On the physics of intranuclear organization

by

Amogh Sood

Bachelor of Engineering
Nagoya University

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMISTRY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Authored by:	Amogh Sood Department of Chemistry May 28, 2024
Certified by:	Bin Zhang Pfizer-Laubach Career Development Associate Professor of Chemistry Thesis Supervisor
Accepted by:	Adam P. Willard Department of Chemistry Graduate Officer Professor of Chemistry

This doctoral thesis has been examined by a committee of the
Department of Chemistry as follows:

Professor Arup K. Chakraborty: .....
Chairman, Thesis Committee
John M. Deutch Institute Professor,
Professor of Chemical Engineering, Physics & Chemistry,
Massachusetts Institute of Technology,
Core Faculty Member and former Founding Director,
Institute for Medical Engineering and Science (IMES), MIT,
Founding Steering Committee Member,
Ragon Institute of MGH, MIT & Harvard

Professor Bin Zhang: .....
Supervisor, Thesis Committee
Pfizer-Laubach Career Development Associate Professor of Chemistry,
Massachusetts Institute of Technology

Professor Mehran Kardar: .....
Member, Thesis Committee
Francis Friedman Professor of Physics,
Massachusetts Institute of Technology

On the physics of intranuclear organization

by

Amogh Sood

Submitted to the Department of Chemistry
on May 28, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMISTRY

ABSTRACT

Eukaryotic nuclei, despite their diverse and crowded chemical milieu, can achieve precise spatiotemporal organization of their contents and chemistry, despite lacking access to membrane-bound organelles. It has recently become apparent that the cells accomplish this feat by leveraging physical processes such as liquid-liquid phase separation driven by multivalent macromolecular interactions to form biomolecular condensates which can serve as membrane-less organelles for the precise, vectorial organization of intranuclear contents. In particular, the hierarchical and functional packaging of DNA into chromatin is mediated by phase separation. Epigenetic modifications of histone proteins, which DNA wraps around to form nucleosomes, are key determinants of nucleosomes' condensability and chromatin's higher-order structure. Chromatin structure, by regulating access of transcriptional machinery to the genome, in turn, has broad implications for cellular processes such as gene regulation and cellular differentiation. Furthermore, there exists a bi-directional feedback between 1D epigenomic sequence and 3D chromatin structure as the former is spread and maintained by enzymes that have a "reader-writer" functionality that allows them to similarly modify nucleosomes close to each other in sequence but not necessarily in space. Recent advances suggest chromatin has the properties of a viscoelastic network and exhibits non-trivial dynamics. Therefore, the dynamics of chromatin structure and the spread and maintenance of epigenetic marks are intimately and inextricably linked yet poorly understood. Part I of this thesis is devoted to understanding the complex interplay between chromatin structural dynamics and stochastic reaction networks describing histone modifications. Furthermore, given the prominent role phase separation plays in intranuclear organization, we devote Part II of this thesis to study the impact of competition between specific and non-specific interactions on liquid-liquid phase separation coupled to percolation and thereby attempt to elucidate the molecular grammar of phase separating biomolecules and evolutionary pressures that shape them.

Thesis supervisor: Bin Zhang
Title: Pfizer-Laubach Career Development Associate Professor of Chemistry

Acknowledgments

Throughout my educational journey, I have been incredibly fortunate to receive the guidance, mentorship, friendship, and support of an outstanding group of teachers, advisors, colleagues, friends, and family. Although it is nearly impossible to fully articulate and enumerate their contributions, I dedicate this chapter to acknowledging the profound impact they have had on my academic and personal growth.

First of all, I am deeply grateful to Prof. Bin Zhang who has been a phenomenal teacher and advisor. I thoroughly enjoyed our scientific discussions during the course of my PhD and I hope to emulate his scientific enthusiasm and rigor in my future career. He deftly fostered a welcoming and collaborative environment within the group that allows his students, myself included, to find their scientific voice and excel in their graduate studies. His impact on my personal and professional development cannot be overstated. I am also grateful to the other members of my thesis committee, Prof. Arup Chakraborty and Prof. Mehran Kardar for their insightful comments and discussions during the course of my PhD. Our committee meetings have been a source of great learning for me.

Prior to graduate school, I was fortunate enough to have teachers and mentors that inspired me to pursue and complete my graduate study. In particular, I would like to thank Prof. Eiji Yashima, Prof. John Wojdylo, Prof. Peter Butko and Prof. Jiyoung Shin.

I would also like to thank the students and postdocs who have been in the Zhang Group during my time there including Zhongling Jiang, Yifeng Qi, Andrew Latham, Wenjun Xie, Xingcheng Lin, Xinqiang Ding, Kartik Kamat, Joe Paggi, Shuming Liu, Yumeng Zhang, Cong Wang, Zhuohan Lao, Greg Schuette, Advait Athreya, Justin Airas, Camryn Carter, Ivan Riveros, Jared Zheng and Abraham Park. I have benefited a lot from our scientific discussions during group meetings and I have also enjoyed thoroughly our many (sometimes) non-scientific debates and discussions during lunch time. I am similarly grateful to other members of the MIT theoretical chemistry community and residents (current and former) of the zoo including Changhae Andrew Kim, Alexandra McIsaac, Hong-Zhou Ye, Nathan Ricke, Dina Sharon, Henrik Heelweg, Eric Alt, Amr Dodin, Kyaw Myint, Michiel Niesen, Yizhi Shen, Philipp Stegmann, Ilia Tutunnikov and Evan Piephoff.

In addition, I also want to thank the wonderful people in the Chemistry Education office and the administrative staff who go out of their way to make lives easier for graduate students including Jennifer Weisman, William McCoy Angelina Toro, Madavin Vong, Michelle Contos, Rachael Fuller, Rebecca Teixeira, Jay Matthews, Grace Kimball, Mitch Moise, Peter Giunta, Li Miao, and Betty Lou McClanahan.

Additionally, my graduate studies and personal well-being benefited greatly from a tremendous group of friends at MIT and beyond who have always offered their support and encouragement. In particular I would like to thank Aditya Nandy, Nathan Faialaga, Ameya Rao, Ruofei Jia, Ardavan Farahvash, Maria Alejandra Castellanos Morales, Daniel Harper, Jet Lem, Calvin Leung, Mikail Khona, Dhruva Ganapathy, Arkya Chatterjee, Madhav Sankaranarayanan, Henry Tran, Natalie Golota, Vyshnavi Vennelakanti, Clifford Allington, Mikaila Hoffman, Bryan Tang, Jeong Min Park, Jess Xu, Christian Gomez, Erica Tsai, Andrew Lew, Genki Kubota, Adrien Levacic, Kitty Sompiyachoke and Yukina Chiba. I am also thankful for the friends and community I found at Ashdown House. I am grateful to

Xiaowei Ou, Wentao Cui, Zheng Dai, Tzu-An Sheng, Yanina Pankratova, Akhilan Boopathy, Tejas Jayashankar, Charvi Gopal, Ignacio Arzuaga, Ruoxuan Yang, Ivan Paskov, Lisa Lin and Narumi Nagaya for their camaraderie.

Finally, I am very grateful for the encouragement, love and support of my parents Atul Sood and Aradhna Sood, and my younger sister, Ambica Sood.

Dedication

To my father, Atul Sood, and my mother, Aradhna Sood, whose unconditional love and sacrifice made it possible for me to pursue my dreams.

Contents

List of Figures

Chapter 1

Introduction

Eukaryotic cells typically contain a membrane-bound organelle called the cell nucleus, which is home to the cell's genome. The genome is often hailed as the blueprint of life as it instructs the development and operation of an entire organism. The cell nucleus occupies a prominent role in cell biology as it controls gene expression and mediates the replication of DNA during the cell cycle [1].

A remarkable feature of nuclei is that despite their diverse and crowded chemical milieu, nuclear processes such as DNA replication, transcription, and RNA processing, all of which involve the concerted choreography of several biomolecules, are able to proceed with a high degree of specificity and selectivity. The specific, vectorial, and hierarchical organization of chemistry within the nucleus, unlike the cytoplasm, proceeds in the absence of membrane-bound organelles. Recent studies have elucidated that many such processes occur within bio-molecular condensates. It has recently become apparent that the cell leverages physical processes such as liquid-liquid phase separation driven by multivalent macromolecular interactions to exercise precise control over biomolecular condensates [2–5]. However, details concerning the precise molecular grammar are only beginning to emerge [6–12].

Another astonishing feat accomplished by higher organisms is the ability of every cell to express only a limited and highly specific set of genes that defines its cell type, despite all cells sharing the same genome. Robust heritability of not only genetic information but also cell type, via stable and inheritable changes in gene expression during cell division is accomplished through epigenetic control, such as post-translational modifications of nucle-

osomes' constituent histone proteins. Once again, cells deftly leverage physical principles, such as the universal behavior of polymers, general features of dynamical systems, and the electrostatic, physicochemical, and mechanical properties of DNA, histone proteins and their chemical modifications, alongside the physics of phase separation to achieve robust heritability of epigenetic states.

Interestingly, nucleosomal arrays and chromatin also have an intrinsic ability to phase separate, and further epigenetic modifications of constituent nucleosomes lead to the segregation of chromatin into blocks [13]. Furthermore, liquid phase condensation has been known to direct nucleosomal epigenetic modifications [14]. Chromatin, thus, shows characteristic behaviors of a viscoelastic polymer gel embedded in a proteinaceous liquid and consequently exhibits non-trivial relaxation dynamics [13, 15]. Although the interplay between one-dimensional epigenomic sequence and three-dimensional chromatin structure and organization has been well recognized [16–18], the explicit coupling between the chromatin's structural dynamics and stochastic epigenetic modification reactions and the consequences of dynamical asymmetry on the behavior of the coupled system have remained underappreciated in contemporary literature.

In this work, motivated by the biological observations previously outlined, we seek to investigate through a physical lens the mechanisms of epigenetic multistability and the concomitant impact of dynamical asymmetry between histone modification reaction networks and chromatin's structural dynamics. In Chapter 1.1 we briefly review the biology of the spread and maintenance of epigenetic histone modifications and their intimate connections with chromatin structure. We then provide a brief survey of theoretical models describing these processes in contemporary literature and succinctly motivate some unanswered questions that this thesis attempts to address. Furthermore, given the prominent role that liquid-liquid phase separation occupies in key nuclear processes such as chromatin structure, epigenetic control, and transcriptional regulation [19], in this work, we additionally carry out theoretical investigations into the interaction and sequence determinants of liquid-liquid phase separation coupled to gelation. In Chapter 1.2 we provide a brief review of biomolecular condensates and in particular emphasize their role in the context of regulation of subnuclear processes alongside a short overview of current theoretical approaches

describing sequence determinants of liquid-liquid phase separation coupled to percolation.

Finally, Chapter 1.3 provides an outline for the remainder of this thesis.

1.1 The interplay between chromatin structure and epigenetic histone modifications

1.1.1 An overview of biological mechanism

In human cells, a 2-meter-long strand of DNA is carefully stored in a nucleus approximately 10 micrometers wide [20]. Eukaryotic cells compact this long strand of DNA into chromatin that consists mostly of nucleosomes formed by DNA wrapping around histone proteins. The core histones that make up the nucleosome are subject to many different types of modifications, including acetylation, methylation, phosphorylation, and ubiquitination, which occur primarily at specific positions within the amino-terminal histone tails [21–23]. Specific additions and removals of these nucleosome modifications are carried out by classes of enzymes, including histone acetyltransferases (HATs), histone methyltransferases (HMTs), histone deacetylases (HDACs), and histone demethylases (HDMs).

From a chromatin structural point of view, eukaryotic genomes can generally be divided into two distinct environments [24, 25]. The first is a relatively open environment, containing most of the active genes and undergoing cyclical changes during the cell cycle, referred to as euchromatin. In contrast, other regions are relatively compact structures containing mostly inactive genes and are refractive to cell-cycle cyclical changes, and are referred to as heterochromatin.

The intimate connection between chromatin structure and histone modifications is highlighted by the observation that heterochromatin and euchromatin are each enriched with, and also lacking in, characteristic histone modifications[25]. Charge-altering modifications such as acetylation and phosphorylation directly alter the physical properties of the chromatin fiber [26–31]. For example, they can reduce the positive charge of histones, and this can disrupt the electrostatic interactions between histones and DNA. This presumably leads to a less compact chromatin structure, thereby facilitating DNA access by protein machin-

ery such as those involved in transcription. Multiple histone acetylations are enriched at enhancer elements and particularly in gene promoters, where they presumably facilitate the transcription factor access [21].

Another more indirect paradigm involves interactions with proteins via specific domains that allow them to recognize particular histone modifications, and effect structural changes [25, 32–38]. For example, histone-acetylated lysines are bound by bromodomains, which are often found in HATs and chromatin-remodeling complexes [25, 32]. For example, Swi2/Snf2 contains a bromodomain that targets it to acetylated histones. In turn, this recruits the SWI/SNF remodeling complex, which functions to 'open' the chromatin [33].

Positive feedback often arises in the spread and maintenance of histone modifications, when nucleosomes that carry a particular modification recruit (directly or indirectly) enzymes that catalyze similar modification of neighboring nucleosomes. A well known example exists in *S. pombe* between H3K9 methylated histones and the Swi6 chromodomain which help maintain the heterochromatic state through cell division [39]. This involves a positive feedback mechanism in which H3K9 methylation recruits Swi6 via its chromodomain, and Swi6 in turn recruits the H3K9 methyltransferase Clr4, which then modifies H3K9 on other histones in the vicinity. Indeed, many other HATs, HDACs, and HMTs are known to associate *in vitro* or *in vivo* with histones of the type that they are capable of producing [40–42]. Local chromatin structure also has an effect on the effectiveness of these modifying enzymes, hence an inherent feedback loop exists between structure and histone modification. For example, genomic regions localized inside compact three-dimensional domains often exhibit epigenetic profiles that are more extended compared to those of loci inside weakly compacted regions [43].

Histone modifications and corresponding structural changes in chromatin can be associated with active and repressed chromatin states as they play a role in modulating gene expressions by controlling the openness of chromatin [44–48], and thereby affecting the accessibility of binding sites to transcription factors and other regulatory proteins [49–53]. This means chromatin is not an inert structure, but rather an instructive DNA scaffold [44, 46, 54–56]. This makes understanding these systems crucial for a wide range of important biological processes such as cellular differentiation, cellular reprogramming, and tumori-

genesis [57–59].

1.1.2 A brief survey of theoretical approaches

In the theoretical literature, multistability in chromatin states formed by various histone modifications has been attributed to the presence of these positive feedback loops [16, 43, 60–74]. This feedback is born out of the ability of existing epigenetic marks to recruit enzymes to confer similar marks at new nucleosomes [39–42, 75]. The three-dimensional structure of the genome facilitates these processes by bringing nucleosomes far apart in sequence into spatial proximity facilitating long-range spreading of histone marks [76–80]. These models integrate feedback between one-dimensional epigenomic sequence and three-dimensional chromatin structure by modeling chromatin as an array of nucleosomes whose local state can fluctuate between 2 (or more) values indicating the presence (or absence) of particular histone modifications. The stochastic switching between these states is controlled by either "random" conversions accounting for non-feedback mediated processes such as the exchange of nucleosomes with nucleoplasm or via "recruited" conversions accounting for the histone modifying enzymes to be recruited by specific histone marks and to catalyze the addition of the same mark (or the erasure of an antagonistic mark) on proximal nucleosomes.

Early efforts in coupling chromatin structure and epigenetic reaction networks primarily relied on mean-field models, either assuming fully connected nucleosomal sites or incorporating a power law contact probabilities between sites to allow for long-range spreading of marks [60, 67, 74]. These zero-dimensional kinetic models capture multi-stability in histone modification states and demonstrate that local interactions between nucleosomes are not sufficient to maintain a stable epigenetic state and that cooperativity and interactions beyond the nearest neighbor are required to establish robust bistability. However, the explicit contributions from the chromatin's three-dimensional structure are obviously neglected. Many groups have addressed these shortcomings by incorporating polymeric effects either via explicit polymer simulation or leveraging poylmer-based Monte-Carlo approaches when modeling histone modification kinetics [81–91]. Typically, it is assumed that the spreading process is directly linked to the current 3D configuration by positing that

the transition rate of a monomer towards an epigenetically modified state is proportional to the current number of other modified monomers that are in spatial proximity. Similarly, the pairwise interactions driving polymer dynamics are directly dependent on the current epigenomic state of each monomer. These models have offered some insights. They display a first-order-like transition between a swollen, epigenetically disordered phase and a compact, epigenetically coherent chromatin globule [81, 83]. Theoretical support for this notion of first-order phase transitions has been found in analytical studies of equilibrium, mean-field "magnetic polymer models" [92–98] which support two types of phases–a phase in which the polymer chain is extended and histone marks are incoherent, and another where the chain is collapsed and the histone marks are coherent.

Although these explicit polymer simulations have been insightful, they are often computationally expensive and aren't particularly convenient for studying rare-event transitions between stable minima, which again is of interest if one wishes to make connections with biological processes like cellular differentiation and reprogramming. Furthermore, in polymer simulation schemes, the marks and the polymer dynamics are in fact not updated in concert as authors typically poll the polymer and attempt a metropolis Monte-Carlo recoloring move periodically [81, 83]. Such schemes are thus implicitly making assumptions about the relative relaxational dynamics of chromatin and marks. Recent investigations have revealed chromatin exhibits nontrivial rheology and viscoelastic properties, with multiple, disparate relaxation timescales, and organizes into regions of varying mobility [15, 99–106]. In particular, *in vivo* studies of chromatin report solid-like behavior [107–109] and structural relaxation occurring on the timescale of hours, comparable to the rate of enzyme-mediated histone modifications [43, 110, 111]. However, explicit coupling between the chromatin structure and epigenetic modifications, and a systematic interrogation of the concomitant impact of time scale separation between their dynamics is underexplored in existing literature, necessitating further investigation.

Furthermore, the nature of steady states produced by existing models requires additional deliberation. While they support the multistability of distinct histone modification patterns, the coherently modified states often adopt identical, compact chromatin conformations [62, 73, 74, 81–83, 90, 93, 112]. However, numerous studies that probed chro-

matin organization with various techniques have revealed dramatic structural differences among chromatin with distinct modifications [113, 114]. A more biologically relevant outcome would correspond to two states that support an open, unmarked (euchromatin) and collapsed, marked chromatin (heterochromatin).

1.2 The role of biomolecular condensates as membrane-less organelles in intranuclear organization

1.2.1 An overview of biological mechanism

The spatiotemporal organization of cellular material is a necessary prerequisite for the complex biochemical processes that allow cellular life to exist. Cells can accomplish this either via membrane-bound organelles or via biomolecular condensates which act like membrane-less organelles (MLOs) [4, 115–120]. The latter are non-stochiometric assemblies of biomolecules that can form via spontaneous or driven processes and share hallmarks of phase separation and percolation [2, 121, 122]. The chemical milieu and the physicochemical environment of condensates are distinct from their surrounding environment [123–127]. This allows cells to selectively partition biomolecules, avoid interference between various biochemical pathways and buffer against chemical noise [6, 128–131]. Intrinsically disordered proteins have been implicated as key constituents in many biomolecular condensates and have a demonstrated propensity to phase-separate at physiological concentrations due to the prevalence of multivalent protein-protein (or protein-nucleic acid) interactions [132–135].

Remarkably, even though the cell nucleus is a crowded chemical environment wherein multiple biochemical processes essential to life are orchestrated with a high degree of specificity, it is strikingly lacking in membrane-bound organelles. The cell nucleus deftly leverages the physics of liquid-liquid phase separation for the hierarchical organization of its contents across length scales, from super-enhancers ($\sim$ 10 kb) to nucleoli ($\sim 0.2 - 3.5$ micron) [2, 13]. Thus, bio-molecular condensates mediate and regulate a wide array of nuclear processes including, regulation of chromosome structure and maintenance [13, 34, 35,

136–138], transcription [36, 53, 139–142], RNA processing [143, 144], DNA replication [145], and DNA repair [146, 147].

In particular, nucleosomal arrays and chromatin display intrinsic ability to phase separate [13, 136]. This ability is regulated by the ability of histone tails to participate in weak multivalent interactions mediated by their post-translation modifications, contributing to the establishment of different chromatin states associated with varying levels of gene activity [23, 148]. Furthermore, as discussed in Chapter 1.1, histone modifications are spread and maintained via proteins that will bind to (or be excluded by) specifically modified residues. These proteins that weakly bind to histones based on their modification state can, in turn, recruit enzymes that contribute to local gene activity or repression [54, 55]. Additionally, the multivalent interactions among these proteins and histone post-translational modifications are associated with condensate formation and dissolution. Therefore, one can imagine genomic loci with specific "chromatin marks" might in fact represent different dynamic condensate states [23, 136, 138]. Concretely, it has been observed that BRD4 molecules, readers of acetylated nucleosomes, concentrate in active transcriptional condensates [36]. Similarly, HP1α proteins bind nucleosomes marked by H3K9me3 in heterochromatin condensates. Moreover, components of heterochromatin such as nucleosomes and DNA preferentially partition into the HP1α droplets, but transcription factors do not show such preference [19, 28].

Beyond mediating gene regulation at the level of chromatin structure and epigenetic control, biomolecular condensates and liquid-liquid phase separation are also implicated in transcriptional regulation. For example, transcriptional condensates comprising transcription factors have been known to form at enhancers and promoters [139, 141, 142]. Mutations localized within the low complexity domains of transcription factor can disrupt the propensity to phase separate and thereby disrupt normal functioning. [140, 149]. Phase separation-mediated mechanisms have also been proposed as a way for transcription factors to efficiently find and bind to their targets in eukaryotic genomes [150–153]

1.2.2 A brief survey of theoretical approaches

In recent years, biomolecular condensates have garnered significant prominence and attracted substantial research interest [3, 4]. Notably, many theoretical and computational approaches have come to the fore to interpret, predict, and explain experimental outcomes [5]. Prominent among them are data-driven and machine learning approaches [7, 154, 155], molecular simulations [27, 156–162] and physical theory [3, 4, 131, 163–174]. There exists a natural synergy between these approaches. Physical simulation and data-driven approaches often enable quantitative comparisons with experiments and can help make context-specific predictions. Computationally derived data can also be used for validation and development of physical theories. On the other hand, physical theories give us a way to tease out overarching physical principles governing condensate behavior, without regard to specific realizations (either *in vivo, in vitro* or *in silico*) of systems.

As a first approximation, mean field theories can successfully capture core phenomenology, such as spinodal and binodal decomposition, associated with multi-component, phase separating homopolymer solutions [131, 166–169]. However, phase-separating proteins are often finite-sized heteropolymers, and homopolymer-inspired mean-field theories alone fail to adequately capture the sequence and structural heterogeneities and the hierarchy of anisotropic interactions encoded by the multiway interplay among heteropolymers and the solvent. Therefore, recently the “stickers and spacers" model, originally developed within the context of associative polymers and polymer gelation theory [163–165], has gained popularity for modeling protein condensates [3, 4, 9, 175–178]. In this model, biomolecules within condensates are envisioned as possessing two distinct functional components: “stickers" and “spacers". Stickers represent specific molecular domains or motifs with a high affinity for one another, facilitating interaction and bringing molecules into close proximity, thereby contributing to the condensate’s cohesive, networked structure. Conversely, spacers act as flexible linkers connecting the stickers, enabling the dynamic and transient nature of interactions within the condensate, and modulate phase separation via their effective solvation volume [179–182]. This model offers a conceptual framework that assists in interpreting experimental observations regarding condensate stability and

material properties [183–185].

The simplicity inherent in the stickers and spacers model, which contributes to its theoretical elegance, faces challenges when applied to realistic biomolecules. In particular, when dealing with IDPs, the identification of stickers and spacers becomes a non-trivial task [4, 116, 186, 187]. Unlike the model's assumption that only two amino acid types or segments can straightforwardly map onto stickers and spacers, most proteins exhibit a more complex structure. Treating a substantial fraction of proteins as identical spacers oversimplifies the intricate nature of protein sequences and the diverse interactions they facilitate. Therefore, it becomes pertinent to explicitly consider spacer heterogeneity. Such inclusion can offer valuable insights into the impact of sequence diversification on condensate formation, contrasting with predictions from the stickers and spacers model. This deliberation may provide a clearer understanding of whether IDPs tend to favor sequences with low or high complexity. Moreover, it could shed light on the evolutionary optimization of IDPs toward specific sequence patterns.

