the x-axis shows the activity of signaling species, and the y-axis indicates FC or the target gene expression for all the panels. Here, solid lines and circles represent the theoretical and simulation results, respectively.

E.coli that also comprises a genetic FL as they contain the gene for arabinose and xylose transporters, respectively (Siegele and Hu, 1997; Ozbudak et al., 2004; Choi et al., 2008). Nevertheless, they play an essential role in the control mechanism and follow a specific biomolecular network of interactions, where the population of two TFs regulates each other transcription in the FL (Gardner et al., 2000). The feedback mechanism emerges from the direct activity of a regulator toward itself or indirectly via downstream products of its target gene (English et al., 2021). We mention a minimal model (See Figure 3.2) that consists of two genes G_X and G_Y. They produce two proteins, namely, TF_Y and TF_X, forming four possible complex assemblies for the feedback loop, namely, a) Positive feedback loop (PFL), b) two Negative feedback loops (NFL), and c) mutually inhibiting

or Fully negative feedback loop (FNFL) in the previous section. We analyzed these minimal models and calculated the population of such complexes on the promoter region of G_X and G_Y. We control the FC by introducing two signaling molecules, s_X and s_Y and present those results as FC as a function of their activity (λ_{s_X} and λ_{s_Y}) in the Figure 3.7. Our results show that tuning the interaction strength between TFs and signaling molecules controls the complexes' population. Further, nearest-neighbor interactions play a vital role in simultaneous activation and repression, which we have shown in our analysis.

Our analysis clearly shows that the binding of TF_Y activates the transcription of G_Y and produces another TF_X that further activates G_X and produces again TF_Y in the PFL. Note that here, both of the TFs are activators. The binding and unbinding of TFs and RNAP to the promoter region of the genes formed this specific network of interactions at thermal equilibrium. We find the FCs on G_X and G_Y increase simultaneously as the activity of the signaling molecules, s_Y and s_X increases. The FC reaches to a saturation point until all R molecules form the complexes with the TFs. We further show that such FC can be modulated by increasing interaction strength between TFs and signaling molecules. This analysis reveals that the cooperative interactions among TFs increase the FC functions on both genes. We consider the same network of interactions to construct the NFL, but one of the TFs must be a repressor. The repressor binds to the gene's promoter region by competing with RNAP or replacing RNAP. As mentioned before, we consider excluded volume interaction for the repressor binding to the DNA binding site. However, we consider TF_Y and TF_X act as activators and repressors or vice versa for constructing the NFL1 and NFL2, respectively. In the case of NFL1, the binding of TF_Y to the promoter region of G_Y inhibits expression of G_Y gene, which is reflected from our FC analysis. As a result, we observed an activation on G_X and repression on G_Y from our analysis. We find an opposite effect for NFL2, where TF_Y and TF_X act as an activator and a repressor, respectively. Considering both of the TFs are repressors, the GRN forms an FNFL. Both repressors reduce the population of R from the promoter regions of G_X and G_Y as both of the TFs are activated. As a result, we find repressions from both genes as we increase the activities of s_Y and s_X.

We chose only a few identical sets of free-energy parameters for the G_X and G_Y

to demonstrate the role of binding free energy in complex biological processes. For this reason, we observe symmetric response functions from both genes. Our results show that the PFL enhances the FC functions significantly and can act as a switching function in a biological system (English et al., 2021). However, we find an asymmetric distribution for the NFL. Both NFLs could improve the robustness and stability as the FC functions simultaneously show activation and repression on both genes (English et al., 2021). The analyses of the population of complexes under the NFL further give a clear indication for producing a uniform signal in response to cellular noise (Kim et al., 2008). However, for FNFL, we observe a narrow distribution of complexes at all values of λ_s, a clear signature of sustained oscillations and homeostasis (Kim et al., 2008). In the case of FNFL, both genes become silent as both repressors are activated. These signatures are quite common and found in various biological systems and used combinatorially to construct synthetic gene circuits (Kim et al., 2008). An example of such coupled positive and negative feedback motif is the galactose uptake network of Saccharomyces cerevisiae of yeast cells (Avendaño et al., 2013). The assembly of this dual feedback is composed of two proteins: a dimeric Gal4P, which acts as an activator, and Mig1P, which acts as a repressor to all the GAL genes. Our method allows one to calculate the population of that assembly at thermal equilibrium and correlate it with the functional response.

3.3.1.3 Feedforward Loops (FFL)

We analyzed a series of model FFL network motifs in this section. We present the biomolecular assembly and the network topology (Mangan and Alon, 2003) that consists of two regulatory genes, G_X and G_Y and a target or structural gene G_A in the previous section. The transcriptional output, TF_Y and TF_A, of both genes, G_X and G_Y that regulate the gene expression of G_A. Both TF_Y and TF_A act either as an activator or a repressor in the network, and they influence the recruitment or replacement of R on the promoter of G_A through a competitive mechanism. Such recruitment or replacement of R happens whether the TF is an activator or a repressor, and the specific free energy of interactions among various biomolecules present in the network determines that fact. Here, we show that the genetic output of the G_A as a function of the activities of signaling molecules, λ_{s_Y} and λ_{s_A} that bind specifically to the TF_Y and TF_A and control the
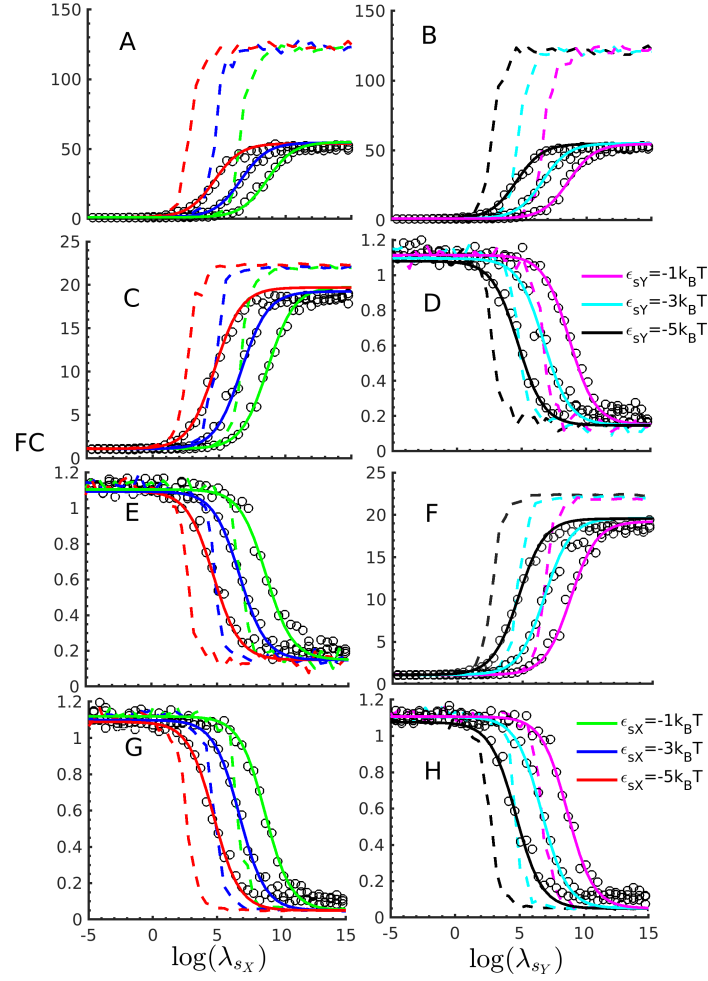
Figure 3.7: This figure contains the results for different cases of the feedback circuit. Panel A shows the results for PFL, panel B and C show the results for NFL1 and NFL2, respectively, and panel D shows the results for the FNFL. Here, the x-axis and y-axis indicate the activity of signaling species and the FC or the target gene expression. The solid lines show the results obtained without considering nearest-neighbor interactions among biomolecules. The dashed lines are used to show the results obtained by considering nearest-neighbor interactions. Circles show simulation results, and the calculations are performed for the different values of ϵ_{sX}.

activation or repression of G_A in a dose-dependent manner. Without the loss of any generality, the transcriptional output of G_A is nothing but the equilibrium probability of binding R molecules at the promoter region of G_A (Buchler et al., 2003). Finally, these network motifs are common and known to exist in *E.Coli* and yeast (Shen-Orr et al., 2002; Milo et al., 2002; Mangan and Alon, 2003).

We show the simulation results for coherent and incoherent FFLs in Figures 3.8 and 3.9. We show that the DNA looping and nearest-neighbor interactions increase the sharpness of the response functions, as evident from our analysis. The presented results for coherent and incoherent FFLs show that the population of R enhances or impedes as the active TFs recruit or replace R through their binding free energy. Therefore, the activity of ligands can modulate the gene expression for FFL networks in both positive and negative manners. Overall, we find that if TF_Y and TF_A behave as an activator, the FC will reach its maximum value at the larger values of activities of ligands. However, if TF_Y and TF_A act as a repressor, then FC comes as the minimum value for smaller values of activities of ligands. The origin of the distinct two-dimensional map is related to the network topology and the nature of TFs found in our analysis.

In the case of coherent FFL, we find a clear correlation between the strength of the input signal and the binding of R at the promoter region of G_A. For the type 1 coherent FFL, we find TAC spans a specific region in the 2-d map at large values of activities of the signaling molecules, λ_{s_Y} and λ_{s_A}. Here, both the TF_Y and TF_A act as an activator for this case. An enhancement in occupancy at the promoter regions will result in more glue-like interactions with the R, thereby recruiting more R for the pre-initiation complex (Buchler et al., 2003). Similarly, we can explain the obtained 2-d maps for the types 2, 3, and 4 coherent FFLs. However, at least one or both TFs are repressors that bind strongly with the promoter region and impede the binding of R via excluded-volume interaction. Therefore, it becomes conceivable that the activity of ligands can modulate the gene expression for FFL networks in both positive and negative manners. Overall, we find that if TF_Y and TF_A behave as an activator, the FC will reach its maximum value at the larger values of activities of ligands. However, if TF_Y and TF_A act as a repressor, then FC comes as the minimum value for smaller values of activities of ligands. The TF_Y acts as a repressor when it binds with the promoter region of G_Y and G_A, whereas TF_A binds as an activator with the promoter region of G_A for the type 2 coherent FFL. In the case of type 3 coherent FFL, the TF_Y binds as an activator for the G_Y, but it acts as a repressor for G_A, and TF_A binds as a repressor for G_A. Similarly, both the TF_Y and TF_A bind as a repressor to the promoter region of G_Y and G_A, but TF_Y binds as an activator and facilitates the recruitment of R at the

Figure 3.8: Two-dimensional response functions (FC) for different coherent feedforward Loops are shown here. Panels A, B, C, and D show the results for types 1, 2, 3, and 4 coherent FFLs. Here, the x-axis and y-axis indicate the activities of signaling species bound to Y and A TF. The color bar along the Z-axis shows the fraction of RNAP that is bound to structural gene A, indicating the mRNA production.

promoter region of G_A for the type 4 coherent FFL. Therefore, the origin of the distinct two-dimensional map is related to the network topology and the nature of TFs found in our analysis.

We also find a similar correlation between input signals and the genetic output employing TAC formation at the promoter region of G_A for the incoherent FFLs. For type 1 incoherent FFL, we notice that the population of TF_Y is enhanced as the values of λ_{s_Y} increase. These activated TF_Ys increase their binding probability with the promoter region of G_Y and G genes. The binding of TF_Y to the promoter region of G_Y produces TF_A, which is a repressor, and undergoes binding with the promoter region of G_A. As the values of λ_{s_A} are increased, the occupancy of TF_A impedes R to bind with the pro-

moter region of G_A through excluded-volume interaction, thereby decreasing the gene expression. Thus, we find a controlled expression at high values of λ_{s_Y} and low values of λ_{s_A} for the type 1 incoherent FFL. We can also similarly explain the output expression for types 2, 3, and 4, which are incoherent FFL. The biomolecular network assembly of the Gal operon system of *E. Coli* bacteria shares a type 1 IFFL network that encodes the amphibolic pathway for the metabolism of D-galactose (Mangan et al., 2006). Immediately, we can map this system with our framework and calculate the population of a complex formed between cAMP and the dimeric CRP that activates and represses two promoters simultaneously.

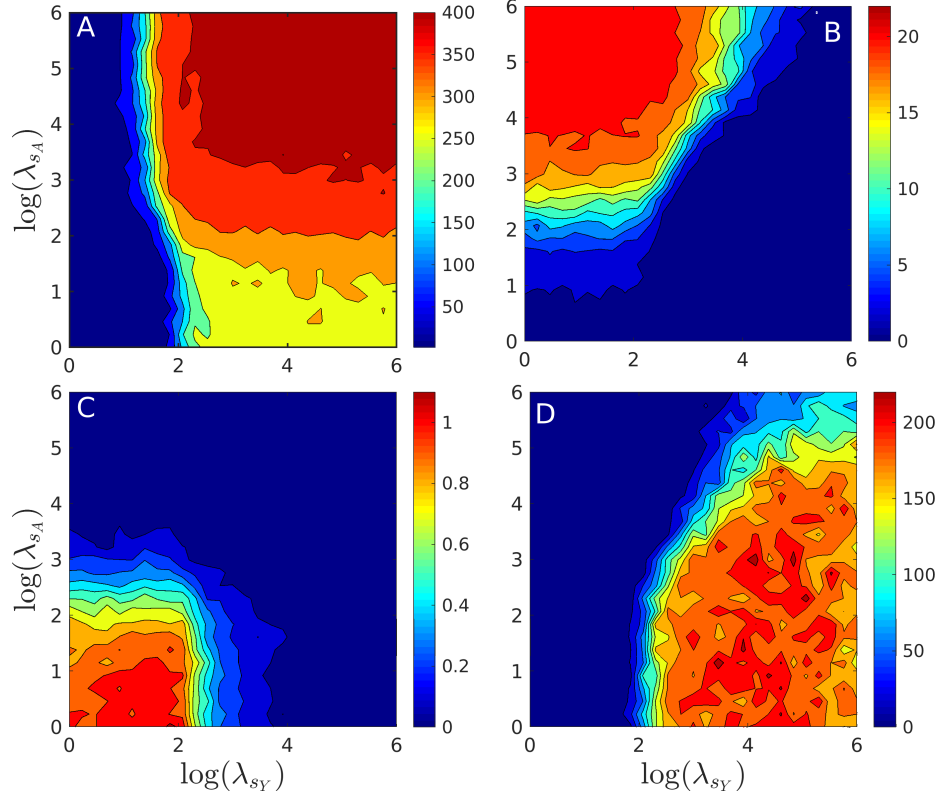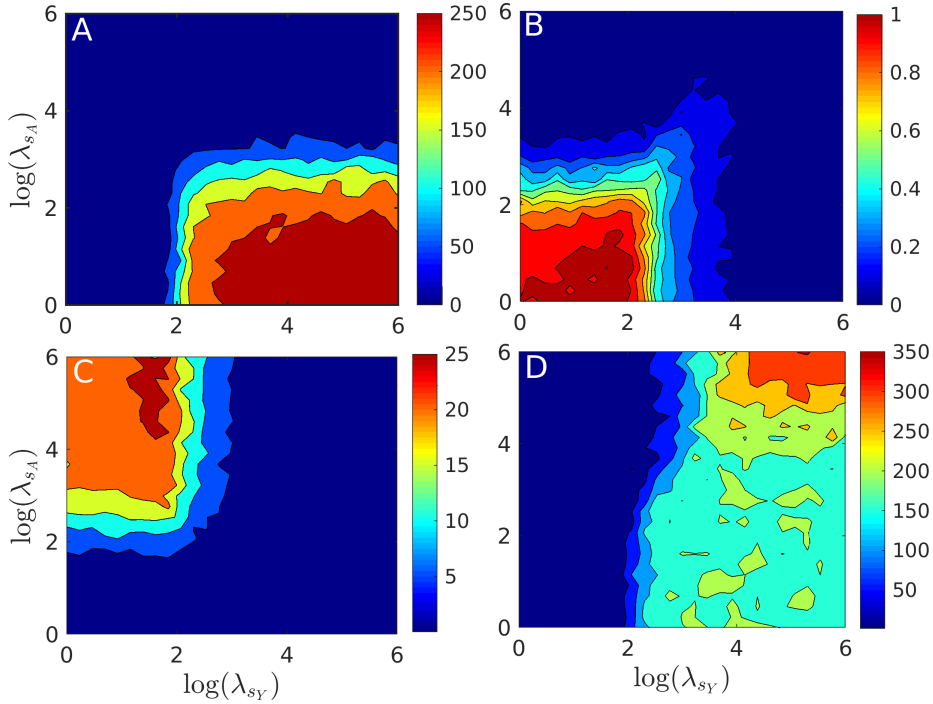Figure 3.9: Two-dimesional response functions (FC) for different incoherent feedforward Loops are shown here. Panels A, B, C, and D show the results for types 1, 2, 3, and 4 incoherent FFLs. Here, the x-axis and y-axis indicate the activities of signaling species bound to Y and A TF. The color bar along the Z-axis shows the fraction of RNAP that is bound to the structural gene A, indicating the mRNA production.

3.3.2 Galactose Utilization Pathway in Yeast Cell:

Here, we apply our proposed modeling scheme to explore the galactose uptake pathway in *Yeast* cell (Venturelli et al., 2012; Avendaño et al., 2013). A collection of genes that follows a galactose-responsive pathway (GAL) for deciding whether the galactose will be metabolized in the presence or absence of other sugars such as glucose (Douglas and Hawthorne, 1966; Johnston, 1987; Escalante-Chong et al., 2015). The biomolecular assembly network for this system forms a dual feedback loop consisting of positive and negative feedback loops. Here, a dimeric protein, Gal4p, acts as an activator, and Mig1p represses all the GAL genes. The cell prefers to uptake glucose from a mixture of sugars, followed by less-preferred carbon sources, a phenomenon commonly known as diauxic growth (Monod, 1949). A common consensus in this problem is that if glucose is available in the environment, the expression of genes associated with the metabolism of alternative sugars is switched off via catabolite expression (Gancedo, 1998). However, a recent study suggests that regulating GAL genes does not solely depend upon glucose depletion; instead, they respond to the concentration ratio between galactose and glucose, a phenomenon named ratio-sensing (Escalante-Chong et al., 2015). Thus, this galactose: glucose ratio reaches a threshold value, determining whether the gene will turn on or off. However, the mechanism that governs this ratio-sensing is quite complex and yet needs to be clarified, which requires detailed knowledge of the biomolecular network that links the regulation and fine-tuning of cellular decisions for yeast cells.

The figure shows that the activities of glucose and galactose control the expression of the GAL1 gene. We further notice that the GAL1 gene is not turning off because of the presence of glucose; instead, the ratio of galactose: glucose decides the cell fate by choosing the suitable sugar and thereby sets a threshold limit known as ratio-sensing as shown in Figure 3.10. It is clear from that figure that glucose alone does not control the GAL1 gene; instead, a combination of galactose and glucose is required to turn it on or off. Specifically, we find that the GAL1 is turned off at large values of glucose activities, but the gene becomes active at large values of galactose. We further notice that our modeling scheme captures the diauxic growth, i.e., the GAL1 gene becomes active as both activities, λ_{Glu} and λ_{Gal}, are at large values. Therefore, our modeling scheme successfully captures the complex mechanism of ratio-sensing for the GAL genes

in yeast cells.
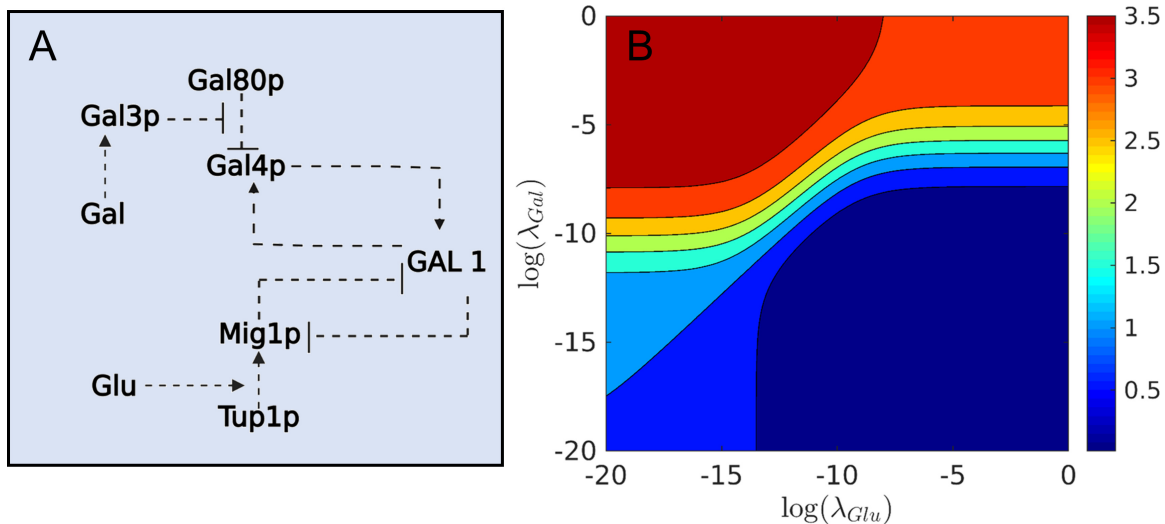
Figure 3.10: Schematic of the yeast GAL signaling pathway. The dual feedback mechanism of this signaling pathway is shown in panel A). In panel B, we show the expression of the GAL1 gene in response to glucose and galactose. The diauxic effect in the presence of glucose: galactose is explored.

3.3.3 Logical Decisions of NF-κB Signaling System

In synthetic biology, various digital circuits can be fabricated by exploiting biomolecular self-assembly (Qiu and Chiechi, 2022; Kimchi et al., 2020; Chen et al., 2020; Yin et al., 2008) as the basis of molecular-level computations. Such studies suggest that self-assembly can be harnessed as a unique property to achieve the desired manipulation for biological function. Moreover, it has been recognized that programmable DNA loops play a crucial characteristic in designing digital circuits. One can employ the CRISPR-based DNA looping method (Hao et al., 2017) to customize or manipulate DNA loops. This method offers massive flexibility for DNA manipulation for various cell types, refines the knowledge of various loops, and helps design flexible digital circuits. Drawing from such studies, we explored a biological system for NF-κB as a building block and how self-assembly and programmable DNA loops lead to biocomputing machines. In Figure 3.11, we also show how the switching among various gates is regulated by stimuli-TF interaction and DNA looping energies that generate various self-assembled

species. Thus, we explore the origin of such modularity of the self-assembly and the cooperativity among regulatory components, coupled with the ability to design various protein complexes, enabling the design of sophisticated transcriptional logic furnishing a vast range of biological functions.
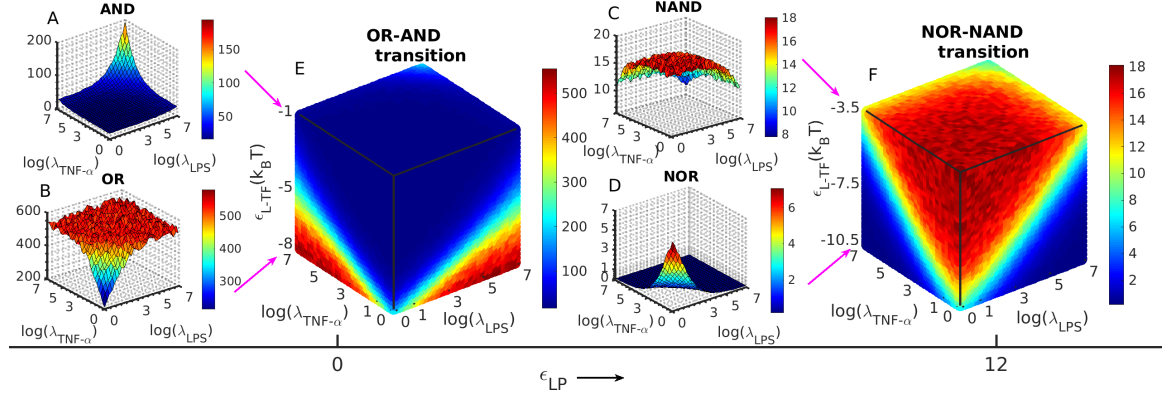
Figure 3.11: Evolution of the various logic gates in the parameter space. (A) Panel A, B, C, and D refer to the population of the active assemblies for the AND, OR, NAND, and NOR logic gates as a function of stimuli activities ($\lambda_{TNF-\alpha}$ and λ_{LPS}). (B) Panel E and F are the logic gates switching between AND to OR and NAND to NOR. Note that the switching only happens as the strength of interaction between stimuli and TF (ϵ_{L-TF}) varies. The switching from OR-AND to NOR-NAND happens through the variation of DNA stiffness parameter (ϵ_{LP}) as a function of stimuli activities.

Figure 3.11 (panel A, B, C, D) presents various logic gate results obtained from partition function calculations and MC simulations in the parameter space. Here, we show that the formation of different logic gate configurations evolves upon variation of free energy parameters and stimuli activities. The weak activation of TF limits the population of active TFs, but a significant population of active TFs on enhancers is observed at high stimuli activities. The binding of the active TFs to the enhancer elements does not necessarily mean that the configuration is active unless they reach the promoter region and alter the mRNA production. As discussed above, various mechanisms can reach the promoter region: a) by long-distance protein-protein interaction through DNA looping and b) by diffusion-like mode. In the case of AND configuration, binding of active NF-κB and IRF to the enhancer elements promotes the form of DNA loops between enhancer and promoter through a cooperative interaction among the bound NF-κB, IRF, and

p300 molecules. Note that an active TF binding to the promoter rarely happens; therefore, the long-distance interaction through DNA looping and cooperative interaction among protein molecules are essential for an active AND assembly, as revealed from our study. Since TFs access the promoter region by this mechanism, they become an active assembly because they can produce mRNA or participate in gene regulation. Therefore, we find a narrow region in the parameter space where the active complex assemblies of AND-like configurations are observed when a large quantity of stimuli weakly induces TF. Overall, two DNA loops formed at high stimuli activity values and weak stimuli induced TF activation characterizes AND assembly. Similarly, the origin of the OR-like gate's configurations relies upon tuning the free energy of activation between the TF and stimuli. As we increase the free energy of activation between TF and stimuli, the strong interaction between them allows TFs to reach proximity to the promoter region by single or double DNA loops. Therefore, we obtain a broad region of responses as both stimuli' activities increase.