1.3 Outline of this thesis

The thesis is organized into two parts. Part I is motivated by the discussion in Chapter 1.1. We focus on the interplay between chromatin structural dynamics and stochastic reaction networks describing histone modifications. In Chapter 2, we introduce a spatial-mean field model in the spirit of seminal models of epigenetic regulation [60, 67] alongside a novel second-quantization based approach that can be used to analyze discrete stochastic models with a fixed, finite number of particles using a representation of the $SU(2)$ algebra. We apply the approach to a kinetic model of chromatin states that captures the feedback between nucleosomes and the enzymes conferring histone modifications. Using a path integral expression for the transition probability, we compute the epigenetic landscape that helps identify the emergence of bistability and the most probable path connecting the two steady states. In Chapter 3, we propose kinetic models to investigate the dynamic fluctuations of histone modifications and the spatial interactions between nucleosomes. We explicitly incorporate the influence of chemical modifications on the structural stability of

chromatin and the contribution of chromatin contacts to the cooperative nature of chemical reactions. Through stochastic simulations and analytical theory, we demonstrate distinct steady-state outcomes in different kinetic regimes and uncover the hallmarks of a dynamical phase transition. Importantly, we validate that the emergence of this transition, which occurs on biologically relevant timescales, is robust against variations in model design and parameters. Our findings suggest that the viscoelastic properties of chromatin and the timescale at which it transitions from a gel-like to a liquid-like state significantly impact dynamic processes that occur along the one-dimensional DNA sequence.

In Chapter 4 we introduce a minimal kinetic model for gene regulation that combines the impact of both histone modifications and transcription factors. We further develop the field-theoretic methodology first introduced in Chapter 2 and propose an approximation scheme based on variational principles to solve the corresponding master equation in the second quantized framework. By analyzing the steady-state solutions at various parameter regimes, we demonstrate that histone modification kinetics can significantly alter the behavior of a genetic network, resulting in qualitative changes in gene expression profiles. The emerging epigenetic landscape captures the delicate interplay between transcription factors and histone modifications in driving cell-fate decisions.

Motivated by the prominent role of biomolecular condensates in regulating chromatin's structure, epigenetic memory, and transcriptional control (see Chapter 1.2), we devote Part II of this thesis to the study of liquid-liquid phase separation coupled to percolation. Intrinsically disordered proteins (IDPs) often serve as primary components of these biomolecular condensates due to their flexibility and ability to engage in multivalent interactions, leading to spontaneous aggregation. Theoretical advancements are critical to connecting IDP sequences with condensate emergent properties to establish the so-called molecular grammar. Therefore, in Chapter 5 we propose an extension to the stickers and spacers model, incorporating heterogeneous, non-specific pairwise interactions between spacers alongside specific interactions among stickers. Our investigation reveals that while spacer interactions contribute to phase separation and co-condensation, their non-specific nature leads to disorganized condensates. Specific sticker-sticker interactions drive the formation of condensates with well-defined networked structures and molecular composition.

We further discuss how evolutionary pressures might emerge to affect these interactions, leading to the prevalence of low-complexity domains in IDP sequences. These domains suppress spurious interactions and facilitate the formation of biologically meaningful condensates.

Part I

The dynamical interplay between chromatin structure, epigenetic reaction networks and gene regulation

Chapter 2

Quantifying Epigenetic Stability with Minimum Action Paths [1]

2.1 Introduction

A remarkable achievement of multicellular organisms is the formation of distinct cell types with identical genomes. Covalent modifications of histone proteins, of which DNA wraps around to form chromatin, are expected to be crucial for the emergence of cellular diversity [22]. These epigenetic marks can regulate the output of the genome by promoting or restricting the accessibility of the DNA sequence. They are known to impact the openness of chromatin and global genome organization, though the molecular mechanisms are only beginning to emerge [44–47]. Therefore, multistability in chromatin states formed by various histone modifications or combinations thereof can potentially give rise to distinct patterns of gene expression and inheritable phenotypes [16, 110, 188, 189]. Evidence for bistable and inheritable epigenetic marks has indeed been found that can be attributed to the presence of positive feedback loops wherein nucleosomes that carry a particular modification recruit, either directly or indirectly, enzymes that catalyze the same modification on neighboring nucleosomes [48, 51, 52, 60, 61].

[1] Based on work published in Ref. 69.

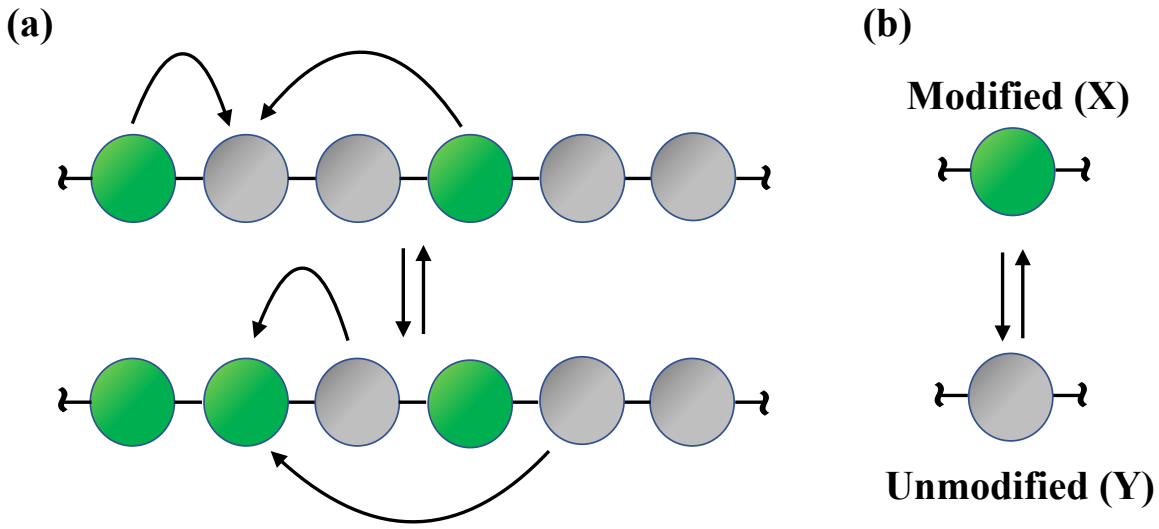

Figure 2.1: Illustration of the kinetic model for the interconversion between modified (green, X) and unmodified nucleosomes (grey, Y). (a) Recruited conversion defined in Eq. (2.1) that requires a pair of (un-)modified nucleosomes to alter the state of a nucleosome. (b) Noisy conversion (Eq. (2.2)) with first order kinetics.

2.2 Model

Mathematical modeling of the reaction networks of histone modifications can help determine factors that are crucial for epigenetic stability. Distilling the essence of feedback mechanisms, Dodd and coworkers introduced a simplified kinetic model with bistable chromatin states [67]. One envisions a system of N nucleosomes, where a nucleosome can exist in either a modified or an unmodified state (see Fig. 2.1). As a first approximation, spatial organization of chromatin is neglected, and the kinetics of the system can be described with the non-linear dynamics given below

$$X + X + Y \xrightarrow{c_1} 3X, \qquad Y + Y + X \xrightarrow{c_1} 3Y, \tag{2.1}$$

$$X \xrightarrow{c_2} Y, \qquad Y \xrightarrow{c_2} X. \tag{2.2}$$

The reactions governing the dynamics in Eq. (2.1) and (2.2) represent recruited and thermalized, noisy conversions respectively. From analysis of regulatory circuits, we know that

in deterministic systems, bistability requires not only positive feedback but also nonlinearity in the feedback loop [60, 190–192]. However, sufficient noise can produce bistability with fewer constraints on the deterministic system [193, 194]. The model presented is nonlinear, since the recruited conversion of Y to X (or X to Y) is bimolecular in X (or Y) and unimolecular in Y (or X). Thus the rate of production for a given nucleosome type responds to increases in its own concentration in a nonlinear fashion. This produces bistability in the determinstic system, wherein noise can allow for transitions between the two stable attractors.

Far from being a trifle toy-model, the kinetic scheme above is not unlike the mating type silencing in *S. Cerevisiae* [67, 195–198]. Generalizing the above model to more than 2 epigenetic states has been attempted as well [63–66, 198]. Their elegance and biological relevance have inspired numerous theoretical studies of these models [16]. A popular approach used in these studies to investigate epigenetic stability is to posit deterministic rate equations followed by bifurcation analysis. Insight into the switching among chromatin states is missed in such deterministic analyses, however. To study the rare transition events between steady states, Dodd and coworkers introduced an approximate Fokker-Planck equation, from which an epigenetic landscape can be constructed.

2.3 Results

In this work, we present an alternative way to analyze such zero-dimensional models. We turn to the original master equation, which is an exact stochastic description of the underlying process describing the temporal evolution of the system's configurational probability, and reformulate it using second-quantization (or Fock-Space) methods (Doi-Peliti approach) [199–203]. While canonical approaches rely on *bosonic* creation and annihilation operators, we employ operators that are a representation of the $SU(2)$ algebra, in order to treat the constraint that fixes the total number of nucleosome types ($X+Y \equiv N$) in a more mathematically natural fashion. The Doi-Peliti method has been successfully employed in the study of reaction-diffusion processes [204], gene switches [205, 206], and other systems [207]. We outline the main results here and the detailed derivations are consigned to

Appendix A.1. The Doi-Peliti approach allows us to reformulate the time evolution of the original master equation as an *imaginary* time Schrödinger equation

$$\partial_t \left|\psi(t)\right\rangle = -H \left|\psi(t)\right\rangle, \tag{2.3}$$

where we have introduced *formally* a state vector $\left|\psi(t)\right\rangle$ as a superposition of all possible occupation number configurations weighted with their corresponding probabilities (a *generating* function),

$$\left|\psi(t)\right\rangle = \sum_{\{n_x,n_y\}} P(\{n_x,n_y\};t) \left|n_x,n_y\right\rangle. \tag{2.4}$$

Following standard procedures [205–207], H is usually expressed in a second-quantized form

$$\begin{aligned} H = \frac{c_1}{(2j)^3}[a_x^{\dagger 3}a_x^2 a_y - a_x^{\dagger 2}a_x^2 a_y^\dagger a_y] + \frac{c_1}{(2j)^3}[a_y^{\dagger 3}a_y^2 a_x - a_y^{\dagger 2}a_y^2 a_x^\dagger a_x] \\ + \frac{c_2}{2j}[a_y^\dagger a_x - a_x^\dagger a_x] + \frac{c_2}{2j}[a_x^\dagger a_y - a_y^\dagger a_y]. \end{aligned} \tag{2.5}$$

a_i and $a_i^\dagger$ for $i \in \{x,y\}$ are bosonic creation and annihilation operators that obey the canonical commutation relations

$$[a_i, a_j^\dagger] = \delta_{ij} \quad \text{and} \quad [a_i, a_j] = 0 = [a_i^\dagger, a_j^\dagger]. \tag{2.6}$$

The action of $a_x, a_x^\dagger$ on ket vectors $\left|n_x,n_y\right\rangle$ is given by $a_x \left|n_x,n_y\right\rangle = n_x \left|n_x-1,n_y\right\rangle$, $a_x^\dagger \left|n_x,n_y\right\rangle = \left|n_x+1,n_y\right\rangle$, and $a_x^\dagger a_x \left|n_x,n_y\right\rangle = n_x \left|n_x,n_y\right\rangle$. Similar operations can be defined for $a_y, a_y^\dagger$. Since for a vacuum state $|0,0\rangle$, $a_i |0,0\rangle = 0$, one can obtain any arbitrary ket state as $\left|n_x,n_y\right\rangle = (a_x^\dagger)^{n_x}(a_y^\dagger)^{n_y}|0,0\rangle$.

While it is straightforward to apply the standard formalism up until this point, one notices that the total number of nucleosomes in our system $N = \hat{n}_x + \hat{n}_y = a_x^\dagger a_x + a_y^\dagger a_y$ is constant. This is evident from the fact that our Hamiltonian commutes with the total number operator, $a_x^\dagger a_x + a_y^\dagger a_y$. One might also find it slightly philosophically troubling, to use bosonic ladder operators to describe a system with a large, but still finite number

of particles. Moreover, not all combinations of states X and Y are allowed, but only those satisfying $n_x + n_y = N$. In conventional quantum mechanics, bosonic ladder operators allow for neither an exclusion principle, nor a cap on the total particle number. However, since the Hamiltonian conserves particle number, we aren't remiss in our formalism. We can take any equilibrium solution of the master equation and project down to the subspace where the conserved quantity takes a fixed value and we will get another equilibrium solution. In probability theory we would say we are conditioning on the conserved quantity taking a definite value. Thus if our initial state $|\psi(0)\rangle$ is a configuration such that $\hat{n}_x + \hat{n}_y = N$ holds true, and then given our prescription of stochastic Hamiltonian all subsequent states will meet this condition as well. However, given that there is essentially only one independent variable, namely n_x, we can perhaps phrase this problem in a more natural framework. We develop this framework in the sections that follow.

The starting point of this reformulation is the Jordan-Schwinger map [208, 209], where we introduce (see Appendix A.1, Eqs. (A.3), (A.4))

$$J_+ = a_x^\dagger a_y \qquad J_- = a_y^\dagger a_x. \tag{2.7}$$

For notational convenience we set $N = 2j$. Here the operators satisfy the commutation relations of $SU(2)$ algebra

$$[J_z, J_\pm] = \pm J_\pm \qquad [J_i, J_j] = i\varepsilon_{ijk} \quad \forall i,j,k \in \{x,y,z\}, \tag{2.8}$$

where the structure constant ε_{ijk} is the Levi-Civita symbol. Re-defining the ket $\left|n_x, n_y\right\rangle = |n_x, 2j - n_x\rangle \equiv |n\rangle$, their action is given by

$$\begin{aligned} J_+ |n\rangle &= (2j - n)\,|n+1\rangle\,, \\ J_- |n\rangle &= n\,|n-1\rangle\,, \\ J_z |n\rangle &= (n - j)\,|n\rangle\,, \end{aligned} \tag{2.9}$$

$$\hat{n}_x |n\rangle = (J_z + j) |n\rangle = n |n\rangle , \tag{2.10}$$
$$\hat{n}_y \left|n_y\right\rangle = (j - J_z) |n\rangle = (2j - n) |n\rangle .$$

The Hamiltonian can be reformulated as

$$H = \frac{c_1}{(2j)^3}[J_+ \hat{n}_x^2 - \hat{n}_x^2 \hat{n}_y] + \frac{c_1}{(2j)^3}[J_- \hat{n}_y^2 - \hat{n}_y^2 \hat{n}_x] \tag{2.11}$$
$$+ \frac{c_2}{2j}[J_- - \hat{n}_x] + \frac{c_2}{2j}[J_+ - \hat{n}_y],$$

where we have used $\hat{n}_i^{\underline{2}} = \hat{n}_i(\hat{n}_i - 1))$ to denote the falling factorial to write down the Hamiltonian in a more compact form.

A great advantage of the second quantization approach is its relative convenience for deriving analytical solutions. For example, a formal solution to Eq. (2.3) is given by

$$|\psi(t)\rangle = \exp(-Ht) |\psi(0)\rangle . \tag{2.12}$$

In addition, the transition probability of starting in a state with particle number n_i at time $t = 0$ and ending up in a state with particle number n_f at t_f can be defined as

$$\left\langle \psi(t_f)\right| e^{-tH} \left|\psi(0)\right\rangle \equiv \left\langle n_f\right| e^{-tH} \left|n_i\right\rangle . \tag{2.13}$$

We next seek for a path integral expression of the transition probability that is useful for finding steady states and transition pathways between them. We discretize the time interval $[0, t_f]$ into N_t time slices, and then insert a resolution of identity between each time slice. Finally, taking the limit $N_t \to \infty$ we get,

$$\begin{aligned} \mathbb{P}(\rho_f; t_f | \rho_i; 0) &= \left\langle n_f\right| e^{-tH} \left|n_i\right\rangle \\ &= \int \mathscr{D}[z_i, z_f] \left\langle n_f | z_f\right\rangle \left\langle z_f\right| e^{-tH} \left|z_i\right\rangle \left\langle z_i | n_i\right\rangle \\ &= \int \mathscr{D}[\overline{z}_i, z_f] \int \mathscr{D}[\overline{z}, z] e^{-S}, \end{aligned} \tag{2.14}$$

where we have introduced, $\rho = n/2j$. By definition, ρ represents an order parameter that

quantifies the fraction of modified nucleosomes. After performing the integration over $z_i, \bar{z}_i, z_f, \bar{z}_f$. we introduce one final re-parametrization in terms of the density ρ. Using

$$\rho = \frac{\langle z|j + J_z|z\rangle}{2j} = \frac{z\bar{z}}{1+z\bar{z}}, \tag{2.15}$$

we can rewrite

$$z = \frac{\rho}{1-\rho}\exp(-\tilde{\rho}), \qquad \bar{z} = \exp(\tilde{\rho}) \tag{2.16}$$

with $\rho(0) = \rho_i$ and $\rho(t) = \rho_f$ and $\tilde{\rho}(t)$, $\tilde{\rho}(0)$ unconstrained. Making these substitutions, the action finally reads,

$$S[\tilde{\rho}, \rho] = 2j \int_0^{t_f} \mathrm{d}t \left[\tilde{\rho}\partial_t \rho - H(\tilde{\rho}, \rho)\right], \tag{2.17}$$

where

$$\begin{aligned} H(\tilde{\rho}, \rho) = -\frac{1}{2j^2} e^{-\tilde{\rho}} \left(e^{\tilde{\rho}} - 1\right) \left[c_1 (j-1)(2j-1)\right. \\ \left. \left(\rho - 1\right)\rho \left(\rho + \rho e^{\tilde{\rho}} - 1\right) + 2c_2 j^2 \left(\rho + (\rho - 1)e^{\tilde{\rho}}\right)\right]. \end{aligned} \tag{2.18}$$

Correspondingly, the time-dependent transition probability (propagator) can be expressed as

$$\mathbb{P}(\rho_f; t_f | \rho_i; 0) = \int \mathscr{D}[\tilde{\rho}, \rho] \exp(-S[\tilde{\rho}, \rho]). \tag{2.19}$$

Eq. (2.17) is the main result of this paper. It allows the computation of both steady state and kinetic results for the model in terms of the order parameter ρ. In particular, for $2j \gg 1$, i.e., the small noise regime with many nucleosomes, the path integral in Eq. (2.19) will be dominated by contributions from the minimum action path [210–214]. The

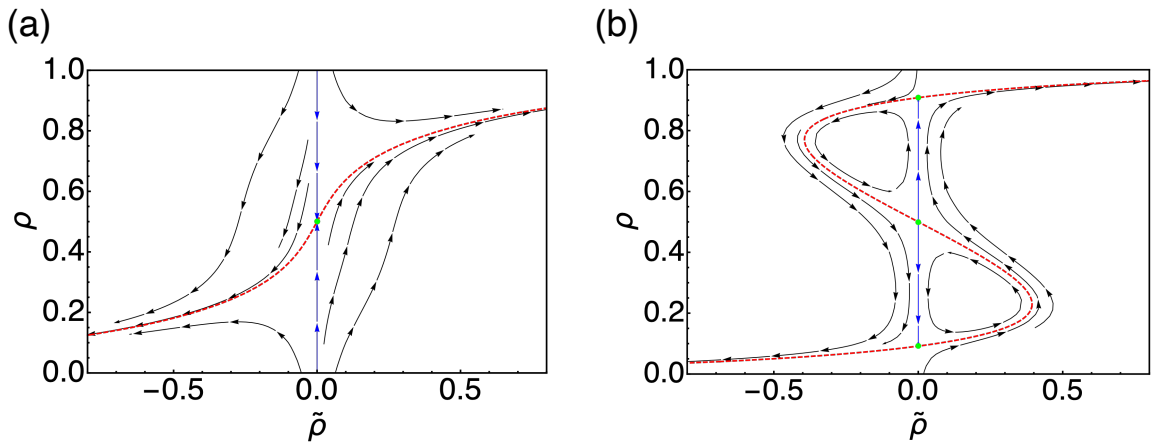

Figure 2.2: Phase portrait determined using Eq. (2.20) with kinetic parameters $c_1/c_2 = 3$ (a) and 12 (b). The red dashed lines are zero-energy paths and green dots are steady state solutions. The blue paths represent deterministic trajectories. The number of nucleosomes was held fixed at $N = 60$.

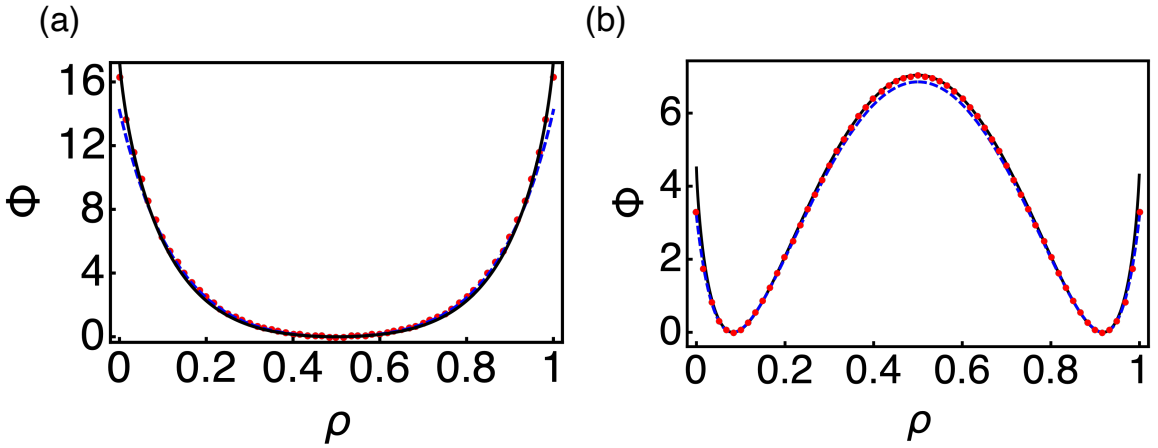

Figure 2.3: Comparison between the steady state distribution ($-\log P_{\text{eq}}$, red dots) and the quasi potential (Φ) computed using Eq. (2.21) (black solid line) and the Fokker-Planck equation (blue dashed line) for $c_1/c_2 = 3$ (a) and 12 (b). The number of nucleosomes was held fixed at $N = 60$.

variational derivatives that minimize the action yield the *classical Hamiltonian equations*

$$\dot{\rho} = \frac{\partial H}{\partial \tilde{\rho}},$$
$$\dot{\tilde{\rho}} = -\frac{\partial H}{\partial \rho}. \tag{2.20}$$

We note that $\tilde{\rho} \equiv 0$ is always a solution to the above Hamiltonian equation. As shown in Figure 2.2 in blue color, these paths correspond to deterministic dynamics flowing towards steady-state solutions (green dots). The resulting deterministic equation (see Eq. (A.17)) is identical to that presented in Ref. [67], which was obtained using phenomenological arguments. For the Hamiltonian paths with non-zeros $\tilde{\rho}$, the zero-energy path with $H \equiv 0$ is of particular interest as it represents fluctuations away from the steady states. In bistable regimes, the zero-energy path connects the two steady states via the saddle point (see Fig. 2.2b red), and corresponds to the maximum likelihood transition path [215, 216].

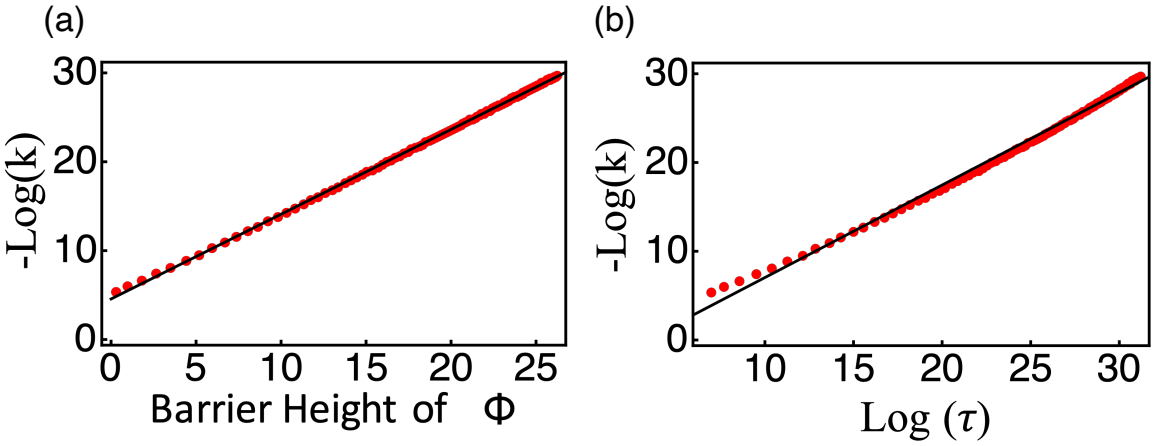

Figure 2.4: Correlation between the exact transition rates (k) computed from diagonalizing the transition matrix and the barrier height of the quasi-potential (a) or the mean first passage time (τ) estimated using the Fokker-Planck equation (b). Each data point corresponds to an independent calculation for integer values of the parameter c_1/c_2 between 5 and 120. The total nucleosome number was fixed as $N = 60$.