In contrast to the above two gates, the NAND and NOR gates rely on the stiffness of DNA. We increased the stiffness of DNA by increasing the elastic parameter for the DNA chain and observed complementary AND and OR logical responses. We show that binding TFs to the enhancer elements allows only their translation along with DNA. Such translation motion of TFs along DNA is crucial since it allows them to access the promoter region for activation or repression of the gene regulation. As mentioned in the previous sections, the nature of binding between TFs and DNA creates NAND and NOR complexes in the configurational space. We find that the weak activation of TFs and moderate levels of their activities produce NAND-like gates. However, one can control the accessibility of TFs to the promoter region of DNA by tuning the activation free energy between TF and stimuli. As we increase the free energy of activation, the TFs rarely visit the promoter region. As a result, we find the NOR responses only at high values of stimuli activities. Further, it is evident from our analysis that the movement of TFs is restricted at very high values of stimuli activities that appear as inactive NAND and NOR-like responses. We have observed fair consistency between our theoretical and simulation results for the NF-κB signaling system.

It is clear from the above analysis that the diverse range of logical computations

by the interactions between protein and DNA through the formation of specific complexes is quite possible in the parameter space. We show a specific complex responsible for a specific logical response in a narrow range of parameters. Such parameter variations are common in cellular biology as the activities of biomolecules change through many biological processes such as cooperative binding, post-translational modifications, oligomerization, etc. Therefore, the observed output patterns for various logic expressions switch among themselves because of the existence of such parameter variations in cellular systems. We discuss them below.

The reason behind switching between OR to AND-like logical responses is the formation of DNA loops that vary for AND and OR assemblies. The formation of two DNA loops promotes cooperative interactions among proteins in the locally formed complexes on the DNA. On the other hand, the OR-like gate requires either or both TFs to interact with the promoter to form the DNA loops. It happens because of the strong stimuli-induced TF activation, which increases the occupancy of active TFs to enhancers even at low values of stimuli activities. As the population of active TFs increases at their low values, the probability of the formation of DNA loops increases in the absence of TF-TF cooperative interaction. Our analysis revealed the importance of cooperative interaction, which is crucial for an AND-like response. However, forming an OR-like gate requires no cooperative effect through the TF-TF interaction. As a result, either single or double DNA loops at low stimuli values activate the promoter region and provide an OR-like response.

The NOR-NAND is also observed upon variation of ϵ_{L-TF}. In this case, we first set the high DNA looping energy (ϵ_{LP}). Then we performed a continuous variation of ϵ_{L-TF}, λ_{LPS} and $\lambda_{TNF-\alpha}$ on the active complex. The high values of ϵ_{L-TF} disfavor the long-distance interaction through DNA looping; TFs only access the promoter region by translation mode along the DNA. However, the movement of TFs is controlled by the binding between TF and DNA: the enhancement of interactions produces NOR, and its suppression produces NAND logical responses. Therefore, the origin of the NOR-NAND switching again lies in the variation of the ϵ_{L-TF} parameter. Modulation of interaction between stimuli and TF controls the movement of the TFs on the DNA, which is the origin of the NOR and NAND logical responses. As we increase the λ_{LPS} and $\lambda_{TNF-\alpha}$,

the population of activated TFs is enhanced. Such enhancement of the population of TFs increases the TF-DNA interactions that stop their accessibility to the promoter region of the regulatory motif. As demonstrated earlier, we find the NAND response with a low to moderate level of stimuli activities. In contrast, the NOR response is observed only at low values of λ_{LPS} and $\lambda_{TNF-\alpha}$. We find a narrow red region for the NOR response spreads over and transforms to NAND response upon variation of ϵ_{L-TF} and the stimuli activities.

3.3.3.1 D_{KL} Analysis:

We characterize the switching between two logical responses by calculating the D_{KL} function to explore the behavior space for the complete set of available configurations of assemblies. It measures the similarity between the results obtained from MC simulations and the theoretical logical functions obtained from our partition function calculation. Plotting D_{KL} as a scatter revealed that AND and OR Boolean-like computations are contained in flexible DNA, whereas rigid DNA can compute NAND and NOR responses. We find a divergence region for both the $AND-OR$ and $NAND-NOR$ switching. It is a clear sign of the inter-conversion between the OR and AND, like logical responses, which are detectable in the parameter space. The signature of wide divergence for the $AND-OR$ switching suggests that they are distinguishable, and the exclusive AND and OR-like logical responses are detectable in the behavior space from our analysis. The conversion between $NAND$ and NOR switching is less detectable, as found from the D_{KL} analysis, since the divergence for the NOR-NAND switching is narrow. We further vary the degree of oligomerization of NF-κB to achieve robustness of gate switching. We find that the oligomerization of NF-κB does not enhance switching robustness; their monomers show robust switchings from AND to OR or NAND to NOR. We find from the analysis that the formation of the higher-order oligomers perturbs the logic gate operations, a signature that moves away from the precise computation.
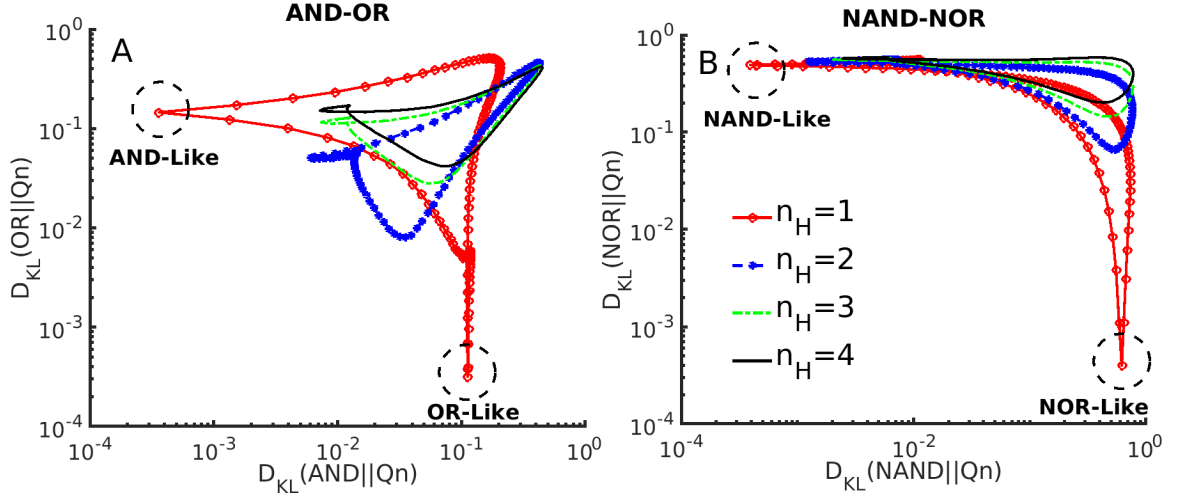
Figure 3.12: The behavior space for the complete set of available assembly configurations is plotted as K-L divergence (D_{KL}): similarity between theoretical model computed output surfaces and the Boolean surfaces obtained from Monte Carlo simulations. The regions are marked for the exclusive AND, OR, NAND, and NOR logic gates. The calculations are done by varying degrees of oligomerization (n_H) to explore the robustness of the switching among gates.

3.4 Conclusions

This work employs theoretical and computational methods to characterize the cellular networks formed by self-assembling proteins and RNAPs on DNA. We show that the functional responses of a GRN rely on the network topology of an assembly. We apply statistical thermodynamics to estimate the population of a specific assembly. One can characterize those assemblies using the cryo-EM technique as demonstrated in recent studies (Liu et al., 2017; Bashor et al., 2019). One key advantage of the thermodynamic model is that it considers only a few free energy interaction parameters, allowing us to predict whether the desired complex will form at thermodynamic equilibrium. Overall, these networks can respond to information encoded in the binding of signaling molecules to the TFs. In particular, we characterized some of the most common GRN motifs like activation, repression, FL, and FFL using our theoretical and simulation methods (Buchler et al., 2003; Mangan and Alon, 2003; Freire-Rios et al., 2020; Bashor et al., 2019; Kaplan et al., 2008). We validate our model by revisiting the published results for the gene expression of the Gal promoter system of yeast cells (Escalante-Chong et al.,

2015).

We further establish the link between complex assembly and network motifs for the information processing of GRNs. We consider each GRN motif composed of unique genetic architecture and a specific set of protein-DNA, protein-protein, and protein-ligand interactions (Buchler et al., 2003; Bashor et al., 2019). We assume that the TFs and RNAPs assemble at the gene promoter regions and form a functional complex at thermodynamic equilibrium. We show that they produce precise functional responses in the presence of molecular noise. We calculate those response functions using a specific complex assembly's conditional occupation number at the various promoter regions. Our calculated FC for various GRN motifs qualitatively agree with the reported experimental data (Avendaño et al., 2013; Shen-Orr et al., 2002; Milo et al., 2002; Mangan and Alon, 2003). In the present analysis, we considered GRNs for the activation, repression, feedback, and feed-forward loops. We chose these networks because of the large class of experimental data available for them (Aviziotis et al., 2015; Siegele and Hu, 1997; Jenkins and Macauley, 2017; Nevozhay et al., 2009). We control activation and repression by increasing the free energy of interactions between RNAP and TFs. In the case of the feedback loop, we propose a network of assembly that shows functional responses with significant variations. The effect of network topology on functional response is crucial for higher-order complex assemblies (Buchler et al., 2003; Liu et al., 2017; Bashor et al., 2019). Incorporating other biophysical effects such as DNA looping, TF-ligand interactions, and nearest-neighboring interaction further alters the shape of response functions, as reported extensively in many literatures (Buchler et al., 2003; Vilar and Leibler, 2003). Depending on the nature of TFs and interaction parameters, the RNAPs are recruited or replaced on the DNA to form a functional assembly, and their population determines the functional responses (Bintu et al., 2005b; Phillips, 2015; Crews and Pearson, 2009). As evident, alteration of the free energy of interaction parameters perturbs the assembly and further disrupts the information processing in the GRN motifs. The proposed technique may help to predict the response functions for partially known or unknown GRN motifs.

Using our theoretical calculations, we also demonstrate the possibility of creating Boolean logic in the NF-κB system. We explored the building blocks of such logic op-

erations by considering the programmed DNA loop and the NF-κB-stimuli interactions. We show that various active self-assemblies formed under two input conditions. In our calculation, we define an active assembly when a TF interacts with the promoter. DNA looping or the translational movement of TFs are a few such mechanisms that move a TF from a nonspecific binding region to the promoter region on DNA for forming active configurations. Each of these active configurations produces a unit of mRNA molecule. We manipulate the stiffness of DNA that allows us to create programmable DNA loops, a crucial factor for AND, OR, NAND, and NOR Boolean operations. Since we can control DNA flexibility externally, gates are interconvertible in the parameter space. Integration of such logic gates may offer high-level biomolecular computation in a cellular system. They have the potential to identify and analyze disease-related genes (Benenson et al., 2004).

The work demonstrates that biomolecular self-assemblies have the potential to capture digital information in the form of mRNA molecules. Our findings demonstrate that individual cells can execute molecular arithmetic functions using modulated self-assembly. This feature has been demonstrated by Bashor *et. al.*, showing how a complex signal is processed in synthetic gene circuits using cooperative regulatory assemblies (Bashor et al., 2019). Many studies have shown that mRNA-based logic operations can detect disease indicators, including mRNA of genes linked with lung cancer and prostate cancer (Benenson, 2009; Leisner et al., 2010; Benenson et al., 2004; Gil et al., 2011). Here, we describe it for NF-κB as a building block and how self-assembly and programmable DNA loop lead to forming a typical biological microprocessor. This device takes stimuli as input information and then rewires the gene regulatory networks through a modulated self-assembly that produces the Boolean output as the population of mRNA.

Stochastic dynamics of gene regulatory networks driven by intrinsic molecular noise

4.1 Introduction

In the previous chapter, we have demonstrated that proteins can perform multiple tasks in a cell by interacting with a short DNA sequence in the genome. The final result appears as a gene expression by which the cell executes various operations ranging from signaling to immune response. In this chapter, we explore the dynamics of gene regulatory networks driven by intrinsic molecular noise. It has been reported previously that gene expression is a tightly regulated process and inherently stochastic (Ko, 1991; Elowitz et al., 2002; Golding et al., 2005). Many modeling schemes are developed in this direction to account for stochasticity in gene expression. Most systems experience two types of noises at the cellular level, one intimately related to the random molecular events governed by the elementary reactions (internal noise), and the other one is the external noise governed by the changes in environmental conditions (Van Kampen, 1992). Cellular systems experience both noises naturally, but their understanding using experiments is quite challenging (Gebhardt et al., 2013; Bratsun et al., 2005). Single-molecule imaging is a powerful tool that captures such signals from a noisy environment in the live cell, but their theoretical modeling remains a difficult problem (Gebhardt et al., 2013). In this regard, stochastic simulation algorithms (SSA) provide a brute-force method to model elementary realized reactions for gene regulation (Gillespie, 1977). The modeling schemes are system-specific, and often, they experience a broad spectrum of noises ranging from delta function correlated white noise to the correlated colored noise (Gardiner,

2009).

The mass action kinetics-based deterministic models offer to understand the existence of alternative steady states of a biochemical network (Alon, 2019). These dynamical models are typically nonlinear. Therefore, their analytical treatments work well for a small network containing few molecular components. However, obtaining a closed-form analytical solution is challenging for a complex gene regulatory system owing to its nonlinearity. The bifurcation theory can help to study the stability of dynamic nonlinear systems (Strogatz, 2018). However, one can also develop stochastic models of this network by incorporating external or internal noises. The noise can drive transitions from one steady state to another alternative state, a feature absent in deterministic systems. These stochastic models have broad applicability, from ecological to gene regulatory systems operating at diverse timescales (Black and McKane, 2012; Tian and Burrage, 2006). Since the system experiences broad-spectrum noises, one must carefully incorporate them for appropriate modeling. The noise sources are typically from various stochastic events linked with gene expression, cellular responses to external stimuli, or immune responses.

Generally, two approaches are there for the stochastic time evolution of these networks. These are (a) solving the coupled stochastic differential equations (extrinsic) and (b) solving a chemical master equation whose joint distribution is governed by a set of elementary reactions of the network (intrinsic) (Kaern et al., 2005; Satija and Shalek, 2014). The analytical solutions of both of these models are challenging. Therefore, stochastic simulations have been widely used for their analyses. We employ Kinetic Monte Carlo (KMC) simulations by following the Gillespie algorithm that samples the Markov process as realized by a set of elementary reactions (Gillespie, 1976). Both of these models do not require as many assumptions, but the dimensions and the number of parameters increase as the complexity of the network increases (Bailey, 2001; Brown et al., 2004). In this regard, suitable approximations such as the time scale separation can reduce their dimension (Golikeri and Luss, 1974; Holehouse et al., 2020). Another serious difficulty arises while incorporating the correlated noise into the stochastic models. The existence of correlated noise is common in the biochemical network (Sompolinsky et al., 2001; Liu et al., 2001; Holehouse et al., 2020). In the case of correlated noise,

the system takes information from its past for its dynamic evolution. The governing chemical master equation for dynamic evolution becomes non-Markovian (Zhang and Zhou, 2019). One can adopt unified color noise approximation (UCNA) to explore their dynamics analytically (Jung and Hänggi, 1987; Holehouse et al., 2020). Of particular interest, we follow work by (Spanio et al., 2017), where they studied the population dynamics in which birth and death rates fluctuate over time. However, the KMC simulations are alternative approaches for taking into account correlated extrinsic or intrinsic noises (Spanio et al., 2017; Cao and Grima, 2018; Sarkar et al., 2019).

In this chapter, we study the dynamics of network motifs in the presence of noise. In particular, we propose the master equation for each protein-DNA assembly for their dynamic evolution. We characterize their dynamical feature using stochastic simulation and obtain a signature of multistability in a higher-order assembly. We further explore how a time-varying rate parameter impacts protein production. We incorporate the time variation into the rate parameter for protein degradation as the newly synthesized protein has a finite lifetime. To analyze this model, we use the UCNA, which is valid in the limits of very short-range correlation (Jung and Hänggi, 1987; Hänggi, 1994) and explore the validation in the limit of white-noise approximation as obtained from the KMC simulations using the modified Gillespie algorithm (Gibson and Bruck, 2000). We show that our proposed modeling framework could be employed for studying the dynamics of such complex systems with colored noise. Specifically, we aim to develop a general theoretical framework that one can employ to obtain the analytical form for the steady state distribution of produced protein in a network motif (Alon, 2019), and to observe the effect of the time-varying rate parameter that introduces the correlated noise. Furthermore, we explore the quantitative distinction between typical and super-enhancers mediated gene expression.

4.2 Models and methods

4.2.1 Deterministic model for Gene Transcription:

We propose the dynamic model for each of the networks using mass action kinetics, where the state variables σ_i's undergoes a transition from σ to $(1 - \sigma)$ with a reaction

rate $r_i = k_{on}(1-\sigma_i) + k_{off}\sigma_i$. The k_{on} and k_{off} are two rate constants for the association and dissociation processes. The edges of the network are an elementary reaction and propose their deterministic kinetic equations, which can be written as.

$$\frac{d\mathbf{x}}{dt} = SF(\mathbf{x}), \tag{4.1}$$

where S is the stoichiometry matrix, $\mathbf{x}$ is a vector whose length is equal to the number of components present in the network, $F(\mathbf{x})$ is the vector of rate laws. The explicit form of the equations for each network is presented in the table in Appendix 1. In general, we consider four types of elementary reactions: a) basal expression that happens upon binding of only RNAP, b) activation of promoter state that happens upon forming a complex with TFs, c) controlled gene expression mediated through the interactions between RNAP and TF, and d) degradation of proteins in our modeling scheme. We also describe that the number of reactions increases as the components and complexity of the networks increase. Based on the table containing biomolecular reactions, one can write a set of above-coupled differential equations containing kinetic parameters. These coupled equations can be solved directly and obtain the deterministic dynamics of each network. The state of the system changes depending on kinetic parameters but remains unchanged unless any molecular perturbation is applied to the network. We took the parameter values from the literature for kinetic calculations (Cao and Grima, 2018, 2020).

Immediately, we apply our proposed theoretical scheme to a gene regulatory system that exploits protein assembly on DNA for a specific function. In this regard, we chose the NF-κB signaling system to apply our proposed scheme. Here, it follows the gene activation/deactivation mechanism that encodes IκBα mediated via NF-κB. Experimental techniques have confirmed that the NF-κB forms dimer and then associate with IκBα protein to form hetero dimeric IκBα-NF-κB complexes that participate for signal transduction (Ngo et al., 2020; Hayden and Ghosh, 2014). The GRN exhibits a negative feedback loop and a delayed degradation of IκBα protein (Lee and Schiemann, 2011; Zambrano et al., 2016). Protein degradation occurs through a complex proteolytic pathway, which may be assumed as delay degradation (Krishna et al., 2006; Mather et al., 2009). Note that negative feedback in the regulatory circuit dampens the effect

of noise. We use this simple model as a guideline here, but it maintains a high degree of biological relevance for gene regulation in other similar systems of interest. The dynamics of this system were analyzed deterministically as well as stochastically. In the modeling process, we write the biochemical reaction scheme for this system below.

$$G \xrightarrow{k_1 f_A(NF-\kappa B)} G^*, \quad G^* \xrightarrow{k_2 f_R(I\kappa B\alpha)} G, \tag{4.2}$$

$$G^* \xrightarrow{k_P} G^* + I\kappa B\alpha, \quad I\kappa B\alpha \xrightarrow{k_d} \phi \tag{4.3}$$
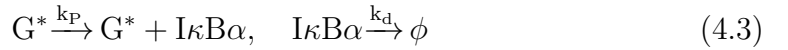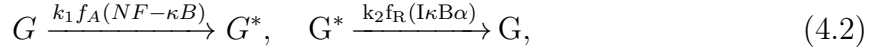
where G and G^* are the genes in inactive and active states encoding IκBα protein; k_1, k_2, k_P and k_d are rate constants of those reactions. Our analysis refers to NF-κB and IκBα as N and I, respectively. The deterministic rate equations for this model system are given by

$$\frac{dG^*}{dt} = k_1 f_A(N)G - k_2 f_R(I)G^*, \text{and} \tag{4.4}$$

$$\frac{dI}{dt} = k_P G^* - k_d I \tag{4.5}$$

Here, $G + G^* = 1$, $f_A(N) = \frac{(N)^{n_{H_1}}}{K_{m_1}^{n_{H_1}} + (N)^{n_{H_1}}}$ and $f_R(I) = \frac{K_{m_2}^{n_{H_2}}}{K_{m_2}^{n_{H_2}} + (I)^{n_{H_2}}}$, where K_{m_1}, K_{m_2} stand for the binding affinity of NF-κB and IκBα and n_{H_1} and n_{H2} are the Hill's coefficients given for the binding of NF-κB and IκBα to their respective enhancer regions. One can realize the degree of the binding of these two proteins to the enhancer regions embedded in n_{H_1} and n_{H2}, which, in other words, considers the cooperative effect. To carry out our analysis, we perform the linear stability analysis by means of drawing a bifurcation diagram over a range of parameter values. Under quasi-steady state approximation (QSSA), the corresponding deterministic rate equation for mean IκBα protein number (I) reduces to the form:

$$\frac{dI}{dt} = \frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)} - k_d I \tag{4.6}$$

4.2.2 Stochastic Approach

4.2.2.1 Markovian Dynamics

The analysis based on the above kinetic equations is limited in its ability to characterize probabilistic events, as they do not capture multi-modality in gene expression that arises from slow promoter binding (Vellela and Qian, 2009). Therefore, one can consider the birth and death processes of the networks' elementary reactions. Generally, it is described by the time evolution of a grand probability function governed by a chemical master equation.

$$\frac{dp(\mathbf{x_0}, t_0, \mathbf{x}, t)}{dt} = \sum_{i=1}^{m} \Big[h_i(\mathbf{x} - S^{(i)}, c_i) p(\mathbf{x_0}, t_0, \mathbf{x} - S^{(i)}, t) - h_i(\mathbf{x}, c_i) p(\mathbf{x_0}, t_0, \mathbf{x}, t), \Big] \quad (4.7)$$

where $p(\mathbf{x_0}, t_0, \mathbf{x}, t)$ is grand probability function, the subscript 0 is for the initial condition, m the total number of elementary reactions of the network, c_i's are the stochastic rate constants, and the associated rate law for each of the reactions, $h_i(\mathbf{x}, c_i)$ and S is the stoichiometry matrix. We present the explicit of the S, h_i and the values of c_i's in Appendix 1. However, solving the above equation with an analytical approach is limited as the complexity of the network increases. Several attempts have been made to solve the chemical master equation using the generating function method, self-consistent proteomic field, binomial moment, etc. (Walczak et al., 2005; Gardiner, 2009; Barzel and Biham, 2011).

For example, the CME for the NF-κB mediated gene activation and deactivation under Markovian approximation may be written as

$$\frac{dP(I, t)}{dt} = S^+(I - 1)P(I - 1, t) + S^-(I + 1)P(I + 1, t) - (S^+(I) + S^-(I))P(I, t)$$

$$(4.8)$$

where $P(I, t)$ is the grand the probability function that there exist I proteins in the system at time t; S^+ and S^- are the propensities of IκBα protein production and

degradation, respectively. These propensities have the form:

$$S^+(I) = \frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)}, \text{and} \qquad (4.9)$$

$$S^-(I) = k_d I \qquad (4.10)$$

The above equation can be reduced to Fokker Planck equation (FPE) by the Stratonovich sense (Risken and Risken, 1996), which allows one to calculate the probability, $P(I, t)$

$$\frac{\partial}{\partial t} P(I, t) = -\frac{\partial}{\partial I}[a_1(I) + a_2'(I) a_2(I) P(I, t)] + \frac{1}{2} \frac{\partial^2}{\partial I^2}[a_2(I) P(I, t)], \qquad (4.11)$$

where $a_1(I)$ and $a_2(I)$ are the first two jump moments, written as $a_1(I) = S^+(I) + S^-(I)$ and $a_2(I) = S^+(I) - S^-(I)$ respectively. The steady-state solution of the FPE has the form (Van Kampen, 1992):

$$P(I) = \frac{Q}{\sqrt{S^+(I) + S^-(I)}} exp\left(2 \int^I \frac{S^+(I) - S^-(I)}{S^+(I) + S^-(I)}\right) dI, \qquad (4.12)$$

where Q is a normalization constant.

Despite their importance, analytical methods cannot handle complex regulatory systems involving many non-linear functioning promoters. Therefore, one can perform stochastic simulation by considering each elementary reaction of each of the networks. The stochastic dynamics directly provide a signature on the robustness and stability of each network motif. We use the Gillespie algorithm to perform our stochastic simulations (Gillespie, 1976). It provides a realistic view of where the fluctuations in the abundance of the molecules in living cells affect their growth. Geometrically, the number of molecules change is the random walk on a multidimensional state space (Van Kampen, 1992). Stochastic simulation has an advantage over its deterministic version because the former method takes care of the system's intrinsic fluctuations, allowing the state of the system to switch from one to another. The trajectories obtained from stochastic simulations sample the Markov process as realized by a set of elementary reactions whose joint distribution is described by the master equation.