Quantitative results of the kinetic model can be obtained with the definition of a quasi-potential, Φ, in the Friedlin-Wentzell sense [217] in terms of the least-action path (denoting

$\varphi = \{\rho, \tilde{\rho}\}$)

$$\Phi(x_1, x_2) = \inf_{t>0} \inf_{\varphi \in \gamma_{x_1}^{x_2}(0,T)} S(\varphi), \tag{2.21}$$

where $\gamma_{x_1}^{x_2}(0,t)$ is the set of continuous curves f connecting two points x_1, x_2 in configuration space, such that $f(0) = x_1, f(t) = x_2$. The minima of the Friedlin-Wentzell quasi-potential correspond to attractors of the dynamical system, and the height of the barrier corresponds to the ease of transition between two stable fixed points. As shown in Fig. 2.3, the quasi-potential correctly captures the emergence of bistability as the parameter c_1/c_2 varies from 3 to 12. c_1 and c_2 are the rate coefficients for recruited and random nucleosome conversions defined in Eqs. (2.1) and (2.2). In addition, we found that in both cases, the quasi-potential agrees well with the negative-log of the steady state probability distribution determined using the zero eigenvalue eigenvector of the transition rate matrix (see Appendix A.2 for details). When compared to the transitions rates between the two steady states, the barriers of the saddle point determined from the quasi-potential strongly correlate with the numerical values over a wide range of parameters (see Fig. 2.4).

2.4 Discussion

In this communication, we applied the Doi-Peliti approach to a reaction network that captures the emergence of epigenetic stability from histone modifications [67]. Together with a transformation enabled by the $SU(2)$ algebra, it allowed for the derivation of analytical results that rigorously account for the constraints imposed by a fixed number of particles. The semi-classical treatment of the path integral expression for transition probability further provided a fresh view of the stochastic reaction network in the guise of a "pseudo-mechanical system". Hamilton's equation of motion and the quasi-potential, much like their counterparts in classical mechanics, provide intuition regarding the dynamics and landscape of the reaction network.

We note that the steady state distribution and transition rate for the chromatin state model can be obtained from the Fokker–Planck equation introduced in Ref. [67] as well.

The Fokker–Planck equation may, in fact, seem more appealing compared to the Doi-Peliti approach used here due to its mathematical simplicity. As shown in Figs. 3 and 4, the accuracy of the two methods is comparable. The Doi-Peliti approach slightly outperforms in reproducing solutions obtained from diaogonalizing the master equation, due to its improved treatment of larger deviations.

The advantage of the Doi-Peliti approach will become more evident for more complex problems with coupled reaction networks, such as a chromatin switch coupled to a self-activating gene. There, due to the high dimensionality, analytical/numerical solutions of the Fokker-Planck equation cannot be readily obtained. On the other hand, robust algorithms have been introduced to compute the minim action paths of complex networks [218]. These most probable paths could be used to quantify epigenetic stability and investigate the mechanism for switching between steady states [206]. We note that both the Fokker-Planck equation and the minimum action approximation are only valid at the small-noise limit. When the number of nucleosomes is small (<30), significant deviations from the analytical results presented here may occur. The small number regime, though physically interesting, is of less biological interest as epigenetic domains are large and can cover a long stretch of DNA (10 $\sim$ 100 Kb). In addition, estimations based on the minimum action can, in principle, be improved by including higher-order terms from a functional Taylor expansion of Eq. (2.19) [219].

Chapter 3

Dynamical phase transition in models that couple chromatin folding with histone modifications [1]

3.1 Introduction

Eukaryotic cells compactly package their genome into chromatin that consists primarily of nucleosomes formed by DNA wrapping around histone proteins [20, 221]. These core histones are often subject to post-translational marking, including acetylation and methylation, [21, 22] which partitions chromosomes into distinct domains with differential transcription activity [24, 25, 32, 33, 45, 48–53, 189, 205, 206, 222–224], providing active (euchromatic) and inactive (heterochromatic) regions with characteristic chemical signals. Co-existence of stretches of chromatin enriched and depleted of specific histone marks implies multistability in epigenetic regulation [16, 48].

Many theoretical models have been introduced to study the stability of histone marks [43, 60, 63–71]. Positive feedback underpins these models since existing marks recruit enzymes to confer similar marks at new nucleosomes [39–42, 75]. 3D structure of the genome births this feedback as chromatin loops bring nucleosomes far apart in sequence

[1] Based on work published in Ref. 220.

into spatial proximity facilitating long-range spreading of histone marks [76–79]. Additionally, chemical modifications also affect nucleosome-nucleosome interactions, either by directly altering the physicochemical properties of amino acids or by recruiting additional protein molecules [26–31], impacting chromatin organization. Therefore, chromatin is an instructive scaffold inextricably linked to epigenetic regulation [44, 46, 56].

Explicitly accounting for chromatin organization when studying histone marks is becoming increasingly important in recent studies [62, 73, 74, 81–83, 90, 93, 112, 225]. Early efforts in coupling chromatin structure and epigenetic reaction networks primarily relied on mean-field models, either assuming fully-connected nucleosomal sites or incorporating a power law contact probabilities between sites to allow for long-range spreading of marks [60, 67, 74]. Many groups expanded on these ideas by extracting nucleosome contacts with explicit polymer simulations when modeling histone modification kinetics [81–83, 85–89, 92]. Furthermore, recent work has revealed that chromatin exhibits nontrivial rheology and viscoelastic properties, with multiple, disparate relaxation timescales, and organizes into regions of varying mobility [15, 99–106]. In particular, *in vivo* studies of chromatin report solid-like behavior [107–109] and structural relaxation occurring on the timescale of hours, comparable to the rate of enzyme-mediated histone modifications [43, 110, 111]. However, explicit coupling between the chromatin structure and epigenetic modifications, and a systematic interrogation of the concomitant impact of time scale separation between their dynamics is under-explored in existing literature, necessitating further investigation. Furthermore, the nature of steady states produced by existing models requires additional deliberation. While they support the multistability of distinct histone modification patterns, the modified states often adopt identical, compact chromatin conformations [62, 73, 74, 81–83, 90, 93, 112, 225]. However, numerous studies that probed chromatin organization with various techniques have revealed dramatic structural differences among chromatin with distinct modifications [113, 114]. A more biologically relevant outcome would correspond to two states that support an open, unmarked (euchromatin) and collapsed, marked chromatin (heterochromatin). Models supporting structural changes between steady states would better represent biological systems and provide deeper insight into chromatin stability.

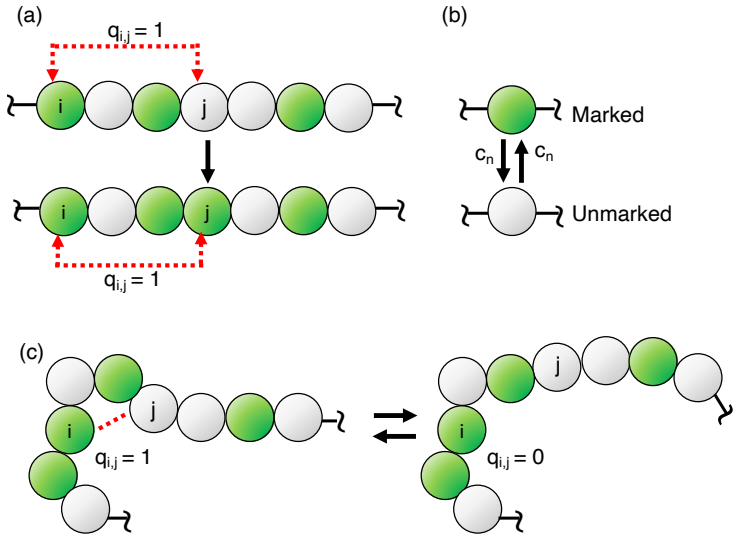

Figure 3.1: A schematic illustration of the salient features of a kinetic model explicitly accounting for the interdependence between changes in histone marks and chromatin contacts. Green and grey circles indicate marked and unmarked nucleosomes, respectively. (a) Marks can be added (or removed) via an enzyme-mediated recruited process wherein two sites that are in contact become similarly modified (Eq. (3.2)). $q_{ij} = 1$ indicates a direct contact in 3D space between two nucleosomes (i and j) separated in a linear sequence. (b) Nucleosomes can also be marked (or unmarked) via random conversions occurring independent of chromatin contacts (Eq. (3.1)). (c) Chromatin conformational dynamics are modeled as stochastic transitions in contact space, where contact formation (breaking) rates depend on polymer topology and nucleosome marks.

3.2 Model

We present a theoretical model with explicit coupling between chromatin conformational dynamics and histone modifications (Figure 3.1). The vector $\boldsymbol{n}(t) \equiv \{n_i(t)\}$ for $i \in [1, \mathbf{N}]$ of length $\mathbf{N}$ denotes the chemical state of chromatin at any given time, t. The binary variable, $n_i \in \{0, 1\}$, indicates the presence (or absence) of a histone mark at nucleosome i. Inspired by protein folding literature [226–229], we adopt a contact space representation of the chromatin conformation. A vector of size $\mathbf{M}$, $\boldsymbol{q}(t) \equiv \{q_{ij}(t)\}$ for $i, j \in [1, \mathbf{N}]$ and $j - i > 1$ represents the chromatin conformation at time t. $q_{ij} \in \{0, 1\}$ is again a binary variable denoting the presence (or absence) of 3D contacts between a pair of nucleosomes (i, j). Neighboring nucleosomes are always assumed to be in contact (i.e., $q_{i,i+1} = 1$). Similar to explicit polymer simulations, this model allows the coupling of histone chemical kinetics with instantaneous chromatin structural changes. Importantly, no assumptions about the timescale separation are needed, and rigorous stochastic simulation algorithms can be employed to examine the dynamical coupling across a wide range of timescales.

Following previous studies [60, 67, 69], two types of reactions that drive changes in histone marks are considered. First is an on-site, random conversion that arises from exchanging histone proteins with the nucleoplasm or reactions catalyzed by non-cooperative enzymes. For example, an unmarked nucleosome i with $n_i = 0$ (0_i) can become marked with $n_i = 1$ (1_i) at a basal rate c_n independent of chromatin conformation and the state of other nucleosomes. Similarly, marked nucleosomes can be converted back to unmarked ones. The corresponding reaction schemes are

$$0_i \xrightarrow{c_n} 1_i, \qquad 1_i \xrightarrow{c_n} 0_i. \tag{3.1}$$

The second is recruited conversions, a measure of cooperativity in the system, ensuring nucleosomes in spatial proximity are similarly marked. These reactions can arise due to the transfer of enzymes among nucleosomes in contact. We consider the cooperative effect for both addition and removal enzymes. Therefore, for a pair of contacting nucleosomes

(i, j) in different chemical states, either the mark at site j is removed, or a new mark is introduced to site i via recruited conversions as denoted below

$$0_i + 1_j \xrightarrow{c_{1\to 0}} 0_i + 0_j, \qquad 0_i + 1_j \xrightarrow{c_{0\to 1}} 1_i + 1_j. \tag{3.2}$$

Both reactions occur with rate c_r unless otherwise specified.

We treat chromatin conformational dynamics analogously as stochastic transitions in contact space. The rates of contact formation and dissolution are influenced by the interplay between the attraction among modified nucleosomes and the entropic effects stemming from homopolymer dynamics. However, due to the lack of precise expressions for polymer entropy within contact space, and consequently for these rates, we investigated three different approximations of increasing simplicity.

Firstly, in Section 3.3.1, we devised an Ising-like Hamiltonian to replicate contact statistics obtained from molecular dynamics simulations of a homopolymer model [230–233]. This Hamiltonian provides a microscopic framework for describing the stochastic transition of individual contacts, incorporating both the entropic costs associated with contact formation and the pairwise correlation between contacts. Using this schema we observe, that in transitioning from a slow to a fast chromatin regime, the system undergoes a dynamic phase transition, marked by three key signatures. First, the system's steady state probability distribution for fraction of contacts made shifts from being monostable to bistable. Simultaneously, the probability distribution for fraction of nucleosomal sites marked becomes skewed and asymmetric. Lastly, there is a concomitant divergence in the average lifetimes of marked and unmarked states. Subsequently, in Section 3.3.2, we introduced a mean field expression for polymer entropy to examine the impact of specific parameters of the Ising Hamiltonian on our findings. Our analysis confirmed that the mean field model qualitatively reproduces the key results outlined in Section 3.3.1. Finally, in Section 3.3.3, we proposed a phenomenological master equation that does not rely on explicit expressions for polymer entropy. Here, we treat contact formation (and disruption) akin to a birth-death process and consider the marks as a two-level system, while explicitly addressing the interconnected dynamics of chromatin contacts and histone modifications. These simplifica-

tions make the model partially analytically tractable. Using this phenomenological model, we show that the observed dynamical phase transition in the preceding models emerges from the coupling of structure and sequences, rather than from specific treatments of free energy functionals in contact space. Additionally, we estimate that the average lifetime near the critical birth-rate of contacts aligns approximately with predictions in contemporary literature [106, 107, 234].

3.3 Results

3.3.1 An Ising-like Hamiltonian for chromatin contacts

To describe the microscopic dynamics of chromatin conformation in contact space, we introduce the following Hamiltonian, $\mathscr{H}(\boldsymbol{q})$, defined as

$$\mathscr{H}(\boldsymbol{q}) = \sum_{j>i} h_{ij} q_{ij} + \sum_{j>i} \sum_{l>k} J_{ijkl} q_{ij} q_{kl} + \lambda \sum_{j>i} q_{ij} \sum_{k\neq i,j} (q_{ik} + q_{kj}). \tag{3.3}$$

The linear term, h_{ij}, accounts for the entropic penalty of bringing nucleosomes i and j into contact [230]. The symmetric term $J_{ijkl} = J_{klij}$ accounts for the coupling of contact formation between distinct pairs of nucleosomes [233]. This coupling emerges whenever the existence of contact (i,j) affects the configurational entropy penalty associated with forming contact (k,l) as a result of polymer topological effects [231, 232]. We followed a preexisting pseudolikelihood maximization approach [235] to obtain all parameters, h_{ij} and J_{ijkl}, that most likely to reproduce the statistical distribution of a set of homopolymer configurations that resemble *in vivo* chromatin organization. We refer readers to Appendix B.1 for details. The λ term in Eq. (3.3) accounts for the excluded volume effect by penalizing the formation of multiple contacts with the same nucleosome. Similar models have been applied successfully to study protein folding mechanisms [228, 229].

From the above Hamiltonian, we defined the rate of breaking and forming a contact between nucleosomes i and j as $k_c \exp\left(-\beta \varepsilon n_i n_j \Delta q_{ij}\right)$ and $k_c \exp(-\beta \Delta \mathscr{H})$, respectively.

These terms account for the energetic and entropic costs of contact breaking and formation, assuming contact formation is diffusion limited [236, 237]. Here k_c is the basal rate constant and is representative of thermal nucleosomal motions. $\beta = 1/k_BT$ and k_B is the Boltzmann constant. $\varepsilon = -2.5k_BT$ measures the interaction energy between marked nucleosomes. This attraction is meant to account for the effects of architectural proteins associated with epigenomic states that aid 3D chromatin organization [238]. The order of magnitude of this parameter was chosen to be comparable to values estimated from force spectrometry experiments [239]. $\Delta\mathscr{H} = \mathscr{H}(\boldsymbol{q}|q_{ij}=1) - \mathscr{H}(\boldsymbol{q}|q_{ij}=0)$.

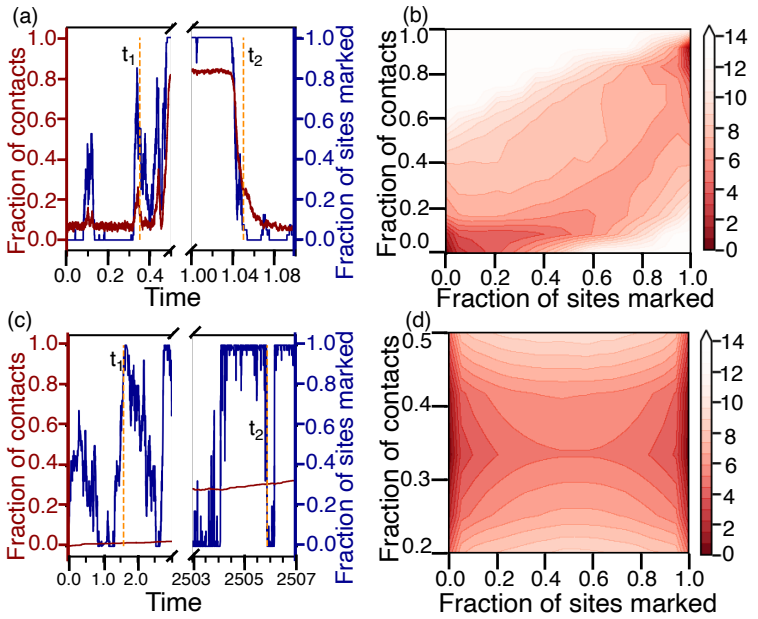

Figure 3.2: The model exhibits distinct kinetic and steady-state behaviors in fast and slow chromatin dynamics regimes. (a, c) Time evolution of the fraction of chromatin contacts (red) and the fraction of marked sites (blue) along representative simulation trajectories initialized from a state with zero histone marks and chromatin contacts in the fast ($k_c = 10^3$, a) and slow ($k_c = 10^{-1}$, c) chromatin regimes. (b, d) The negative logarithm of the steady state distributions as a function of the fraction of marked sites and the fraction of chromatin contacts for $k_c = 10^3$ (b) and $k_c = 10^{-1}$ (c). We hold fixed $\mathbf{N} = 40$, $\lambda = 0.01, \varepsilon = -2.5$.

Without explicit chromatin conformational dynamics and $q_{ij} = 1 \ \forall \ i,j \in [1,\mathbf{N}] \ \forall \ t$, the above model reduces to an extensively studied mark-only version [60, 63, 69, 73, 83] which exhibits bistability with two steady states where the fraction of marked nucleosomes is either close to 1 or to 0 for large c_r values. Hereon, the values for all rate constants are reported in the unit of c_n. We focus on a strongly cooperative regime wherein $c_r = 100$.

Following contemporary literature, we first explore the regime where chromatin conformational dynamics are fast and choose $k_c = 10^3$. We interrogate the system using stochastic simulations performed with the the Gillespie algorithm [240]. Since c_n has been estimated to be around 0.6 h^{-1} [61], this value for k_c correspond to nucleosome motions on the second timescale, which matches well with experimental estimations from live cell imaging [234]. Fig. 3.2a shows an example trajectory initialized with zero chromatin contacts and marked nucleosomes. The blue and red traces depict the time evolution of the fraction of marked nucleosomes ($\overline{n} \equiv \sum_i n_i/\mathbf{N}$) and the fraction of contacts formed ($\overline{q} \equiv \sum_{ij} q_{ij}/\mathbf{M}$), respectively. Initially, transitions to the fully marked state are unsuccessful without sufficient contacts (Fig. 3.2a). However, as contacts build, they endow the system with greater cooperativity and facilitate the spreading of marks through the recruited conversion pathway (Fig. 3.2a). Moreover, since marks confer attraction between sites, their establishment drives further collapse of the chromatin structure, and both contacts and marks increase in concert, culminating in the formation of the collapsed marked state.

We computed the steady-state probability distributions as a function of $\overline{n}$ and $\overline{q}$, i.e., $\mathbb{P}_{\text{ss}}(\overline{n},\overline{q})$ to examine the long-time behaviors of this model. The negative logarithm of this distribution, which can be interpreted as a pseudo-potential quantifying the landscape of the stochastic system [110, 210–214], is shown in Fig. 3.2b. Two distinct steady states, a collapsed marked state and an open unmarked state resembling heterochromatin and euchromatin, respectively, are evident. The bistable behavior is consistent with the two-state switching kinetics shown in Fig. 3.2a. Therefore, our model, unlike existing ones, produces steady states that naturally account for changes in chromatin organization upon changing histone modifications.

Since the precise value of chromatin dynamics would be affected by a multitude of factors, such as the specifics of the polymer model and the environment the chromatin is embedded in, we modulated k_c from 10^3 to 10^{-1} to explore the phenomenology of the system. Strikingly, we observed signatures of a dynamical phase transition. Fig. 3.3a shows the negative logarithm of the steady-state probability distributions as a function of the fraction of sites marked. The result from a model neglecting explicit dynamics of chromatin contacts by setting $q_{ij} = 1 \ \forall \ i,j \in [1,\mathbf{N}] \ \forall \ t$ is included for comparison. As

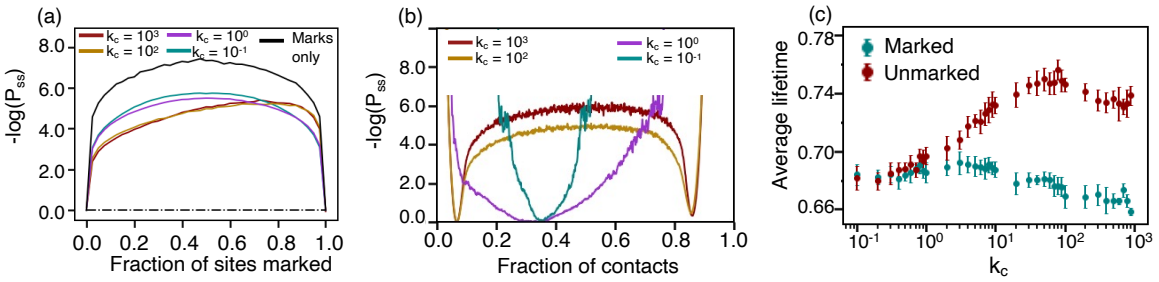

Figure 3.3: Coupling between histone marks and chromatin contacts introduces an asymmetry in the epigenetic landscape and stabilizes euchromatin in the fast chromatin regime. (a) Negative logarithm of the steady state distribution for the fraction of marked nucleosomes computed with $k_c = 10^3$ (red), $k_c = 10^2$ (yellow), $k_c = 10^0$ (purple) and $k_c = 10^{-1}$ (cyan). The result from a mark-only system without explicit chromatin conformational dynamics with $q_{ij} = 1\ \forall i, j$ is provided as a reference (black). (b) Negative logarithm of the steady state distribution for the fraction of contacts made computed with $k_c = 10^3$ (red) $k_c = 10^2$ (yellow), $k_c = 10^0$ (purple) and $k_c = 10^{-1}$ (cyan). (c) Variation in the average lifetime of marked (cyan) and unmarked states (red) with the basal chromatin contact rate constant k_c. We hold fixed $\mathbf{N} = 40$, $\lambda = 0.01, \varepsilon = -2.5$.

the reactions for histone marks are symmetric by design, the steady-state distribution is symmetric in the mark-only model. In Fig. 3.3a, we note an asymmetry in the landscape at high $k_c = 10^3$, and the entire landscape is tilted towards the unmarked state. Strikingly, this asymmetry vanishes below a critical value for k_c, denoted as k_c^o. Similarly, we plot the negative logarithm of the steady-state probability distributions as a function of the fraction of contacts made in Fig. 3.3b and note that the landscape transitions from a bistable regime at high k_c to monostable at low k_c.

After the transition, the system's kinetic behavior deviates significantly from those shown in Fig. 3.2a-b. As an illustrative example, we show a simulation trajectory with $k_c = 10^{-1}$ in Fig. 3.2c-d, but similar trends can be expected for other values below k_c^o. While the histone marks transition between completely marked and unmarked states, as in the fast chromatin case, chromatin contacts vary much slower. Consequently, the intimate coupling between structure and sequence has disappeared. This is clear at the beginning of the trajectory, where the formation of even only a handful of non-backbone contacts seeds the spread of marks and supports cooperative transitions (Fig. 3.2c). After the initial equilibration ($\approx 10^3\ \tau$), the dynamics occur on a network with relatively fixed connectivities for the duration of the simulated trajectory ($\approx 10^5\ \tau$). This is reflected in the steady-state

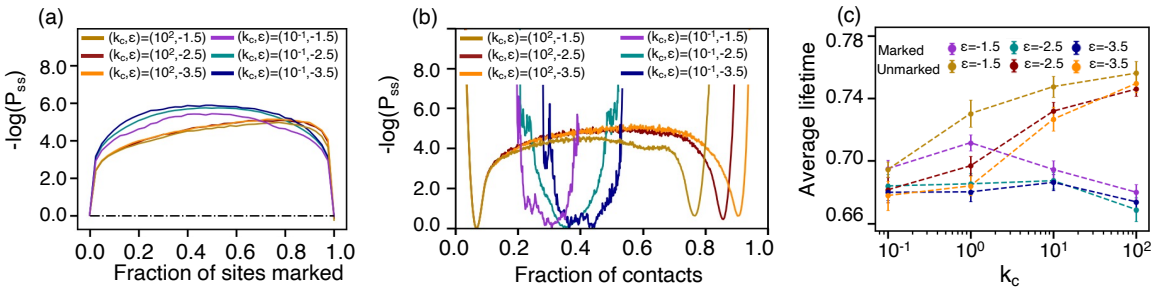

Figure 3.4: The existence of the dynamical phase transition and the qualitative behavior of the steady states in the two regimes is insensitive to the parameters of the model. For a $\mathbf{N} = 40$ bead system and fixed $\lambda = 0.01$, qualitatively similar results to Fig. 3.3 can be recovered for different ε values. (a) Negative logarithm of the steady state distribution for the fraction of marked nucleosomes computed with $(k_c, \varepsilon) = (10^2, -1.5)$ (yellow), $(k_c, \varepsilon) = (10^{-1}, -1.5)$ (purple), $(k_c, \varepsilon) = (10^2, -2.5)$ (red), $(k_c, \varepsilon) = (10^{-1}, -2.5)$ (cyan), $(k_c, \varepsilon) = (10^2, -3.5)$ (orange), $(k_c, \varepsilon) = (10^{-1}, -3.5)$ (blue). (c) Negative logarithm of the steady state distribution for the fraction of contacts made computed with $(k_c, \varepsilon) = (10^2, -1.5)$ (yellow), $(k_c, \varepsilon) = (10^{-1}, -1.5)$ (purple), $(k_c, \varepsilon) = (10^2, -2.5)$ (red), $(k_c, \varepsilon) = (10^{-1}, -2.5)$ (cyan), $(k_c, \varepsilon) = (10^2, -3.5)$ (orange), $(k_c, \varepsilon) = (10^{-1}, -3.5)$ (blue). (c) Variation in the average lifetime of marked states for $\varepsilon = -1.5$ (purple), $\varepsilon = -2.5$ (cyan), $\varepsilon = -3.5$ (blue) and unmarked states for $\varepsilon = -1.5$ (yellow), $\varepsilon = -2.5$ (red), $\varepsilon = -3.5$ (orange) with the basal chromatin contact rate constant k_c.

behavior plotted in Fig. 3.2d, as the slow chromatin regime exhibits a partially collapsed marked and a partially collapsed unmarked state.

To further our understanding of the dynamical system, we determined the lifetime of both marked and unmarked states by partitioning the simulation trajectories into the two states. As shown in Fig. 3.3c, while the lifetime of marked states remains largely unchanged, the lifetime of unmarked states increases significantly as k_c increases. For small k_c, transitions between marked and unmarked states happen at rates much faster than the chromatin structural relaxation and are dictated mainly by the symmetric reaction network, producing comparable average lifetimes for both states. As chromatin contacts become more responsive to histone modifications at larger k_c, fewer contacts are expected for the unmarked states. In contrast, more will form for the marked one, driving the monostability to bistability transition (Fig. 3.3a). The decrease in contacts makes transitioning out of the unmarked state harder due to a lack of recruited conversions, leading to the observed increase in lifetime. On the other hand, the enhanced presence of contacts for the marked state facilitates cooperative reactions that erase the marks. The imbalance between the life-

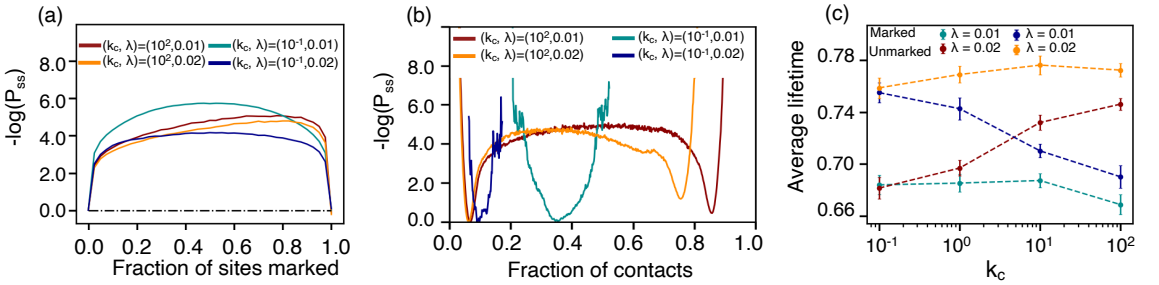

Figure 3.5: The dynamical phase transition and the qualitative behavior of the stead state persists if we hold $\varepsilon = -2.5$ fixed and vary λ and for a $\mathbf{N} = 40$ bead system, qualitatively similar results to Fig. 3 can be recovered for different λ values. (a) Negative logarithm of the steady state distribution for the fraction of marked nucleosomes computed with $(k_c, \lambda) = (10^2, 0.01)$ (red), $(k_c, \lambda) = (10^{-1}, 0.01)$ (cyan), $(k_c, \lambda) = (10^2, 0.02)$ (orange), $(k_c, \lambda) = (10^{-1}, 0.02)$ (blue). (c) Negative logarithm of the steady state distribution for the fraction of contacts made computed with $(k_c, \lambda) = (10^2, 0.01)$ (red), $(k_c, \lambda) = (10^{-1}, 0.01)$ (cyan), $(k_c, \lambda) = (10^2, 0.02)$ (orange), $(k_c, \lambda) = (10^{-1}, 0.02)$ (blue). (c) Variation in the average lifetime of marked states for $\lambda = 0.01$ (cyan), $\lambda = 0.02$ (blue) and unmarked states for$\lambda = 0.01$ (red), $\lambda = 0.02$ (orange) with the basal chromatin contact rate constant k_c.

times of steady states produces the asymmetry in the landscape seen in Fig. 3.3a. We point out that the existence of the dynamical phase transition and the qualitative behavior of the steady states in the two regimes are insensitive to the parameters of the model, including ε (Fig. 3.4a-c), λ (Fig. 3.5a-c).

3.3.2 A mean-field expression for the contact space Hamiltonian

In the previous section, we used an Ising-like Hamiltonian to describe the stochastic transition of various contacts in chromatin. To examine whether our findings are sensitive to the functional form and parameters in the Hamiltonian, we next introduce a mean-field expression for the free energy of total contacts in the system, $\mathscr{F}(\overline{q})$. We designed the mean-field expression to capture two prominent features of polymer systems: (1) in the absence of marks, the free energy has a singular minima which corresponds to entropically favored configuration with few contacts (low $\overline{q}$). (2) as we titrate marks into the system the self-attraction between marked sites results in a secondary minima in the free energy corresponding to a more collapsed configuration (high $\overline{q}$). A simple example would be to construct $\mathscr{F}(\overline{q})$ as a quartic polynomial over $[0,1]$, where $\mathscr{F} = \sum_{r=0}^{4} a_r \overline{q}^r + \varepsilon_{ij} q_{ij}$, where a_r are the polynomial coefficients of $\overline{q}^r$, and ε_{ij} is a small attraction between two marked sites

(i,j). One choice of parameters for this quartic polynomial is $a_0 = 0$, $a_1 = 150.0$, $a_2 = 664.043$, $a_3 = -6312.1$, $a_4 = 15000.0, \varepsilon = -0.55$. However, we do not anticipate our results to be sensitive to these specific parameter values as long as conditions (1) and (2) outlined previously are met.

We again performed stochastic simulations for the reaction network using an implementation of the Gillespie stochastic simulation algorithm [240]. The rate for contact breaking and formation for a pair of nucleosomes (i,j) was again defined as $k_c \exp\left(-\beta \varepsilon n_i n_j\right)$ and $k_c \exp(-\beta \Delta \mathscr{H})$, respectively. Now, $\mathscr{H}(\overline{q})$ is defined as $\mathscr{H}(\langle q \rangle) \equiv \sum_{r=0}^{4} a_r \overline{q}^r + \log\left(\binom{\mathbf{M}}{\overline{q}\mathbf{M}}\right)$. The second term is needed in the microscopic model to account for the degeneracy in different configurations that yield the same $\overline{q}$ so that the macroscopic expression for $\mathscr{F}$ simplifies to a simple quartic polynomial. Simulations were carried out for $\mathbf{N} = 40$ $(\mathbf{M} = 741)$ sites. In Fig. 3.6 we recover qualitatively similar results to the ones demonstrated in Fig. 3.2, Fig. 3.3, Fig. 3.4 and Fig. 3.5.

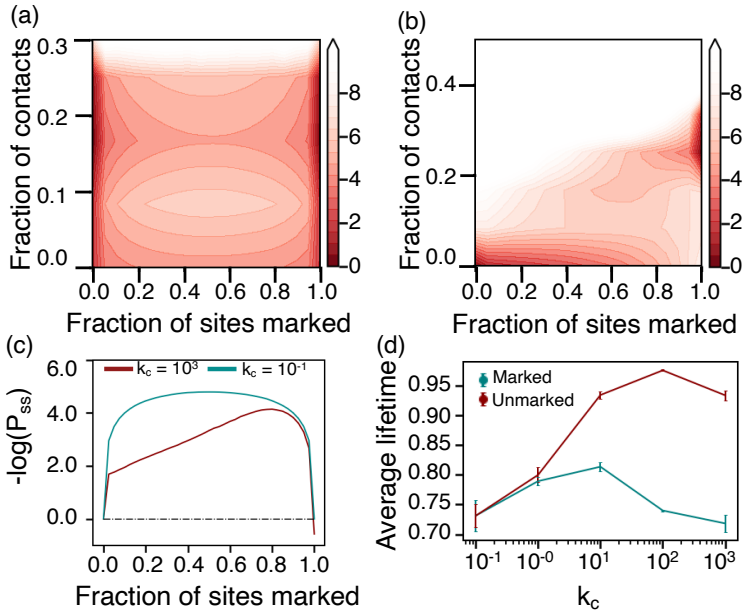

Figure 3.6: A minimal mean field model recapitulates the main results presented in Fig. 3.2, Fig. 3.3, Fig. 3.4 and Fig 3.5 for a $\mathbf{N} = 40$ bead system. Steady state probability distributions for the (a) slow chromatin regime with $k_c = 10^{-1}$ and (b) fast chromatin regime with $k_c = 10^3$. (c) $-\log(\mathbb{P}_{ss})$ for marks plotted as a function of the fraction of marked sites for $k_c = 10^3$ (red) and $k_c = 10^{-1}$ (cyan). (d) Variation in average lifetime of marked (cyan) and unmarked states (red) with k_c.

Therefore, the dynamical phase transition in this paper is also insensitive to the explicit

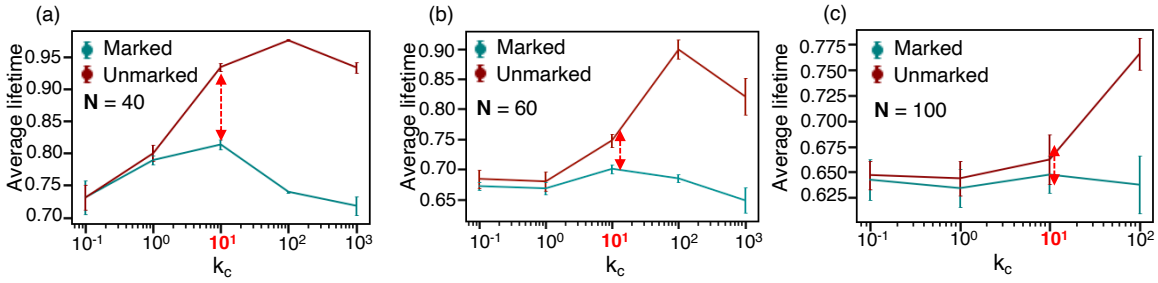

Figure 3.7: k_c^o increases with system size. We plot the the average lifetimes in the marked (red) and unmarked states (cyan) against k_c for systems of size (a) $\mathbf{N} = 40$, (b) $\mathbf{N} = 60$ and (c) $\mathbf{N} = 100$. We note that at $k_c = 10$, the smaller ($\mathbf{N} = 40$) is the system is already asymmetric, however, there is no appreciable difference in the average lifetimes at at $k_c = 10$ for the larger ($\mathbf{N} = 100$) system. Simulations performed using the mean-field approach to the polymer model discussed in Section 3.3.2 and Fig 3.6

form of the Hamiltonian defined in Eq. (3.3), indicating that the results presented are both robust and of general interest.

3.3.3 An analytically tractable phenomenological model with coupled structure and sequence changes

In the full kinetic model, we found that the transition from the monostable to bistable regime in the probability of steady state distribution of contacts and the concomitant emergence of asymmetry in the average lifetimes of the marked and unmarked state begins to emerge around $k_c^o \sim 1$, which corresponds to a timescale of 1 h for nucleosome diffusion. While this number may seem too slow compared to experimental values on the order of seconds [234], it does not immediately exclude the biological relevance of the dynamical phase transition. Though the existence of this transition is insensitive to model details, the numerical value for k_c^o is not. For example, as shown in Fig. 3.7, k_c^o can increase by order of magnitude as we change the system size, N, from 40 to 100. Furthermore, the specific Hamiltonian chosen here, parameterized with homopolymer simulations, may be insufficient for reproducing complex viscoelastic behavior of chromatin *in vivo*, underestimating k_c^o. To provide more insight into the determining factors of the dynamical phase transition, in this section, we introduce a phenomenological model that captures the essence of the full kinetic model but is now analytically tractable.

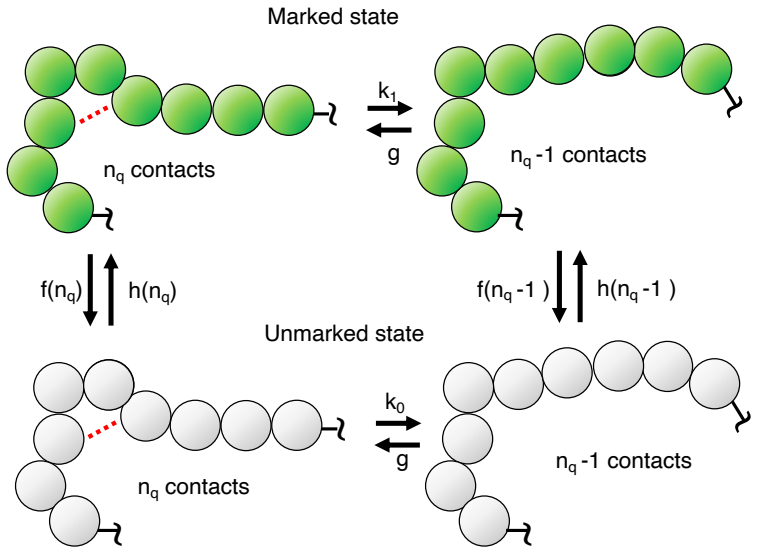

Figure 3.8: A schematic illustration of the salient features of a phenomenological model described by Eq. (3.4). The system can transition between fully marked (green) and fully unmarked (grey) states with rate $f(n_q), h(n_q)$ respectively. In either state, precise topological, polymeric effects are ignored, and the number of contacts (n_q) is incremented at rate g. Furthermore, n_q is decremented at rate k_1, k_0 in the fully marked and unmarked states. $k_1 < k_0$ accounts for the attraction conferred between marked nucleosomes.

The model presented herein captures the stochastic fluctuations in chromatin contacts and histone marks as in the full kinetic model presented prior and also accounts for the coupling between changes in chromatin structure and sequence (Fig. 3.8). We treat contact formation (and breaking) like a birth-death process. Furthermore, we approximate the marks as a two-state system, transitioning between fully marked ($s = 0$) and fully unmarked ($s = 1$) states. The dynamics of this analytical model can be described by the following master equation,

$$\begin{aligned} \partial_t \boldsymbol{P}(n_q,t) &= g\mathbb{1}\left\{\boldsymbol{P}(n_q-1,t)-\boldsymbol{P}(n_q,t)\right\} \\ &+ \begin{pmatrix} k_1 & 0 \\ 0 & k_0 \end{pmatrix}\left\{(n_q+1)\boldsymbol{P}(n_q+1,t)-n_q\boldsymbol{P}(n_q,t)\right\} \\ &+ \begin{pmatrix} -h & f \\ h & -f \end{pmatrix}\boldsymbol{P}(n_q,t). \end{aligned} \tag{3.4}$$

Where, $\boldsymbol{P}(n_q,t) = \begin{pmatrix} P_1(n_q,t) \\ P_0(n_q,t) \end{pmatrix}$ represents the probability of finding the system in fully marked (s=1) or unmarked (s=0) state with total n_q number of chromatin contacts. k_s and g represent the rate of contact breaking or formation. The transition rates between marked and unmarked states, h and f, are assumed equal to $k_m \exp(-\Delta V(\overline{n},\overline{q}))$. $V(\overline{n},\overline{q})$ acts as a pseudo-potential and is derived based on an analytical theory of epigenetic stability [67]. The pseudo-potential accounts for both stochastic fluctuations of histone marks and the impact of chromatin structures. The transition from $s=0$ to $s=1$ occurs with rate $h(\overline{n},\overline{q}) = k_m \exp(-(V(\overline{n}=0.5,\overline{q}) - V(\overline{n}=1,\overline{q})))$, while the transition from $s=1$ to $s=0$ occurs with rate $f(\overline{n},\overline{q}) = k_m \exp(-(V(\overline{n}=0.5,\overline{q}) - V(\overline{n}=0,\overline{q})))$. We estimate $k_m \sim N^{-1} c_n$, where c_n is the rate for random removal of histone modifications as introduced in the text prior.

Following Ref. [67], we derive the pseudo-potential $V(\overline{n},\overline{q})$ ((3.5)). The details of the derivation are outlined in the Appendix B.2.

$$V(\overline{n}) = 2\mathbf{N}\overline{n}(1-\overline{n}) + \left(1 - \frac{4\mathbf{N}}{Fq^2}\right) \log[Fq^2\overline{n}(1-\overline{n}) + 1] \tag{3.5}$$

Following a second quantization approach [205, 224], we rewrite the master equation as an imaginary time Schrödinger equation,

$$\partial_t |\Psi(t)\rangle = \Omega |\Psi(t)\rangle . \tag{3.6}$$

Here, the state vector $|\Psi(t)\rangle = \begin{pmatrix} \Psi_1(t) \\ \Psi_0(t) \end{pmatrix}$ is a superposition of all possible configurations weighted with their corresponding probabilities such that $\Psi_i(t) = \sum_{n_q} P_i(\{n_q\};t) \left|n_q\right\rangle$ for $s = 0, 1$. The "Hamiltonian" operator Ω is defined as

$$\Omega = \boldsymbol{g}(a^\dagger - 1) + \boldsymbol{k}(a - a^\dagger a) + \begin{pmatrix} -h(\overline{n},\overline{q}) & f(\overline{n},\overline{q}) \\ h(\overline{n},\overline{q})) & -f(\overline{n},\overline{q}) \end{pmatrix}. \tag{3.7}$$

The operator $a^\dagger$ serves to create a contact, i.e it acts on a state with n_q contacts ($\left|n_q\right\rangle$), $a^\dagger \left|n_q\right\rangle = \left|n_q+1\right\rangle$. Similarly a serves to decrement contacts, $a\left|n_q\right\rangle = n_q\left|n_q-1\right\rangle$. Correspondingly, $a^\dagger a\left|n_q\right\rangle \equiv \hat{n}_q\left|n_q\right\rangle = n_q\left|n_q\right\rangle$, simply returns the number of contacts in a given state.

We note that the imaginary time Schrödinger equation is equivalent to the functional variation of the action (Γ) with respect to Φ, i.e. $\frac{\delta\Gamma}{\delta\Phi} = 0$, where $\Gamma = \int \mathrm{d}t \left\langle\Phi\right|\partial_t - \Omega\left|\Psi\right\rangle$. Thus, for appropriate trial functions for Φ and Ψ parameterized with $\boldsymbol{\alpha}_L = \alpha_L^1, \alpha_L^2, \cdots, \alpha_L^K$ and $\boldsymbol{\alpha}_R = \alpha_R^1, \alpha_R^2, \cdots, \alpha_R^K$, minimizing the action leads to a set of ordinary differential equations,

$$\sum_{l=1}^{K}\left[\left\langle\frac{\partial\Phi}{\partial\alpha_L^m}\middle|\frac{\partial\Psi}{\partial\alpha_R^l}\right\rangle\frac{\mathrm{d}\alpha_R^l}{\mathrm{d}t} - \left\langle\frac{\partial\Phi}{\partial\alpha_L^m}\middle|\Omega\middle|\Psi\right\rangle\right]_{\alpha_L^m=0} = 0, \tag{3.8}$$

Using a variational *ansatz* we obtain the following set of variational equations,

$$\frac{\mathrm{d}c_1}{\mathrm{d}t} = c_0\left\langle f(\hat{n}_q)\right\rangle_0 - c_1\left\langle h(\hat{n}_q)\right\rangle_1 \tag{3.9a}$$

$$\frac{\mathrm{d}c_1}{\mathrm{d}t}\overline{q}_1 + \frac{\mathrm{d}\overline{q}_1}{\mathrm{d}t}c_1 = c_1 g_1 - c_1 k_1\overline{q}_1 + c_0\left\langle f(\hat{n}_q)\hat{n}_q\right\rangle_0 \tag{3.9b}$$

$$- c_1\left\langle h(\hat{n}_q)\hat{n}_q\right\rangle_1 \tag{3.9c}$$

$$\frac{\mathrm{d}c_0}{\mathrm{d}t}\overline{q}_0 + \frac{\mathrm{d}\overline{q}_0}{\mathrm{d}t}c_0 = c_0 g_0 - c_0 k_0\overline{q}_0 - c_0\left\langle f(\hat{n}_q)\hat{n}_q\right\rangle_0 \tag{3.9d}$$

$$+ c_1\left\langle h(\hat{n}_q)\hat{n}_q\right\rangle_1 \tag{3.9e}$$

The angular brackets represent averaging over number of contacts using a Poisson distribution, i.e., $\langle\cdot\rangle_s = \sum_{n_q}\cdot\frac{e^{-\overline{q}_s}}{n_q!}\overline{q}_s^{n_q}$. We also make a simplifying assumption that $\langle F(x)\rangle \approx F(\langle x\rangle)$. We assume steady-state solutions of Eqs. (3.9) take the form $\overline{q}_1 = \overline{q}^h + \delta$ and $\overline{q}_0 = \overline{q}^h - \delta$,

where $\overline{q}^h = 2g/(k_1+k_0)$. Plugging this into, Eqs. (3.9) we obtain,

$$g - k_1(\overline{q}^h + \delta) + 2\left\langle h(\overline{q}^h + \delta)\right\rangle \delta = 0 \tag{3.10}$$

We linearise $h(\overline{q}^h + \delta)$ around δ, and then solve for δ,

$$\delta = \frac{\frac{2gk_1}{k_1+k_0} - g}{-k_1 + 2e^{N/2}\left(1 + \frac{Fg^2}{(k_1+k_0)^2}\right)^{\frac{N(k_1+k_0)^2}{Fg^2}-1} k_m} \tag{3.11}$$

Finally, we observe $\delta \to 0$ as $g/k_m \to 0$ and $\delta \to \infty$ as $g/k_m \to \infty$. In the former case, when the rate of contact creation is low relative to mark turnover we are in the monostable regime. However, as the rate of contact formation is appreciably large then we transition to a bistable regime.