4.2.2.2 Incorporation of colored noise using UCNA

The IκBα protein exhibits a delay degradation (Krishna et al., 2006). We introduce the delay degradation into our model by introducing a time lag τ. Here, we define delay degradation as if an IκBα is produced at time t then it will degrade at $t + \tau$. The functions f_A and f_R are for the activation and repression. In our analysis, we have not done any formal delay dynamic analysis; rather, we introduce it by short-range correlated noise into the degradation parameter k_d. We include the effect of delay on protein degradation by introducing short-range correlated noise to the degradation parameter, k_d. Technically, we consider fluctuating $\tilde{k}_d$ driven by the Ornstein-Uhlenbeck (OU) process (Holehouse et al., 2020; Spanio et al., 2017). We define $\tilde{k}_d$ as $k_d + k_d\epsilon$, where k_d is the mean value and ϵ is the Gaussian correlated noise (color) defined by the OU process, $\frac{d\epsilon}{dt} = -\frac{\epsilon}{\tau} + \sqrt{\frac{2}{\tau}}\xi(t)$, where $\xi(t)$ is a zero-mean Gaussian white noise with $< \xi(t)\xi(t') > = 2D\delta(t - t')$. With this definition, the Gaussian color noise, $\epsilon(t)$ has zero mean and correlator $< \epsilon(t)\epsilon(t') > = \frac{D}{\tau}e^{-\frac{|t-t'|}{\tau}}$. By using the propensities given in the equations (4.9) and (4.10), the corresponding Langevin equation to the chemical FPE equation (4.11) is given as

$$\frac{dI}{dt} = \frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)} - k_d I + \sqrt{\frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)} + k_d I}.\Gamma(t) \qquad (4.13)$$

where $\Gamma(t)$ is the Gaussian white noise with zero mean and follows $< \Gamma(t)\Gamma(t') > = \delta(t - t')$. At this point we replace k_d by $\tilde{k}_d = k_d(1 + \epsilon(t))$ that takes care the degradation of protein that happens at $t + \tau$ time. The resulting equation becomes

$$\frac{dI}{dt} = \frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)} - (k_d + k_d\epsilon(t))I + \sqrt{\frac{k_1 k_P f_A(N)}{k_1 f_A(N) + k_2 f_R(I)} + k_d I}.\Gamma(t) \quad (4.14)$$

Note that $\epsilon(t)$ is the correlated noise generated by the OU process. However, it is to be noted that in the argument of the square root in equation (4.14), we have replaced $\epsilon(t)$ by its mean of zero, which comprises a mean-field approximation. This approximation is helpful because we can now solve (4.14) and (4.16) equations analytically. For clarity,

we rewrite the equation (4.14) together with the OU process as

$$\frac{dI}{dt} = h(I) + g_1(I)\epsilon + g_2(I)\Gamma(t) \tag{4.15}$$

$$\frac{d\epsilon}{dt} = -\frac{\epsilon}{\tau} + \sqrt{\frac{2}{\tau}}\xi(t) \tag{4.16}$$

where, $h(I) = \frac{k_1 k_P f_A(N)}{k_1 f_A(N)+k_2 f_R(I)} - k_d I$, $g_1(I) = -k_d I$, and $g_2(I) = \sqrt{\frac{k_1 k_P f_A(N)}{k_1 f_A(N)+k_2 f_R(I)} + k_d I}$ and $\xi(t)$ is a zero-mean Gaussian white noise with $< \xi(t)\xi(t') > = 2D\delta(t-t')$. We employ the UCNA method to obtain the approximate solution of equations (4.15) and (4.16). The method eliminates the noise adiabatically from the equation (4.16) (Jung and Hänggi, 1987). After elimination of noise from the equation (4.16), the equation (4.15) may be rewritten in the limit $\tau \to 0$ and $\tau \to \infty$:

$$\dot{I} \approx \frac{h(I)}{Z} + \frac{\sqrt{2\tau}g_1(I)\xi(t) + g_2(I)\Gamma(t)}{Z}, \tag{4.17}$$

where, $Z = 1 + \tau\left(\frac{g_1'(I)h(I)-h'(I)g_1(I)}{g_1(I)}\right)$. To obtain the simplified Langevin equation, we redefine (4.17) such that we only have one effective Gaussian white noise term. We write:

$$g(I)\tilde{\Gamma}(t) = \sqrt{2\tau}g_1(I)\xi(t) + g_2(I)\Gamma(t) \tag{4.18}$$

where, $\tilde{\Gamma}(t)$ is Gaussian white noise with mean zero and correlator $< \tilde{\Gamma}(t)\tilde{\Gamma}(t') > = 2\delta(t-t')$. We assume zero correlation between $\tilde{\Gamma}(t)$ and $\xi(t)$ (i.e., $< \tilde{\Gamma}(t)\xi(t') > = < \tilde{\Gamma}(t')\xi(t) > = 0$) to find the $g(I)$. Use of these correlators, we obtain $g(I)^2 < \tilde{\Gamma}(t)\tilde{\Gamma}(t') > = 2\tau g_1(I)^2 < \xi(t)\xi(t') > + g_2(I)^2 < \Gamma(t)\Gamma(t') >$ that finally gives us, $g(I) = \sqrt{2\tau D g_1(I)^2 + \frac{g_2(I)^2}{2}}$. By proposing $g(I)\tilde{\Gamma}(t) = \sqrt{2\tau}g_1(I)\xi(t) + g_2(I)\Gamma(t)$, the reduced Langevin equation becomes:

$$\dot{I} = \frac{h(I)}{Z} + \frac{g(I)\tilde{\Gamma}(t)}{Z}, \tag{4.19}$$

The FPE associated with the SDE can be written in the Stratonovich form (Risken and

Risken, 1996; Holehouse et al., 2020) as

$$\frac{\partial}{\partial t}P(I,t) = -\frac{\partial}{\partial I}[(\tilde{h}(I) + \tilde{g}(I)\tilde{g}'(I))P(I,t)] + \frac{\partial^2}{\partial I^2}[\tilde{g}(I)^2 P(I,t)] \qquad (4.20)$$

Here, $\tilde{h}(I) = h(I)/Z$ and $\tilde{g}(I) = g(I)/Z$. Immediately, the corresponding steady-state solution is written as:

$$P(I) = \frac{Q}{\tilde{g}(I)^2}exp\left(\int^I \frac{\tilde{h}(I) + \tilde{g}(I)\tilde{g}'(I)}{\tilde{g}(I)^2}\right)dI, \qquad (4.21)$$

where the Q is the normalization constant.

4.3 Results

4.3.1 Stochastic Dynamics of Network Assembly:

We perform the stochastic simulation for each network motif, as discussed in Chapter 3. The stochastic dynamics of network assembly allow switching from one state space to another for a given set of kinetic parameters. We fixed the kinetic parameter values of the network throughout our dynamic analysis and presented them in Appendix 1. We analyze the effect of binding and unbinding, together with a few other reactions, in the production dynamics. Our analyses revealed that the protein production obtained from various network motifs fluctuates around a mean, which follows the deterministic dynamics. We find oscillations of protein number in the trajectory, which is a clear signature of multistability (Cao and Grima, 2018; Duddu et al., 2020; Zhu et al., 2022; Li et al., 2018). We calculate the steady state probability distribution (P_{ss}) of the protein number for each network motif and, after that, calculate the stochastic potentials, $F = -\log(P_{ss})$ to establish the presence of multistability. We show the results in Figure 4.2. These results correspond between the dynamic and the previously explored thermodynamic results.

In the case of the activation network of assembly, we find that protein production follows a Gaussian distribution for both the simple and induced activations. The derived potential from the distribution function deepens as the ligand activates TFs. Moreover, it is evident from the figure that the depth of potential is increased for the induced
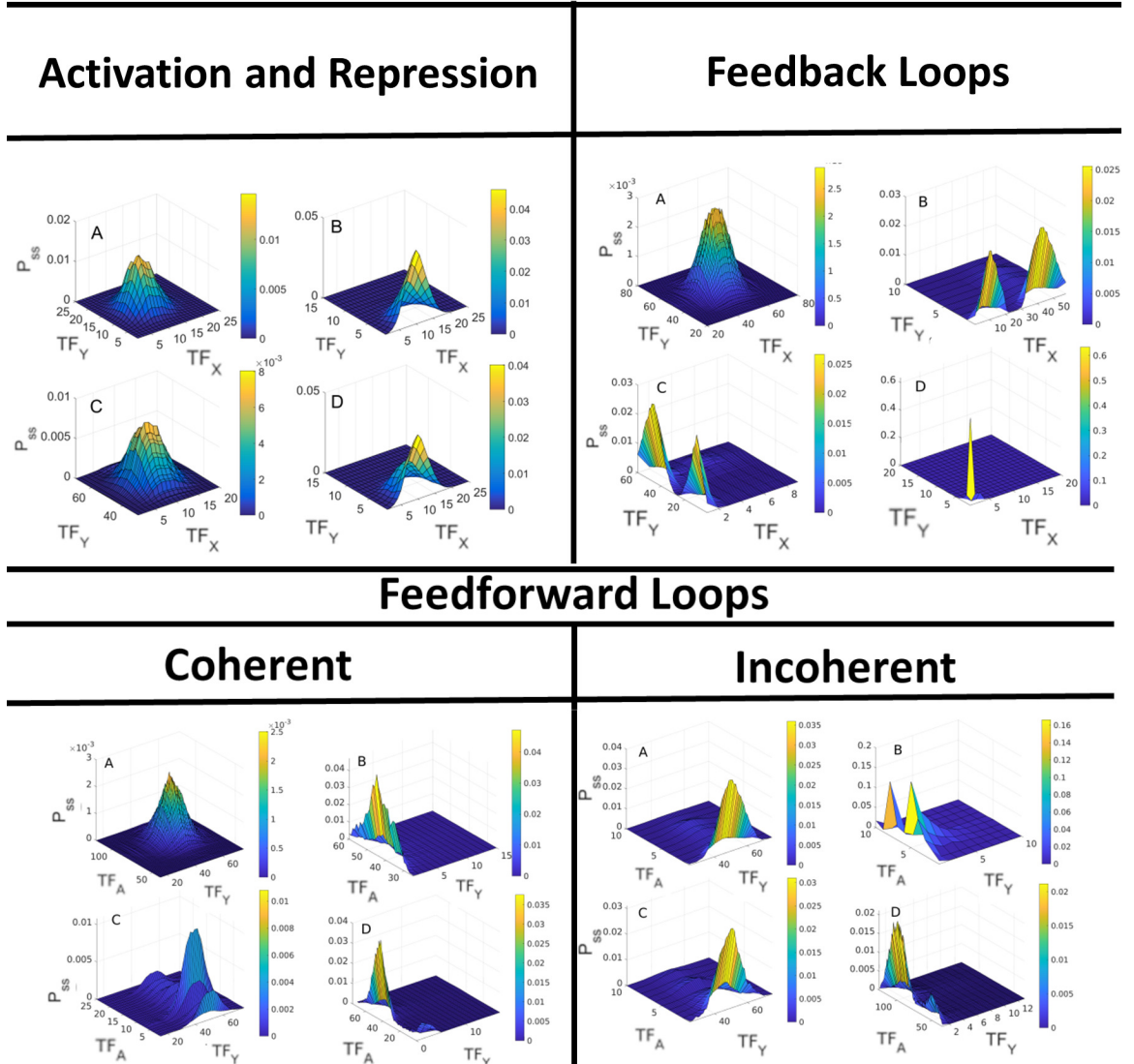
Figure 4.1: The probability distribution of proteins obtained from stochastic simulation is shown here. The presented results are for the activation and repression, feedback loops, and feedforward loops. A clear signature of multimodality for higher-order network assembly is visible from these graphs.

expression. These signatures suggest the protein assembly tightens its regulation as the ligand stabilizes the protein-DNA interaction network. We also notice that there is a mono-stable region for the repression. However, the binding of ligands to TFs enhances repression, as evidenced by the broad, deep potential well observed in our analysis. Overall, the obtained potential diagram reveals that the dynamics of these networks are

relatively simple, showing the unimodal distribution in protein production, and have a clear sign of mono-stability.

In contrast to simple activation and repression, we find different features for the feedback loops. In the case of the positive feedback loop, we find the unimodal distribution of protein molecules for both genes G_X and G_Y. Therefore, the production of proteins from the G_X and G_Y amplify each other mutually for this case. However, we notice the bimodal distribution of produced protein for the negative feedback loops and exhibit bistability. However, the stochastic potential obtained from the gene expression for mutually inhibiting the feedback loop exhibits frustration in their dynamics. Once again, we find a clear correspondence between our dynamic and thermodynamic results obtained from the assembly networks.

The probability distribution of proteins and stochastic potential functions obtained from the dynamics of FFLs are also shown in Figure 4.1 and 4.2, respectively. It is clear from the figure that the FFLs exhibit multistability. Such signatures of FFLs are consistent with the previous studies (Cao and Grima, 2018; Duddu et al., 2020). In the case of incoherent loops, the system exhibits oscillation in the dynamics, a typical feature usually found in FFLs (Zhang et al., 2016). Our analysis also reveals that the coherent FFLs also show multimodal behavior. Thus, the multimodality feature is quite common for the FFLs. The multimodality or multistability features are the origin of a wide range of phenotype diversity of a cell as reported in many literatures (Cao and Grima, 2018; Ochab-Marcinek and Tabaka, 2010). In the following section, we apply this method to the NF-κB system since it uses the protein assembly to process a signal.

4.3.2 NF-κB system

4.3.2.1 Bifurcation analysis

The deterministic dynamical equations for the NF-κB are described in the method section. Since the system exhibits nonlinearity, we first conduct the bifurcation analysis to extract parameter sensitivity. A bifurcation diagram quantifies how a dynamical system's long-time behavior changes as a function of a parameter. In other words, it describes a change in the stability or existence of fixed points as the system parameters change. In a biological system, such perturbation in a parameter is common since we
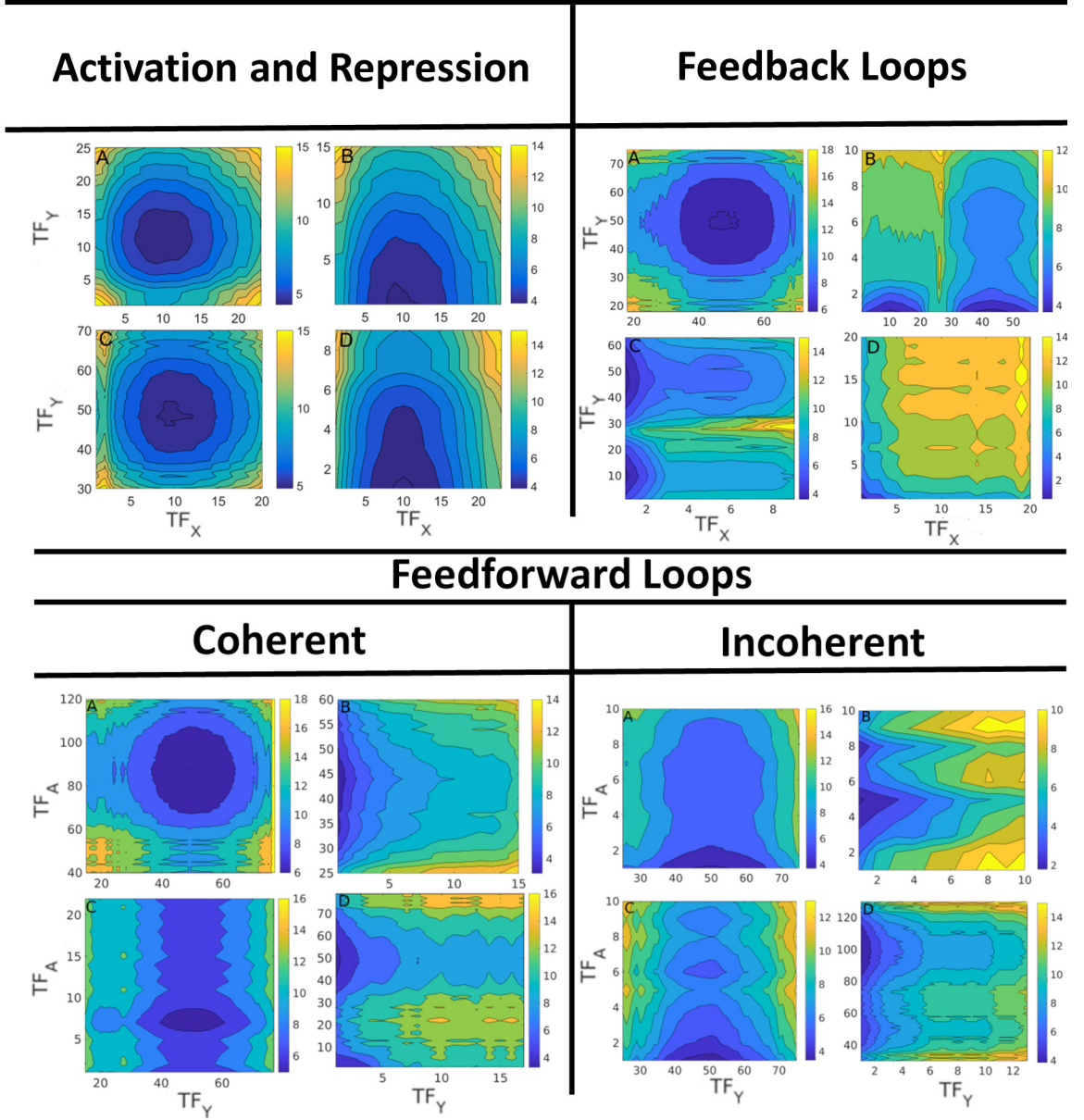
Figure 4.2: The stochastic potentials for various GRNs are shown in the 2D contour maps. The presented results are for the activation and repression, feedback loops, and feedforward loops. A clear signature of multimodality for higher-order network assembly is visible from these maps.

observe sudden switching among stable, steady states.

We performed a bifurcation analysis of our model system, showing the population of IκBα as a function of bifurcation parameters, k_d and NF-κB. We show the results in subfigures A and B of Figure 4.3. We marked stable and unstable regions by the

magenta colored line and green colored circles in those subfigures. The system exhibits bistability, as evident from the saddle-node or fold bifurcation shown in the figure. The existence of saddle-node bifurcation is also used to infer the hysteretic behavior of the system, as evident from Figure 4.3. From subfigure A, we see an increase of k_d from low to high values (for subfigure A), resulting in picking the upper branch in the bistable regime, whereas decreasing from high to low values (for subfigure A) takes us to the lower branch. A similar effect is also evident for the variation of NF-κB as shown in subfigure B. It is clear from the figure that the system exhibits bistability only over a range of parameter values, and it experiences monostability in other regions. Literature suggests that bistability or, in general, multistability is an essential recurring theme for understanding various cell signaling or cellular functioning, which includes decision-making biophysical processes such as cell cycle progression, cell differentiation, and apoptosis (Angeli et al., 2004; Eissing et al., 2004; Sobie, 2011). These observations from subfigure A and B shed light on the presence of the two stable, steady states and the existence of the bimodal distribution in the population of IκBα influenced by the rate constant k_d and NF-κB in our system. This presence of bistability may be correlated with the switching between two functional states or the phenotype diversity of a GRN. The NF-κB system tends to settle into a stable state marked by high IκBα concentration at high values of NF-κB and low concentration of IκBα for low values of NF-κB. We also find similar behavior for k_d where it shows bistability for a specific range of NF-κB.

We further performed a two-parameter bifurcation analysis of equilibria for our model system. The main findings from this analysis are that we can figure out the region corresponding to the monostable and bistable, a line separating them, and an identification of the cusp point, if any. The cusp point in a 2-D bifurcation diagram shows a signature of catastrophic change in the system. It is a bifurcation of equilibria in a two-parameter family of dynamical equations at which the critical equilibrium has one zero eigenvalue, and the quadratic coefficient for the saddle-node bifurcation vanishes. At the cusp bifurcation point two branches of saddle-node bifurcation curve meet tangentially, forming a semicubic parabola. We find from our two-dimensional bifurcation diagram in subfigure c, which delineates a bistable region in the k_d−NF-κB parameter space. Here, we observed a cusp point at the origin. We find another cusp point in the k_1 and k_d parameter

space, a signature of catastrophic change we could observe in the system.
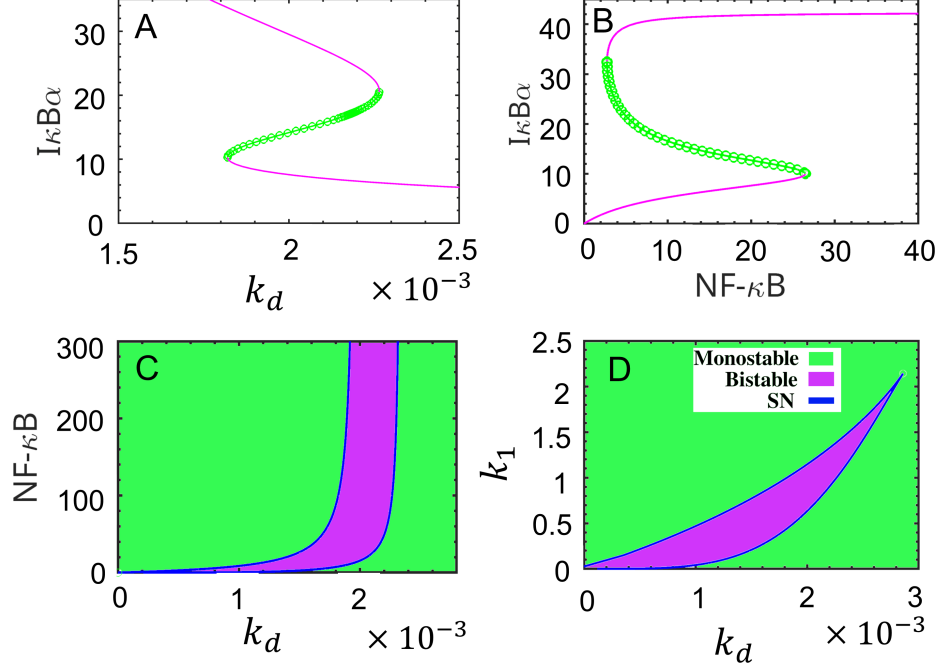
Figure 4.3: This figure shows the concentration of NF-κB and rate of degradation of IκBα protein (k_d) induced saddle-node (SN) bifurcation for the number of IκBα species. Also, we have shown the presence of a cusp point for these parameters, which shows a boundary (marked by blue colour) between monostable and bistable states.

4.3.2.2 Stochastic potentials

Since the system exhibits bistability, exploring the dynamics in the presence of noise will be interesting. We perform stochastic dynamics of this system in the presence of white and color noises. Specifically, we calculate the steady-state probability distributions (SSPD) and stochastic potentials using our theoretical analysis. Figure 4.4 are the SSPD for non-fluctuating protein degradation rate parameters obtained from equation (4.12) and kinetic Monte Carlo simulations. We consider three different values of k_d, i.e., $10^{-3}min^{-1}$, $2 \times 10^{-3}min^{-1}$, and $3 \times 10^{-3}min^{-1}$ chosen from bifurcation analysis in the Figure 4.4. We observe a fair correlation between the results obtained from the reduced master equation and the KMC simulation. We find bimodality in the protein's production over a range of protein degradation parameters, k_d. We find similar results

by varying the concentration of NF-κB. The results correlate well with the bifurcation analysis in the previous section. Our analysis shows that the degradation rate parameter can induce bimodality that may be related to the *cellular decision-making* for this regulatory network. It thus becomes a significant step to identify the parameters space, which can control the phenotypic states and thus demands precise attention toward a better understanding of complex degradation mechanisms associated with such subtle gene regulatory architectures.
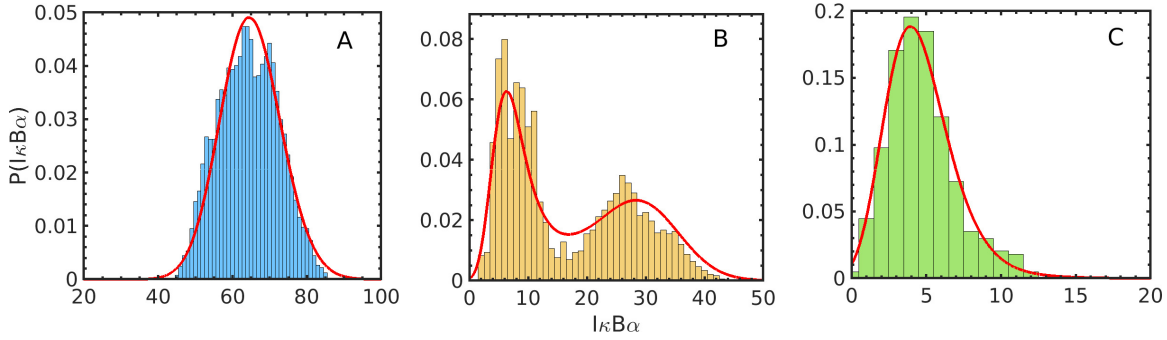
Figure 4.4: Here, we compare the solution of the reduced master equation from equation (4.12) (solid line) and Gillespie's algorithm (histogram). Panel A and C are responsible for the monostable regions, and B indicates the bistable region, as can be well correlated with the bifurcation diagram. The value of the degradation parameter k_d for subfigure A, B and C are $10^{-3}min^{-1}$, $2 \times 10^{-3}min^{-1}$, and $3 \times 10^{-3}min^{-1}$ respectively.

A close comparison between the probability distribution functions obtained from UCNA analysis in the regime of color and white noises is presented in Figure 4.5. We find that a strongly correlated noise ($\tau =1000$) increases the sharpness of the probability distribution function, which signifies its importance in biological systems. Multiple sharp peaks in probability distribution function may be related to the multiple phenotypes observed in biological systems. In general, biological systems are associated with correlated color noise, which is responsible for different phenotype behaviors, e.g., noise can affect the cell fates by randomly turning on either latency or reactivation. The randomness associated with biological reaction events gives rise to biological noise, which may be one of the origins of cell variability (Kellogg and Tay, 2015; Wang et al., 2022). Thus, exploring the noise-induced phenotypes at the single-cell level is essential to understand them better.
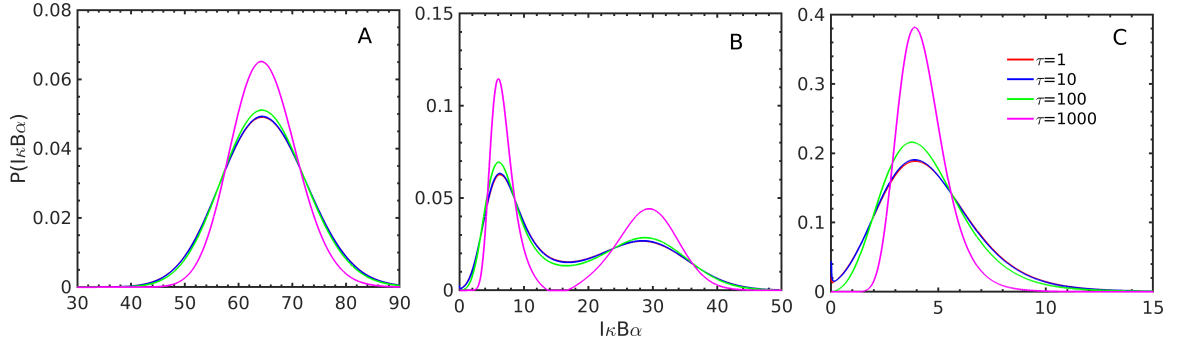
Figure 4.5: Panel A and C are responsible for the monostable regions, and B indicates the bistable region, which can be well correlated with the bifurcation diagram. The value of the degradation parameter k_d for subfigure A, B and C are $10^{-3}min^{-1}$, $2 \times 10^{-3}min^{-1}$, and $3 \times 10^{-3}min^{-1}$ respectively.