Therefore, the phenomenological model reproduces the dynamical phase transition as well. Importantly, it clarifies that the rate of contact formation (birth rate g) determines the transition between the two regimes. g can be impacted by nucleosome diffusion (k_c), polymer topology and the nucleoplasm. Using the condition $\overline{q}^h/\delta >> 1$ we can bound the transition birth-rate, $(g^o)^2 \lesssim \frac{4e^{N/2}(k_0+k_1)^2 k_m}{F(3k_1-k_0)}$. For a birth-death process with n_q contacts, the probability of $P(n_q-1|n_q) \sim \left(\frac{\frac{k_1+k_0}{2}}{g^o+\frac{k_1+k_0}{2}}\right)$. We approximate $k_1 \sim e^{-2.5}$, $k_0 \sim 1, N \sim 40$, $k_m \sim N^{-1}c_n$, $F \sim 10^2$. We estimate $c_n \sim 0.6\,\mathrm{h}^{-1}$. Using these we estimate the lifetime of contacts for near the critical point $\left(\frac{\frac{k_1+k_0}{2}}{g^o+\frac{k_1+k_0}{2}}\right)^{-1} \sim 10$ s.

3.4 Conclusions

In summary, our study demonstrated that incorporating the interplay between chromatin structural dynamics and histone modification kinetics can give rise to a dynamical phase transition. We verified the validity of this transition by reproducing it in multiple models: a comprehensive kinetic model encompassing microscopic chromatin contacts, a mean-field model, and a phenomenological model. This extensive validation underscores the robustness of our findings. Moreover, the behavior observed in the fast chromatin dynamics limit

aligns with well-established observations regarding the influence of histone modifications on chromatin structure. Additionally, experimental evidence for slow chromatin relaxation further highlights the significance of our results in the opposite limit [38, 109]. By introducing the concept of a dynamical phase transition, we provide a cohesive framework that reconciles observations across different limits. Future experiments specifically designed to explore chromatin viscoelasticity under various conditions hold the potential to validate our theoretical predictions further. Moreover, some of the methods and ideas are broadly applicable to the study of dynamical processes and problems in general where there exists a coupling between 3D network structure and 1D sequence information, and where there is the presence of dynamical asymmetry in the relaxation rates for the two.

Chapter 4

Quantifying the stability of coupled genetic and epigenetic switches: A variational approach [1]

4.1 Introduction

A little more than five decades ago, Waddington introduced the metaphor to view cellular differentiation into distinct lineages and cell types as a sequence of transitions among basins in a landscape, wherein basins indicate stable phenotypes [241]. The appeal of this metaphor to intuition has inspired efforts of theoretical formulation at the molecular level by studying genetic networks formed by transcription factors (TF) [22, 188, 205, 206, 222, 242–254]. These studies highlighted the importance of gene expression noise in driving the transition among steady states. Noise is a manifestation of the inherent stochasticity of chemical reactions and arises in gene regulatory networks as a result of protein production/degradation and TF binding/unbinding. Noise, or fluctuation, is nonnegligible due to the finite number of protein molecules and the single molecule nature of DNA. Stochastic noise and network topology together define the epigenetic landscape, much like the one envisioned by Waddington, that quantifies the stability of various cell-defining gene

[1] Based on work published in Ref. 224.

expression levels or patterns.

For eukaryotic organisms, in addition to transcription factors, epigenetic marks such as DNA methylation and histone modifications also play essential roles in regulating gene expression [49–51, 53, 255–259]. They are known to affect local chromatin packaging and global genome organization [44, 45, 260–263], which in turn can regulate DNA accessibility to regulatory proteins. Furthermore, DNA methylation directly impacts the DNA binding affinity of transcription factors [264–266]. Importantly, the chemical modifications themselves may give rise to steady states independent of the TF-centric genetic network. For example, modification of nucleosomes recruits enzymes affecting the neighboring nucleosomes, causing them to be similarly modified [25]. Many elegant theoretical attempts have demonstrated how such interactions can bring about collective changes of many nucleosomes and allow them to exhibit distinct multistable states [60, 63–67, 69, 198]. Therefore, it is crucial to account for the dynamics and regulation of epigenetic modifications when constructing the landscape for cellular differentiation in eukaryotes.

Many researchers have studied the interplay between genetic and epigenetic switches in regulating gene expression. For instance, generalized genetic networks that couple each gene to a binary or ternary variable representing the collective histone states have been used as models for stem cells to account for epigenetic degrees of freedom, albeit in a coarse grained fashion [189, 256, 267–270] . These studies found a significant dependence of the probability landscape of protein expression computed from stochastic simulations on chromatin state dynamics. Similarly coarse-grained treatment of epigenetic switches was shown to introduce hysteresis [223] and homeorhesis [271] to the dynamics of gene regulatory networks. Notably, [110] explicitly considered the modification of individual nucleosomes and studied the impact of such modifications on the probability landscape of a single self-activating gene and a pair of mutually repressive genes. However, the lack of analytical results has made the sensitivity analysis of the computed landscape with respect to parameter values, which may vary along cell differentiation, numerically challenging.

In this work, we investigate the combined impact of TF binding and epigenetic modifications in regulating the expression of a self-activating gene. Rather than coarse-graining the epigenetic switch into a binary or ternary variable, we explicitly account for the dynam-

ical modification of individual nucleosomes. The variational approach [205, 272] was used to compute steady-state probability distributions from deterministic equations and avoid computationally intensive stochastic simulations. Moreover, we generalize the typically used Poisson ansatz to better treat systems with particle conservation constraints, such as our epigenetic switch, that are more naturally described using SU(2) than Bosonic operators [69]. The approach enabled a convenient exploration of the model's steady-state behavior across a wide range of parameters. Our study suggests that fast, random perturbations to individual histone modifications lead to the formation of a poised, uncommitted chromatin state, which in turn can drive noisy gene expression seen in stem cells. As the rate of such random perturbations decreases and the role of cooperative modifications of nucleosome prevails, the system transitions to a bistable regime resembling a differentiated state. The transition goes through an activated state with high gene expression, highlighting the robustness of the network in activating gene expression due to the feedback between genetic and epigenetic switches. We further compared variational results with stochastic simulations and discussed potential improvements in the accuracy of the variational method.

4.2 Model

We consider a simplified model of eukaryotic gene regulation that accounts for TF binding/unbinding as well as histone modifications. The model couples the regulatory network of a self-activating gene with an epigenetic switch that can lead to active and repressive chromatin states.

For self-activating genes, their protein products bind with the promoter to upregulate the transcription rate. As illustrated in Figure 4.1, proteins are produced and destroyed with rates of g and k, respectively. The protein production rate is further dependent on whether the gene's promoter is bound by TF (state 0) or not (state 1), and we have $g_1 < g_0$ since the proteins are activators. Here TFs correspond to gene transcription products, and they bind to the promoter with rate h as dimers. The corresponding unbinding rate is f. Binding rate depends on protein copy number n_p as well as the number of modified nucleosomes n_x as detailed in Eq. (4.3) below. Self-activating genes are known to occur both as isolated

entities [273–276] and as common motifs of larger interacting networks [277–279]. They have been the subject of extensive theoretical study as models of cellular differentiation [22, 188, 205, 206, 222, 242–246, 249, 250]. The epigenetic switch concerns a cluster of $N = 60$ nucleosomes, each of which can exist in a modified (X) or unmodified (Y) state. The kinetics of chromatin system can be described with the non-linear dynamics given below

$$X + X + Y \xrightarrow{s_1, s_0} 3X, \qquad Y + Y + X \xrightarrow{z} 3Y, \tag{4.1}$$

$$X \xrightarrow{q} Y, \qquad Y \xrightarrow{q} X. \tag{4.2}$$

The inter-conversion between modified and unmodified nucleosomes can either proceed via Eq. (4.1) that requires a pair of similarly modified nucleosomes to alter the state of a nucleosome, or via noisy conversion (Eq. (4.2)) with first-order kinetics. The former is meant to account for nucleosomes being actively interconverted by modifying and removing enzymes recruited by the similarly modified nucleosomes in their vicinity. It is this recruitment that forms the positive feedback in the system [48, 60, 67, 69]. s, z, and q are the rate constants of the corresponding reactions.

The coupling between the genetic and epigenetic switch is achieved by introducing a dependence of protein binding rate on the number of modified nucleosomes, i.e.

$$h(n_p, n_x) = h_o \frac{n_p(n_p - 1)}{1 + \exp(-0.5(n_x - 35))}. \tag{4.3}$$

This dependence is motivated by the realization that actively modified chromatin ($n_x > 35$) exists in a more open state that is more accessible to regulatory proteins. The particular expression $[1 + \exp(-0.5(n_x - 35))]^{-1}$ as the probability for chromatin being open is typical of a two state system, assuming that the energetic difference between open and closed chromatin depends linearly on the number of modified nucleosomes. Furthermore, the recruited conversion rate of unmodified to modified nucleosomes depends on TF binding with $s_0 > s_1$, assuming that TFs can attract modification enzymes to chro-

matin. The values for the kinetic parameters were set relative to the degradation rate k as $g_1 = 4, g_0 = 65, h_o = 1, f = 100, s_1 = 8, s_0 = 10s_1, z = 8$. The random histone modification rate, q, was varied over a wide range of values as detailed below. We used $k = 1s^{-1}$, though changing this value will not affect the steady state distributions and only renormalizes the timescale in the model.

We carried out stochastic simulations of the kinetic model using the Gillespie algorithm [240]. Each plot shown in Figure 2 was obtained from averaging over 100 independent 10^5-second-long simulations. These trajectories were initialized with random configurations, and the number of modified nucleosomes and protein molecules along each trajectory was recorded at every second. We then combined the values from all trajectories to estimate the steady state probability distributions, P_{ss}. For the plots shown in Figure 3 we used $q = 10$ and set $n_x = 40$ and $n_p = 20$ at $t = 0$. 200 independent trajectories were performed to produce the average numbers recorded at every 0.5 second.

We reformulated the master equation describing the dynamical evolution of the kinetic network as an *imaginary* time Schrödinger equation

$$\partial_t \left|\Psi(t)\right\rangle = \Omega \left|\Psi(t)\right\rangle . \tag{4.4}$$

The state vector $\left|\Psi(t)\right\rangle = \begin{pmatrix} \Psi_1(t) \\ \Psi_0(t) \end{pmatrix}$ is a superposition of all possible configurations weighted with their corresponding probabilities such that $\Psi_i(t) = \sum_{n_p,n_x} P_i(\{n_p,n_x\};t) \left|n_p,n_x\right\rangle$ for $i = 0,1$. The two components correspond to the DNA state with regulatory proteins unbound (state 1) or bound (state 0), respectively. This reformulation makes use of a second quantization based method (the Doi-Peliti approach), which has been successfully employed in the study of reaction-diffusion processes [204], gene switches [205, 206], and other systems [207]. In previous work, we applied the Doi-Peliti approach to the epigenetic switch using operators that are a representation of the SU(2) algebra [69]. The SU(2) algebra allows us to treat the constraint of conservation of particle in number in a mathematically elegant and convenient way. When coupled to the self-activating gene, the stochastic

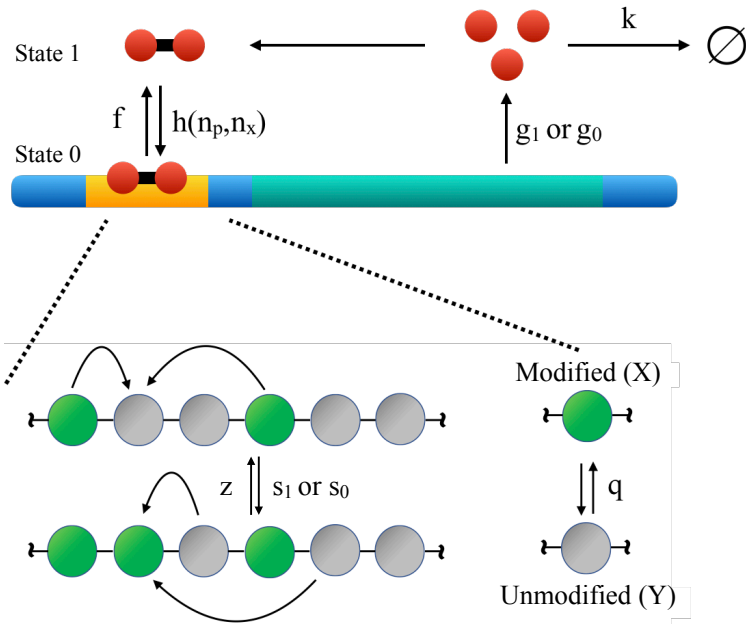

Figure 4.1: Illustration of the kinetic model that couples the regulatory network of a self-activating gene with the reaction network of histone modifications. The gene is auto-regulatory as the protein produced by the gene (red circles) binds to the promoter region (yellow) with rate h and unbinds with rate f. Depending on whether the regulatory protein is bound (State 0) or unbound (State 1), the rate of protein production is g_0 or g_1. Proteins degrade with rate k. Conversions between modified (X) and unmodified (Y) nucleosomes can occur "randomly" (irrespective to the status of other nucleosomes) with a basal rate q. Nucleosome modifications can also occur more cooperatively with rate of z and s.

Hamiltonian for the system described in Figure 4.1 is given by

$$\begin{aligned}\Omega = \boldsymbol{g}(a_p^\dagger - 1) + \boldsymbol{k}(a_p - a_p^\dagger a_p) + \boldsymbol{s}[J_+\hat{n}_x^2 - \hat{n}_x^2\hat{n}_y] + \boldsymbol{z}[J_-\hat{n}_y^2 - \hat{n}_y^2\hat{n}_x] \\ + \boldsymbol{q}[J_- - \hat{n}_x] + \boldsymbol{q}[J_+ - \hat{n}_y] + \begin{pmatrix} -h(\hat{n}_p,\hat{n}_x) & f \\ h(\hat{n}_p,\hat{n}_x) & -f \end{pmatrix},\end{aligned} \tag{4.5}$$

where $\boldsymbol{g} = \begin{pmatrix} g_1 & 0 \\ 0 & g_0 \end{pmatrix}$, $\boldsymbol{s} = \begin{pmatrix} s_1/N^3 & 0 \\ 0 & s_0/N^3 \end{pmatrix}$, $\boldsymbol{z} = \begin{pmatrix} z/N^3 & 0 \\ 0 & z/N^3 \end{pmatrix}$, $\boldsymbol{k} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$, $\boldsymbol{q} = \begin{pmatrix} q/N & 0 \\ 0 & q/N \end{pmatrix}$, and $h(\hat{n}_p,\hat{n}_x) = \frac{\hat{n}_p(\hat{n}_p-1)}{1+\exp(-0.5(\hat{n}_x-35))}$. The operator $a_p^\dagger$ creates a protein molecule when it acts on a state, $a_p^\dagger |n_p,n_x\rangle = |n_p+1,n_x\rangle$, whereas a_p serves to remove a protein molecule when acting on the same state, $a_p |n_p,n_x\rangle = n_p |n_p-1,n_x\rangle$. J_+ converts an unmodified nucleosome to a modified one by acting on a state, $J_+ |n_p,n_x\rangle = (N -$

$n_x)\left|n_p, n_x+1\right\rangle$, while J_- acts to convert a modified nucleosome to an unmodified one, $J_-\left|n_p, n_x\right\rangle = n_x\left|n_p, n_x-1\right\rangle$. $\hat{n}_p$ denotes the number operator, as its action on a ket gives the number of protein molecules, $\hat{n}_p\left|n_p, n_x\right\rangle = n_p\left|n_p, n_x\right\rangle$. In a similar fashion, $\hat{n}_x$ gives the number of modified nucleosomes when it acts on a ket, $\hat{n}_x\left|n_p, n_x\right\rangle = n_x\left|n_p, n_x\right\rangle$, and $\hat{n}_y$ gives the number of unmodified nucleosomes, $\hat{n}_y\left|n_p, n_x\right\rangle = (N-n_x)\left|n_p, n_x\right\rangle$. $n^{\underline{2}} = n(n-1)$ denotes the falling factorial.

Exact solutions to Eq. (4.4) are difficult to obtain. Instead, we make use of an approximate, yet succinct and powerful, variational approach originally introduced by Eyink [272, 280]. First, we realize that the imaginary time Schrödinger equation is equivalent to the functional variation of the following action Γ with respect to Φ, i.e. $\frac{\delta\Gamma}{\delta\Phi} = 0$ for

$$\Gamma = \int \mathrm{d}t \, \langle\Phi|\partial_t - \Omega|\Psi\rangle \,. \tag{4.6}$$

By designing trial functions for Φ and Ψ parameterized with $\boldsymbol{\alpha}_L = \alpha_L^1, \alpha_L^2, \cdots, \alpha_L^K$ and $\boldsymbol{\alpha}_R = \alpha_R^1, \alpha_R^2, \cdots, \alpha_R^K$, minimizing the action leads to a set of ordinary differential equations,

$$\sum_{l=1}^{K}\left[\left\langle\frac{\partial\Phi}{\partial\alpha_L^m}\middle|\frac{\partial\Psi}{\partial\alpha_R^l}\right\rangle\frac{\mathrm{d}\alpha_R^l}{\mathrm{d}t} - \left\langle\frac{\partial\Phi}{\partial\alpha_L^m}\middle|\Omega\middle|\Psi\right\rangle\right]_{\alpha_L^m=0} = 0, \tag{4.7}$$

$$\text{for } m = 1, \cdots, K. \tag{4.8}$$

Also, we demand (to stay true to the probabilistic interpretation) $\langle\Phi(\boldsymbol{\alpha}_L = 0)|\Psi(\boldsymbol{\alpha}_R)\rangle = 1$. The variational approach was first applied with great success to stochastic gene regulatory networks by Sasai and Wolynes [205]. In its original formulation, Poisson distributions were used as trial functions, with the Poisson mean being the variational parameter. Since protein molecules can be approximately treated as products of a birth-death process, the probability distribution to find n_p molecules should be Poisson at large t [205]. Furthermore, the stochastic Hamiltonian for genetic networks consists of only Bosonic operators, the coherent states of which correspond to Poisson distributions. In this work, we exploit the symmetry imposed on the system by particle number constraints to derive a new variational trial function for the chromatin switch. As shown in the Appendix C, an excellent candidate is the binomial distribution function since the coherent states for the SU(2) op-

erators in our stochastic Hamiltonian are binomial [281, 282]. Taken together, we can thus use the following ansatz as variational functions for the coupled genetic and epigenetic swtich

$$|\Psi\rangle = \begin{pmatrix} c_1 \exp\big(p_1(a_p^\dagger - 1)\big)(1-\theta_1)^N \exp\Big(\frac{\theta_1}{1-\theta_1} J_+\Big)\,|0,0\rangle \\ c_0 \exp\big(p_0(a_p^\dagger - 1)\big)(1-\theta_0)^N \exp\Big(\frac{\theta_0}{1-\theta_0} J_+\Big)\,|0,0\rangle \end{pmatrix}, \tag{4.9}$$

and

$$\langle\Phi| = \Big(\langle 0,0|\, e^{a_p} e^{J_-} \exp\Big(\alpha_1 + \lambda_1^{(p)} a_p + \lambda_1^{(x)} J_-\Big)\;\; \langle 0,0|\, e^{a_p} e^{J_-} \exp\Big(\alpha_0 + \lambda_0^{(p)} a_p + \lambda_0^{(x)} J_-\Big)\Big). \tag{4.10}$$

The set of variational parameters is $\boldsymbol{\alpha}_R = \{c_1, c_0, p_1, p_0, \theta_1, \theta_0\}$. Here $c_1(c_0)$ represents the probability of the DNA being in state 1 (state 0), while $p_1(p_0)$ and $N\theta_1(N\theta_0)$ represent the mean number of proteins and modified nucleosomes when DNA is in state 1 (state 0). $\boldsymbol{\alpha}_L = \{\alpha_1, \alpha_0, \lambda_1^{(p)}, \lambda_0^{(p)} \lambda_1^{(x)}, \lambda_0^{(x)}\}$ are the corresponding conjugate variables.

Plugging (4.10) and (4.9) into (4.7), we obtain the following set of variational equations

$$\frac{\mathrm{d}c_1}{\mathrm{d}t} = c_0 f - c_1 \left\langle h(n_p, n_x) \right\rangle_1 \tag{4.11a}$$

$$\frac{\mathrm{d}c_0}{\mathrm{d}t} = -c_0 f + c_1 \left\langle h(n_p, n_x) \right\rangle_1 \tag{4.11b}$$

$$c_1 \frac{\mathrm{d}p_1}{\mathrm{d}t} + p_1 \frac{\mathrm{d}c_1}{\mathrm{d}t} = c_1 g_1 - c_1 k p_1 + c_0 f p_0 - c_1 \left\langle n_p h(n_p, n_x) \right\rangle_1 \tag{4.11c}$$

$$c_0 \frac{\mathrm{d}p_0}{\mathrm{d}t} + p_0 \frac{\mathrm{d}c_0}{\mathrm{d}t} = c_0 g_0 - c_0 k p_0 - c_0 f p_0 + c_1 \left\langle n_p h(n_p, n_x) \right\rangle_1 \tag{4.11d}$$

$$\begin{aligned} N\theta_1 \frac{\mathrm{d}c_1}{\mathrm{d}t} + N c_1 \frac{\mathrm{d}\theta_1}{\mathrm{d}t} =& c_1 \left(\frac{s_1}{N^3}\right) \left\langle n_x^{\underline{2}} (N - n_x) \right\rangle_1 - c_1 \left(\frac{z_1}{N^3}\right) \left\langle (N-n_x)^{\underline{2}} (n_x) \right\rangle_1 \\ & + c_1 \frac{q}{N} (- \left\langle n_x \right\rangle_1 + \left\langle N - n_x \right\rangle_1) + c_0 f N \theta_0 - c_1 \left\langle n_x h(n_p, n_x) \right\rangle_1 \end{aligned} \tag{4.11e}$$

$$\begin{aligned} N\theta_0 \frac{\mathrm{d}c_0}{\mathrm{d}t} + N c_0 \frac{\mathrm{d}\theta_0}{\mathrm{d}t} =& c_0 \left(\frac{s_0}{N^3}\right) \left\langle n_x^{\underline{2}} (N - n_x) \right\rangle_0 - c_0 \left(\frac{z_0}{N^3}\right) \left\langle (N-n_x)^{\underline{2}} (n_x) \right\rangle_0 \\ & + c_0 \frac{q}{N} (- \left\langle n_x \right\rangle_0 + \left\langle N - n_x \right\rangle_0) - c_0 f N \theta_0 + c_1 \left\langle n_x h(n_p, n_x) \right\rangle_1 . \end{aligned} \tag{4.11f}$$

The angular brackets represent ensemble averaging over protein numbers and modified nucleosomes, i.e., $\langle \cdot \rangle_i = \sum_{n_p, n_x} \cdot \frac{e^{-p_i}}{n_p!} p_i^{n_p} \binom{N}{n_x} \theta_k^{n_x} (1-\theta_k)^{N-n_x}$. We also make the simplyfing approximation for the average binding rate as $\left\langle h(n_p, n_x) \right\rangle = \left\langle \frac{n_p (n_p - 1)}{1 + \exp(0.5(n_x - 35))} \right\rangle \approx \frac{\left\langle n_p (n_p - 1) \right\rangle}{1 + \exp(0.5(\langle n_x \rangle - 35))}$. Numerical integration of Eq. (4.11) yields the time evolution of the variational parameters α_R, from which the probability distributions can be determined using Eq. (4.9).

We solved Eq. (4.11) using scipy.integrate.odeint() module in python with a time step of 0.01 s. The initial conditions were varied and individual trajectories were integrated for 10^5 s till convergence to obtain the steady state results.