The obtained SSPD from the UCNA method is presented in Figure 4.5 for a range of correlation time τ. We observed that the SSPD becomes narrow as the correlation time increases (panels A and C). Moreover, in panel B, we observed that colored noise influences the bimodality in the population of IκBα at large correlation time ($\tau = 1000$). These results reveal the role of noise in regulating bimodality. Thus, it is crucial to determine the role of colored noise near the boundary between monostable and bistable regions. We chose the value for the protein degradation parameter, k_d=$2\times10^{-3}min^{-1}$, which takes into account the boundary between monostable and bistable region, as shown in Figure 4.5.
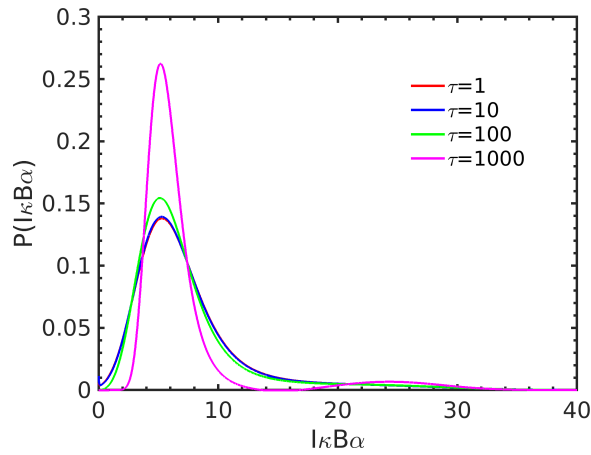
Figure 4.6: The results for *noise induced bimodality* for the fluctuating rate parameters for the degradation rate parameter, k_d are shown. The values of the parameters used here are N=50 and $k_d = 0.0023min^{-1}$.

In Figure 4.6, we vary the correlation time at the boundary point. We observe bimodality for strongly correlated noise (for $\tau = 1000$), which suggests the existence of the *noise induced bimodality* (Holehouse et al., 2020). In the white noise limit, the system exhibits monostability, but it exhibits bistability for strongly correlated noise. This is a clear signature of the noise-induced bistability, as revealed from the appearance of a shallow peak at large τ value. The origin behind such change can be correlated with the large fluctuations in the degradation rate parameter.

4.3.3 Stochastic analysis for the SE and TE-mediated expression

Our proposed model considers the switching of the promoter states between active (G^*) and inactive (G) through the binding of NF-κB. The model considers oligomerization through Hill's coefficients (n_H) present in Hill's function, f_A. The strength of binding or the binding affinity between the NF-κB and promoter region of DNA is characterized by the parameter, K_{m1}. The amount of total NF-κB is given by the parameter N. Note that an increase in N results in the binding of a cluster of proteins at the promoter region. The binding of clusters of proteins to the gene's promoter region enhances the expression abruptly, which is quite distinct from the regular binding of TF to the promoter region. The binding of such a cluster of proteins to the promoter region is defined as a super-enhancer (SE), and the regular protein-promoter interaction is called a typical enhancer (TE). Gene expression via SE produces higher fold change than the expression controlled by the typical enhancers (TEs) (Brown et al., 2014; Michida et al., 2020).

Literature suggests (Michida et al., 2020) that the typical values of n_H for the expressions mediated by TE and SE have values ≈ 1 and 4, respectively. Therefore, we use these numbers to model them as a function of NF-κB concentration. The results are presented in Figure 4.7, where we plot the maximum probability value of the produced protein, $P_{max}(I\kappa B\alpha)$. We observe an abrupt switching in the expression level for the SE at a very low concentration of NF-κB. The results are shown in the panels A and B. Thus, when NF-κB binds as a cluster to promoter regions, a significant difference in gene expression level is exhibited for a specific range of NF-κB concentrations. Our observed results correlate well with the earlier reported results (Michida et al., 2020).

Michida *et. al.* reported that interactions among enhancer region and NF-κB molecules could be cooperative as well as non-cooperative. The choice of these modes is based on the amount of NF-κB present in the system and chromatin accessibility (Michida et al., 2020). Based on a similar argument, Figure 4.7 supports the role of NF-κB as SEs and TEs, where SE produces a high expression level relative to TE. We also observed the existence of bistability in this system. Bistable or multistable features in a GRN are essential for *cellular decision* making and often correlate well between genotype and phenotype relations. Thus, we infer from this analysis that the genes can tune their expression by forming clusters of transcription-regulating proteins at the enhancer elements of a gene.
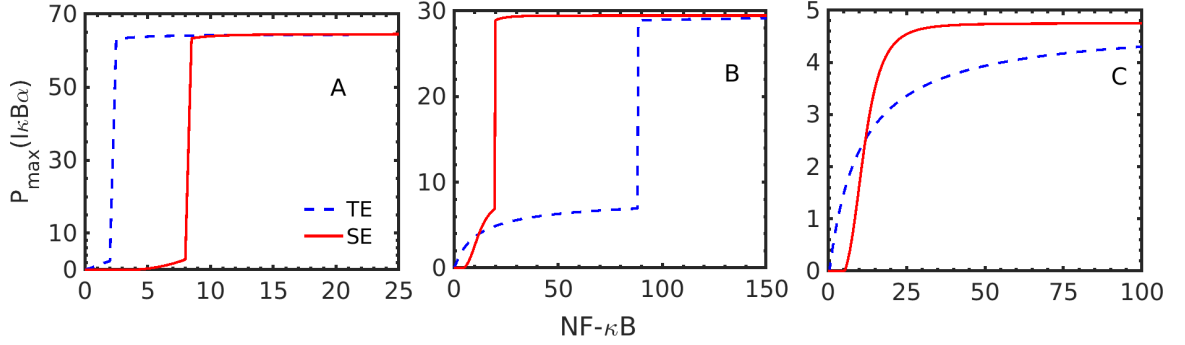
Figure 4.7: Maximum probability value of the IκBα protein produced via SE-and TE-mediated gene expression: The dashed blue and solid red color lines indicate the TE- and SE-mediated gene expression, respectively. The value of the degradation parameter k_d for subfigure A, B and C are $10^{-3}min^{-1}$, $2 \times 10^{-3}min^{-1}$, and $3 \times 10^{-3}min^{-1}$, respectively.

4.4 Discussion

In summary, we have studied theoretical and computational models for the dynamics of gene regulatory networks. Our dynamic analysis shows that each complex assembly's functional response strongly depends on the network's topology. We find that the network topology associated with the complex assembly plays a vital role in the production of protein molecules. We calculate the stochastic potentials from their trajectories. We find the multi-stability features in the higher-order networks. Recent studies have also demonstrated the existence of such multi-stable features in higher-order GRNs

(Duddu et al., 2020; Zhu et al., 2022). In the case of activation or repression, we find mono-stability as evident from the single potential well. However, the assembly shows multi-stable features for feedback and feed-forward loops. Our analysis shows that the apparent structure of the bio-molecular complexes is quite similar. However, their functional responses strongly depend on the network's topology.

We further analyze the stochastic dynamics of the NF-κB since it forms homo and hetero-oligomers on the DNA. We analyze this regulatory system using analytical methods and kinetic Monte Carlo simulation. We first identify the sensitive parameter from bifurcation analysis. We find k_d, k_1 and the input concentration of NF-κB are sensitive parameters as evident from the observed fold bifurcation. Based on this analysis, we chose the k_d parameter and studied the stochastic dynamics when it fluctuates. We first perform our analysis by approximating the system that follows in the limit of white noise. We chose a static rate parameter to model it. We solve the corresponding master equation and obtain the steady-state probability distribution of the produced protein. However, including the protein lifetime requires a fluctuating rate parameter driven by the correlated noise. The introduction of short-range correlated noise to the rate parameter effectively deals with the degradation protein at $t + \tau$ if it is produced at t. We apply the UCNA method to analyze the system analytically and obtain the steady state probability distribution function. We performed the KMC simulation and found our results matched well with the analytical method.

Our analysis shows that the bimodality of a gene regulatory network arises from various factors such as the fluctuating rate parameter k_d, the rate of change of gene from inactive to active state k_1, and Noise-induced bimodality. This bimodal behavior can result from a negative feedback mechanism akin to the toggle switch in a gene regulatory system. Numerous such systems exist in natural systems (Ochab-Marcinek and Tabaka, 2010; Venturelli et al., 2012), so it becomes essential to understand their origin. We observed noise-induced bimodality by employing a theoretical and simulation approach and put effort into understanding the mechanism of such behavior. Our analyses establish a close correspondence between the system's nonlinearities and inherent noise in a biological system. We find a close correlation between an input signal and the multi-modal distribution in protein production as determined by the interaction network. We also

observed that the clustering of transcription-regulating proteins influences the switch from a broad distribution to a narrow distribution of the produced proteins.

Our analysis for the typical and super-enhancers mediated gene expression correlates well with the experimental findings. We made a correspondence between our theoretical and computational results and presented their biological relevance. These combined theoretical and simulation methods can be applied to various other biological systems for anticipating gene expression.

Appendix: 1) Activation and Repression

Table 4.1: This table contains various elementary reactions for the activation(A), repression(R), Induced activation(IA), Induced repression(IR) network motifs. Here, G_X and G_Y represent the genes with the basal expression, G_Y^* represents the activated gene, and G_Y^Θ represents the repressed gene. Also, TF_X and TF_Y represent the proteins that are expressed by G_X and G_Y, respectively. The specific reaction rates for each reaction are shown on the marked arrows.

A	R	IA	IR
1) $G_X \xrightarrow{\rho_X} G_X + TF_X$	1) $G_X \xrightarrow{\rho_X} G_X + TF_X$	1) $G_X \xrightarrow{\rho_X} G_X + TF_X$	1) $G_X \xrightarrow{\rho_X} G_X + TF_X$
2) $G_Y \xrightarrow{\rho_Y} G_Y + TF_Y$	2) $G_Y \xrightarrow{\rho_Y} G_Y + TF_Y$	2) $G_Y \xrightarrow{\rho_Y} G_Y + TF_Y$	2) $G_Y \xrightarrow{\rho_Y} G_Y + TF_Y$
3) $G_Y + TF_X \xrightarrow{\sigma_{PYY}} G_Y^*$	3) $G_Y + TF_X \xrightarrow{\kappa_{YY}} G_Y^\Theta$	3) $TF_X + L_X \xrightarrow{\sigma_X} TF_X^*$	3) $TF_X + L_X \xrightarrow{\sigma_X} TF_X^*$
4) $G_Y^* \xrightarrow{\sigma_{PYY}'} G_Y + TF_X$	4) $G_Y^\Theta \xrightarrow{\kappa_{YY}'} G_Y + TF_X$	4) $TF_X^* \xrightarrow{\sigma_{X'}} TF_X + L_X$	4) $TF_X^* \xrightarrow{\sigma_X'} TF_X + L_X$
5) $G_Y^* \xrightarrow{\rho_{YY}'} G_Y^* + TF_Y$	5) $TF_X \xrightarrow{k_{dX}} \phi$	5) $G_Y + TF_X^* \xrightarrow{\sigma_{PY}} G_Y^*$	5) $G_Y + TF_X^* \xrightarrow{\kappa_Y} G_Y^\Theta$
6) $TF_X \xrightarrow{k_{dX}} \phi$	6) $TF_Y \xrightarrow{k_{dY}} \phi$	6) $G_Y^* \xrightarrow{\sigma_{PY}'} G_Y + TF_X^*$	6) $G_Y^\Theta \xrightarrow{\kappa_Y'} G_Y + TF_X^*$
7) $TF_Y \xrightarrow{k_{dY}} \phi$		7) $G_Y^* \xrightarrow{\rho_Y'} G_Y^* + TF_Y$	7) $TF_X \xrightarrow{k_{dX}} \phi$
		8) $TF_X \xrightarrow{k_{dX}} \phi$	8) $TF_Y \xrightarrow{k_{dY}} \phi$
		9) $TF_Y \xrightarrow{k_{dY}} \phi$	

The **x** vector corresponding to Equation (4.1) that is associated with the reaction

network for activation is written as $\mathbf{x} = [G_Y \; G_Y^* \; TF_X \; TF_Y]'$ and the stoichiometry matrix S for the same is given as follows

$$
S_A = \begin{pmatrix}
0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 \\
1 & 0 & -1 & 1 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & -1
\end{pmatrix}
$$

Also, $F(\mathbf{x})$ is written as $[\rho_X G_X \; \rho_Y G_Y \; \sigma_{PYY} G_Y TF_X \; \sigma'_{PYY} G_Y^* \; \rho'_{YY} G_Y^* \; k_{dX} TF_X \; k_{dY} TF_Y]'$. Here, stochastic rate constants (c_i) and the associated rate law for each of the reactions, $h_i(\mathbf{x}, c_i)$ taking place in a volume of size Ω are written as:

$c_i = [\rho_X \; \rho_Y \; \frac{\sigma_{PYY}}{\Omega} \; \sigma'_{PYY} \; \rho'_{YY} \; k_{dX} \; k_{dY}]$, and,

$h_i(\mathbf{x}, c_i) = [\rho_X g_X \; \rho_Y g_Y \; \frac{\sigma_{PYY}}{\Omega} g_Y n_X \; \sigma'_{PYY} g_Y^* \; \rho'_{YY} g_Y^* \; k_{dX} n_X \; k_{dY} n_Y]$, where g_i^* and g_i represent the deterministic mean number of bounded and unbounded i^{th} genes, respectively, and n_X and n_Y are the mean protein number production from the respective X and Y gene.

The $\mathbf{x}$ vector associated with the reaction network for repression is $\mathbf{x} = [G_Y \; G_Y^\Theta \; TF_X \; TF_Y]'$ and stoichiometry matrix S for the same is given as follows

$$
S_R = \begin{pmatrix}
0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 \\
1 & 0 & -1 & 1 & -1 & 0 \\
0 & 1 & 0 & 0 & 0 & -1
\end{pmatrix}
$$

Also, here $F(\mathbf{x})$ in this case is written as $[\rho_X G_X \; \rho_Y G_Y \; \kappa_{YY} G_Y TF_X \; \kappa'_{YY} G_Y^\Theta \; k_{d_X} TF_X \; k_{d_Y} TF_Y]'$, with

$c_i = [\rho_X \; \rho_Y \; \frac{\kappa_{YY}}{\Omega} \; \kappa'_{YY} \; k_{d_X} \; k_{d_Y}]$ and,

$h_i(\mathbf{x}, c_i) = [\rho_X g_X \; \rho_Y g_Y \; \frac{\kappa_{YY}}{\Omega} g_Y n_X \; \kappa'_{YY} g_Y^\Theta \; k_{d_X} n_X \; k_{d_Y} n_Y]$

where g_i^Θ represents the deterministic mean number of repressed genes.

The $\mathbf{x}$ vector associated with the reaction network for Induced activation is $\mathbf{x} = [G_Y \; G_Y^* \; TF_X \; L_X \; TF_X^* \; TF_Y]'$ and stoichiometry matrix S for the same is given as follows

$$S_{IA} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

Also, $F(\mathbf{x})$ is written as

$[\rho_X G_X \ \rho_Y G_Y \ \sigma_X TF_X L_X \ \sigma'_X TF_X^* \ \sigma_{PY} G_Y TF_X^* \ \sigma'_{PY} G_Y^* \ \rho'_Y G_Y^* \ k_{dX} TF_X \ k_{dY} TF_Y]'$,

$c_i = [\rho_X \ \rho_Y \ \frac{\sigma_X}{\Omega} \ \sigma'_X \ \frac{\sigma_{PY}}{\Omega} \ \sigma'_{PY} \ \rho'_Y \ k_{dX} \ k_{dY}]$ and,

$h_i(\mathbf{x}, c_i) = [\rho_X g_X \ \rho_Y g_Y \ \frac{\sigma_X}{\Omega} n_X l_X \ \sigma'_X n_X^* \ \frac{\sigma_{PY}}{\Omega} g_Y n_X^* \ \sigma'_{PY} g_Y^* \ \rho'_Y g_Y^* \ k_{dX} n_X \ k_{dY} n_Y]$

where, l_i represents the mean ligand number binding and activating the ith protein produced from i^{th} gene.

The $\mathbf{x}$ vector associated with the reaction network for induced repression is, $\mathbf{x} = [G_Y \ G_Y^\Theta \ TF_X \ L_X \ TF_X^* \ TF_Y]'$ and stoichiometry matrix S for the same is given as follows

$$S_{IR} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Now, $F(\mathbf{x})$ is written as $[\rho_X G_X \ \rho_Y G_Y \ \sigma_X TF_X L_X \ \sigma'_X TF_X^* \ \kappa_Y G_Y TF_X^* \ \kappa'_Y G_Y^\Theta \ k_{dX} TF_X \ k_{dY} TF_Y]'$,

$c_i = [\rho_X \ \rho_Y \ \frac{\sigma_X}{\Omega} \ \sigma'_X \ \frac{\kappa_Y}{\Omega} \ \kappa'_Y \ k_{dX} \ k_{dY}]$, and

$h_i(\mathbf{x}, c_i) = [\rho_X g_X \ \rho_Y g_Y \ \frac{\sigma_X}{\Omega} n_X l_X \ \sigma'_X n_X^* \ \frac{\kappa_Y}{\Omega} g_Y n_X^* \ \kappa'_Y g_Y^\Theta \ k_{dX} n_X \ k_{dY} n_Y]$.

Appendix: 2) Deterministic Model

Table 4.2: The deterministic equations for the A, R, IA, and IR network motifs are shown here.

A	R
$$\frac{dG_Y}{dt} = -\sigma_{P_{YY}} G_Y TF_X + \sigma'_{P_{YY}} G_Y^*$$ $$\frac{dG_Y^*}{dt} = \sigma_{P_{YY}} G_Y TF_X - \sigma'_{P_{YY}} G_Y^*$$ $$\frac{dTF_X}{dt} = \rho_X G_X - \sigma_{P_{YY}} G_Y TF_X + \sigma'_{P_{YY}} G_Y^* - k_{dX} TF_X$$ $$\frac{dTF_Y}{dt} = \rho_Y G_Y + \rho'_{YY} G_Y^* - k_{dY} TF_Y$$	$$\frac{dG_Y}{dt} = -\kappa_Y G_Y TF_X + \kappa'_Y G_Y^\Theta$$ $$\frac{dG_Y^\Theta}{dt} = \kappa_Y G_Y TF_X - \kappa'_Y G_Y^\Theta$$ $$\frac{dTF_X}{dt} = \rho_X G_X - \kappa_Y G_Y TF_X + \kappa'_Y G_Y^\Theta - k_{dX} TF_X$$ $$\frac{dTF_Y}{dt} = \rho_Y G_Y - k_{dY} TF_Y$$
IA	IR
$$\frac{dG_Y}{dt} = -\sigma_{P_Y} G_Y TF_X^* + \sigma_{P'_Y} G_Y^*$$ $$\frac{dG_Y^*}{dt} = \sigma_{P_Y} G_Y TF_X^* - \sigma'_{P_Y} G_Y^*$$ $$\frac{dTF_X}{dt} = \rho_X G_X - \sigma_X TF_X L_X + \sigma'_X TF_X^* - k_{dX} TF_X$$ $$\frac{dTF_Y}{dt} = \rho_Y G_Y + \rho'_Y G_Y^* - k_{dY} TF_Y$$ $$\frac{dTF_X^*}{dt} = \sigma_X TF_X L_X - \sigma'_X TF_X^* - \sigma_{P_Y} G_Y TF_X^* + \sigma'_{P_Y} G_Y^*$$ $$\frac{dL_X}{dt} = -\sigma_X TF_X L_X + \sigma'_X TF_X^*$$	$$\frac{dG_Y}{dt} = -\kappa_Y G_Y TF_X^* + \kappa'_Y G_Y^\Theta$$ $$\frac{dG_Y^\Theta}{dt} = \kappa_Y G_Y TF_X^* - \kappa'_Y G_Y^\Theta$$ $$\frac{dTF_X}{dt} = \rho_X G_X - \sigma_X TF_X L_X + \sigma'_X TF_X^* - k_{dX} TF_X$$ $$\frac{dTF_X^*}{dt} = \sigma_X TF_X L_X - \sigma'_X TF_X^* - \kappa_Y G_Y TF_X^* + \kappa'_Y G_Y^\Theta$$ $$\frac{dTF_Y}{dt} = \rho_Y G_Y - k_{dY} TF_Y$$ $$\frac{dL_X}{dt} = -\sigma_X TF_X L_X + \sigma'_X TF_X^*$$

Role of network topology in controling the cellular fate under stressed condition: A tumor enigma

5.1 Introduction

In this Chapter, we develop a thermodynamic model for the p53 signaling network to explore the role of network topology in controlling cell fate. The p53 tumor suppressor protein can regulate the expression of nearly 100 genes in response to DNA damage (Wright and Dyson, 2015; Sullivan et al., 2018). The tumor suppressor gene products frequently regulate selected signal transduction pathways and monitor the efficiency of cellular duplication by checkpoints in cell division. The p53-signaling pathway responds to a wide variety of stress signals and regulates the process of tumorigenesis. These expressions are associated with determining cell fate, e.g., cell cycle arrest, DNA repair, senescence, apoptosis, etc. (Levine, 1997; Bieging et al., 2014). Perturbations along the pathway could compromise p53 activity and consequently promote tumor development.

Genotoxic stress created by DNA damage induces the transcriptional activity of the p53 protein. Unstressed cells maintain low levels of p53 by continuous proteasomal degradation mediated by a protein, MDM2 (Kubbutat et al., 1997; Haupt et al., 1997). However, the genome with lost integrity produces a mutant p53 protein with intense phosphorylation and acetylation at sites common for stabilizing wild-type p53, thereby facilitating the collection of defective mutant p53 inside the nucleus, ultimately ending up as an oncogene (Bode and Dong, 2004; Vousden and Prives, 2009). Cascading events in post-translational modifications regulate the transcriptional activity of p53 that inhibits binding with MDM2 and promotes the expression of the p21 gene. Studies suggest that

phosphorylation of p53 significantly affects its binding affinity to DNA. Single-site phosphorylation of p53 at Thr18 results in a twofold increase in affinity to DNA, double-site phosphorylation at Ser15, Thr18 leads to an approximately five-fold increase, and triple-site phosphorylation at Ser15, Thr18, and Ser20 causes more than Ten-fold increase in binding affinity to DNA (Lee et al., 2010; Ferreon et al., 2009). The degree of phosphorylation affects the phenotypical changes differentially in a cell, a signature accepted widely for inhibiting the cancer progression over the last few decades (Lee et al., 2010; Olsson et al., 2007). This fact is verified in mice, where a single-site phosphorylation of serine or threonine of p53 affects little to their activity. In contrast, their simultaneous phosphorylation impacts mice's activities significantly, indicating that multisite phosphorylation has synergistic effects on p53 response (Chao et al., 2006; Lee et al., 2010). In humans, simultaneous phosphorylation on threonine and serine also synergizes, increasing p53 responses (Gatti et al., 2000; Mayo et al., 2005). However, the mechanism linked with the phosphorylation and especially multisite phosphorylation in controlling p53 activation remains unclear (Vousden and Prives, 2009; Olsson et al., 2007).

Under stressed conditions, cells produce an uncontrolled amount of p53 suppressor that may not degrade during the cell cycle. The excess p53 undergoes phosphorylation once it senses external stimuli and participates in a series of transcriptional activation processes. Lack of participation of p53 for transcription regulation initiates tumor progression. The outcome of this series of events is unique, which we can relate to the fate of a cell (Hafner et al., 2019). It is found that few genes trigger the fate of a cell. Expressions of them determine the unique phenotype of the cell. In our study, we consider the unique phenotype for the p53 repressor system to be one of the possibilities of cell cycle arrest, DNA repair, inhibition of angiogenesis, and apoptosis (Levine and Oren, 2009; Kastenhuber and Lowe, 2017). These four events are governed by the binding of phosphorylated tetramers of the p53 repressor (p53*) to the promoter region of four genes such as p21 (Rokudai et al., 2009), p48 (Williams and Schumacher, 2016), PAI (Teodoro et al., 2006), and Bax (Farkas et al., 2021; Chao et al., 2006) in the genome.

The mechanism behind p21-dependent apoptosis is still unclear, but it is believed that the p53* promotes the production of p21 protein that binds with CDK-cyclin complex for the cell cycle arrest (Yeap et al., 2015). Further, few studies suggest that p21 plays a vital role in inhibiting apoptosis in a p53-dependent manner (Attardi et al., 1996; Bissonnette and Hunting, 1998). Therefore, one can ask a pertinent question on how a cell opts between p21-dependent cycle arrest and apoptosis once it receives signals from the damaged DNA. Few studies report that high levels of p21 expressions result in cell cycle arrest and promote apoptosis in a p53-dependent manner (Choi et al., 2016; Macleod et al., 1995). Moreover, p21 is also known for regulating a few other genes that cause inhibition of angiogenesis and DNA repair (Furuta et al., 2006; Kuljaca et al., 2009). Few studies also show that the PAI regulates apoptosis in cells (Chen et al., 2004).

Since the phosphorylated tetramer of p53 binds differentially to the response elements of the four cell fate-determining genes, they produce differential gene expression and consequently determine the state of a cell. Further, these four genes are not independent in a cell; they are often connected through protein-protein interaction. Upon consideration of the internal networks or the network topology among these four genes, the expression further differentiated that plays a crucial role in determining the state of a cell. Therefore, it is critical to understand that the binding of p53* to these four genes is necessary to determine a cell's fate. Moreover, various cancer cell lines have different internal gene regulatory networks among these four genes, further introducing differential gene expression that leads to a cell's fate (Fan et al., 2020; Georgakilas et al., 2017; Mirzayans et al., 2015).

The mechanism behind p53 suppressor-mediated regulation is complex and requires a system-level analysis (Braithwaite et al., 2005; Vousden and Prives, 2009). However, a minimal regulatory network model can be constructed based on four cell fate-determining genes for this system (Figure 5.1). This model considers how the p53 suppressor is produced under the oxidative stress of DNA and its binding with the promoter region of these genes. According to our proposed model, the p53 gene transcribes p53 suppressors that undergo degradation under normal conditions (Lavin et al., 2006). It happens through a negative feedback mechanism in which the p53 suppressor targets the MDM2 gene that produces the negative regulator MDM2 protein, and then MDM2 selectively

promotes p53 degradation through a p53-MDM2 interaction (Lahav, 2008). Therefore, the negative feedback loop maintains a low p53 level and shows oscillatory dynamics in their population (Lahav, 2008). However, if DNA is damaged under stress, it activates various other protein kinases, such as ATM, ATR, Chk1, and Chk2, which promote the phosphorylation of p53 and MDM2. As a result, the interaction between p53 and MDM2 is disrupted, stabilizing p53 and enhancing its transcriptional activity. The excess p53 undergoes phosphorylation by ATM protein that controls the populations of p53* and p53. The p53* inhibits binding with MDM2 but activates several stress response programs, including cell cycle arrest, DNA repair, and programmed cell death (Lahav, 2008). The p53* binds with the promoter region of four genes, Bax, p21, p48, and PAI, competitively, determining the cell fate (Saramäki et al., 2006; Bieging et al., 2014; Harris and Levine, 2005).

Moreover, other oncogenes such as RAS, MYC, and β-catenin have a common regulatory element, namely E2F, that produces ARF protein, further regulating the population between phosphorylated and unphosphorylated MDM2 (Zindy et al., 1998; Weber et al., 1999). However, the production of ARF is inhibited by a p53* suppressor. A small molecule, Nutlin, binds with the MDM2, prevents the binding with p53, and inhibits the degradation of the p53 suppressor (Vassilev et al., 2004; Tovar et al., 2006). The p53* suppressor simultaneously activates WIP and miR genes, producing WIP1 protein (Fiscella et al., 1997; He et al., 2007).

Moreover, these four cell fate-determining genes are often interconnected but vary among cell lines. We focus on three cancer lines: a) breast cancer (MCF-7), b) colorectal cancer (HCT116), and c) blood cancer (K562) to compare their expressions (Tor et al., 2015; He et al., 2020; Georgakilas et al., 2017; Ehrhardt et al., 2013; Yen et al., 2020; Harris and Levine, 2005; Drullion et al., 2012). The network topology among the four genes of these three cell lines is different, producing qualitatively different gene expressions and determining the cell fate (Fan et al., 2020; Mirzayans et al., 2015). In MCF-7, the p21 activates both p48 and Bax genes along with the binding of p53* to the promoter regions of four genes (Tor et al., 2015; He et al., 2020). This effect finally leads to the cell cycle arrest and induces apoptosis in a drug-dependent manner mediated via twist regulation (Ranganathan et al., 2015). In the case of HCT116 cell lines, the p21 represses

both p48 and Bax genes (Ehrhardt et al., 2013; Georgakilas et al., 2017). Since this interaction network differs from the MCF-7, it produces different expressions. However, the p21 activates p48 and represses Bax genes for the K562 cell line (Yen et al., 2020; Fan et al., 2020; Drullion et al., 2012). Such difference in interconnectivity among the genes enhances differentiability in gene expressions, which requires careful attention to repair a cell.

In this work, we aim to explore three crucial questions through our theoretical analysis: a) Whether binding of the p53* to the promoter regions of the four cell fate-determining genes, Bax, p21, p48, and PAI, show phase transition characteristics. b) To investigate the role of internal networks among different cell fate-determining genes that add an extra layer of control on expression. c) To study the role of such internal networks for different cancer cell lines associated with breast, colon, and blood cancer cell lines, i.e., MCF-7, HCT116, K562. Our proposed thermodynamic model for the biological network revealed that the differential gene expression that decides a cell's fate is predominantly associated with internal networks among different cell fate genes and the free energy of binding among protein-DNA and protein-protein interaction. Our study considers only the p53* that participates in binding events and determines a cell's fate. We organize this chapter in the following way. We first discuss the proposed minimum free energy model and grand partition function for the biological network. Then, we discuss the results, and finally, we conclude our findings and explain the role of network topology among four genes for different cancer cell lines in regulating the gene's expressions in a p53*-dependent manner.

5.2 Model

5.2.1 Minimal Free Energy Model to Calculate Equilibrium DNA-Transcription Factor Binding:

Here, we adopt the minimum free energy model for protein binding to a large DNA molecule (Teif et al., 2002; Lando and Teif, 2000). In our model system, p53* proteins bind to the promoter regions of the cell fate-determining genes. Specifically, the p53* repressor can reach all promoter regions of those genes, forming a protein-DNA interac-
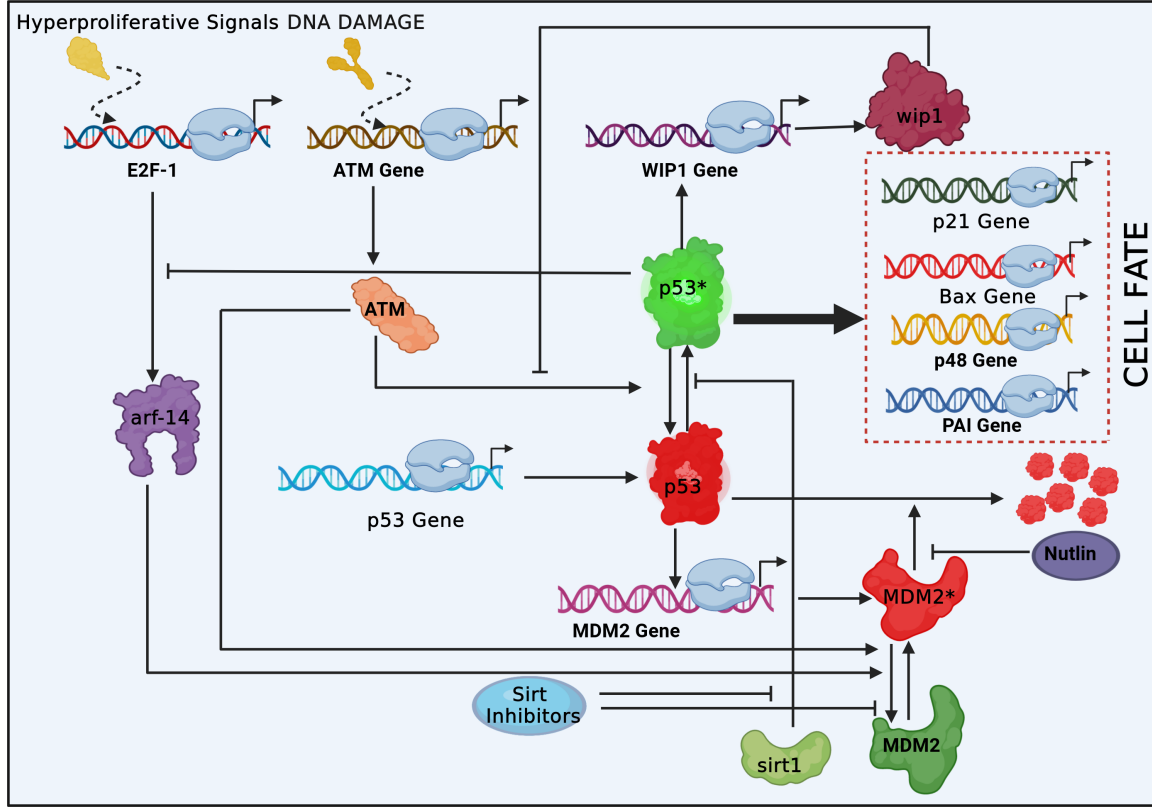
Figure 5.1: The schematic figure shows how a cell responds under environmental stress in the p53-mediated signaling pathway. The stress triggers cascading events ranging from DNA damage to synthesizing p53* to the cell's fate. Various labeled cartoons are used to depict different proteins and genes. The tetramer of p53* that binds to the four cell fate genes is also marked in the figure. The binding of p53* to the promoter regions of four cell fate-determining genes is competitive, as shown in the dotted box. The internal connectivity among the cell fate genes, which may be linked with various cell lines, also plays a critical role in binding the tetrameric p53* to the promoter regions of these genes. The relative population of these protein-DNA complexes determines the fate of a cell.

tion network responsible for cell fate determination. The minimum free energy model allows us to determine whether this protein-DNA interaction provides a signature of phase transitions in a cell. With this aim, we propose a free energy function, $F(k, b, \sigma)$, for binding k proteins to the DNA. We consider a total of N lattice sites on DNA, and the initial concentration of p53* is c_0.

$$F(k, b, \sigma) = k \times \epsilon_0 - k \times G(c) + b \times \epsilon_1 + \sigma \times \epsilon_2 - R \times T \times \log W \qquad (5.1)$$

The first term contributes the free energy (ϵ_0) for the direct binding of k molecules of p53* repressor with DNA. The second term, $G(\frac{k}{N} = c)$, is associated with the free energy of interaction of a bound protein with all other bound proteins. Such interaction may originate from changes in DNA structure that alter the topology of the protein-DNA interactions network. In the third term, we include the nearest neighboring interactions ($b \times \epsilon_1$) among all the bound p53 on the lattice. Here, b is the number of contacts formed by the binding of k p53* to the DNA, and the free energy associated with each contact is ϵ_1. The values of b lie within the interval $0 \leq b \leq k - 1$. We include the free energy contribution for the phosphorylation bound p53 through a factor as given in the fourth term. Here, σ and ϵ_2 indicate the degree of phosphorylation and the free energy associated with each interaction and the values of σ lie in the interval $0 \leq \sigma \leq k$. The fifth term is associated with the entropy contribution due to the number of ways to bind k p53 repressor, which form b p53-p53 contacts in the DNA of lattice sites N. If a p53* covers m lattices of the DNA, then one can calculate the combinatorial factor, W, by the following.

$$W = \frac{(N - m \times k)!}{[N - (m+1) \times k + b)]!(k - 1 - b)!} \times \frac{(k-1)!}{(k-1-b)!b!} \times \frac{k!}{(k-\sigma)!\sigma!} \qquad (5.2)$$

Since the initial concentration of the p53* is c_0 in the solution, the change in free energy upon the addition of DNA to the solution may be written as

$$\Delta F(k, b, \sigma) = F(k, b, \sigma) - k \times [\mu_0 + RT \log(c_0)] \qquad (5.3)$$

The last term above the equation is the loss of free energy that appears due to the binding of k p53* from the solution. Here, the chemical potential of the p53* in the solution is $\mu = \mu_0 + RT \log(c_0)$, where μ, μ_0 are the chemical potential of the p53 in solution, and its standard state, and R, and T are the gas constant and absolute temperature. Once we define the free energy function for the protein-DNA network, we can optimize it to obtain its equilibrium properties. Upon applying the Stirling approximation for the large values of k, b, σ and N, we obtain their most probable or the equilibrium values k^* or $c^* = k^*/N$, b^* and σ^*.

$$\exp\left(\frac{\mu_0 - \epsilon_0}{RT}\right) \cdot \exp\left(\frac{2wc^*}{RT}\right) = \frac{(1 - mc^*)^m}{1 - (m+1)c^* + z_1} \frac{(c^* - z_1)^2}{c^*} \frac{(c^* - z_2)}{c_0 c^*} \tag{5.4}$$

$$\exp\left(-\frac{\epsilon_1}{RT}\right) = (1 - (m+1)c^* + z_1)\frac{z_1}{(c^* - z_1)^2} \tag{5.5}$$

$$\exp\left(-\frac{\epsilon_2}{RT}\right) = \frac{z_2}{(c^* - z_2)} \tag{5.6}$$

Where, we define $K = \exp(\frac{\mu_0 - \epsilon_0}{RT})$ is the binding constant, $A(c^*) = \exp\left(\frac{2wc^*}{RT}\right)$ is related to the long-range interactions, $a_1 = \exp(-\frac{\epsilon_1}{RT})$ and $a_2 = \exp(-\frac{\epsilon_2}{RT})$ are the factor of contact interactions of bound ligands and phosphorylation: $z_1 = b^*/N$, $z_2 = \sigma^*/N$ are the relative concentration of ligand-ligand contacts and phosphorylation. Note that we apply the Stirling approximation on N, k, b and m during the optimization. The above equations ((5.4), (5.5) and (5.6)) are solved simultaneously and then insert into the equation (5.3) $\Delta F(c^*, b^*, \sigma^*)$ for a given c_0.

$$
\begin{aligned}
\frac{\Delta F^*}{NRT} = &-c^* \log(K \times B^* \times c_0 \times (c^*)^2) - z_1 \log(a_1) - z_2 \log(a_2) - \\
&(1 - mc^*) \times \log(1 - m \times c^*) + 2(c^* - z1)\log(c^* - z1) + \\
&(c^* - z2)\log(c^* - z2) + z_1 \log(z_1) + z_2 \log(z_2) \\
&+ (1 - (m+1)c^* + z_1)\log(1 - (m+1)c^* + z_1)
\end{aligned}
$$

where, $B^* = \exp(G(c^*)/RT) = \exp(wc^*/RT)$.

5.2.2 Grand Partition Function For the Network

The above model provides an understanding of cell transitions, but it poorly considers the topology of the network. We use the average values of the topology parameter, w, obtained from our simulations. Therefore, we propose a network model that explicitly considers the topology of the protein-DNA interactions. A gene regulatory network (GRN) is a directed graph in this model. The elements of this graph are proteins or transcription factors, and they bind to the promoter regions of genes. In this problem, four genes, p21, p48, PAI, Bax, and p53*, form a small network. Since a cell's fate is

determined by the binding of p53* to the four genes, p21, p48, PAI, and Bax, we consider this small network to identify the binding patterns and infer the gene expression. The wiring of the networks is critical because they vary for different cell lines (Fan et al., 2020; Mirzayans et al., 2015). Thus, one can expect heterogeneous cell fate for different cell lines.

We define the network as a directed graph (G), where we consider genes as vertices (V), and the protein-promoter and promoter-promoter interactions form the edges (E). The p53* binds to the promoter region of these four genes, and the interaction among the genes exists either in a protein occupancy-dependent or independent manner. In particular, one can realize the protein occupancy-dependent edges if proteins occupy both participating promoters. We further consider that a saturated level of RNAP is present in the system. With this aim, we define the total energy function or Hamiltonian of the network.

$$H(G) = \sum_i \sum_{s_i=0}^{1} \epsilon_i s_i + \sum_{i \neq j} \sum_{\sigma_{ij}=0}^{1} \omega_{ij} \sigma_{ij} \tag{5.7}$$

Here, the ϵ_i and ω_{ij} are the free energy of protein-DNA and promoter-promoter interactions; s_i takes a value of 1 if there exists a direct protein-DNA interaction; otherwise, it takes value 0. The σ_{ij} is the adjacency matrix for the interaction among genes, which takes a value of either 1 or 0 depending on connections. If the total no of proteins is N, then $\sum_i s_i = N$. If we assume s_i and σ_{ij} are independent, then the Grand partition function for a network is

$$\begin{aligned}
\Xi &= \sum_{N=0}^{\infty} \sum_G \exp(-\beta H) \lambda^N \\
&= \sum_{N=0}^{\infty} \sum_G \exp(-\beta(\sum_i \sum_{s_i=0}^{1} \epsilon_i s_i + \sum_{i \neq j} \sum_{\sigma_{ij}=0}^{1} \omega_{ij} \sigma_{ij})) \lambda^{\sum_i s_i} \\
&= \sum_{\{s_i, \sigma_{ij}\}} \exp(-\beta(\sum_i \sum_{s_i=0}^{1} \epsilon_i s_i)) \lambda^{s_i} \exp(-\beta \sum_{\sigma_{ij}=0}^{1} \omega_{ij} \sigma_{ij}) \\
&= \prod_i (1 + \lambda \exp(-\beta \epsilon_i)) \prod_{i \neq j} (1 + \exp(-\beta \omega_{ij}))
\end{aligned}$$

where λ is the $\exp(\beta\mu)$ and μ is the chemical potential of p53*. One can immediately calculate the average occupancy of p53* on each of the genes by taking the $\lambda\frac{\partial\ln\Xi}{\partial\lambda}$. We express our results in terms of fold change (FC), which is defined as the ratio between the occupancy of RNAP in the presence of p53* and the occupancy of only RNAP (basal) in the network. However, in practice, genes are influenced by each other through protein-protein interaction, which immediately restricts the application of the independence between s_i and σ_{ij}. To explore all the effects together, we first list all possible network configurations as shown in figure 5.2. We then identify all the configurations where at least one p53* is bound to a particular promoter. The ratio of statistical weights between those configurations and the total number of configurations provides a signature of the probability of gene expression of that particular gene. Note that the counting process becomes exceedingly complex as the network size increases. Therefore, one can employ Monte Carlo (MC) simulations that offer an elegant approach to identifying those relevant configurations for expressing a particular gene in a large network (Gautam and Kumar Sinha, 2021; Gautam and Sinha, 2023).
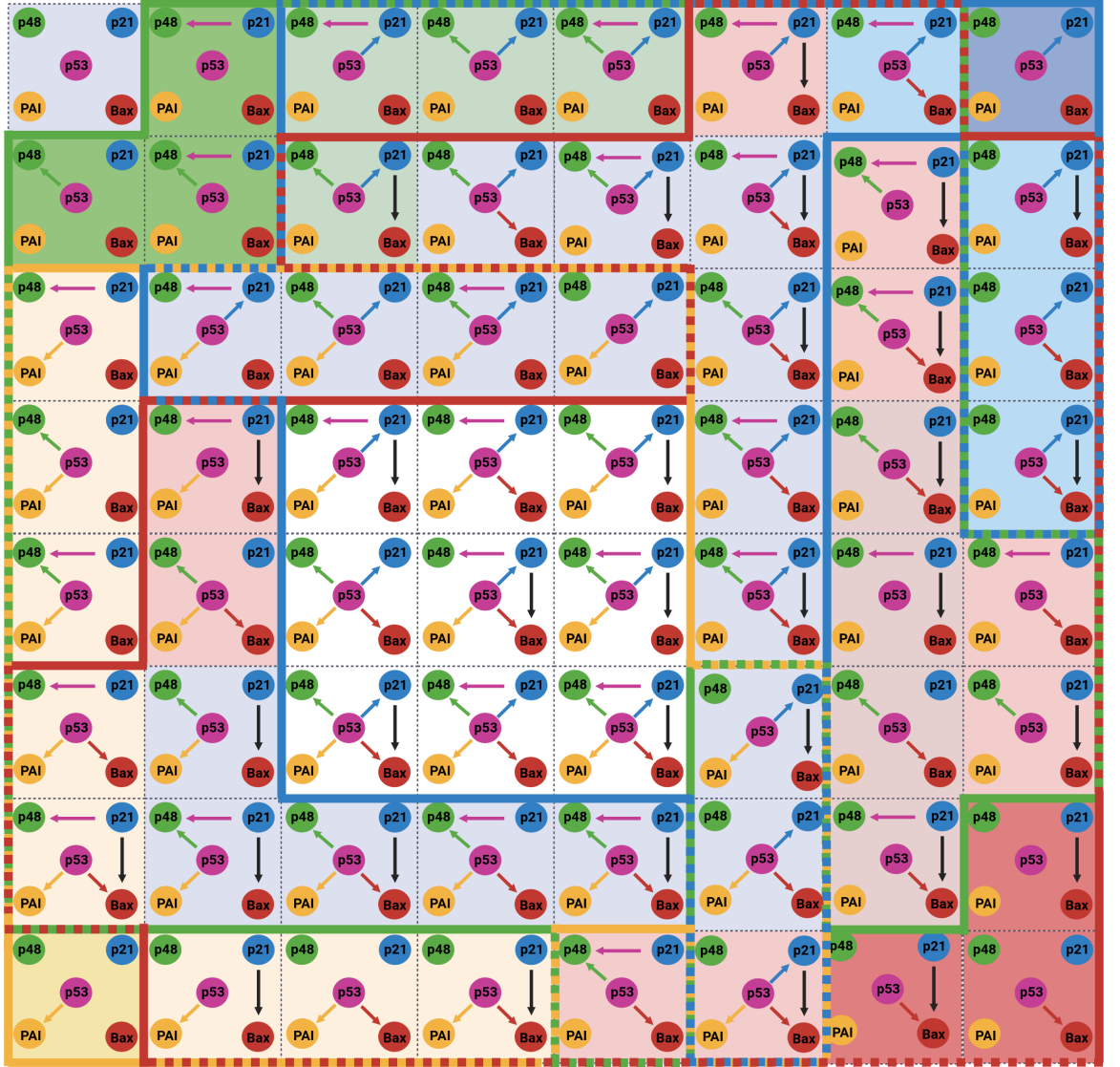
Figure 5.2: The figure demonstrates all sixty-four possible microstates for the gene regulatory network of the MCF-7 cancer cell line. We calculate the statistical weights, thereby the population of each configuration, using the grand partition function. We marked green, blue, yellow, and red background color regions for the configurations associated with cell fate genes, p48, p21, PAI, and Bax genes, respectively. The solid and shed backgrounds are used to show the exclusive and shared configurations for a specific gene. Similarly, the solid and dashed lines of different colors are used to show the border for a specific gene's exclusive and shared configurations. The configurations located at the middle white background region are common to all genes.

5.3 Results

We start with binding the p53* to the promoter regions of the four cell fate-determining genes, Bax, p21, p48, and PAI. We further study the importance of the topology of a network of interactions that alters a cell's fate. We rationalized the network topology with different cell lines, MCF-7, HCT116, and K562, and found that the gene expressions are quite different for these three lines (Tor et al., 2015; He et al., 2020; Georgakilas et al., 2017; Ehrhardt et al., 2013; Yen et al., 2020; Harris and Levine, 2005; Drullion et al., 2012), triggering diverse cell fate. We discuss them in detail below.

5.3.1 p53-Dependent Cell Transitions from Thermodynamic Models:

We first aim to identify if there are any phase transitions on the binding of p53* to the promoter regions of the genes. Since the topology of the network of interactions varies from cell to cell, we study this behavior for three different cell lines that we consider in this study. We probe two critical questions: a) Will the binding of p53* to the Bax, p21, p48, and PAI genes show phase transition characteristics in a cell? b) if it happens, what will be the role of the topology of the interaction network for these transitions? Understanding phase transitions for this system at a fundamental level is crucial for the preventive measurements. We chose the fraction of bound p53* at equilibrium, c^* as our order parameter to study the phase transition behaviour (Callen, 1991). We further vary the topology parameter, w , defined as the average number of edges of the network, to mimic the cellular variation. We use the w values obtained from our GCMC simulations for the three different cell lines ($w =2.02$, 2.97, and 4.03) as considered in our study.

We calculate c^* as function of their initial concentration c_0 of p53* shown in the Figure 5.3. We also consider three cell lines by varying the w parameter in this study. We notice an S-shaped curve for each cell line, a signature of phase transition exhibited in a cell upon binding the p53*. To confirm whether this phase transition is continuous or discontinuous, we calculate the equilibrium free energy of interaction ΔF^* as included in the inset of the figure. The free energy analysis reveals that the system

exhibits bistability that passes through a barrier. We mark the free energy data on the order parameter c^*, which reveals a clear signature of abrupt transitions. Therefore, we characterize this system as exhibiting first-order phase transitions upon binding the p53* using the minimal free energy model.

The concentration of p53* at which the model network exhibits a transition is called the potency point, and the point at which the p53* binding to the promoter saturates is called the efficacy point (Martins et al., 2006). We draw a line between these two points to compare a qualitative signature of these transitions among the three cell lines. They are different, as revealed by our analysis. We notice that the potency points for the HCT116, K562, and MCF-7 cell lines are at $2.2\mu M$, $1.33\mu M$, and $0.81\mu M$, showing that HCT116 requires more p53* to saturate the promoter. However, MCF-7 takes the least amount of p53* to saturate the promoter. We also notice that the saturation level for HCT116 is relatively low compared to the other two cell lines. The origin of differential behavior in these three cell lines' phase transitions primarily arises from the difference in network topology, which we included through the w parameter, as discussed earlier (Teif et al., 2002; Lando and Teif, 2000; Teif, 2005).

5.3.2 Role of Network Topology

p53 is a crucial transcription factor that determines cellular fate by interacting with the promoter regions of various genes with a wide range of binding affinity. Here, four genes, p21, Bax, p48, and PAI, participate in p53* binding, and the degree of its binding determines a cell's fate (Harris and Levine, 2005; Hafner et al., 2019). The differential expressions from these four genes upon binding p53* lead to the initiation of cell cycle arrest, apoptosis, DNA repair, and inhibition of angiogenesis. In other words, such binding events control the cancer progression. These differential expressions promote the cellular transitions and determine the actual state of the cell as a consequence. In other words, it determines the fate of a cell. However, the binding of the p53* may be influenced by many other factors, such as a) internal rewiring of the GRN, b) interfering with other proteins, c) abrupt external perturbation, etc. Since we restrict our study to an isolated GRN within an equilibrium regime, we ignore all other factors except internal rewiring. To demonstrate the importance of internal rewiring among genes, we
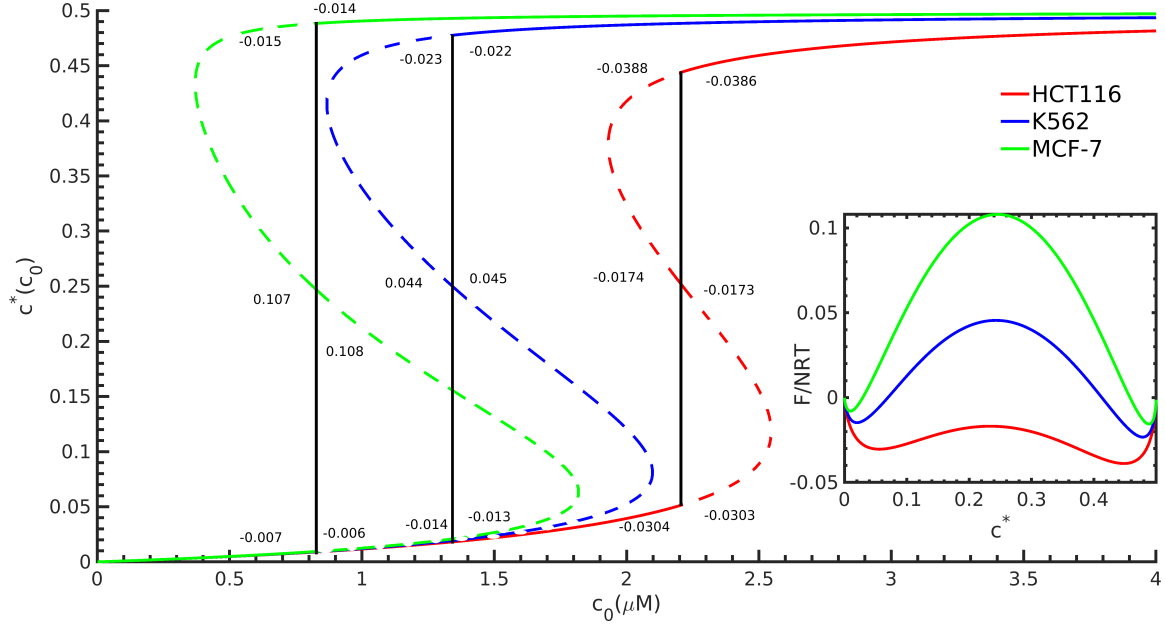
Figure 5.3: We compare binding curves among various cell lines based on their network topology. Here, the values of the parameters used are $m = 2$, $a_1 = e^3$, $a_2 = e^9$, $K = e^9$. The connectivity parameters (w) we obtained from GCMC simulations for three cell lines are 2.02, 2.97, and 4.03, respectively. The dotted lines are for the unstable region obtained from the free energy analysis. The network topology governing the cell lines influences the first-order cellular transition.

first perform a comparative study between gene expressions upon binding p53* to the promoter regions of these four genes in the absence and the presence of an arbitrary internal network. Note that the p53* binds to the four genes independently through their free energy of interactions without wiring among genes.