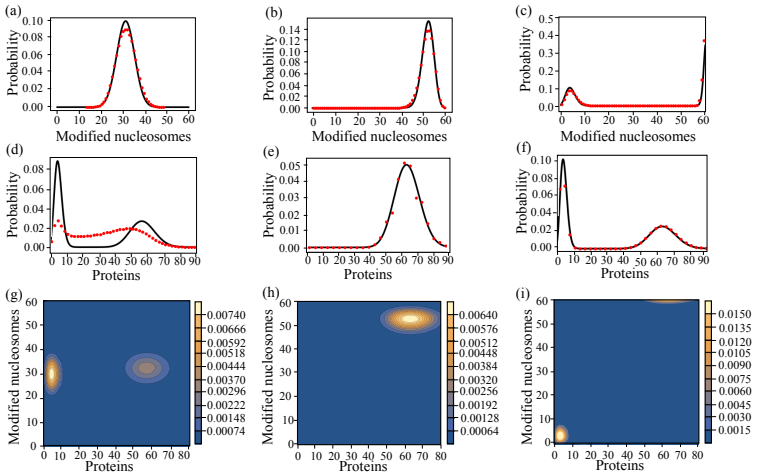

Figure 4.2: Comparison between the probability distributions obtained from the variational approach and from stochastic simulations. (a-c) Steady state probability distributions for the number of modified nucleosomes computed using the variational method (black solid line) and from stochastic simulations (red dots) for $q = 100$ (a), 10 (b), and 0.5 (c). (d-f) Steady state probability distributions for the number of protein molecules computed using the variational method (black solid line) and from stochastic simulations (red dots) for $q = 100$ (d), 10 (e), and 0.5 (f). (g-i) Steady state probability distributions as a function of both number of proteins and modified nucleosomes computed using the variational method for $q = 100$ (g), 10 (h), and 0.5 (i), showing two, one and two fixed points respectively.

4.3 Results

Using the variational equations, we studied the dependence of steady-state solutions on the rate of noisy histone mark modification, q. For comparison, we carried out stochastic simulations of the kinetic network using the Gillespie algorithm [240] at selected q values. The noisy modification rate and, in particular, its relative value to the rate for recruited conversions is an important parameter for cell differentiation. For example, recruited conversions arise due to the diffusion of histone-modifying enzymes from modified nucleosomes to the nearby unmodified ones. The more open chromatin conformation seen in stem cells with larger inter-nucleosome distances [283, 284] will, therefore, suppress recruited conversions in favor of the noisy ones. As cells differentiate, chromatin will become more compact, and the importance of noisy conversions will decline. Previous studies of isolated epigenetic switches [60, 67, 69] also found q as an important parameter that controls the onset and maintenance of bistability in the epigenetic switch.

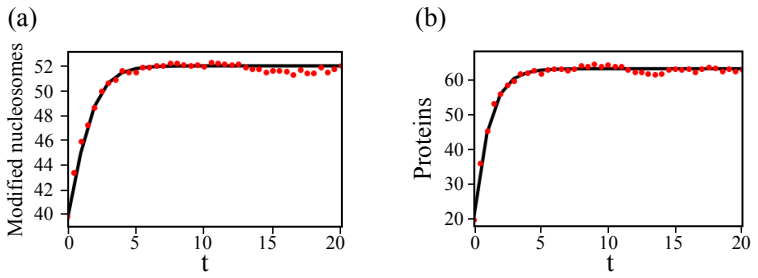

Figure 4.3: Dynamical trajectories determined from the variational approach agree well with stochastic simulations in favorable regimes. (a) Time evolution of the average number of modified nucleosomes computed using the variational method (black solid line) and stochastic simulations (red dots). (b) Time evolution of the average number of modified nucleosomes computed using the variational method (black solid line) and stochastic simulation (red dots). We used $q = 10, M = 60$ and set $c_1 p_1 = 0, c_0 p_0 = 20, c_1 t_1 = 0, c_1 t_0 = 0.66$ as the initial values when solving the deterministic equations. (Eq.(4.11)).

In Fig. 4.2 we show the probability distributions obtained from stochastic simulations and from the variational approach at $q = 100, 10$ and 0.5. We notice that the Binomial ansatz introduced in the *Theory* section captures the distribution for the number of modified nucleosomes with quantitative accuracy (Fig. 4.2a-c). The Poissonian ansatz also performs well for the distribution of protein numbers at small and medium q values, though deviations from stochastic simulations are apparent at large q (Fig. 4.2d-f). The inconsistency between the two distributions in that regime is mainly due to underestimating the population of intermediate states that bridge the high and low gene expression values by the variation method.

In addition to steady-state solutions, the time evolution of observables, such as the mean number of proteins and modified nucleosomes, can be determined using the variational approach as well. As shown in Fig. 4.3, in parameter regimes where the effect of fluctuations is not too drastic, the dynamical trajectories determined using Eq.(4.11) are in quantitative agreements with those computed using stochastic simulations.

Given its reasonable performance, we next applied the variational approach to study the network model's steady-state behavior at a broader range of q values. As already mentioned, q is an important variable that might be tuned along the developmental axis for cell differentiation. For large q values, chromatin stabilizes in an undecided state with roughly half the nucleosomes modified (active) and the other half carrying no modification (repres-

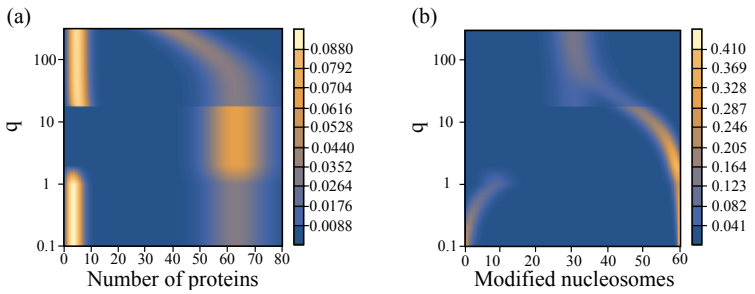

Figure 4.4: Variation of the steady state probability distribution for the number of proteins (a) and modified nucleosomes (b) as a function of the noisy histone modification rate, q.

sive). The corresponding protein expression is noisy with a broad probability distribution. Stochastic simulations further support a significant mixing between "on" and "off" gene states, and an unambiguous assignment of either state is not warranted (Fig. 4.2d). When the value for q is quenched, we observe the emergence of a coherent epigenetic state along with coherent gene expression. Therefore, both switches are turned on and the combined system exhibits a single attractor. At even lower values of q, both the epigenetic and gene switch exhibit bistability.

We note that the chromatin state changes described above differs from that of an isolated epigenetic switch studied previously [69]. There, we saw a shift from a unimodal probability distribution indicating an equal admixture of modified and unmodified nucleosomes to a symmetric bimodal probability distribution as the value for q is quenched. The appearance of a single coherent epigenetic state in Fig. 4.4 is a result of the coupling with the gene switch in our model, which breaks the symmetry between active and repressive chromatin states. The coupling works both ways. In an isolated gene switch, a single state with high gene expression is not expected either. Modulating the kinetics of TF binding to the promoter only resolves a broad probability distribution exhibiting no coherent gene expression to a bistable state with high and low levels of gene expression [249].

4.4 Discussion

We introduced a kinetic model that couples a genetic network with an epigenetic switch to study the combined role of transcription factors and histone modifications in regulating

gene expression. An approximation scheme based on the variational approach was further developed to obtain steady-state solutions. This method is unencumbered by the complexity associated with numerical simulations and more detailed analytical calculations. It would be a useful tool for exploratory studies of the parameter space and identifying regions of interest. While we focused our analysis on a single gene, the variational method can be relatively easily generalised to networks with multiple interacting genetic and epigenetic switches that provide more realistic modeling of stem cell differentiation [206].

We explored the behavior of the network model across a wide range of parameters. Our model exhibits a poised state for the gene switch at high q, where the chromatin system contains an equal admixture of modified and unmodified nucleosomes. The network in this parameter regime appears to qualitatively capture the behavior of chromatin and gene expression in undifferentiated stem cells. In particular, stem cells are known to exhibit bivalent chromatin with both activating and repressive marks [285, 286] and noisy gene expression profiles [287]. We point out that the exact definition of bivalent chromatin remains controversial, and multiple mechanisms have been proposed for its formation [288–290]. Additional studies are needed to determine whether the stochastic conversion observed here is the key driver for the observed chromatin bivalency.

Upon quenching q, the gene is activated along with a concomitant resolution of the chromatin state. The coupling between the two switches reinforces the stability of the active state and can lead to more robust upregulation of gene expression upon cell differentiation. It also ensures that the genetic and epigenetic switches are always in sync. We observe at most two steady states representing active chromatin with high gene expression and repressive chromatin with low gene expression. We note that the inactive state only becomes stable at minimal q values, arguing for strong noise suppression for gene silencing. Its limited stability may explain the presence of DNA methylation on top of histone modifications to safeguard the silent state against perturbations that might arise from fluctuation in protein concentration or histone marks during cell division.

The strong dependence of the landscape tomography on q shown in Fig. 4.4 suggests that the histone modification rate may act like a knob to be tuned along the developmental axis to facilitate cellular differentiation. Of course, the presented landscape is probably

too crude a simplification to be termed the Waddington landscape since many additional factors that contribute to the stability of gene expression patterns could be varied along the developmental axis as well.

In favorable regimes, the variational approach produces results of quantitative accuracy. The discrepancy between the probability distribution obtained from stochastic simulations and the variational method in the high q region can be attributed to the fact that the Poisson ansatz does not sufficiently account for the variance and the effect of fluctuations which become increasingly important as the value for q increases. This situation can be remedied by going beyond the Poisson ansatz, and utilising the superposition ansatz as described in [291]. Mathematically, this would mean to modify our ansatz as follows,

$$|\Psi\rangle = \begin{pmatrix} \int_0^\infty \mathrm{d}p_1 \; \mathscr{F}(p_1;\{\lambda_j^{(1)}\})\; c_1 \exp\big(p_1(a_p^\dagger - 1)\big)(1-\theta_1)^N \exp\Big(\frac{\theta_1}{1-\theta_1} J_+\Big)\,|0,0\rangle \\ \int_0^\infty \mathrm{d}p_0 \; \mathscr{F}(p_0;\{\lambda_j^{(0)}\})\; c_0 \exp\big(p_0(a_p^\dagger - 1)\big)(1-\theta_0)^N \exp\Big(\frac{\theta_0}{1-\theta_0} J_+\Big)\,|0,0\rangle \end{pmatrix}. \tag{4.12}$$

This new "superposition ansatz" is constructed by the superposition of the coherent states (i.e., Poisson distribution) as defined in (4.12), where now $\mathscr{F}$ serves as the variational function. Hence, the real probability distribution is obtained by the superposition of the Poisson distributions of mean p_i weighed by the distribution $\mathscr{F}$ with parameters $\{\lambda_j^{(i)}\}$. We anticipate that doing so can not only improve the agreement between theory and simulation but can in principle allow for the computation of time evolution of other interesting quantities such as variance, and covariance in addition to means. However, in general the choice of an appropriate $\mathscr{F}$ is a non-trivial problem, and thus has been avoided in this text in favour of a clearer exposition. The choice of appropriate variational functions can be guided by the work done on exact solutions of the master equations of genetic switches [242, 246, 292].

Part II

The role of sequence and interaction heterogeneity on phase seperation

Chapter 5

Preserving condensate structure and composition by lowering sequence complexity [1]

5.1 Introduction

Biomolecular condensates serve as important mechanisms for the hierarchical and vectorial organization of chemistry within cells [3, 4, 115–120]. They are non-stoichiometric assemblies of biomolecules that can form via spontaneous or driven processes and exhibit characteristics of phase separation and percolation [2, 121, 122]. The chemical environment and physicochemical conditions within condensates are distinct from their surroundings [123–127], enabling cells to selectively partition biomolecules [6, 128], prevent unwanted cross-talk and interference between various biochemical pathways [130], and buffer against chemical noise [129, 131].

In recent years, these condensates have gained significant prominence and attracted substantial research interest. They have been found to primarily form through the involvement of intrinsically disordered proteins (IDPs) [132–135]. These unique proteins lack a fixed three-dimensional structure, enabling them to be highly flexible and engage in multivalent

[1] Based on work published in Ref. 293.

interactions. Multivalent interactions involve the binding of multiple partners simultaneously. Many research groups have attempted to establish what is known as the “molecular grammar”, [7–12] which connects amino acid sequences with protein phase behaviors. Additionally, considerable effort has also been dedicated to understanding the physics of multi-component systems [166–169] to better understand the collective physical properties of condensates.

Establishing a theoretical framework is crucial for advancing our understanding of biomolecular condensates. The “stickers and spacers" model, originally developed within the context of polymer gelation theory [163–165], has gained popularity for modeling protein condensates [3, 4, 9, 175–178]. In this model, biomolecules within condensates are envisioned as possessing two distinct functional components: “stickers" and “spacers". Stickers represent specific molecular domains or motifs with a high affinity for one another, facilitating interaction and bringing molecules into close proximity, thereby contributing to the condensate’s cohesive, networked structure. Conversely, spacers act as flexible linkers connecting the stickers, enabling the dynamic and transient nature of interactions within the condensate, and modulate phase separation via their effective solvation volume [179–182]. This model offers a conceptual framework that assists in interpreting experimental observations regarding condensate stability and material properties [183–185].

The simplicity inherent in the stickers and spacers model, which contributes to its theoretical elegance, faces challenges when applied to realistic biomolecules. In particular, when dealing with IDPs, the identification of stickers and spacers becomes a non-trivial task [4, 116, 186, 187]. Unlike the model’s assumption of only two amino acid types or segments that can straightforwardly map onto stickers and spacers, most proteins exhibit a more complex structure. Treating a substantial fraction of proteins as identical spacers oversimplifies the intricate nature of protein sequences and the diverse interactions they facilitate. Therefore, it becomes pertinent to explicitly consider spacer heterogeneity. Such an inclusion can offer valuable insights into the impact of sequence diversification on condensate formation, contrasting with predictions from the stickers and spacers model. This deliberation may provide a clearer understanding of whether IDPs tend to favor sequences with low or high complexity. Moreover, it could shed light on the evolutionary optimization

of IDPs toward specific sequence patterns.

We expanded the stickers and spacers model to include explicit non-specific, heterogeneous pairwise interactions between spacers, alongside the specific interactions among sticker motifs. This extension allowed us to systematically explore the intricate interplay between specific and non-specific interactions in determining the structural and compositional properties of condensates. Our investigation revealed that spacer interactions contribute to phase separation and the co-condensation of multiple molecules. However, the non-specific nature of these interactions results in disorganized condensates with undefined molecular compositions. In contrast, specific sticker-sticker interactions drive the formation of condensates with robust contacts and precise compositions. Subsequently, we discussed the implications of our theory for the evolution of protein sequences, asserting the existence of evolutionary constraints even on segments of protein sequences that interact non-specifically. These constraints ensure the functionality of condensates. This evolutionary pressure naturally favors the emergence of low complexity domains to suppress spurious interactions, facilitating the formation of biologically meaningful condensates.

5.2 Theory

5.2.1 Stickers and random spacers model

We present a generalized version of the stickers and spacers model to investigate the phase behaviors of associative polymers (Fig. 5.1). These polymers consist of N monomers, with f privileged monomers that exhibit specific attractive interactions, denoted by a strength of $-u_a$, leading to the formation of non-covalent, physical bonds. We refer to these privileged monomers as stickers. For simplicity, we consider stickers composed of a single monomer. However, it's important to note that in biological sequences, stickers may consist of multiple amino acids. The remaining monomers are designated as spacers.

In a departure from the traditional model, we consider the interaction energy between a pair of spacers, or a spacer-sticker pair, to be a random variable, ε_i, drawn from a normal distribution, $\mathcal{N}(\bar{\varepsilon}, \Delta\varepsilon^2)$, with mean $\bar{\varepsilon}$ and variance $\Delta\varepsilon^2$. The contribution to the total

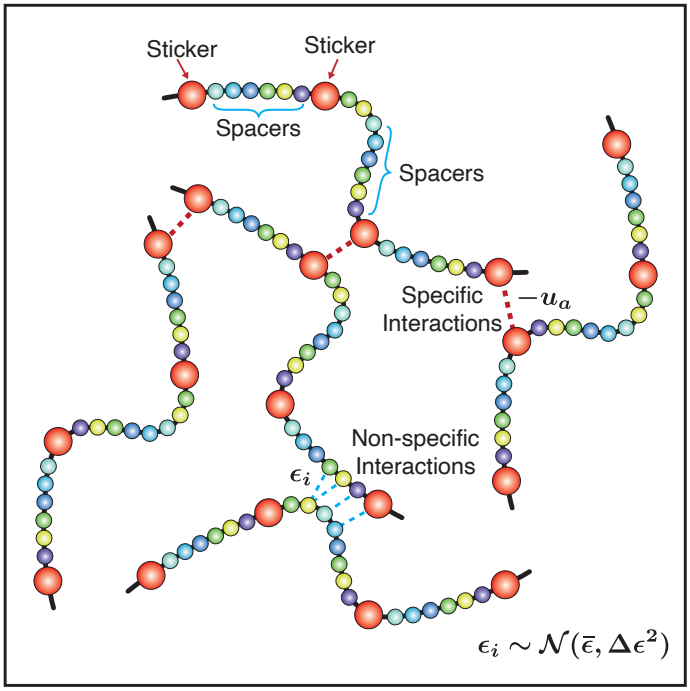

Figure 5.1: A schematic illustration of the stickers and random spacers model. Red spheres indicate stickers that interact specifically, and the strength for sticker-sticker interactions is a well-defined number, $-u_a$. We indicate the random spacers using shades of blue-green. These contribute non-specific interactions, and the pairwise interactions between adjacent pairs of spacers are chosen from a normal distribution, $\mathcal{N}(\bar{\varepsilon}, \Delta\varepsilon^2)$.

energy for a given configuration on the lattice from non-specific interactions is

$$-E = \sum_{i \in \text{n.n.}} \varepsilon_i. \tag{5.1}$$

The summation denotes a sum over neighboring spacer-spacer pairs and non-bonded sticker and spacer pairs. Our use of random energy follows the tradition of protein folding theory [294, 295], allowing the derivation of expressions with a mean field theory [296] that are not specific to considerations of any particular sequence.

When introducing interactions among spacers, the line between stickers and spacers becomes less distinct. Specifically, we may explore scenarios where spacer interactions rival those between stickers in strength. However, in accordance with existing literature, stickers in our model still exhibit restricted valence for establishing physical crosslinks among themselves. Conversely, spacers undergo multivalent, non-specific interactions.

In the following, we explore two systems to investigate the influence of spacer interactions on the organization and compositions of condensates. Initially, we examine a homotypic system consisting of a single polymeric species encompassing both specifically interacting sticker moieties and non-specifically interacting spacer residues. In this homotypic system, we formulate a mean-field free energy and investigate percolation coupled with phase separation. Our analysis reveals how the distribution of non-specific spacer interactions impacts the critical temperature, critical concentration, gel point, and degree of conversion. Subsequently, we extend this model to a heterotypic system involving two polymeric species, denoted as A and B. We highlight the necessity of finely tuning non-specific interactions to ensure robust composition in the dense phase for $A-B$ mixtures.

5.2.2 Phase behaviors of a single component system

We begin by examining a single-component system comprising n_p identical polymer chains on a lattice containing n sites. Let $\varphi \equiv n_p N/n$ represent the fraction of sites occupied by the polymer. The formation of $n_p m$ bonds occurs between $2n_p m$ stickers, and we define the degree of conversion as $p = \frac{2m}{f}$. Here, the f sticker segments are assumed to be uniformly distributed, effectively partitioning the chain into $f+1$ segments. We denote the expected length of each individual segment as l. By symmetry, $(f+1)l = N-f$. Consequently, $l = \frac{N-f}{f+1} \approx \frac{N}{f}$ for $N \gg f$ and $f \gg 1$.

The partition function for forming $n_p m$ bonds between $2n_p m$ stickers is,

$$\mathscr{Z} = \int \mathrm{d}E P(E) \Omega_{\mathrm{ref}} \Omega_{\mathrm{st}} \exp(\beta u_a n_p m) \exp(-\beta E). \tag{5.2}$$

Here, the probability distribution of the energy, $P(E)$, is Gaussian, since E is a sum of independent random variables following the Gaussian distribution (Eq. (5.1)). The mean and variance of $P(E)$ are given by

$$-n^{-1}\bar{E}[p,\varphi] = \frac{(z-p/l)\varphi^2\bar{\varepsilon}}{2}, \tag{5.3a}$$

$$n^{-1}\Delta E^2[p,\varphi] = \frac{(z-p/l)\varphi^2\Delta\varepsilon^2}{2}. \tag{5.3b}$$

Here, z denotes the coordination number of the lattice. The term p/l appears to account for the fact that once a sticker forms a physical bond with another sticker, it is no longer available to interact with spacers. $\beta = (k_B T)^{-1}$ where k_B is the Boltzmann's constant. For convenience, we set $k_B = 1$ and report all parameters in natural units.

When regarding $P(E)$ as a Gaussian distribution, we are overlooking the impact of chain connectivity and sequence correlations on the polymer configuration and total energy [170–174]. While this simplification might influence the phase diagram quantitatively, we expect our qualitative predictions to remain robust despite these assumptions [297].

Ω_{ref} is the number of ways of placing the n_p polymers on the lattice and is a standard result in polymer physics [298]. Its expression is given by,

$$\log(\Omega_{\text{ref}}) = -n\left[\frac{\varphi}{N}\log\left(\frac{\varphi}{Ne}\right) + (1-\varphi)\log(1-\varphi)\right]. \tag{5.4}$$

Ω_{st} corresponds to configurational entropy of forming sticker-sticker bonds. Following [163], we derive the expression as

$$\Omega_{\text{st}} = \frac{(n_p f)!}{(n_p f - 2n_p m)!(n_p m)!2^{n_p m}}\left(\frac{z-2}{n}\right)^{n_p m}. \tag{5.5}$$

In deriving the above expression, we assumed that the chains are strongly overlapping ($\varphi >> \varphi_{\text{overlap}} \sim N^{-1/2}$) and inter-chain bonds between stickers dominate over intra-chain ones [163]. Thus, to arrive at Ω_{st}, we first count the number of ways of choosing $2n_p m$ stickers out of $n_p f$ stickers to form bonds. We then multiply this number by the number of ways of pairing $2n_p m$ stickers together which is given by $(2n_p m - 1)!!$. Finally, we multiply this by the probability that all of the chosen stickers are neighboring each other. We also used the relation $(2k-1)!! = \frac{2k!}{2^k k!}$ to further simplify the expression finally yielding Eq. (5.5).

The integration in Eq. (5.2) is readily performed, and we use $-\beta F = \log(\mathscr{Z})$ to obtain the free energy. The free energy must be minimised with respect to the degree of conver-

sion, p, yielding the condition

$$\frac{p}{(1-p)^2} = \frac{(z-2)\varphi}{l}\exp\left(\beta u_a - \varphi\left(\beta\bar{\varepsilon} + \frac{\beta^2}{2}\Delta\varepsilon^2\right)\right). \tag{5.6}$$

Remembering that $p \in [0,1]$, we derive

$$p = 1 - \frac{\sqrt{\frac{4\varphi\lambda[\varphi]}{l}+1}-1}{\frac{2\varphi\lambda[\varphi]}{l}}, \tag{5.7}$$

where $\lambda[\varphi] \equiv (z-2)\exp\left(\beta u_a - \varphi\left(\beta\bar{\varepsilon} + \frac{\beta^2\Delta\varepsilon^2}{2}\right)\right)$. We substitute the condition, Eq. (5.6), into the original expression for βF, to obtain the free energy density $\beta\mathscr{F} \equiv n^{-1}\beta F|_{\partial_p F=0}$ as

$$\beta\mathscr{F} = \frac{\varphi}{N}\log\left(\frac{\varphi}{Ne}\right) + (1-\varphi)\log(1-\varphi) + \frac{\varphi}{l}\left[\frac{p}{2} + \log(1-p)\right] - \frac{z\varphi^2\beta\bar{\varepsilon}}{2} - \frac{z\varphi^2\beta^2\Delta\varepsilon^2}{4} \tag{5.8}$$

Eq. (5.8) provides the starting point for deriving equilibrium properties of the system. For example, to determine the critical behaviour we compute the chemical potential, $\mu = \partial_\varphi\mathscr{F}$. The critical point is then the intersection of the nullclines $\partial_\varphi\mu = 0$ and $\partial_\varphi^2\mu = 0$. Additionally, we also consider the concentration of chains with all f stickers free, $\mathscr{C}_{\text{free}}$, defined as,

$$\mathscr{C}_{\text{free}} = \frac{\varphi}{N}(1-p)^f. \tag{5.9}$$

The gel line is obtained by the condition $\partial_\varphi\mathscr{C}_{\text{free}} = 0$.

We also compute φ_{dense}, the concentration in the dense phase, by using $\Pi \approx 0$, where Π is the osmotic pressure given by the relation,

$$\beta\Pi = \beta(\varphi\partial_\varphi\mathscr{F} - \mathscr{F}). \tag{5.10}$$

The concentration in the dilute phase (φ_{dilute}) can then be obtained by equating the the

chemical potentials in the two phases.