In Figure 5.4, we present the response function in the absence of any internal networks among the promoters of cell fate-determining genes, and we observed that the binding of p53* on genes takes place sequentially as a function of its increasing concentration values owing to its wide range of binding affinities. However, we find a significant change in the gene expression pattern when considering a few arbitrary links among these genes. We present the response functions for both cases in Figure 5.4. The rationale behind adding a few links among these four genes is that they vary from one cell line to another. Since this system is responsible for cancer progression, various cancer cell lines are found with different internal links (Jangili et al., 2022; Georgakilas et al., 2017; Han et al., 2002;

Fan et al., 2020; Li et al., 2015; Rodriguez and Meuth, 2006; Tor et al., 2015; He et al., 2020).

Without internal links, we observe a sequential expression for these genes, but all are expressed at the same saturation level at large values of λ_{p53}. In other words, we infer that these four genes express differentially at low values of λ_{p53} , but their expressions are almost identical at large values of λ_{p53}. The differential expressions at low values of λ_{p53} appear due to the different binding affinities of the promoter regions of these four genes to the p53*. Therefore, if a cancer cell consists of these four genes without internal links, the cell reaches equiprobably to any of these four states. However, internal links show differential expressions, even at large λ_{p53} values. The results indicate the importance of internal links that determine a differential expression. For example, all four genes express differentially upon introducing two inhibitory loops. We also notice that the cell primarily initiates the inhibition of angiogenesis as the p48 gene is preferentially expressed for this network. Therefore, it is clear from our demonstration that the internal networks among genes play a crucial role in determining the state of a cell. The internal networks for different cancer cell lines significantly differ. We discuss them in detail below (Mirzayans et al., 2015; Mayo et al., 2005; Fan et al., 2020).

5.3.3 Comparison of Different Cancer Cell Lines

The immortal cancer cells uncontrollably divide and multiply with the progression of time. Human cancer cell lines have received considerable attention in the biology community to understand their origin and develop a method that inhibits cancer progression (Cortes-Ciriano et al., 2016; Abdullah et al., 2009; Guo et al., 2019). Understanding their origin and the mechanism behind cancer progression is crucial to exploring the potency and efficacy of existing cancer treatments and finding new efficient treatments (Martins et al., 2006). Since cancer can progress through a p53-dependent pathway (Powell et al., 2014), it is essential to have a better view or understanding of the underlying mechanisms that regulate the binding of p53 to various cell fate-determining genes. As described in the introduction, for a cell under stress, these four genes are part of the concerted gene expression programs responsible for various fundamental biological processes and the underlying cell fate determination, orchestrated by complex gene regu-
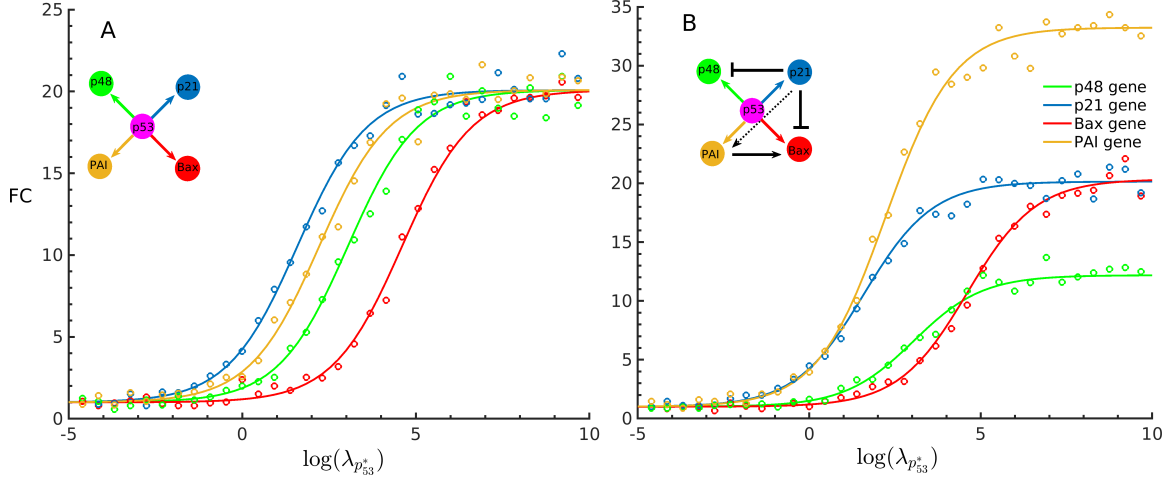
Figure 5.4: The FC for different genes in the absence and the presence of internal connectivity. The binding of p53* to the promoter regions of the four cell fate genes is different under these two situations. We show the respective networks in the inset of the figure. We notice differential gene expressions for systems with internal connectivity. Such heterogeneous expression is absent for the system with simple binding events based on their affinity. Here, solid lines are for the theoretical results, and circles represent the simulation results.

latory networks centered around p53 binding (Wilkinson et al., 2017; Harris and Levine, 2005). However, the genetic mutations in these genes are associated with abnormal expression, which prolongs the tumor growth (Pryczynicz et al., 2014). Various cancer cell lines are reported in the literature where the internal links among these four genes are quite different. Therefore, we focus on three cancer lines: a) breast cancer (MCF-7), b) colorectal cancer (HCT116), and c) blood cancer (K562) to compare their expressions. The internal links among the four genes of these three cell lines are different, producing qualitatively different gene expressions and determining the cell fate.

We performed partition function-based analysis and GCMC simulation for each of the networks. It is evident from the figure that the internal connectivity among genes alters the expressions significantly upon binding of p53* to the promoter regions of four cell fate-determining genes. Our analysis shows that these four genes play an essential role in repairing a malignant cell by different modes of action (Torgovnick and Schumacher, 2015; Alhmoud et al., 2021; Wang et al., 2021a). Our analysis reveals that the MCF-7 cell line predominantly expresses the p48 and Bax genes, which either initiate the process of repairing DNA or follow apoptosis. We find significantly different expressions

between the HCT116 and the MCF-7 cell lines. As mentioned above, the origin behind differential expression for this cell line is related to the network's topology. As observed from our analysis, the repression effect of the p21 gene triggers the activation of PAI and p21 genes, which either inhibit angiogenesis or arrest the cell cycle. However, the p48 gene expresses exclusively for the K562 cell line, as evidenced by our analysis. Thus, it is clear from our analysis that the impact of p53* in controlling the expression for the cell fate-determining genes by forging the connection between an input signal and the network topology (Benstead-Hume et al., 2022; Hafner et al., 2019).
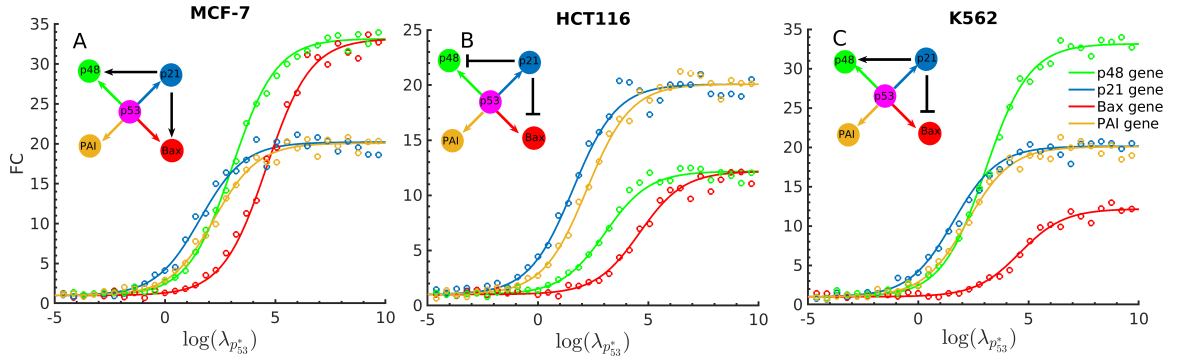
Figure 5.5: The Gene expression FC of different genes obtained for three cell lines are shown here. Since the internal connectivity among the genes for the cell lines is different, we show them in the figure. The binding of p53* to the promoter regions of the four genes is heterogeneous, exploring the possibility of differential gene expression and a cell's fate. Here, the solid lines and circles represent the theoretical and simulation results, respectively.

5.4 Discussion

In this work, we put an effort to encapsulate multiple biophysical processes and gene networks to understand the fine-tuned expression and cell fate. Our modeling scheme and findings establish a predictive protocol that can be employed to understand transcription patterns for vast possibilities of cellular fates. Our analyses for the different cell lines based on the network topologies are known for four genes, p21, Bax, p48, and PAI (Tor et al., 2015; He et al., 2020; Georgakilas et al., 2017; Ehrhardt et al., 2013; Yen et al., 2020; Harris and Levine, 2005; Drullion et al., 2012). The p53 maintains its low level in a healthy cell by a continuous degradation mechanism through an interaction

with a protein MDM2 (Alarcon-Vargas and Ronai, 2002; Kubbutat et al., 1997; Haupt et al., 1997). A cell under stress damages DNA that produces a p53*, which inhibits binding with MDM2, and thus, the concentration of p53* is stabilised (Chehab et al., 1999). The tetramer of p53* controls many cellular processes, including cell fate, i.e., cell cycle arrest, apoptosis, senescence, inhibition of angiogenesis, DNA repair, metastasis, etc. (Harris and Levine, 2005). For example, the programmed cell death or cell cycle arrest is triggered by activating a p21 gene upon binding tetrameric p53* that produces the p21 protein. The p21 protein binds with another cyclin-dependent kinase complex (CDK/cyclin) that promotes cell cycle arrest (Bissonnette and Hunting, 1998; Rokudai et al., 2009). Failing to bind p53* to these cell fate-determining genes promotes human cancer progression. Therefore, it is urgent to understand the mechanisms behind p53* dependent cellular functionality for curing early-stage cancer progression (Marei et al., 2021).

As mentioned above, the system exhibits multiple layers of control, ranging from phosphorylation of p53 protein that alters its biological activity and has a wide range of binding affinity to the promoter regions of the cell-fate-determining genes (Chao et al., 2006; Mayo et al., 2005). These genes, often connected, form a network of interaction to determine the cell fate. Here, we seek to understand the role of p53* in determining the cell fate of a malignant cell through the lens of the minimum free energy model. Our model offers an understanding of the role of network topology in regulating the expression of various cell fate-determining genes. We show that the degree of binding of p53* to the promoter regions of these genes shows phase transition characteristics. The transition is discontinuous or first order, as revealed by the sign of free energy data at equilibrium (Callen, 1991). We further explore the phase transition behavior for different cancer cell lines, MCF-7, HCT116, and K562, as they have the same constituents, but the network topologies differ. The results reveal that phase transition behavior strongly depends on network topology, which is also explored by a few other studies (Teif et al., 2002; Lando and Teif, 2000).

To further pin down the actual cellular state of the cell, we propose a grand partition function for the biological networks and perform GCMC simulations of them to explore the microscopic origin of the cell fate (Hill, 1986). We find a strong dependence between

cell fate and the network topology. The MCF-7 cell line presents an example where the p21 protein regulates the expression of p48 and Bax genes, which are associated with initiating the DNA repair process and apoptosis. However, p21 represses p48 and Bax genes in the HCT116 cell line, ultimately activating p21 and PAI genes, inhibiting angiogenesis, or arresting the cell cycle. However, the p48 gene expression exclusively promotes the initiation of DNA repair process in the K562 cell line. Therefore, the diverse range of outcomes obtained from the different cell lines governed by the network topology provides a clear understanding of the actual state of an early-stage malignant cell (Benstead-Hume et al., 2022; Rashid et al., 2022).

Cell phenotype is often the result of key transcription factors that regulate the expression and are inherently related to the cascading event of the cell fate decisions (Jia et al., 2017). Sometimes, a transcription factor can decide the cellular fate based on the cell's context. As the tetrameric p53* decides to activate one of the genes among many genes based on the sequence-specific response element (Kern et al., 1991). A prime question in this regard is how p53 decides the cellular fate based on a wide range of binding affinity dependent on the promoters and internal networks for the cell fate-determining genes. We explored this question here in this direction. However, there is a possibility that some other factors can dominate over network topology, such as the entanglement of these genes with some common genes, which can either elevate or bring their expression down, dysregulated transcriptional programs that result from genetic mutations, epigenetic regulation, RNA stability, protein translation, post-translational control, etc. Understanding each factor offers the opportunity to cure malignant to healthy cells. Our proposed theoretical method could include those factors in the model, which we are currently exploring.

Chapter **6**

Conclusions

6.1 Key Findings

This thesis provides a detailed understanding of protein-DNA interaction and its impacts on gene regulation. Specifically, we investigate various biomolecular mechanisms that regulate genes in various organisms, ranging from prokaryotes to eukaryotes. We explored gene regulation for a system under an equilibrium regime and dynamically evolving biological systems. While dealing with biological systems under equilibrium, we made a partition function-based thermodynamic model and counter-validated its calculations with our in-house Grand Canonical Monte Carlo (GCMC) based simulations. While considering dynamically evolving systems, we analyzed the master equations as well as performed Kinetic Monte Carlo (KMC) simulations. We explored the gene regulation of various biological systems. We established quantitative relationships between the shape of the response curve and the underlying biophysical mechanism that occurs at a molecular level.

The quantitative prediction of the shape of the genetic response curve from the underlying molecular mechanism needs careful attention. To explore it, we first perform *in silico* binding studies of transcription factors to the promoter regions of genes for various gene regulatory systems. The detailed mechanism of protein-DNA interactions in gene regulatory systems relates to the ordered genetic response of a complex and noisy biological cell. We show that the TFs bind to multiple DNA sites with high cooperativity and spread from one non-specific binding site to an adjacent specific binding site of DNA. Furthermore, the DNA undergoes loop formation through the long-distance protein-protein interaction. Therefore, our calculation considers various controlling fac-

tors, such as TF oligomerization, TF-ligand interactions, and DNA looping. We apply our developed method to an isolated gene and a system containing multiple interacting genes. We validate our model against the published data for the bacterial *lac operon* system. In short, our partition function-based theoretical frameworks and GCMC simulation can predict the shape of the response curve by following the detailed protein-DNA interactions in a gene regulatory system. We infer that the shape of the gene expression curves can be altered by modulating TF-ligand interactions and the looping energy of DNA.

Cooperative protein-protein and protein-DNA interactions form programmable complex assemblies. These assemblies produce a precise functional response that can strongly depend on the topology of those networks. We demonstrated how the coordinated self-assembly creates gene regulatory network motifs that corroborate the existence of a precise functional response at the molecular level using thermodynamic analysis. We performed partition function-based calculations and Monte Carlo simulations to show that a complex network of interactions, such as feedback and feedforward circuits, can form a decision-making loop only by a few molecular mechanisms. We characterize each possible network of interactions by systematic variations of free energy parameters associated with the binding among biomolecules and DNA looping. We further show that the self-assembly of proteins on DNA differentially promotes logic operations such as AND, NAND, NOR, and OR. We can account for the origin of various logic gates formed in gene regulatory networks through the quantitative thermodynamic model of transcriptional regulation and systematic variation of promoter-enhancer interaction modes. We further show that the inter-conversion or switching among various logic gates yielded under the systematic variations of the stimuli-TF binding and DNA looping parameters. Our calculations establish that the network topology is vital in phenotype diversity in regulatory circuits.

We find a strong correlation between our dynamic and thermodynamic analysis. Our dynamic analysis shows that the functional response for gene regulatory networks strongly depends on the network's topology. We calculate the protein production from the solution of the chemical master equation. We further use the stochastic trajectories to calculate the stochastic potentials. We found that multi-stability is a common

feature in GRN, often linked with a biological system's phenotype diversity. Our dynamical analysis of the biomolecular assembly between protein and DNA shows that the functional responses strongly depend on the network's topology. We find a close correspondence between the system's nonlinearities and the noise of a biological system. We find noise-induced bimodality upon introducing short-range correlated noise defined by the OU process. The noise-induced bimodality for a gene regulatory network may be linked with the different phenotypes that appear due to the prolonged degradation of a protein molecule. Our analysis shows that an input signal is processed through a noisy network of interactions that shifts the distribution from unimodal to bimodal. Our analyses establish a close correspondence between the system's nonlinearities and inherent noise in a biological system. We find a close correlation between an input signal and the multi-modal distribution in protein production as determined by the interaction network. We also observed that the clustering of transcription-regulating proteins enhances the sharpness in the distribution of the produced protein.

Lastly, we study the role of tumor suppressor p53 to understand the origin behind the fate of a damaged cell. The tumor suppressor p53 responds to stress by selectively triggering one among many potential transcriptomes that influence cellular fate decisions. We develop thermodynamic models for this purpose. Our model explores the fate of a damaged cell governed by the molecular interactions present in the tumor suppressor p53 signaling network. The control involves many layers of input signal modification, ranging from phosphorylation of p53, the wide range of binding affinity of p53 with the promoters of various genes, to internal connectivity among cell fate genes. We show that minimum free energy is a fundamental property of biological networks that establishes a connection between cell fate and network topology. This model offers a robust understanding of the cell fate and unriddles the enigma between the network topology and the cell transition. We apply our developed model to various cancer cell lines ranging from breast cancer (MCF-7), colon cancer (HCT116), and leukemia (K562) that exhibit different network topologies. We find that the network topology of these different cell lines determines the fate of a damaged cell. Our developed model correlates the biological relevance of these mechanisms and suggests that they represent general archetypal designs for developmental decisions.

6.2 Future Directions

Our findings highlight the understanding of the connection between biomolecular mechanisms and characteristic response functions for various complex biological systems ranging from prokaryotes to eukaryotes. However, many questions remain unanswered. The following questions could be of interest to the research community in this direction:

We have employed the coarse-grained modeling scheme to understand gene regulation using statistical thermodynamics. Nevertheless, a system-level multi-scale approach will enhance our understanding of these biophysical phenomena. Thus, comparing atomistic, coarse-grained, and continuum scales will give a better view of gene regulation for various complex biological systems. Therefore, developing such multi-scale models and mapping them over experimental data will be interesting. In this direction, Molecular Dynamics simulations can estimate the free energy of interactions between protein and DNA. One can plug these derived interaction energies into our developed model for understanding the *in silico* studies of gene regulation.

We explore gene regulation primarily using thermodynamic models and a few stochastic models in this thesis. However, spatiotemporal-based models could be an immediate extension of our work, which could help us understand many other relevant biological phenomena, such as tumor growth. Such models may establish a correlation between the underlying biophysical processes that promote cellular organization and tumor growth at the gene level.

Stochastic fluctuations are significant in various biochemical reactions when the number of biomolecules is low inside a living cell. Many reactions are not instantaneous; instead, they have a natural time delay during the evolution of cell states. Thus, a possible extension of our work could be to include such time delays in the dynamic studies of gene regulation in our modeling scheme. It will be interesting to systematically explore stochastic dynamics and time delays and include their combined effects in our modeling schemes.

Finally, theoretical and computational models that enhance our understanding of the correlation between biomolecular mechanisms and characteristic response curves at the genomic level will undoubtedly be a step toward developing disease-specific therapies.

References

Abdullah, N.M., Rosania, G.R., and Shedden, K. (2009). Selective targeting of tumorigenic cancer cell lines by microtubule inhibitors. *PLoS One*, 4(2), e4470.

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the national academy of sciences*, 79(4), 1129–1133.

Adhya, S. (1989). Multipartite genetic control elements: communication by dna loop. *Annual review of genetics*, 23(1), 227–250.

Ahnert, S.E. and Fink, T.M. (2016). Form and function in gene regulatory networks: the structure of network motifs determines fundamental properties of their dynamical state space. *Journal of the Royal Society Interface*, 13(120), 20160179.

Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014). A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. *BMC biology*, 12(1), 1–23.

Alarcon-Vargas, D. and Ronai, Z. (2002). p53–mdm2—the affair that never ends. *Carcinogenesis*, 23(4), 541–547.

Albert, F.W. and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 16(4), 197–212.

Alberts, B. (2017). *Molecular biology of the cell*. Garland science.

Alhmoud, J.F., Woolley, J.F., Al Moustafa, A.E., and Mallei, M.I. (2021). Dna damage/repair management in cancers. *Advances in Medical Biochemistry, Genomics, Physiology, and Pathology*, 309–339.

Ali Al-Radhawi, M., Del Vecchio, D., and Sontag, E.D. (2019). Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS computational biology*, 15(2), e1006784.

Alon, U. (2019). *An introduction to systems biology: design principles of biological circuits*. CRC press.

Angeli, D., Ferrell Jr, J.E., and Sontag, E.D. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences*, 101(7), 1822–1827.

References

Attardi, L.D., Lowe, S.W., Brugarolas, J., and Jacks, T. (1996). Transcriptional activation by p53, but not induction of the p21 gene, is essential for oncogene-mediated apoptosis. *The EMBO journal*, 15(14), 3693–3701.

Avendaño, M.S., Leidy, C., and Pedraza, J.M. (2013). Tuning the range and stability of multiple phenotypic states with coupled positive–negative feedback loops. *Nature communications*, 4(1), 1–8.

Aviziotis, I., Kavousanakis, M., Bitsanis, I., and Boudouvis, A. (2015). Coarse-grained analysis of stochastically simulated cell populations with a positive feedback genetic network architecture. *Journal of mathematical biology*, 70(7), 1457–1484.

Bailey, J.E. (2001). Complex biology with no parameters. *Nature biotechnology*, 19(6), 503–504.

Barzel, B. and Biham, O. (2011). Binomial moment equations for stochastic reaction systems. *Physical Review Letters*, 106(15), 150602.

Bashor, C.J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J.J., and Khalil, A.S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science*, 364(6440), 593–597.

Baumann, C.G., Smith, S.B., Bloomfield, V.A., and Bustamante, C. (1997). Ionic effects on the elasticity of single dna molecules. *Proceedings of the National Academy of Sciences*, 94(12), 6185–6190.

Bazett-Jones, D.P., Leblanc, B., Herfort, M., and Moss, T. (1994). Short-range dna looping by the xenopus hmg-box transcription factor, xubf. *Science*, 264(5162), 1134–1137.

Benenson, Y. (2009). Rna-based computation in live cells. *Current opinion in biotechnology*, 20(4), 471–478.

Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E. (2004). An autonomous molecular computer for logical control of gene expression. *Nature*, 429(6990), 423–429.

Benstead-Hume, G., Wooller, S.K., Renaut, J., Dias, S., Woodbine, L., Carr, A.M., and Pearl, F.M. (2022). Biological network topology features predict gene dependencies in cancer cell-lines. *Bioinformatics Advances*, 2(1), vbac084.

Bentovim, L., Harden, T.T., and DePace, A.H. (2017). Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development*, 144(21), 3855–3866.

Berg, J. (2008). Out-of-equilibrium dynamics of gene expression and the jarzynski equality. *Physical review letters*, 100(18), 188101.

Biddle, J.W., Nguyen, M., and Gunawardena, J. (2019). Negative reciprocity, not ordered assembly, underlies the interaction of sox2 and oct4 on dna. *Elife*, 8, e41017.

Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. *Nature Reviews Cancer*, 14(5), 359–370.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development*, 15(2), 125–135.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2), 116–124.

Bissonnette, N. and Hunting, D. (1998). p21-induced cycle arrest in g1 protects cells from apoptosis induced by uv-irradiation or rna polymerase ii blockage. *Oncogene*, 16(26), 3461–3469.

Black, A.J. and McKane, A.J. (2012). Stochastic formulation of ecological models and their applications. *Trends in ecology & evolution*, 27(6), 337–345.

Bocci, F., Jia, D., Nie, Q., Jolly, M.K., and Onuchic, J. (2023). Theoretical and computational tools to model multistable gene regulatory networks. *arXiv preprint arXiv:2302.07401*.

Bode, A.M. and Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer*, 4(10), 793–805.

Boedicker, J.Q., Garcia, H.G., and Phillips, R. (2013). Theoretical and experimental dissection of dna loop-mediated repression. *Physical review letters*, 110(1), 018101.

Boettiger, A.N., Ralph, P.L., and Evans, S.N. (2011). Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS computational biology*, 7(5), e1001136.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *cell*, 122(6), 947–956.

Braithwaite, A.W., Royds, J.A., and Jackson, P. (2005). The p53 story: layers of complexity. *Carcinogenesis*, 26(7), 1161–1169.

Bratsun, D., Volfson, D., Tsimring, L.S., and Hasty, J. (2005). Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences*, 102(41), 14593–14598.

References

Brown, J.D., Lin, C.Y., Duan, Q., Griffin, G., Federation, A.J., Paranal, R.M., Bair, S., Newton, G., Lichtman, A.H., Kung, A.L., et al. (2014). Nf-κb directs dynamic super enhancer formation in inflammation and atherogenesis. *Molecular cell*, 56(2), 219–231.

Brown, K.S., Hill, C.C., Calero, G.A., Myers, C.R., Lee, K.H., Sethna, J.P., and Cerione, R.A. (2004). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical biology*, 1(3), 184.

Buchler, N.E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, 100(9), 5136–5141.

Burz, D.S. and Ackers, G.K. (1996). Cooperativity mutants of bacteriophage λ ci repressor: temperature dependence of self-assembly. *Biochemistry*, 35(10), 3341–3350.

Bustamante, C., Guthold, M., Zhu, X., and Yang, G. (1999). Facilitated target location on dna by individualescherichia coli rna polymerase molecules observed with the scanning force microscope operating in liquid. *Journal of Biological Chemistry*, 274(24), 16665–16668.

Callen, H.B. (1991). *Thermodynamics and an Introduction to Thermostatistics*. John wiley & sons.

Cao, Z., Filatova, T., Oyarzún, D.A., and Grima, R. (2020). A stochastic model of gene expression with polymerase recruitment and pause release. *Biophysical Journal*, 119(5), 1002–1014.

Cao, Z. and Grima, R. (2018). Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nature communications*, 9(1), 1–15.

Cao, Z. and Grima, R. (2020). Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proceedings of the National Academy of Sciences*, 117(9), 4682–4692.

Cech, T.R. (2000). The ribosome is a ribozyme. *Science*, 289(5481), 878–879.

Chao, C., Herr, D., Chun, J., and Xu, Y. (2006). Ser18 and 23 phosphorylation is required for p53-dependent apoptosis and tumor suppression. *The EMBO journal*, 25(11), 2615–2622.