5.2.3 Phase behaviors for a two component system

To further understand the influence of spacer interactions on condensate composition, we examine systems composed of two types of chains, denoted as A and B. Here, n_a and n_b represent the number of chains, while N_a and N_b denote the degree of polymerization for chains A and B, respectively. Chain A comprises f_a stickers, whereas chain B consists of f_b stickers. We permit specific $A-B$ sticker-sticker interactions ($-u_{ab}$) while prohibiting self-interactions (i.e., $u_{aa} = u_{bb} = 0$). This assumption mirrors a common scenario involving two proteins with specific binding.

The partition function for the two component system takes the familiar form,

$$\mathscr{Z} = \int \mathrm{d}E P(E) \Omega_{\mathrm{ref}} \Omega_{\mathrm{st}} \exp(\beta k u_{ab}) \exp(-\beta E). \tag{5.11}$$

We once again assume the spacer-spacer and spacer-sticker interactions between $A-A$, $B-B$ and $A-B$ chains are each drawn from a normal distribution $\mathscr{N}(\bar{\varepsilon}_{xx}, \Delta\varepsilon^2_{xx})$ where, $xx \in \{aa, bb, ab\}$. Correspondingly, $P(E)$ is Gaussian with mean and variance given by

$$-n^{-1}\bar{E}[p,\varphi] = \frac{1}{2}\Big[(z-p/l)\,\varphi_a^2\bar{\varepsilon}_{aa} + \left(z - \frac{\varphi_a}{\varphi_b}\frac{p}{l}\right)\varphi_b^2\bar{\varepsilon}_{bb} + \left(2z\varphi_a\varphi_b - \frac{p}{l}\varphi_a(\varphi_a+\varphi_b)\right)\bar{\varepsilon}_{ab}\Big], \tag{5.12a}$$

$$n^{-1}\Delta E^2[p,\varphi] = \frac{1}{2}\Big[(z-p/l)\varphi_a^2\Delta\varepsilon^2_{aa} + \left(z - \frac{\varphi_a}{\varphi_b}\frac{p}{l}\right)\varphi_b^2\Delta\varepsilon^2_{bb} + \Delta\varepsilon^2_{ab}\left(2z\varphi_a\varphi_b - \frac{p}{l}\varphi_a(\varphi_a+\varphi_b)\right)\Big]. \tag{5.12b}$$

The mixing entropy is given by [298],

$$-n^{-1}\log(\Omega_{\mathrm{ref}}) = \frac{\varphi_a}{N_a}\log\left(\frac{\varphi_a}{N_a e}\right) + \frac{\varphi_b}{N_b}\log\left(\frac{\varphi_b}{N_b e}\right) + (1-\varphi_a-\varphi_b)\log(1-\varphi_a-\varphi_b). \tag{5.13}$$

We assume the existence of k pairs of $A-B$ sticker-sticker bonds in the system. Therefore, $\Omega_{\mathrm{st}} = \binom{n_a f_a}{k}\binom{n_b f_b}{k} k! \left(\frac{z-2}{n}\right)^k$. In order to define the degree of conversion we note that

the maximum possible number of bonds is $\min\{n_a f_a, n_b f_b\}$, and without loss of generality we can assume it is chain A [176]. Therefore, we define $p_{ab} \equiv k/n_a f_a$. For algebraic convenience, we assume $\frac{N_a}{f_a} \equiv l_a \approx \frac{N_b}{f_b} \equiv l_b = l$. Consequently, the configurational entropy of the sticker-sticker bonds can be readily computed as:

$$-n^{-1}\log(\Omega_{\mathrm{st}}) = -\frac{\varphi_a}{N_a} p_{ab} f_a \log\left(\frac{(z-2)\varphi_b f_b}{eN_b}\right) + \frac{\varphi_a}{N_a} f_a \left(p_{ab}\log p_{ab} + (1-p_{ab})\log(1-p_{ab})\right) \tag{5.14}$$
$$+ \frac{\varphi_b}{N_b} f_b \left(1 - p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}\right)\log\left(1 - p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}\right).$$

Here we have used the Stirling's approximation (i.e, for $n >> 1$, $\log(n!) \approx n\log n - n + \mathscr{O}(\log n)$) and the following helpful identities, $\frac{k n_a f_a}{n_b f_b n_a f_a} = p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}$, $\varphi_t = N_t n_t/n$, $k/n = p_{ab} n_a f_a/n = p_{ab} f_a \varphi_a/N_a$.

We perform the integration in Eq. (5.11) and use $-\beta F = \log(\mathscr{Z})$ to obtain the free energy. We then minimise it with respect to the degree of conversion, p_{ab}, to yield,

$$\frac{p_{ab}}{(1-p_{ab})\left(1 - p_{ab}\frac{\varphi_a}{\varphi_b}\right)} = \frac{(z-2)\varphi_b}{l}\exp\left(\beta u_{ab} - \frac{\beta}{2}\left[\bar{\varepsilon}_{aa}\varphi_a + \bar{\varepsilon}_{bb}\varphi_b + \bar{\varepsilon}_{ab}(\varphi_a+\varphi_b)\right]\right. \tag{5.15}$$
$$\left. - \frac{\beta^2}{4}\left[\Delta\varepsilon^2_{aa}\varphi_a + \Delta\varepsilon^2_{bb}\varphi_b + \Delta\varepsilon^2_{ab}(\varphi_a+\varphi_b)\right]\right).$$

We substitute the above expression into the original expression for βF to obtain the free energy density $\beta\mathscr{F} \equiv n^{-1}\beta F|_{\partial_p F = 0}$

$$\begin{aligned}\beta\mathscr{F} = {} & \frac{\varphi_a}{N_a}\log\left(\frac{\varphi_a}{N_a e}\right) + \frac{\varphi_b}{N_b}\log\left(\frac{\varphi_b}{N_b e}\right) + (1-\varphi_a-\varphi_b)\log(1-\varphi_a-\varphi_b) \\ & + \frac{\varphi_a}{l}p + \frac{\varphi_a}{l}\log(1-p) + \frac{\varphi_b}{l}\log\left(1 - p\frac{\varphi_a}{\varphi_b}\right) \\ & - \beta z\varphi_a^2\frac{\bar{\varepsilon}_{aa}}{2} - \beta z\varphi_b^2\frac{\bar{\varepsilon}_{bb}}{2} - \beta z\varphi_a\varphi_b\bar{\varepsilon}_{ab} \\ & - \frac{\beta^2}{4}z\varphi_a^2\Delta\varepsilon^2_{aa} - \frac{\beta^2}{4}z\varphi_b^2\Delta\varepsilon^2_{bb} - \frac{\beta^2}{2}z\varphi_a\varphi_b\Delta\varepsilon^2_{ab}.\end{aligned} \tag{5.16}$$

One can recover the expressions derived for the homotypic case from the above equation by replacing $\varphi_a = \varphi_b \equiv \varphi$, up to a $\varphi \to 2\varphi$ transformation.

For the two component system, the spinodal can be obtained from the nullcline of the determinant of the hessian matrix of $\mathscr{F}$, namely, $\det(\mathscr{H}_{\mathscr{F}}) = 0$. We determine the binodal by solving the following system of equations

$$\begin{aligned} \mu_a^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \mu_a^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right), \\ \mu_b^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \mu_b^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right), \\ \Pi^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \Pi^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right). \end{aligned} \tag{5.17}$$

where, $\mu_k = \partial_{\varphi_k}\mathscr{F}$ and $\Pi = \sum_k \varphi_k(\partial_{\varphi_k}\mathscr{F}) - \mathscr{F}$. These equations ensure that the chemical potentials and the osmotic pressure are identical between coexisting phases. We solve the above equations by numerically finding roots to the the function $\mathscr{B}(\{\varphi_x^k\}) \equiv (\mu_a^{(0)} - \mu_a^{(1)}, \mu_b^{(0)} - \mu_b^{(1)}, \Pi^{(0)} - \Pi^{(1)}, 0) = 0$ using standard python libraries. More details on the numerical solutions are provided in Appendix D.

5.3 Results

5.3.1 Non-specific spacer interactions facilitate phase separation

As described in the Theory section, we introduce a novel model designed to investigate the phase behavior of condensate-forming proteins (Fig. 5.1). Expanding upon the stickers and spacers model [3, 4, 163–165], we identify specific chemical groups within proteins that exhibit robust interactions as stickers. However, contrary to prevalent approaches in contemporary literature, we account for heterogeneity in interactions among spacers. For simplicity, we assume that the strength of spacer interactions follows a normal distribution, $\mathscr{N}(\bar{\varepsilon}, \Delta\varepsilon^2)$, characterized by a mean of $\bar{\varepsilon}$ and a variance of $\Delta\varepsilon^2$. This variability accommodates the diversity of amino acids, resulting in weak yet significant interactions [294, 299, 300]. We refer to this model as the “stickers and random spacers" model, or STARS.

We compute the complete phase diagram of the STARS model to investigate the impact

of nonspecific spacer interactions on condensate behaviors. For a protein solution with identical molecules, we derive its free energy density $\mathscr{F}$ (Eq. (5.8)). From this, we calculate the chemical potentials as $\mu = \partial_\varphi \mathscr{F}$, where φ represents the polymer volume fraction. The binodal is obtained by equating the chemical potentials of each component in coexisting phases, while the spinodal is obtained as the nullcline of $\partial_\varphi \mu = 0$. These lines delineate the boundaries among the unstable, meta-stable, and stable regions of the phase diagram. The critical point (T_c, φ_c) denotes the temperature and concentration at which the solution first becomes unstable, aiding our understanding of the macromolecular solubility of the system.

We present the phase diagram for the STARS model in Fig. 5.2a. The sticker interactions are set as $u_a = 5k_BT$ and $\Delta\varepsilon = 2k_BT$ with $\bar{\varepsilon} = 0$. Additional model parameters are included in the caption. For comparison, we compute the phase diagram for a stickers and spacers model with similar parameters (Fig. 5.2b). Similar to the stickers and spacers model [301, 302], the STARS model exhibits both a percolation (gelation) transition and phase separation. In this work, we define gelation in the spirit of prior work on thermoreversible gelation in associative polymers [3, 4, 301], wherein the gel-phase we have a system-spanning, percolated network of sticker-sticker crosslinks. Owing to the transient nature of the crosslinks, the system might display elastic properties only in the short term, transitioning to liquid-like behaviors over longer periods, even after reaching the gelation threshold defined here.

At high temperatures with $T > T_c$, the system can undergo a sol-gel transition without phase separation. We also note that since the effect of random fluctuations enters the free energy via a term proportional to β^2 (Eq. (5.8)), we expect the high temperature behavior to be consistent with the standard stickers and spacers model [164, 301, 302] as this term approaches zero faster than terms linear in β. Decreasing the temperature further leads to phase separation coupled with gelation.

The inclusion of spacer interactions also leads to quantitative alterations in the phase diagram. In Fig. 5.2c, we observe a sharp rise in the critical temperature T_c with increasing $\Delta\varepsilon$, thereby promoting phase separation. The critical concentration, φ_c, follows a similar trend, albeit exhibiting a slight increase at larger values of $\Delta\varepsilon$ before quickly plateauing.

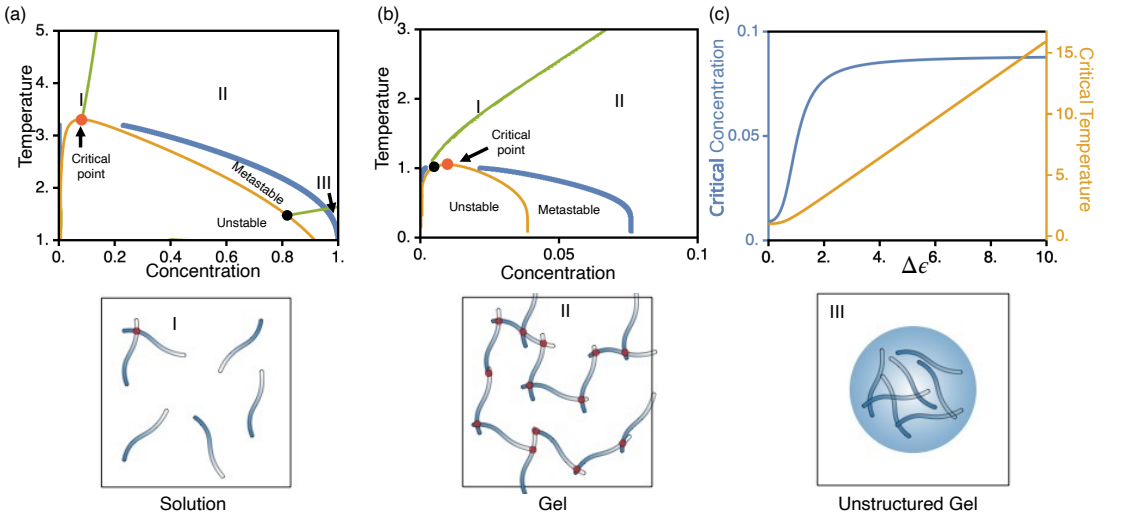

Figure 5.2: Phase behavior of the stickers and random spacers model. (a, b) Phase diagrams for the STARS model with $\Delta\varepsilon = 2$ (a) and the stickers and spacers model with $\Delta\varepsilon = 0$ (b). We plot the spinodal (orange) and binodal (blue) curves that demarcate the boundaries between the stable, meta-stable and unstable regions in the phase diagram. The critical point is highlighted in red. For the STARS model, the gel line (green) crosses the binodal twice, partitioning the stable phase into three regions. Illustrative configurations for the three regions corresponding to the solution phase, the gel phase, and the unstructured gel are shown in the bottom. (c) Dependence of the critical point on the strength of non-specific interaction among spacers, $\Delta\varepsilon$. We set $u_a = 5, \bar{\varepsilon} = 0, l = 10, N = 100$, and $z = 6$ when computing the phase diagrams.

Notably, these trends remain consistent regardless of other system parameters (Figs. D.1, D.2 and D.3). Spacer interactions, while chosen from a normal distribution with a zero mean, can produce contacts with negative energies that are favored by the Boltzmann factor, as defined in the partition function (Eq. (5.2)). Consequently, these contacts stabilize the condensed phase, thereby facilitating phase separation.

5.3.2 Non-specific spacer interactions modulate condensate organization

Fig. 5.2c indicates that diversifying IDP sequences might increase the variation in their interaction energies, potentially contributing positively to phase separation. However, it seems counter-intuitive that many IDPs, known to participate in condensate formation, would evolve sequences featuring low complexity regions [303–306]. These regions of-

ten contain amino acid repeats with suppressed sequence diversification and interaction patterns, leading to smaller $\Delta\varepsilon$ values.

We propose that spacer-spacer interactions could negatively impact condensates' networked structures arising from physical crosslinks between stickers. These crosslinks differentiate biomolecular condensates from simple liquids, generating heterogeneous environments with specific protein-protein interfaces. Robust contacts inside condensates could facilitate fast processing of intermediates, such as in metabolic channeling [307]. Additionally, physical crosslinks influence the viscoelastic and rheological properties crucial for condensate function [3, 4, 301]. For instance, differences in viscoelasticity between nucleolar core and outer layers strongly affect ribosomal assembly [308].

To characterize condensates' networked structure, we introduce the degree of conversion, p, measuring the fraction of bonded stickers. Additionally, we study the dependence of p on $\Delta\varepsilon$ at a constant temperature $T = 1$. For the parameters outlined in the caption of Fig. 5.3, the system undergoes phase separation coupled with gelation at this temperature. We denote the polymer volume fraction of the dense phase as φ_{dense} and use it to determine p from Eq. (5.7). As depicted in Fig. 5.3a, the degree of conversion initially increases slightly, reaching a maximum before decreasing to zero with increasing $\Delta\varepsilon$. This suggests that spacer-spacer interactions not only influence the overall macromolecular solubility but also mediate network properties of the stickers.

Moreover, spacer-spacer interactions can qualitatively change the phase diagram. We study $\mathscr{C}_{\text{free}}$, measuring the concentration of polymer chains with all stickers free (i.e., not bound to other stickers, Eq. (5.9)). As per established theories, for $\Delta\varepsilon = 0$, $\mathscr{C}_{\text{free}}$ first increases with concentration (φ) in the pre-gel regime, achieving a maximum at the gel point (Fig. 5.3b, blue). Subsequently, in the post-gel regime, $\mathscr{C}_{\text{free}}$ monotonically decreases to zero as an increasing number of stickers form crosslinks [163]. Consequently, the maximum of $\mathscr{C}_{\text{free}}$ often defines the gel line ($\partial_\varphi \mathscr{C}_{\text{free}} = 0$), depicted in Fig. 5.2 as the green curves.

Strikingly, we note that the monotonic decrease in $\mathscr{C}_{\text{free}}$ as a function of increasing φ in the post-gel regime for $\Delta\varepsilon \neq 0$ no longer holds true. As shown by the orange line in Fig. 5.3b, $\mathscr{C}_{\text{free}}$ decreases to a minimum before increasing again at higher concentrations.

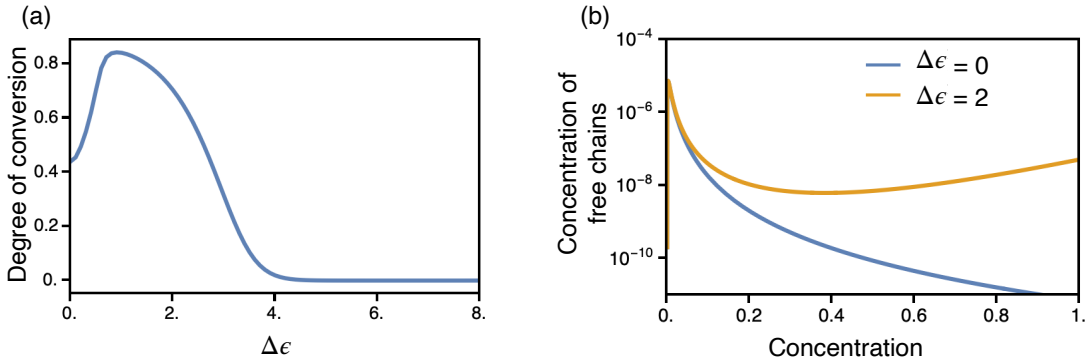

Figure 5.3: Impact of non-specific interactions among spacer on the network properties of condensates. (a) The degree of conversion evaluated at φ_{dense} (the concentration of polymers in the dense phase) shows a moderate increase followed by a subsequent decrease as we widen the spread of the spacer-spacer interaction energy distribution by increasing $\Delta\varepsilon$. (b) The concentration of free chains (with all stickers free) increases with concentration in the pre-gel regime and reaches a maximum at the gel-point. It monotonically decreases in the post-gel regime when $\Delta\varepsilon = 0$, but exhibits non-monotonic behavior for $\Delta\varepsilon \neq 0$. We set $u_a = 5$, $T = 1$, $\bar{\varepsilon} = 0$, $l = 10$, $N = 100$, and $z = 6$.

Once again, these observations remain qualitatively insensitive to the exact choice of system parameters (Figs D.1, D.2).

This non-monotonic behavior results in the first derivative of $\mathscr{C}_{\text{free}}$ crossing the binodal twice in Fig. 5.2a, partitioning the stable regions into three. Regions I and II resemble the typical solution and gel phase found in the stickers and spacers model (see Fig. 5.2b). However, in region III, $\mathscr{C}_{\text{free}}$ increases again due to a drop in sticker-sticker crosslinks. This transition leads to a less structured gel due to the competition between spacer interactions and the formation of sticker-sticker bonds.

5.3.3 Non-specific spacer interactions modulate condensate composition

Until now, our focus has been on single-component systems. However, biomolecular condensates within cells commonly consist of multiple molecules [2, 122, 156, 157, 178, 309–312]. These condensates possess well-defined compositions, allowing only certain molecules to selectively partition into them [313]. Maintaining such a specific composition is crucial for their function in both physiological contexts and potential therapeutic appli-

cations [6, 128, 314, 315]. To explore further, we investigated whether spacer interactions influence condensate composition in a minimal system comprising two components.

We aimed to model a scenario where A serves as the host, incorporating a polymeric guest B. Consequently, we focused on homotypic $A-A$ interactions ($\bar{\varepsilon}_{aa}=1, \Delta\varepsilon_{aa}=0$) to facilitate the demixing of polymer A from B [131]. To simplify, our attention was directed toward sequence diversification in protein B. This diversification results in both self-interactions and cross-interactions with A. Hence, we set both ε_{bb} and ε_{ab} as random variables with variance $\Delta\varepsilon_{bb} \equiv \Delta\varepsilon_{ab}$ and zero-mean (for convenience).

For this two-component system, we derived the free energy expression (Eq. (5.16)) to study its phase behaviors. Fig. 5.4 illustrates the phase behavior at constant temperature ($T=1$) as a function of composition in the $\varphi_a-\varphi_b$ plane, considering varying strengths of specific (u_{ab}) and nonspecific interactions ($\Delta\varepsilon_{bb}, \Delta\varepsilon_{ab}$). The orange lines denote the spinodal, while the blue lines represent the binodal. Additionally, the gray lines represent tie lines connecting coexisting phases.

We observed that the inclusion of nonspecific spacer-spacer interactions facilitates the co-condensation of A and B. Specifically, in Fig. 5.4a-b we consider the scenarios wherein sticker-sticker interactions are sufficiently weak ($u_{ab}=2$). In the absence of any $A-B$ or $B-B$ spacer interactions, the negative slope of tie lines in Fig. 5.4a suggests a demixing behavior. Stable phases at the top left and bottom right exhibit enrichment in only one component without a balanced presence of both. However, upon introducing nonspecific spacer interactions in Fig. 5.4b, tie line slopes become positive, indicating stable phases enriched or depleted in both components simultaneously–signifying co-condensation. A similar behavior was observed by Deviri and Safran [131] using the Flory-Huggins theory. The observed splay in the tie lines in Fig. 5.4b indicates that slight concentration fluctuations in the dilute phase can lead to significantly distinct compositions in the dense phase.

Furthermore, enhancing interaction strength among stickers ($u_{ab}=5$), while maintaining $\Delta\varepsilon_{ab}=\Delta\varepsilon_{bb}=0$ also facilitates co-condensation. As shown in Fig. 5.4c, a new stable phase emerges within the diagram's center. Tie lines connect this stable phase with regions depleted in one component at the top left and lower right corners. In this new stable region, $p_{ab}\approx 1$. Additionally, the stable co-condensation phase has a narrower range of

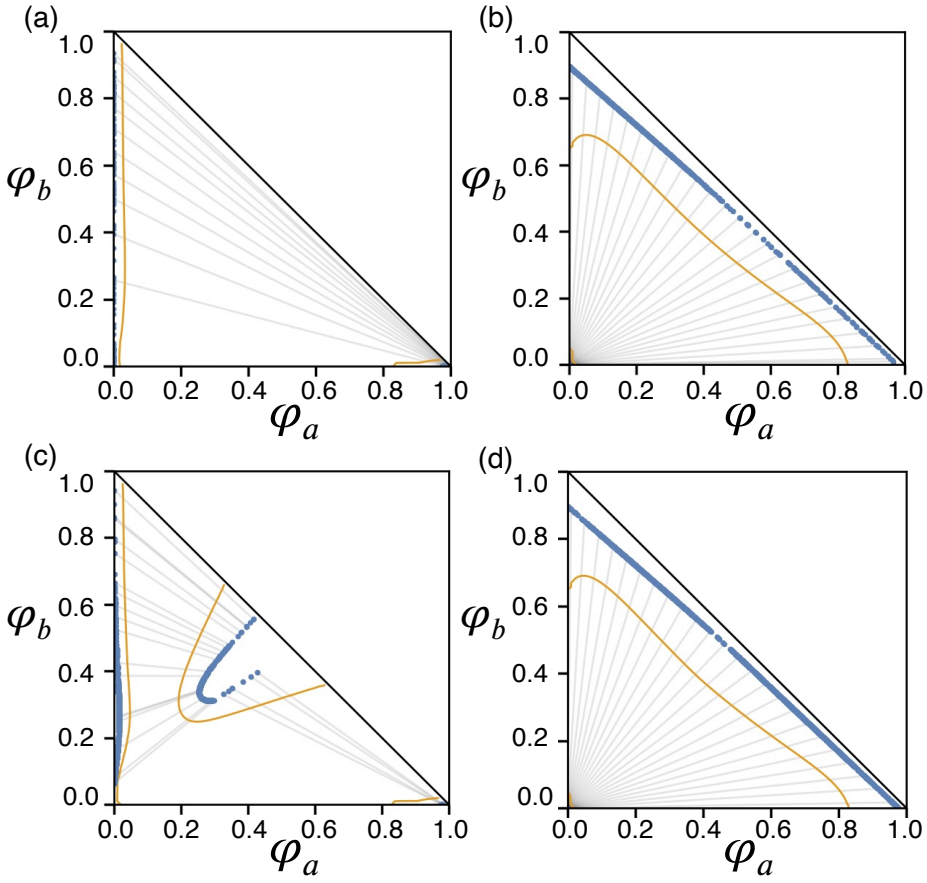

Figure 5.4: Phase diagrams showing the spinodal line (orange, solid) and binodal line (blue, dots) for a two component system with (a) $u_{ab} = 2, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, (b) $u_{ab} = 2, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.1$, (c) $u_{ab} = 5, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, and (d) $u_{ab} = 5, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.1$. The tie-lines (light grey, solid) connects co-existing points on the binodal. We set $N_a = N_b = 10, l = 2, z = 6, \bar{\varepsilon}_{aa} = 1, \Delta\varepsilon_{aa} = 0$, and $\bar{\varepsilon}_{bb} = \bar{\varepsilon}_{ab} = 0$ in all systems.

possible concentrations for A and B. This narrow concentration range is unlike the wider range supported in Fig. 5.4b and d, where a broader spectrum of mixing ratios exists in the stable condensed phase. In our current setup, the stable phase in Fig. 5.4c is centered around $(0.5, 0.5)$ due to symmetric sticker distributions in A and B. Altering the interaction strengths among stickers or the relative abundance of A or B stickers can aid in promoting phases with varied compositions. A detailed survey of titrating $\Delta\varepsilon_{bb}, \Delta\varepsilon_{ab}$ on the phase diagram in both the weak ($u_{ab} = 2$) and strong ($u_{ab} = 5$) sticker-sticker interaction regimes is shown in the Appendix D(Figs D.4 and D.5).