Chehab, N.H., Malikzay, A., Stavridi, E.S., and Halazonetis, T.D. (1999). Phosphorylation of ser-20 mediates stabilization of human p53 in response to dna damage. *Proceedings of the National Academy of Sciences*, 96(24), 13777–13782.

Chen, J. (2016). The cell-cycle arrest and apoptotic functions of p53 in tumor initiation and progression. *Cold Spring Harbor perspectives in medicine*, 6(3), a026104.

Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z., Deng, J., Li, Y., Wang, X., and Zhao, L. (2018). Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget*, 9(6), 7204.

Chen, Y., Kelm Jr, R.J., Budd, R.C., Sobel, B.E., and Schneider, D.J. (2004). Inhibition of apoptosis and caspase-3 in vascular smooth muscle cells by plasminogen activator inhibitor type-1. *Journal of cellular biochemistry*, 92(1), 178–188.

Chen, Z.P., Iyer, J., Bourguet, W., Held, P., Mioskowski, C., Lebeau, L., Noy, N., Chambon, P., and Gronemeyer, H. (1998). Ligand-and dna-induced dissociation of rxr tetramers. *Journal of molecular biology*, 275(1), 55–65.

Chen, Z., Kibler, R.D., Hunt, A., Busch, F., Pearl, J., Jia, M., VanAernum, Z.L., Wicky, B.I., Dods, G., Liao, H., et al. (2020). De novo design of protein logic gates. *Science*, 368(6486), 78–84.

Cheng, C.S., Feldman, K.E., Lee, J., Verma, S., Huang, D.B., Huynh, K., Chang, M., Ponomarenko, J.V., Sun, S.C., Benedict, C.A., et al. (2011). The specificity of innate immune responses is enforced by repression of interferon response elements by nf-κb p50. *Science signaling*, 4(161), ra11–ra11.

Choi, K.W., Suh, H., Oh, H.L., Ryou, C., and Lee, C.H. (2016). p21cip1 induces apoptosis via binding to bcl2 in lncap prostate cancer cells treated with mcs-c3, a novel carbocyclic analog of pyrrolopyrimidine. *Anticancer research*, 36(1), 213–220.

Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900), 442–446.

Choy, J.S., Wei, S., Lee, J.Y., Tan, S., Chu, S., and Lee, T.H. (2010). Dna methylation increases nucleosome compaction and rigidity. *Journal of the American Chemical Society*, 132(6), 1782–1783.

Cortes-Ciriano, I., Van Westen, G.J., Bouvier, G., Nilges, M., Overington, J.P., Bender, A., and Malliavin, T.E. (2016). Improved large-scale prediction of growth inhibition patterns using the nci60 cancer cell line panel. *Bioinformatics*, 32(1), 85–95.

Cossart, P. and Gicquel-Sanzey, B. (1985). Regulation of expression of the crp gene of escherichia coli k-12: in vivo study. *Journal of bacteriology*, 161(1), 454–457.

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nature reviews genetics*, 14(8), 572–584.

Cournac, A. and Plumbridge, J. (2013). Dna looping in prokaryotes: experimental and theoretical approaches. *Journal of bacteriology*, 195(6), 1109–1119.

References

Crews, S.T. and Pearson, J.C. (2009). Transcriptional autoregulation in development. *Current biology: CB*, 19(6), R241.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563.

Dalma-Weiszhausz, D.D. and Brenowitz, M. (1992). Interactions between dna-bound transcriptional regulators of the escherichia coli gal operon. *Biochemistry*, 31(30), 6980–6989.

Dangkulwanich, M., Ishibashi, T., Bintu, L., and Bustamante, C. (2014). Molecular mechanisms of transcription through single-molecule experiments. *Chemical reviews*, 114(6), 3203–3223.

Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, 109(43), 17454–17459.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science*, 315(5817), 1405–1408.

Djordjevic, M., Sengupta, A.M., and Shraiman, B.I. (2003). A biophysical approach to transcription factor binding site discovery. *Genome research*, 13(11), 2381–2390.

Dobrzyński, M. and Bruggeman, F.J. (2009). Elongation dynamics shape bursty transcription and translation. *Proceedings of the National Academy of Sciences*, 106(8), 2583–2588.

Dodd, I.B., Shearwin, K.E., and Egan, J.B. (2005). Revisited gene regulation in bacteriophage λ. *Current opinion in genetics & development*, 15(2), 145–152.

Douglas, H. and Hawthorne, D. (1966). Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. *Genetics*, 54(3), 911.

Drullion, C., Trégoat, C., Lagarde, V., Tan, S., Gioia, R., Priault, M., Djavaheri-Mergny, M., Brisson, A., Auberger, P., Mahon, F., et al. (2012). Apoptosis and autophagy have opposite roles on imatinib-induced k562 leukemia cell senescence. *Cell death & disease*, 3(8), e373–e373.

Duddu, A.S., Sahoo, S., Hati, S., Jhunjhunwala, S., and Jolly, M.K. (2020). Multistability in cellular differentiation enabled by a network of three mutually repressing master regulators. *Journal of the Royal Society Interface*, 17(170), 20200631.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D., and Misteli, T. (2002). A kinetic framework for a mammalian rna polymerase in vivo. *Science*, 298(5598), 1623–1626.

Ehrhardt, H., Pfeiffer, S., Schrembs, D., Wachter, F., Grunert, M., and Jeremias, I. (2013). Activation of dna damage response by antitumor therapy counteracts the activity of vinca alkaloids. *Anticancer Research*, 33(12), 5273–5287.

Einav, T., Duque, J., and Phillips, R. (2018). Theoretical analysis of inducer and operator binding for cyclic-amp receptor protein mutants. *PloS one*, 13(9), e0204275.

Eissing, T., Conzelmann, H., Gilles, E.D., Allgower, F., Bullinger, E., and Scheurich, P. (2004). Bistability analyses of a caspase activation model for receptor-induced apoptosis. *Journal of Biological Chemistry*, 279(35), 36892–36897.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584), 1183–1186.

English, M.A., Gayet, R.V., and Collins, J.J. (2021). Designing biological circuits: synthetic biology within the operon model and beyond. *Annual review of biochemistry*, 90, 221–244.

Escalante-Chong, R., Savir, Y., Carroll, S.M., Ingraham, J.B., Wang, J., Marx, C.J., and Springer, M. (2015). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proceedings of the National Academy of Sciences*, 112(5), 1636–1641.

Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information integration and energy expenditure in gene regulation. *Cell*, 166(1), 234–244.

Fan, Z., Luo, H., Zhou, J., Wang, F., Zhang, W., Wang, J., Li, S., Lai, Q., Xu, Y., Wang, G., et al. (2020). Checkpoint kinase-1 inhibition and etoposide exhibit a strong synergistic anticancer effect on chronic myeloid leukemia cell line k562 by impairing homologous recombination dna damage repair. *Oncology reports*, 44(5), 2152–2164.

Farkas, M., Hashimoto, H., Bi, Y., Davuluri, R.V., Resnick-Silverman, L., Manfredi, J.J., Debler, E.W., and McMahon, S.B. (2021). Distinct mechanisms control genome recognition by p53 at its target genes linked to different cell fates. *Nature Communications*, 12(1), 484.

Faucon, P.C., Pardee, K., Kumar, R.M., Li, H., Loh, Y.H., and Wang, X. (2014). Gene networks of fully connected triads with complete auto-activation enable multistability and stepwise stochastic transitions. *PloS one*, 9(7), e102873.

Ferreon, J.C., Lee, C.W., Arai, M., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2009). Cooperative regulation of p53 by modulation of ternary complex formation with cbp/p300 and hdm2. *Proceedings of the National Academy of Sciences*, 106(16), 6591–6596.

References

Fiscella, M., Zhang, H., Fan, S., Sakaguchi, K., Shen, S., Mercer, W.E., Vande Woude, G.F., O'Connor, P.M., and Appella, E. (1997). Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner. *Proceedings of the National Academy of Sciences*, 94(12), 6048–6053.

Freire-Rios, A., Tanaka, K., Crespo, I., Van der Wijk, E., Sizentsova, Y., Levitsky, V., Lindhoud, S., Fontana, M., Hohlbein, J., Boer, D.R., et al. (2020). Architecture of dna elements mediating arf transcription factor binding and auxin-responsive gene expression in arabidopsis. *Proceedings of the National Academy of Sciences*, 117(39), 24557–24566.

Friedman, L.J., Mumm, J.P., and Gelles, J. (2013). Rna polymerase approaches its promoter without long-range sliding along dna. *Proceedings of the National Academy of Sciences*, 110(24), 9740–9745.

Furuta, T., Hayward, R., Meng, L., Takemura, H., Aune, G., Bonner, W., Aladjem, M., Kohn, K., and Pommier, Y. (2006). p21cdkn1a allows the repair of replication-mediated dna double-strand breaks induced by topoisomerase i and is inactivated by the checkpoint kinase inhibitor 7-hydroxystaurosporine. *Oncogene*, 25(20), 2839–2849.

Gancedo, J.M. (1998). Yeast carbon catabolite repression. *Microbiology and molecular biology reviews*, 62(2), 334–361.

Garcia, H.G. and Phillips, R. (2011). Quantitative dissection of the simple repression input–output function. *Proceedings of the National Academy of Sciences*, 108(29), 12173–12178.

Gardiner, C. (2009). *Stochastic methods*, volume 4. Springer Berlin.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767), 339–342.

Gasic, A.G., Sarkar, A., and Cheung, M.S. (2021). Understanding protein-complex assembly through grand canonical maximum entropy modeling. *Physical Review Research*, 3(3), 033220.

Gatti, A., Li, H.H., Traugh, J.A., and Liu, X. (2000). Phosphorylation of human p53 on thr-55. *Biochemistry*, 39(32), 9837–9842.

Gautam, P. and Kumar Sinha, S. (2021). Anticipating response function in gene regulatory networks. *Journal of the Royal Society Interface*, 18(179), 20210206.

Gautam, P. and Sinha, S.K. (2023). Theoretical investigation of functional responses of bio-molecular assembly networks. *Soft Matter*, 19(21), 3803–3817.

Gebhardt, J.C.M., Suter, D.M., Roy, R., Zhao, Z.W., Chapman, A.R., Basu, S., Maniatis, T., and Xie, X.S. (2013). Single-molecule imaging of transcription factor binding to dna in live mammalian cells. *Nature methods*, 10(5), 421–426.

Georgakilas, A.G., Martin, O.A., and Bonner, W.M. (2017). p21: a two-faced genome guardian. *Trends in molecular medicine*, 23(4), 310–319.

Gerland, U., Moroz, J.D., and Hwa, T. (2002). Physical constraints and functional characteristics of transcription factor–dna interaction. *Proceedings of the National Academy of Sciences*, 99(19), 12015–12020.

Geva-Zatorsky, N., Rosenfeld, N., Itzkovitz, S., Milo, R., Sigal, A., Dekel, E., Yarnitzky, T., Liron, Y., Polak, P., Lahav, G., et al. (2006). Oscillations and variability in the p53 system. *Molecular systems biology*, 2(1), 2006–0033.

Gibson, M.A. and Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9), 1876–1889.

Gil, B., Kahan-Hanum, M., Skirtenko, N., Adar, R., and Shapiro, E. (2011). Detection of multiple disease indicators by an autonomous biomolecular computer. *Nano letters*, 11(7), 2989–2996.

Gillespie, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4), 403–434.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25), 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6), 1025–1036.

Golikeri, S.V. and Luss, D. (1974). Aggregation of many coupled consecutive first order reactions. *Chemical Engineering Science*, 29(3), 845–855.

Gomes, N.P. and Espinosa, J.M. (2010). Differential regulation of p53 target genes: it's (core promoter) elementary. *Genes & development*, 24(2), 111–114.

González, A., Jafari, S., Zenere, A., Alenius, M., and Altafini, C. (2019). Thermodynamic model of gene regulation for the or59b olfactory receptor in drosophila. *PLoS computational biology*, 15(1), e1006709.

Guo, P., Yang, J., Liu, D., Huang, L., Fell, G., Huang, J., Moses, M.A., and Auguste, D.T. (2019). Dual complementary liposomes inhibit triple-negative breast tumor progression and metastasis. *Science advances*, 5(3), eaav5010.

References

Gupta, A. and Khammash, M. (2022). Frequency spectra and the color of cellular noise. *Nature communications*, 13(1), 4305.

Guthold, M., Zhu, X., Rivetti, C., Yang, G., Thomson, N.H., Kasas, S., Hansma, H.G., Smith, B., Hansma, P.K., and Bustamante, C. (1999). Direct observation of one-dimensional diffusion and transcription by escherichia coli rna polymerase. *Biophysical journal*, 77(4), 2284–2294.

Gyorgy, A., Menezes, A., and Arcak, M. (2023). A blueprint for a synthetic genetic feedback optimizer. *Nature communications*, 14(1), 2554.

Hafner, A., Bulyk, M.L., Jambhekar, A., and Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nature reviews Molecular cell biology*, 20(4), 199–210.

Haldane, A., Manhart, M., and Morozov, A.V. (2014). Biophysical fitness landscapes for transcription factor binding sites. *PLoS computational biology*, 10(7), e1003683.

Han, Z., Wei, W., Dunaway, S., Darnowski, J.W., Calabresi, P., Sedivy, J., Hendrickson, E.A., Balan, K.V., Pantazis, P., and Wyche, J.H. (2002). Role of p21 in apoptosis and senescence of human colon cancer cells treated with camptothecin. *Journal of Biological Chemistry*, 277(19), 17154–17160.

Hänggi, P. (1994). Escape over fluctuating barriers driven by colored noise. *Chemical physics*, 180(2-3), 157–166.

Hanson, S., Kim, E., and Deppert, W. (2005). Redox factor 1 (ref-1) enhances specific dna binding of p53 by promoting p53 tetramerization. *Oncogene*, 24(9), 1641–1647.

Hao, N., Shearwin, K.E., and Dodd, I.B. (2017). Programmable dna looping using engineered bivalent dcas9 complexes. *Nature Communications*, 8(1), 1628.

Harada, Y., Funatsu, T., Murakami, K., Nonoyama, Y., Ishihama, A., and Yanagida, T. (1999). Single-molecule imaging of rna polymerase-dna interactions in real time. *Biophysical journal*, 76(2), 709–715.

Harris, S.L. and Levine, A.J. (2005). The p53 pathway: positive and negative feedback loops. *Oncogene*, 24(17), 2899–2908.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. *Nature*, 402(6761), C47–C52.

Hatfield, G.W. and Benham, C.J. (2002). Dna topology-mediated control of global gene expression in escherichia coli. *Annual review of genetics*, 36(1), 175–203.

Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature*, 387(6630), 296–299.

Hawley, D.K. and McClure, W.R. (1980). In vitro comparison of initiation properties of bacteriophage lambda wild-type pr and x3 mutant promoters. *Proceedings of the National Academy of Sciences*, 77(11), 6381–6385.

Hayden, M.S. and Ghosh, S. (2014). Regulation of nf-κb by tnf family cytokines. In *Seminars in immunology*, volume 26, 253–266. Elsevier.

He, X., Samee, M.A.H., Blatti, C., and Sinha, S. (2010). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS computational biology*, 6(9), e1000935.

He, X., He, L., and Hannon, G.J. (2007). The guardian's little helper: micrornas in the p53 tumor suppressor network. *Cancer research*, 67(23), 11099–11101.

He, Y., Wang, Z., Hu, Y., Yi, X., Wu, L., Cao, Z., and Wang, J. (2020). Sensitive and selective monitoring of the dna damage-induced intracellular p21 protein and unraveling the role of the p21 protein in dna repair and cell apoptosis by surface plasmon resonance. *Analyst*, 145(10), 3697–3704.

Heenan, P.R., Wang, X., Gooding, A.R., Cech, T.R., and Perkins, T.T. (2020). Bending and looping of long dna by polycomb repressive complex 2 revealed by afm imaging in liquid. *Nucleic Acids Research*, 48(6), 2969–2981.

Hellman, L.M. and Fried, M.G. (2007). Electrophoretic mobility shift assay (emsa) for detecting protein–nucleic acid interactions. *Nature protocols*, 2(8), 1849–1861.

Hill, T.L. (1986). *An introduction to statistical thermodynamics*. Courier Corporation.

Holehouse, J., Gupta, A., and Grima, R. (2020). Steady-state fluctuations of a genetic feedback loop with fluctuating rate parameters using the unified colored noise approximation. *Journal of Physics A: Mathematical and Theoretical*, 53(40), 405601.

Huang, B., Jia, D., Feng, J., Levine, H., Onuchic, J.N., and Lu, M. (2018). Racipe: a computational tool for modeling gene regulatory circuits using randomization. *BMC systems biology*, 12(1), 1–12.

Huang, K. (2009). *Introduction to statistical physics*. CRC press.

Idriss, H.T. and Naismith, J.H. (2000). Tnfα and the tnf receptor superfamily: Structure-function relationship (s). *Microscopy research and technique*, 50(3), 184–195.

Jangili, P., Kong, N., Kim, J.H., Zhou, J., Liu, H., Zhang, X., Tao, W., and Kim, J.S. (2022). Dna-damage-response-targeting mitochondria-activated multifunctional prodrug strategy for self-defensive tumor therapy. *Angewandte Chemie International Edition*, 61(16), e202117075.

References

Jenkins, A. and Macauley, M. (2017). Bistability and asynchrony in a boolean model of the l-arabinose operon in escherichia coli. *Bulletin of mathematical biology*, 79(8), 1778–1795.

Jia, D., Jolly, M.K., Kulkarni, P., and Levine, H. (2017). Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers*, 9(7), 70.

Johnston, M. (1987). A model fungal gene regulatory mechanism: the gal genes of saccharomyces cerevisiae. *Microbiological reviews*, 51(4), 458–476.

Jung, P. and Hänggi, P. (1987). Dynamical systems: a unified colored-noise approximation. *Physical review A*, 35(10), 4464.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451–464.

Kalir, S. and Alon, U. (2004). Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell*, 117(6), 713–720.

Kang, J., Gemberling, M., Nakamura, M., Whitby, F.G., Handa, H., Fairbrother, W.G., and Tantin, D. (2009). A general mechanism for transcription regulation by oct1 and oct4 in response to genotoxic and oxidative stress. *Genes & development*, 23(2), 208–222.

Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H. (2006). Initial transcription by rna polymerase proceeds through a dna-scrunching mechanism. *Science*, 314(5802), 1144–1147.

Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular systems biology*, 4(1), 203.

Kapuy, O., Barik, D., Sananes, M.R.D., Tyson, J.J., and Novák, B. (2009). Bistability by multiple phosphorylation of regulatory proteins. *Progress in biophysics and molecular biology*, 100(1-3), 47–56.

Kastenhuber, E.R. and Lowe, S.W. (2017). Putting p53 in context. *Cell*, 170(6), 1062–1078.

Kellogg, R.A. and Tay, S. (2015). Noise facilitates transcriptional control under dynamic inputs. *Cell*, 160(3), 381–392.

Kern, S.E., Kinzler, K.W., Bruskin, A., Jarosz, D., Friedman, P., Prives, C., and Vogelstein, B. (1991). Identification of p53 as a sequence-specific dna-binding protein. *Science*, 252(5013), 1708–1711.

Kim, J.R., Yoon, Y., and Cho, K.H. (2008). Coupled feedback loops form dynamic motifs of cellular networks. *Biophysical journal*, 94(2), 359–365.

Kim, S., Beltran, B., Irnov, I., and Jacobs-Wagner, C. (2019). Long-distance cooperative and antagonistic rna polymerase dynamics via dna supercoiling. *Cell*, 179(1), 106–119.

Kimchi, O., Goodrich, C.P., Courbet, A., Curatolo, A.I., Woodall, N.B., Baker, D., and Brenner, M.P. (2020). Self-assembly–based posttranslational protein oscillators. *Science Advances*, 6(51), eabc1939.

Ko, M.S. (1991). A stochastic model for gene induction. *Journal of theoretical biology*, 153(2), 181–194.

Kouno, T., de Hoon, M., Mar, J.C., Tomaru, Y., Kawano, M., Carninci, P., Suzuki, H., Hayashizaki, Y., and Shin, J.W. (2013). Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome biology*, 14, 1–12.

Krishna, S., Jensen, M.H., and Sneppen, K. (2006). Minimal model of spiky oscillations in nf-κb signaling. *Proceedings of the National Academy of Sciences*, 103(29), 10840–10845.

Kristjuhan, A., Jaks, V., Rimm, I., Tooming, T., and Maimets, T. (1998). Oligomerization of p53 is necessary to inhibit its transcriptional transactivation property at high protein concentration. *Oncogene*, 16(18), 2413–2418.

Kubbutat, M.H., Jones, S.N., and Vousden, K.H. (1997). Regulation of p53 stability by mdm2. *Nature*, 387(6630), 299–303.

Kulasiri, D., Nguyen, L.K., Samarasinghe, S., and Xie, Z. (2008). A review of systems biology perspective on genetic regulatory networks with examples. *Current Bioinformatics*, 3(3), 197–225.

Kuljaca, S., Liu, T., Dwarte, T., Kavallaris, M., Haber, M., Norris, M.D., Martin-Caballero, J., and Marshall, G.M. (2009). The cyclin-dependent kinase inhibitor, p21 waf1, promotes angiogenesis by repressing gene transcription of thioredoxin-binding protein 2 in cancer cells. *Carcinogenesis*, 30(11), 1865–1871.

Kumawat, A. and Chakrabarty, S. (2019). A thermodynamic view of dynamic allostery in a pdz domain protein. *Biophysical Journal*, 116(3), 163a.

Lahav, G. (2008). Oscillations by the p53-mdm2 feedback loop. *Cellular Oscillatory Mechanisms*, 28–38.

Landman, J., Brewster, R.C., Weinert, F.M., Phillips, R., and Kegel, W.K. (2017). Self-consistent theory of transcriptional control in complex regulatory architectures. *PloS one*, 12(7), e0179235.

References

Lando, D.Y. and Teif, V.B. (2000). Long-range interactions between ligands bound to a dna molecule give rise to adsorption with the character of phase transition of the first kind. *Journal of Biomolecular Structure and Dynamics*, 17(5), 903–911.

Larson, D.R. (2011). What do expression dynamics tell us about the mechanism of transcription? *Current opinion in genetics & development*, 21(5), 591–599.

Lavin, M., , and Gueven, N. (2006). The complexity of p53 stabilization and activation. *Cell Death & Differentiation*, 13(6), 941–950.

Lee, C.W., Ferreon, J.C., Ferreon, A.C.M., Arai, M., and Wright, P.E. (2010). Graded enhancement of p53 binding to creb-binding protein (cbp) by multisite phosphorylation. *Proceedings of the National Academy of Sciences*, 107(45), 19290–19295.

Lee, T.H. (2019). Physical chemistry of epigenetics: Single-molecule investigations. *The Journal of Physical Chemistry B*, 123(40), 8351–8362.

Lee, T.K., Denny, E.M., Sanghvi, J.C., Gaston, J.E., Maynard, N.D., Hughey, J.J., and Covert, M.W. (2009). A noisy paracrine signal determines the cellular nf-κb response to lipopolysaccharide. *Science signaling*, 2(93), ra65–ra65.

Lee, T.I. and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), 1237–1251.

Lee, Y.H. and Schiemann, W.P. (2011). Fibromodulin suppresses nuclear factor-κb activity by inducing the delayed degradation of ikba via a jnk-dependent pathway coupled to fibroblast apoptosis. *Journal of Biological Chemistry*, 286(8), 6414–6422.

Leisner, M., Bleris, L., Lohmueller, J., Xie, Z., and Benenson, Y. (2010). Rationally designed logic integration of regulatory signals in mammalian cells. *Nature nanotechnology*, 5(9), 666–670.

Levine, A.J. (1997). p53, the cellular gatekeeper for growth and division. *cell*, 88(3), 323–331.

Levine, A.J. and Oren, M. (2009). The first 30 years of p53: growing ever more complex. *Nature reviews cancer*, 9(10), 749–758.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell*, 157(1), 13–25.

Levine, M. and Davidson, E.H. (2005). Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*, 102(14), 4936–4942.

Li, A.G., Piluso, L.G., Cai, X., Wei, G., Sellers, W.R., and Liu, X. (2006). Mechanistic insights into maintenance of high p53 acetylation by pten. *Molecular cell*, 23(4), 575–587.

Li, R.F., Feng, Y.Q., Chen, J.H., Ge, L.T., Xiao, S.Y., and Zuo, X.L. (2015). Naringenin suppresses k562 human leukemia cell proliferation and ameliorates adriamycin-induced oxidative damage in polymorphonuclear leukocytes. *Experimental and therapeutic medicine*, 9(3), 697–706.

Li, T., Dong, Y., Zhang, X., Ji, X., Luo, C., Lou, C., Zhang, H.M., and Ouyang, Q. (2018). Engineering of a genetic circuit with regulatable multistability. *Integrative Biology*, 10(8), 474–482.

Liu, B., Hong, C., Huang, R.K., Yu, Z., and Steitz, T.A. (2017). Structural basis of bacterial transcription activation. *Science*, 358(6365), 947–951.

Liu, F., Hu, B., and Wang, W. (2001). Effects of correlated and independent noise on signal processing in neuronal systems. *Physical Review E*, 63(3), 031907.

Liu, T., Zhang, J., and Zhou, T. (2016). Effect of interaction between chromatin loops on cell-to-cell variability in gene expression. *PLoS computational biology*, 12(5), e1004917.

Luo, R.X. and Dean, D.C. (1999). Chromatin remodeling and transcriptional regulation. *Journal of the National Cancer Institute*, 91(15), 1288–1294.

Ma, J., Bai, L., and Wang, M.D. (2013). Transcription under torsion. *Science*, 340(6140), 1580–1583.

Macleod, K.F., Sherry, N., Hannon, G., Beach, D., Tokino, T., Kinzler, K., Vogelstein, B., and Jacks, T. (1995). p53-dependent and independent expression of p21 during cell growth, differentiation, and dna damage. *Genes and Development*, 9(8), 935–944.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21), 11980–11985.

Mangan, S., Itzkovitz, S., Zaslaver, A., and Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *Journal of molecular biology*, 356(5), 1073–1081.

Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of molecular biology*, 334(2), 197–204.

Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the national academy of sciences*, 107(14), 6286–6291.

Marei, H.E., Althani, A., Afifi, N., Hasan, A., Caceci, T., Pozzoli, G., Morrione, A., Giordano, A., and Cenciarelli, C. (2021). p53 signaling in cancer progression and therapy. *Cancer cell international*, 21(1), 1–15.

References

Martins, C.P., Brown-Swigart, L., and Evan, G.I. (2006). Modeling the therapeutic efficacy of p53 restoration in tumors. *Cell*, 127(7), 1323–1334.

Marzen, S., Garcia, H.G., and Phillips, R. (2013). Statistical mechanics of monod–wyman–changeux (mwc) models. *Journal of molecular biology*, 425(9), 1433–1460.

Mather, W., Bennett, M.R., Hasty, J., and Tsimring, L.S. (2009). Delay-induced degrade-and-fire oscillations in small genetic circuits. *Physical review letters*, 102(6), 068105.

Matthews, K. (1992). Dna looping. *Microbiology and Molecular Biology Reviews*, 56(1), 123–136.

Mayo, L.D., Seo, Y.R., Jackson, M.W., Smith, M.L., Guzman, J.R., Korgaonkar, C.K., and Donner, D.B. (2005). Phosphorylation of human p53 at serine 46 determines promoter selection and whether apoptosis is attenuated or amplified. *Journal of Biological Chemistry*, 280(28), 25953–25959.

Mazal, H., Aviram, H., Riven, I., and Haran, G. (2018). Effect of ligand binding on a protein with a complex folding landscape. *Physical Chemistry Chemical Physics*, 20(5), 3054–3062.

McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*, 287(5456), 1262–1265.

Michida, H., Imoto, H., Shinohara, H., Yumoto, N., Seki, M., Umeda, M., Hayashi, T., Nikaido, I., Kasukawa, T., Suzuki, Y., et al. (2020). The number of transcription factors at an enhancer determines switch-like gene expression. *Cell Reports*, 31(9), 107724.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594), 824–827.

Mirzayans, R., Andrais, B., Scott, A., Wang, Y.W., Weiss, R.H., and Murray, D. (2015). Spontaneous γh2ax foci in human solid tumor-derived cell lines in relation to p21waf1 and wip1 expression. *International journal of molecular sciences*, 16(5), 11609–11628.

Mitrophanov, A.Y. and Groisman, E.A. (2008). Positive feedback in cellular control systems. *Bioessays*, 30(6), 542–555.

Molina, N., Suter, D.M., Cannavo, R., Zoller, B., Gotic, I., and Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proceedings of the National Academy of Sciences*, 110(51), 20563–20568.

Monod, J. (1949). The growth of bacterial cultures. *Annual review of microbiology*, 3(1), 371–394.

Navalkar, A., Paul, A., Sakunthala, A., Pandey, S., Dey, A.K., Saha, S., Sahoo, S., Jolly, M.K., Maiti, T.K., and Maji, S.K. (2022). Oncogenic gain of function due to p53 amyloids occurs through aberrant alteration of cell cycle and proliferation. *Journal of cell science*, 135(15), jcs259500.

Nevozhay, D., Adams, R.M., Murphy, K.F., Josić, K., and Balázsi, G. (2009). Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences*, 106(13), 5123–5128.

Ngo, K.A., Kishimoto, K., Davis-Turak, J., Pimplaskar, A., Cheng, Z., Spreafico, R., Chen, E.Y., Tam, A., Ghosh, G., Mitchell, S., et al. (2020). Dissecting the regulatory strategies of nf-κb rela target genes in the inflammatory response reveals differential transactivation logics. *Cell reports*, 30(8), 2758–2775.

Ngo, T.T., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., and Ha, T. (2016). Effects of cytosine modifications on dna flexibility and nucleosome mechanical stability. *Nature communications*, 7(1), 1–9.

Nick, H. and Gilbert, W. (1985). Detection in vivo of protein-dna interactions within the lac operon of escherichia coli. *Nature*, 313(6005), 795–798.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer–promoter interactions. *Proceedings of the National Academy of Sciences*, 106(48), 20222–20227.

Ochab-Marcinek, A. and Tabaka, M. (2010). Bimodal gene expression in noncooperative regulatory systems. *Proceedings of the National Academy of Sciences*, 107(51), 22096–22101.

Olsson, A., Manzl, C., Strasser, A., and Villunger, A. (2007). How important are post-translational modifications in p53 for selectivity in target-gene transcription and tumour suppression? *Cell Death & Differentiation*, 14(9), 1561–1575.

Ong, K.M., Blackford, J.A., Kagan, B.L., Simons, S.S., and Chow, C.C. (2010). A theoretical framework for gene induction and experimental comparisons. *Proceedings of the National Academy of Sciences*, 107(15), 7107–7112. doi:10.1073/pnas.0911095107.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of escherichia coli. *Nature*, 427(6976), 737–740.

Peck, B., Chen, C.Y., Ho, K.K., Di Fruscia, P., Myatt, S.S., Coombes, R.C., Fuchter, M.J., Hsiao, C.D., and Lam, E.W.F. (2010). Sirt inhibitors induce cell death and

References

p53 acetylation through targeting both sirt1 and sirt2. *Molecular cancer therapeutics*, 9(4), 844–855.

Pedraza, J.M. and Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861), 339–343.

Pennisi, E. (2020). Massive project reveals complexity of gene regulation.

Petrenko, N., Jin, Y., Dong, L., Wong, K.H., and Struhl, K. (2019). Requirements for rna polymerase ii preinitiation complex formation in vivo. *Elife*, 8, e43654.

Phillips, R. (2015). Napoleon is in equilibrium. *Annu. Rev. Condens. Matter Phys.*, 6(1), 85–111.

Phillips, R., Kondev, J., Theriot, J., and Garcia, H. (2012). *Physical biology of the cell*. Garland Science.

Powell, E., Piwnica-Worms, D., and Piwnica-Worms, H. (2014). Contribution of p53 to metastasis. *Cancer discovery*, 4(4), 405–414.

Pryczynicz, A., Gryko, M., Niewiarowska, K., Cepowicz, D., Ustymowicz, M., Kemona, A., and Guzińska-Ustymowicz, K. (2014). Bax protein may influence the invasion of colorectal cancer. *World Journal of Gastroenterology: WJG*, 20(5), 1305.

Ptashne, M. (2004). A genetic switch: phage lambda revisited. *(No Title)*.

Puc, J., Kozbial, P., Li, W., Tan, Y., Liu, Z., Suter, T., Ohgi, K.A., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2015). Ligand-dependent enhancer activation regulated by topoisomerase-i activity. *Cell*, 160(3), 367–380.

Qiu, X. and Chiechi, R.C. (2022). Printable logic circuits comprising self-assembled protein complexes. *Nature Communications*, 13(1), 2312.

Rabier, C.E., Berry, V., Stoltz, M., Santos, J.D., Wang, W., Glaszmann, J.C., Pardi, F., and Scornavacca, C. (2021). On the inference of complex phylogenetic networks by markov chain monte-carlo. *PLoS Computational Biology*, 17(9), e1008380.

Ragnhildstveit, E., Fjose, A., Becker, P.B., and Quivy, J.P. (1997). Solid phase technology improves coupled gel shift/footprinting analysis. *Nucleic acids research*, 25(2), 453–454.

Ranganathan, S., Halagowder, D., and Sivasithambaram, N.D. (2015). Quercetin suppresses twist to induce apoptosis in mcf-7 breast cancer cells. *PloS one*, 10(10), e0141370.

Rashid, M., Hari, K., Thampi, J., Santhosh, N.K., and Jolly, M.K. (2022). Network topology metrics explaining enrichment of hybrid epithelial/mesenchymal phenotypes in metastasis. *PLoS computational biology*, 18(11), e1010687.

Reczek, E.E., Flores, E.R., Tsay, A.S., Attardi, L.D., and Jacks, T. (2003). Multiple response elements and differential p53 binding control perp expression during apoptosis. *Molecular cancer research*, 1(14), 1048–1057.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by rna polymerase involve dna scrunching. *Science*, 314(5802), 1139–1143.

Rhee, K.Y., Senear, D.F., and Hatfield, G.W. (1998). Activation of gene expression by a ligand-induced conformational change of a protein-dna complex. *Journal of Biological Chemistry*, 273(18), 11257–11266.

Risken, H. and Risken, H. (1996). *Fokker-planck equation*. Springer.

Rodriguez, R. and Meuth, M. (2006). Chk1 and p21 cooperate to prevent apoptosis during dna replication fork stress. *Molecular biology of the cell*, 17(1), 402–412.

Rokudai, S., Aikawa, Y., Tagata, Y., Tsuchida, N., Taya, Y., and Kitabayashi, I. (2009). Monocytic leukemia zinc finger (moz) interacts with p53 to induce p21 expression and cell-cycle arrest. *Journal of Biological Chemistry*, 284(1), 237–244.

Rydenfelt, M., Cox III, R.S., Garcia, H., and Phillips, R. (2014). Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. *Physical Review E*, 89(1), 012702.

Sahoo, D.K., Jena, S., Dutta, J., Chakrabarty, S., and Biswal, H.S. (2018). Critical assessment of the interaction between dna and choline amino acid ionic liquids: evidences of multimodal binding and stability enhancement. *ACS Central Science*, 4(12), 1642–1651.

Saiz, L. and Vilar, J. (2008). Protein–protein/dna interaction networks: versatile macromolecular structures for the control of gene expression. *IET systems biology*, 2(5), 247–255.

Saiz, L. (2012). The physics of protein–dna interaction networks in the control of gene expression. *Journal of physics: Condensed matter*, 24(19), 193102.

Saiz, L. and Vilar, J.M. (2006a). Dna looping: the consequences and its control. *Current opinion in structural biology*, 16(3), 344–350.

Saiz, L. and Vilar, J.M. (2006b). Stochastic dynamics of macromolecular-assembly networks. *Molecular Systems Biology*, 2(1).

Saiz, L. and Vilar, J.M. (2007a). Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic acids research*, 36(3), 726–731.

References

Saiz, L. and Vilar, J.M. (2007b). Multilevel deconstruction of the in vivo behavior of looped dna-protein complexes. *PLoS One*, 2(4), e355.

Samee, M.A.H., Lim, B., Samper, N., Lu, H., Rushlow, C.A., Jiménez, G., Shvartsman, S.Y., and Sinha, S. (2015). A systematic ensemble approach to thermodynamic modeling reveals an enhancer's logic. *Cell systems*, 1(6), 396.

Sánchez, Á. and Kondev, J. (2008). Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences*, 105(13), 5081–5086.

Saramäki, A., Banwell, C.M., Campbell, M.J., and Carlberg, C. (2006). Regulation of the human p21 (waf1/cip1) gene promoter via multiple binding sites for p53 and the vitamin d 3 receptor. *Nucleic acids research*, 34(2), 543–554.

Saravanan, B., Soota, D., Islam, Z., Majumdar, S., Mann, R., Meel, S., Farooq, U., Walavalkar, K., Gayen, S., Singh, A.K., et al. (2020). Ligand dependent gene regulation by transient erα clustered enhancers. *PLoS genetics*, 16(1), e1008516.

Sarkar, S., Sinha, S.K., Levine, H., Jolly, M.K., and Dutta, P.S. (2019). Anticipating critical transitions in epithelial–hybrid-mesenchymal cell-fate determination. *Proceedings of the National Academy of Sciences*, 116(52), 26343–26352.

Satija, R. and Shalek, A.K. (2014). Heterogeneity in immune responses: from populations to single cells. *Trends in immunology*, 35(5), 219–229.

Schleif, R. (1992). Dna looping. *Annual review of biochemistry*, 61(1), 199–223.

Segal, E. and Widom, J. (2009). From dna sequence to transcriptional behaviour: a quantitative approach. *Nature Reviews Genetics*, 10(7), 443.

Severin, P.M., Zou, X., Gaub, H.E., and Schulten, K. (2011). Cytosine methylation alters dna mechanical properties. *Nucleic acids research*, 39(20), 8740–8751.

Sevier, S.A. and Levine, H. (2017). Mechanical properties of transcription. *Physical review letters*, 118(26), 268101.

Shea, M.A. and Ackers, G.K. (1985). The or control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of molecular biology*, 181(2), 211–230.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1), 64–68.

Siegele, D.A. and Hu, J.C. (1997). Gene expression from plasmids containing the arabad promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences*, 94(15), 8168–8172.

Smith, S.B., Finzi, L., and Bustamante, C. (1992). Direct mechanical measurements of the elasticity of single dna molecules by using magnetic beads. *Science*, 258(5085), 1122–1126.

Sobie, E.A. (2011). Bistability in biochemical signaling models. *Science signaling*, 4(192), tr10–tr10.

Soboleski, M.R., Oaks, J., and Halford, W.P. (2005). Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *The FASEB journal*, 19(3), 1–20.

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Physical Review E*, 64(5), 051904.

Spanio, T., Hidalgo, J., and Muñoz, M.A. (2017). Impact of environmental colored noise in single-species population dynamics. *Physical Review E*, 96(4), 042301.

Stamatoyannopoulos, J.A. (2012). What does our genome encode? *Genome research*, 22(9), 1602–1611.

Stathopoulos, A. and Levine, M. (2005). Genomic regulatory networks and animal development. *Developmental cell*, 9(4), 449–462.

Stenger, J.E., Tegtmeyer, P., Mayr, G.A., Reed, M., Wang, Y., Wang, P., Hough, P., and Mastrangelo, I. (1994). p53 oligomerization and dna looping are linked with transcriptional activation. *The EMBO Journal*, 13(24), 6011–6020.

Strogatz, S.H. (2018). *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering.* CRC press.

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). Dna looping between sites for transcriptional activation: self-association of dna-bound sp1. *Genes & development*, 5(5), 820–826.

Sullivan, K.D., Galbraith, M.D., Andrysik, Z., and Espinosa, J.M. (2018). Mechanisms of transcriptional regulation by p53. *Cell Death & Differentiation*, 25(1), 133–143.

Suzuki, Y., Shin, M., Yoshida, A., Yoshimura, S.H., and Takeyasu, K. (2012). Fast microscopical dissection of action scenes played by escherichia coli rna polymerase. *FEBS letters*, 586(19), 3187–3192.

Szymczak, F., Colli, M.L., Mamula, M., Evans-Molina, C., and Eizirik, D.L. (2021). Gene expression signatures of target tissues in type 1 diabetes, lupus erythematosus, multiple sclerosis, and rheumatoid arthritis. *Science advances*, 7(2), eabd7600.

References

Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell nf-κb dynamics reveal digital activation and analogue information processing. *Nature*, 466(7303), 267–271.

Teale, W.D., Paponov, I.A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nature reviews Molecular cell biology*, 7(11), 847–859.

Teif, V.B. (2005). Ligand-induced dna condensation: choosing the model. *Biophysical journal*, 89(4), 2574–2587.

Teif, V.B. (2007). General transfer matrix formalism to calculate dna–protein–drug binding in gene regulation: application to o r operator of phage λ. *Nucleic acids research*, 35(11), e80.

Teif, V.B., Haroutiunian, S.G., Vorob'ev, V.I., Lando, D.Y., and Ivanov, V. (2002). Short-range interactions and size of ligands bound to dna strongly influence adsorptive phase transition caused by long-range interactions. *Journal of Biomolecular Structure and Dynamics*, 19(6), 1093–1100.

Teif, V.B. and Rippe, K. (2010). Statistical–mechanical lattice models for protein–dna binding in chromatin. *Journal of Physics: Condensed Matter*, 22(41), 414105.

Teodoro, J.G., Parker, A.E., Zhu, X., and Green, M.R. (2006). p53-mediated inhibition of angiogenesis through up-regulation of a collagen prolyl hydroxylase. *Science*, 313(5789), 968–971.

Thomas, P., Popović, N., and Grima, R. (2014). Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 111(19), 6994–6999.

Tian, T. and Burrage, K. (2006). Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the national Academy of Sciences*, 103(22), 8372–8377.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular cell*, 10(6), 1453–1465.

Tor, Y.S., Yazan, L.S., Foo, J.B., Wibowo, A., Ismail, N., Cheah, Y.K., Abdullah, R., Ismail, M., Ismail, I.S., and Yeap, S.K. (2015). Induction of apoptosis in mcf-7 cells via oxidative stress generation, mitochondria-dependent and caspase-independent pathway by ethyl acetate extract of dillenia suffruticosa and its chemical profile. *PLoS One*, 10(6), e0127441.

Torgovnick, A. and Schumacher, B. (2015). Dna repair mechanisms in cancer development and therapy. *Frontiers in genetics*, 6, 157.

Tovar, C., Rosinski, J., Filipovic, Z., Higgins, B., Kolinsky, K., Hilton, H., Zhao, X., Vu, B.T., Qing, W., Packman, K., et al. (2006). Small-molecule mdm2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. *Proceedings of the National Academy of Sciences*, 103(6), 1888–1893.

Tripathi, S., Brahmachari, S., Onuchic, J.N., and Levine, H. (2022). Dna supercoiling-mediated collective behavior of co-transcribing rna polymerases. *Nucleic acids research*, 50(3), 1269–1279.

Tucureanu, M.M., Rebleanu, D., Constantinescu, C.A., Deleanu, M., Voicu, G., Butoi, E., Calin, M., and Manduteanu, I. (2018). Lipopolysaccharide-induced inflammation in monocytes/macrophages is blocked by liposomal delivery of gi-protein inhibitor. *International journal of nanomedicine*, 63–76.

van Dieck, J., Fernandez-Fernandez, M.R., Veprintsev, D.B., and Fersht, A.R. (2009). Modulation of the oligomerization state of p53 by differential binding of proteins of the s100 family to p53 monomers and tetramers. *Journal of Biological Chemistry*, 284(20), 13804–13811.

Van Kampen, N.G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., et al. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of mdm2. *Science*, 303(5659), 844–848.

Vellela, M. and Qian, H. (2009). Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schlögl model revisited. *Journal of The Royal Society Interface*, 6(39), 925–940.

Venturelli, O.S., El-Samad, H., and Murray, R.M. (2012). Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proceedings of the National Academy of Sciences*, 109(48), E3324–E3333.

Vilar, J.M. and Leibler, S. (2003). Dna looping and physical constraints on transcription regulation. *Journal of molecular biology*, 331(5), 981–989.

Vilar, J.M. and Saiz, L. (2005). Dna looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Current opinion in genetics & development*, 15(2), 136–144.

Vilar, J.M. and Saiz, L. (2011). Control of gene expression by modulated self-assembly. *Nucleic acids research*, 39(16), 6854–6863.

Vilar, J.M. and Saiz, L. (2014). Suppression and enhancement of transcriptional noise by dna looping. *Physical Review E*, 89(6), 062703.

References

Vousden, K.H. and Prives, C. (2009). Blinded by the light: the growing complexity of p53. *Cell*, 137(3), 413–431.

Walczak, A.M., Sasai, M., and Wolynes, P.G. (2005). Self-consistent proteomic field theory of stochastic gene switches. *Biophysical journal*, 88(2), 828–850.

Wang, A.G., Son, M., Kenna, E., Thom, N., and Tay, S. (2022). Nf-κb memory coordinates transcriptional responses to dynamic inflammatory stimuli. *Cell reports*, 40(7).

Wang, D.G., Wang, S., Huang, B., and Liu, F. (2019a). Roles of cellular heterogeneity, intrinsic and extrinsic noise in variability of p53 oscillation. *Scientific reports*, 9(1), 5883.

Wang, F., Redding, S., Finkelstein, I.J., Gorman, J., Reichman, D.R., and Greene, E.C. (2013). The promoter-search mechanism of escherichia coli rna polymerase is dominated by three-dimensional diffusion. *Nature structural & molecular biology*, 20(2), 174.

Wang, J. et al. (2009). Bayespi-a new model to study protein-dna interactions: a case study of condition-specific protein binding parameters for yeast transcription factors. *BMC bioinformatics*, 10(1), 345.

Wang, M., Chen, S., and Ao, D. (2021a). Targeting dna repair pathway in cancer: Mechanisms and clinical application. *MedComm*, 2(4), 654–691.

Wang, N., Lefaudeux, D., Mazumder, A., Li, J.J., and Hoffmann, A. (2021b). Identifying the combinatorial control of signal-dependent transcription factors. *PLoS computational biology*, 17(6), e1009095.

Wang, X., Yang, K., Wu, Q., Kim, L.J., Morton, A.R., Gimple, R.C., Prager, B.C., Shi, Y., Zhou, W., Bhargava, S., et al. (2019b). Targeting pyrimidine synthesis accentuates molecular therapy response in glioblastoma stem cells. *Science translational medicine*, 11(504), eaau4972.

Weber, J.D., Taylor, L.J., Roussel, M.F., Sherr, C.J., and Bar-Sagi, D. (1999). Nucleolar arf sequesters mdm2 and activates p53. *Nature cell biology*, 1(1), 20–26.

Webster, M.W. and Weixlbaumer, A. (2021). The intricate relationship between transcription and translation. *Proceedings of the National Academy of Sciences*, 118(21), e2106284118.

Weinert, F.M., Brewster, R.C., Rydenfelt, M., Phillips, R., and Kegel, W.K. (2014). Scaling of gene expression with transcription-factor fugacity. *Physical review letters*, 113(25), 258101.

Wenta, N., Strauss, H., Meyer, S., and Vinkemeier, U. (2008). Tyrosine phosphorylation regulates the partitioning of stat1 between different dimer conformations. *Proceedings of the National Academy of Sciences*, 105(27), 9238–9243.

Widom, J. (2005). Target site localization by site-specific, dna-binding proteins. *Proceedings of the National Academy of Sciences*, 102(47), 16909–16910.

Wilkinson, A.C., Nakauchi, H., and Göttgens, B. (2017). Mammalian transcription factor networks: recent advances in interrogating biological complexity. *Cell systems*, 5(4), 319–331.

Williams, A.B. and Schumacher, B. (2016). p53 in the dna-damage-repair process. *Cold Spring Harbor perspectives in medicine*, 6(5).

Wong, F. and Gunawardena, J. (2020). Gene regulation in and out of equilibrium. *Annual review of biophysics*, 49, 199–226.

Wright, P.E. and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology*, 16(1), 18–29.

Yasmin, R., Yeung, K.T., Chung, R.H., Gaczynska, M.E., Osmulski, P.A., and Noy, N. (2004). Dna-looping by rxr tetramers permits transcriptional regulation "at a distance". *Journal of molecular biology*, 343(2), 327–338.

Ye, F., Gao, F., Liu, X., Buck, M., and Zhang, X. (2022). Mechanisms of dna opening revealed in aaa+ transcription complex structures. *Science Advances*, 8(51), eadd3479.

Yeap, S., Akhtar, M.N., Lim, K.L., Abu, N., Ho, W.Y., Zareen, S., Roohani, K., Ky, H., Tan, S.W., Lajis, N., et al. (2015). Synthesis of an anthraquinone derivative (dhaqc) and its effect on induction of g2/m arrest and apoptosis in breast cancer mcf-7 cell line. *Drug design, development and therapy*, 983–992.

Yen, J.H., Lin, C.Y., Chuang, C.H., Chin, H.K., Wu, M.J., and Chen, P.Y. (2020). Nobiletin promotes megakaryocytic differentiation through the mapk/erk-dependent egr1 expression and exerts anti-leukemic effects in human chronic myeloid leukemia (cml) k562 cells. *Cells*, 9(4), 877.

Yildirim, N. and Mackey, M.C. (2003). Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophysical journal*, 84(5), 2841–2851.

Yin, P., Choi, H.M., Calvert, C.R., and Pierce, N.A. (2008). Programming biomolecular self-assembly pathways. *Nature*, 451(7176), 318–322.

Zambrano, S., De Toma, I., Piffer, A., Bianchi, M.E., and Agresti, A. (2016). Nf-κb oscillations translate into functionally related patterns of gene expression. *Elife*, 5, e09100.

References

Zhang, C., Tsoi, R., Wu, F., and You, L. (2016). Processing oscillatory signals by incoherent feedforward loops. *PLoS computational biology*, 12(9), e1005101.

Zhang, D.Y., Turberfield, A.J., Yurke, B., and Winfree, E. (2007). Engineering entropy-driven reactions and networks catalyzed by dna. *Science*, 318(5853), 1121–1125.

Zhang, J. and Zhou, T. (2019). Markovian approaches to modeling intracellular reaction processes with molecular memory. *Proceedings of the National Academy of Sciences*, 116(47), 23542–23550.

Zhou, J., Schor, I.E., Yao, V., Theesfeld, C.L., Marco-Ferreres, R., Tadych, A., Furlong, E.E., and Troyanskaya, O.G. (2019). Accurate genome-wide predictions of spatio-temporal gene expression during embryonic development. *PLoS genetics*, 15(9), e1008382.

Zhu, R., del Rio-Salgado, J.M., Garcia-Ojalvo, J., and Elowitz, M.B. (2022). Synthetic multistability in mammalian cells. *Science*, 375(6578), eabg9765.

Zindy, F., Eischen, C.M., Randle, D.H., Kamijo, T., Cleveland, J.L., Sherr, C.J., and Roussel, M.F. (1998). Myc signaling via the arf tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes & development*, 12(15), 2424–2433.

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer–promoter interactions. *Nature*, 604(7906), 571–577.

References

165

List of Publications

Published papers:

- **Gautam, P.**, and Kumar Sinha, S. (2021). ***Anticipating response function in gene regulatory networks*** Journal of the Royal Society Interface, 18(179), 20210206.

- **Gautam, P.**, and Sinha, S. K. (2023). ***Theoretical investigation of functional responses of bio-molecular assembly networks*** Soft Matter, 19(21), 3803-3817.

Manuscripts under preparation:

- **Gautam, P.**, and Sinha, S. K., ***The blueprint of logical decisions of NF-κB inside a cell***, (under review).

- **Gautam, P.**, and Sinha, S. K., ***A unified model for the cell fate determination of a malignant cell***, (under review).

- **Gautam, P.**, Kashyap, D., and Sinha, S. K., ***Frequency spectra and colored noise for the typical and super enhancers in dynamic environment of NF-κB controlled gene***, (under preparation).

Curriculum Vitae

Personal Details:

Name	: **Pankaj Gautam**
Date of Birth	: 02 December, 1994
Present Address	: Department of Chemistry,
	Indian Institute of Technology Ropar,
	Nangal Road, Rupnagar, Punjab - 140 001.
Permanent Address	: S/o, Karam Chand
	Vill + P.O- Dhoun Kothi, Tehsil, Sadar,
	Dist- Bilaspur, Himachal Pradesh - 174012.
Mail ID	: 2017cyz0006@iitrpr.ac.in

Academic Qualifications:

Exam / Degree	Board / University	Year	Subject
M.Sc	Himachal Pradesh University	2017	Chemistry
NET	UGC	2017	Chemistry
GATE	MHRD	2017	Chemistry
Ph.D*	Indian Institute of Technology Ropar	2017-present	Chemistry

*Ph.D Dissertation: *Theoretical Studies of Modulated Protein-DNA Interactions and Their Impacts on Gene Regulation*

Fellowship awarded:

- **UGC** fellowship for Ph.D