Thus, while both sticker-sticker and spacer interactions induce condensation of multiple components, sticker interactions robustly generate condensates with more defined molecu-

lar compositions and structures. This property can be advantageous in biological systems requiring precise rationing of different species to optimize efficiency in specific chemical processes. These observations remain qualitatively insensitive to the exact choice of system parameters (Figs D.6 and D.7).

5.4 Conclusions and Discussion

We have introduced a new theory aimed at establishing connections between the emergent physical properties of condensates and protein sequences. Specifically, by extending from the stickers and spacers model, we investigate how the interplay between specific and non-specific interactions can influence both the structural integrity and compositional specificity of biomolecular condensates. Specific interactions are defined as those exclusively formed between a pair of stickers due to their chemical selectivity, while non-specific interactions can occur among spacers and between spacers and non-bonded stickers. We analytically solved the new stickers and random spacers model to assess its phase behaviors.

Our findings demonstrate that non-specific spacer interactions have the capacity to promote phase separation and the condensation of multi-component systems. However, these non-specific interactions fail to generate condensates with robust networked structure amongst stickers (as determined by the degree of conversion, p), necessary for fine-tuning material properties and molecular compositions. Conversely, specific interactions show a similar ability to promote phase separation while establishing well-defined interaction networks within the condensate, resulting in precise compositions. Further elucidation of biomolecular condensates' structure and composition could significantly enhance our understanding of their impact on critical cellular processes such as transcription, epigenetic regulation and signalling. [35, 69, 136, 142, 220, 224, 316–326].

5.4.1 Revisiting the definition of stickers and spacers.

While it's acknowledged that stickers and spacers are context-dependent, it's common to identify potential stickers by pinpointing residues known to have the most pronounced

impact on saturation concentrations [116, 180, 327]. However, our study suggests that relying solely on critical behavior as a criterion for identifying potential stickers might be less robust. This is because spacers, characterized by non-specific interactions, can also influence both the critical and percolation behavior of the system. Alternatively, from a functional standpoint, we propose that stickers can also be identified as chemical groups that facilitate protein-protein interactions through specific binding. According to this definition, natural candidates for stickers are molecular recognition features (MoRFs) or short linear motifs (SLiMs). These consist of short stretches of adjacent amino acids essential for molecular recognition and protein binding [328–330]. Regions undergoing significant mutations leading to higher levels of binding promiscuity, however, should be classified as spacers. Identifying stickers and spacers based on interaction specificity helps in understanding the sequence features of IDPs and their evolutionary conservation, as we discuss below.

5.4.2 Evolutionary pressure and the rise of low complexity.

Our study offers insight into the evolution of IDP sequences. IDPs are recognized for their high evolvability, characterized by rapid mutation rates [331, 332]. Unlike globular proteins that require well-defined 3D structures, IDPs theoretically possess a broader sequence space for exploration [306]. However, IDPs do not explore all possible unfoldable sequences. Instead, they adopt simplified sequences enriched with low complexity domains, utilizing only a subset of amino acids. Furthermore, despite their poor conservation in alignments, recent studies have revealed that orthologous IDPs share many conserved molecular features [333–335], indicating non-randomness and suggesting evolutionary constraints that favor functionally fit sequence patterns[327, 336, 337].

Yet, the mere formation of biomolecular condensates cannot explain the conserved features in IDP sequences. As discussed in the main text, the diversification of spacer sequences can effectively reduce saturation concentrations and facilitate phase separation. Therefore, to just optimize condensate stability, IDPs would not heavily favor low complexity domains that suppress interactions.

We propose that the formation of condensates with robust structures and compositions

serves as the main pressure for IDP evolution. This pressure strengthens interactions among stickers while concurrently suppressing promiscuous interactions among spacers, leading to the prevalence of low complexity domains in IDPs. Lowering sequence complexity limits the range of possible amino acids within the sequence, naturally reducing the complexity and fluctuation of unwanted interactions. Additionally, for sequences of identical amino acid composition, interaction complexity can be further reduced by creating repeating, alternating aromatic residue sequences, such as FG-repeats in nucleoporins or YG-repeats in FUS-LCD [116, 338, 339]. These alternating sequences are known to exhibit weaker interactions compared to sequences where amino acids of the same type are clustered together [340].

The suggested evolutionary pressure inherently drives the preservation of stickers to maintain condensate compositional specificity. Molecular recognition features and short linear motifs, acknowledged for their evolutionary conservation, support this notion [333]. Conversely, spacers encounter an evolutionary push to diminish their complexity. This pressure doesn't favor any particular sequence, hence contributing to their high mutation rate. Nevertheless, mutations resulting in excessively strong interactions, even when localized within the low complexity domains (i.e., spacers), can disrupt normal function as well [140, 149].

Chapter 6

Conclusions

In this thesis, we explored the effects of dynamical coupling between epigenetic histone modifications and chromatin structure (Part I). Furthermore, motivated by the prominent role biomolecular condensates play in regulating chromatin's structure, epigenetic memory and transcription, we additionally investigated the role of interaction heterogeneity on liquid-liquid phase separation coupled to percolation in an attempt to decipher the molecular grammar of proteins forming biomolecular condensates (Part II).

In Chapter 2, we introduced a spatial-mean field model in the spirit of seminal models of epigenetic regulation alongside a novel second-quantization-based approach in order to analyze discrete stochastic models with a fixed, finite number of particles using a representation of the $SU(2)$ algebra. We applied this approach to a kinetic model of chromatin states that captures the feedback between nucleosomes and the enzymes conferring histone modifications. Using a path integral expression for the transition probability, we computed the epigenetic landscape that helps identify the emergence of bistability and the most probable path connecting the two steady states. Furthermore, the analytic approaches developed in Chapter 2 were further extended to also investigate epigenetic dynamics in concert with chromatin structural dynamics in Chapter 3, and to the study of epigenetic dynamics coupled to gene regulation in Chapter 4.

In Chapter 3, we demonstrated that incorporating the interplay between chromatin structural dynamics and histone modification kinetics can give rise to a dynamical phase transition. The validity of this transition was further verified by reproducing it in multiple

models: a comprehensive kinetic model encompassing microscopic chromatin contacts, a mean-field model, and a phenomenological model. Importantly, we also validated that the emergence of this transition occurs on biologically relevant timescales. Moreover, the behavior observed in the fast chromatin dynamics aligned with well-established observations regarding the influence of histone modifications on chromatin structure. Additionally, experimental evidence for slow chromatin relaxation further highlighted the significance of our results in the opposite limit. By introducing notions of a dynamical phase transition, we provided a cohesive framework that reconciles observations across disparate limits. Future experiments specifically designed to explore chromatin viscoelasticity under various conditions can potentially validate our theoretical predictions further. Moreover, some of the methods and ideas are broadly applicable to the study of dynamical processes and problems in general where there exists a coupling between 3D network structure and 1D sequence information, and where there is the presence of dynamical asymmetry in the relaxation rates for the two.

In Chapter 4, we introduced a minimal kinetic model for gene regulation that combines the impact of both histone modifications and transcription factors. By analyzing the steady-state solutions at various parameter regimes, we demonstrated the impact of histone modification kinetics on the behavior of a genetic network, resulting in qualitative changes in gene expression profiles. The emerging epigenetic landscape captures the delicate interplay between transcription factors and histone modifications in driving cell-fate decisions and highlights the impact of noise in the epigenetic reaction network on the tomography of the landscape.

In Chapter 5, we examined the role of interaction heterogeneity on the phase separation coupled to percolation. We developed an extension to the stickers and spacers model, incorporating heterogeneous, non-specific pairwise interactions between spacers alongside specific interactions among stickers. Our investigation reveals that while spacer interactions contribute to phase separation and co-condensation, their non-specific nature leads to disorganized condensates. Our analysis also revealed how the distribution of non-specific spacer interactions impacts the critical temperature, critical concentration, gel point, and degree of conversion. Subsequently, we extended this model to a heterotypic system in-

volving two polymeric species, denoted as A and B. We showed the necessity of finely tuning non-specific interactions to ensure robust composition in the dense phase for $A-B$ mixtures. We further discussed how evolutionary pressures might emerge to affect these interactions, leading to the prevalence of low-complexity domains in IDP sequences. These domains likely suppress spurious, promiscuous interactions to facilitate the formation of biologically meaningful condensates.

Appendix A

Supporting Information for Chapter 2

A.1 Constructing the coherent states path integral

For completeness let us start with the master equation for our system (with $N = 2j$),

$$\begin{aligned}\partial_t P(n_x,n_y) =& \frac{c_1}{(2j)^3}[(n_x-1)(n_x-2)(n_y+1)P(n_x-1,n_y+1)\\ &- n_x n_y(n_x-1)P(n_x,n_y)]\\ &+ \frac{c_1}{(2j)^3}[(n_y-1)(n_y-2)(n_x+1)P(n_x+1,n_y-1)\\ &- n_x n_y(n_y-1)P(n_x,n_y)]\\ &+ \frac{c_2}{2j}[(n_x+1)P(n_x+1,n_y-1) - n_x P(n_x,n_y)]\\ &+ \frac{c_2}{2j}[(n_y+1)P(n_x-1,n_y+1) - n_y P(n_x,n_y)]. \end{aligned} \tag{A.1}$$

As mentioned in the main text, following the second quantization approach [199, 200, 205–207] and by introducing the state vector $|\psi(t)\rangle$, the time evolution of the original master equation can be recast into an *imaginary* time Schrödinger equation (Eq. (2.3)), and the Hamiltonian is defined in terms of the bosonic creation and annihilation operators as Eq. (2.5). Now we can begin reformulating this system using the Jordan-Schwinger map [208]. For convenience, let us define the following vectors (Eq. (A.2)), to which we then apply the Jordan transformation to obtain (Eq. (A.3)), where σ_μ denotes the usual Pauli

Matrices.

$$\boldsymbol{a} \equiv \begin{pmatrix} a_x \\ a_y \end{pmatrix}, \qquad \boldsymbol{a}^\dagger \equiv \begin{pmatrix} a_x^\dagger \\ a_y^\dagger \end{pmatrix}. \tag{A.2}$$

$$\begin{aligned} Q_\mu &= \boldsymbol{a}^\dagger \boldsymbol{\sigma}_\mu \boldsymbol{a} \\ J_\pm &= Q_1 \pm i Q_2 \\ J_+ = a_x^\dagger a_y \qquad & J_- = a_y^\dagger a_x, \end{aligned} \tag{A.3}$$

Also for convenience, we define the auxiliary operators (Eq. (A.4)). The operators $J_\pm$ satsify the same commutation relations as $SU(2)$ algebra, given in Eq. (2.8)

$$\begin{aligned} J_x &= \frac{(J_+ + J_-)}{2} \\ J_y &= \frac{(J_+ - J_-)}{2i} \\ [J_+, J_-] &= 2J_z. \end{aligned} \tag{A.4}$$

Using these rules, the stochastic pseudo-hamiltonian was reformulated as in Eq. (2.11). In order to now construct a path integral for the transition probability, one introduces the following left and right spin-coherent states [341, 342], and a resolution of identity,

$$\begin{aligned} |z\rangle &= \frac{1}{(1+z\bar{z})^j} e^{zJ_+} |0\rangle = (1+z\bar{z})^{-j} \sum_{0\le n}^{2j} \binom{2j}{n} z^n |n\rangle, \\ \langle z| &= \frac{1}{(1+z\bar{z})^j} \langle 0| e^{\bar{z}J_-} = (1+z\bar{z})^{-j} \sum_{0\le n}^{2j} \langle n| \bar{z}^n, \end{aligned} \tag{A.5}$$

$$\int \frac{2j+1}{\pi} \frac{d^2 z}{(1+z\bar{z})^2} |z\rangle\langle z| = \mathbb{1}. \tag{A.6}$$

The details and subtleties regarding the construction of spin coherent state path integrals

have been discussed in the literature [209, 343–346]. However, for completeness we give a brief overview. Starting with the propagator between two normalized coherent states, $\langle z_f | e^{-tH} | z_i \rangle$, we discretize the time interval $[0,t]$ into N_t time slices, and then insert a resolution of identity of the form (A.6) between each time slice. Finally, taking the limit $N_t \to \infty$ we get,

$$\langle z_f | e^{-tH} | z_i \rangle = \int \mathscr{D}[\bar{z}, z] \exp(-S), \tag{A.7}$$

where,

$$\begin{aligned} S = &- j \log \frac{(1+\bar{z}_f z(t))(1+\bar{z}(0) z_i)}{(1+\bar{z}_f z_f)(1+\bar{z}_i z_i)} \\ &+ 2j \int_0^t \mathrm{d}t \left[\frac{1}{2} \frac{\bar{z}\dot{z} - \dot{\bar{z}} z}{1+\bar{z}z} - H(\bar{z}, z) \right], \end{aligned} \tag{A.8}$$

and $H(\bar{z}, z) = \langle z | H | z \rangle$. Next, we derive an expression for the *physical* propagator between $\langle n_f |$ and $| n_i \rangle$ representing states of fixed initial and final number of particles respectively. This represents the probability of starting in state with particle number n_i at time $t = 0$ and ending up in a state with particle number n_f at t_f. To do so one takes,

$$\begin{aligned} \mathbb{P}(n_f; t_f | n_i; 0) &= \langle n_f | e^{-tH} | n_i \rangle \\ &= \int \mathscr{D}[z_i, z_f] \langle n_f | z_f \rangle \langle z_f | e^{-tH} | z_i \rangle \langle z_i | n_i \rangle . \end{aligned} \tag{A.9}$$

To get the *physical* propagator from (A.7) one needs to subtract $\log \langle z_i | n_i \rangle + \log \langle n_f | z_f \rangle$ from the action, and then integrate over $z_i, \bar{z}_i, z_f, \bar{z}_f$. Then,

$$\mathbb{P}(\rho_f; t_f | \rho_i; 0) = \int \mathscr{D}[\overline{z_i}, z_f] \int \mathscr{D}[\bar{z}, z] e^{-S}, \tag{A.10}$$

where we have introduced, $\rho = n/2j$. Here S is now given by,

$$\begin{aligned} S = &-j\log\Big[(1+\bar{z}_f z(t)(1+\bar{z}(0)z_i)\Big] \\ &+2j\int_0^t \mathrm{d}t\left[\frac{1}{2}\frac{\bar{z}\dot{z}-\dot{\bar{z}}z}{1+\bar{z}z} - H(\bar{z},z)\right] \\ &+2j\Big[-\rho_i\log\bar{z}_i - \rho_F\log\bar{z}_f + \rho_f\log\rho_f + (1-\rho_f)\log\big(1-\rho_f\big) \\ &+\log\big((+\bar{z}_f z_f)(1+z_i\bar{z}_i)\big)\Big]. \end{aligned} \tag{A.11}$$

Now we can integrate over $z_i, z_f, \bar{z}_i, \bar{z}_f$ using the saddle-point method. The derivatives of (A.11) fix the initial and final conditions,

$$\frac{1}{2j}\frac{\partial S}{\partial z_i} = \frac{\bar{z}_i}{1+\bar{z}_i z_i} - \frac{\bar{z}(0)}{1+\bar{z}(0)z_i}, \tag{A.12a}$$

$$\frac{1}{2j}\frac{\partial S}{\partial \bar{z}_f} = \frac{z_f}{1+\bar{z}_f z_f} - \frac{z(t)}{1+z(t)z_f}, \tag{A.12b}$$

$$\frac{1}{2j}\frac{\partial S}{\partial \bar{z}_i} = \frac{z_i}{1+\bar{z}_i z_i} - \frac{\rho_i}{\bar{z}_i}, \tag{A.12c}$$

$$\frac{1}{2j}\frac{\partial S}{\partial z_f} = \frac{\bar{z}_f}{1+\bar{z}_f z_f} - \frac{\rho_f}{\bar{z}_f}. \tag{A.12d}$$

Thus, we get the following four conditions $\bar{z}(0) = \bar{z}_i$, $z(t) = z_f$, $\rho_i = \frac{\bar{z}_i z_i}{1+\bar{z}_i z_i}$ and $\rho_f = \frac{\bar{z}_f z_f}{1+\bar{z}_f z_f}$. After performing the integration the action now reads,

$$\begin{aligned} S =& 2j\left[\frac{z\bar{z}}{1+z\bar{z}}\log\bar{z} - \log(1+z\bar{z})\right]\Bigg|_f^i \\ &+2j\int_0^t \mathrm{d}t'\left[\frac{1}{2}\frac{\bar{z}\dot{z}-\dot{\bar{z}}z}{1+\bar{z}z} - H(\bar{z},z)\right]. \end{aligned} \tag{A.13}$$

We introduce one final re-parametrisation in terms of the density ρ. Using

$$\rho = \frac{\langle z|\, j + J_z\,|z\rangle}{2j} = \frac{z\bar{z}}{1+z\bar{z}}, \tag{A.14}$$

one rewrites

$$z = \frac{\rho}{1-\rho}\exp(-\tilde{\rho}), \qquad \bar{z} = \exp(\tilde{\rho}) \tag{A.15}$$

with $\rho(0) = \rho_i$ and $\rho(t) = \rho_f$ and $\tilde{\rho}(t)$, $\tilde{\rho}(0)$ unconstrained. The Jacobian for the change of variables is $(1-\rho)^2$, and $\int \frac{2j+1}{\pi}\frac{d^2z}{(1+z\bar{z})^2}$ is replaced by $\int \frac{2j+1}{\pi}d^2\rho$. Doing so we finally recover (2.17) and (2.19) of the main text. To get the deterministic rate equations, we evaluate,

$$\dot{\rho} = \left.\frac{\partial H}{\partial \tilde{\rho}}\right|_{\tilde{\rho}=0} = \frac{(1-2\rho)\left(c_1(j-1)(2j-1)(\rho-1)\rho + 2c_2 j^2\right)}{2j^2} \tag{A.16}$$

which for $j \gg 1$, gives

$$\dot{\rho} = (1-2\rho)(c_1(\rho-1)\rho + c_2). \tag{A.17}$$

This equation is in agreement with Micheelsen *et.al.* [67], and it yields a single stable real fixed point, $\rho^* = 0.5$ when $c_1/c_2 < 4$ and an unstable fixed point at $\rho^* = 0.5$ and 2 stable fixed points at $\rho^* = 0.5 \pm 0.5\sqrt{(c_1 - 4c_2)/c_1}$ when $c_1/c_2 > 4$.

A.2 Details of Transition Matrix Calculations

The transition rate matrix corresponding to the reaction network of the chromatin state model is a $N \times N$ tridiagonal matrix with off-diagonal elements defined as

$$\begin{aligned} H_{i,j} = \delta_{i,i+1}&\left(\frac{c_1}{N^3}i(i-1)(N-i) + \frac{c_2}{N}(N-i)\right) \\ &+\delta_{i,i-1}\left(\frac{c_1}{N^3}i(N-i)(N-i-1) + \frac{c_2}{N}i\right). \end{aligned} \tag{A.18}$$

The diagonal elements were defined to ensure that each column of the matrix sums to zero, namely,

$$H_{i,i} = -\sum_{i} H_{i,j}. \tag{A.19}$$

The smallest (in absolute value) non-zero eigenvalue of the matrix corresponds to the transition rate between two steady states. The eigenvector for the zero eigenvalue quantifies the steady state probability distribution.

Appendix B

Supporting Information for Chapter 3

B.1 Parameterizing the free energy functional of chromatin conformations in the contact space

For efficient simulations across a wide range of time scales, we consider chromatin conformational dynamics in the contact space. For a chromatin segment with $\mathbf{N}$ nucleosomes, the total set of contacts is represented with a vector of size $\mathbf{M}$, $\boldsymbol{q}(t) \equiv \{q_{ij}(t)\}$ for $i, j \in [1, \mathbf{N}]$ and $j - i > 1$. The binary variables $q_{ij} \in \{0, 1\}$ denote the presence (or absence) of 3D contacts between a pair of nucleosomes (i, j). We assume that neighboring nucleosomes are always in contact, i.e., $q_{i,i+1} \equiv 1$, and $\mathbf{M} = \mathbf{N}(\mathbf{N}-1)/2 - (\mathbf{N}-1)$.

Recall that the free energy functional in contact space is given by

$$\mathscr{H}(\boldsymbol{q}) = \sum_{ij} h_{ij} q_{ij} + \sum_{ijkl} J_{ijkl} q_{ij} q_{kl} + \lambda \sum_{ij} q_{ij} \Big(\sum_{t \neq i,j} q_{it} + \sum_{t \neq j,i} q_{tj} \Big). \tag{B.1}$$

The linear terms h_{ij} incorporate the entropic penalty of loop formation. We enforce translational symmetry such that contacts with the same sequence separation share identical penalty [230], i.e., $h_{ij} = h_{kl}$ whenever $j - i = l - k$. The correlation between contacts (i, j) and (k, l), which is caused by the polymer's topology, is accounted for by J_{ijkl}. Without loss of generality, the J indices are ordered such that $k_1 \equiv j - i \leq l - k \equiv k_2$. Whenever $k_1 = k_2$, we additionally require $l > j$. Topologically-equivalent contact pairs can be identified by

also defining $L \equiv l - j$ [231, 232]. Naturally, the topology-driven correlations are identical for topologically-equivalent contact pairs. Therefore, we assign identical J values to all contact pairs with identical k_1, k_2, and L values [231, 232]. Furthermore, we set $J_{ijkl} = 0$ for contacts with non-overlapping loops, which occur whenever $j \leq k$ or $l \leq i$, because the polymer's topology has no effect on these contacts' correlation [231, 232]. The third term accounts for the excluded volume effect that limits the number of contacts a given nucleosome can form.

We determined the parameters in Eq. (B.1) such that the free energy functional accurately describes the conformational distribution of polymers. As detailed below, the conformational distribution was produced with molecular dynamics simulations, and we used the pseudolikelihood approach for efficient parameter inference.

Note that each J_{ijkl} parameter affects the total energy only when both of its corresponding contacts are formed. Therefore, negative J values favor contact formation. Meanwhile, the optimization scheme discussed in Appendix B.1.2 yields a model with many more negative than positive J values. Additionally, the negative J values are larger in magnitude than the positive J values. This biases the system towards states in which many beads are involved with an unphysically large number of contacts. The third term of the Hamiltonian corrects this effect. This second-order correction term penalizes individual beads forming multiple contacts. Physically, this represents the excluded volume effect. A λ value of 0.01 was sufficient to prevent the polymer collapse.

B.1.1 Polymer conformations from molecular dynamics simulations

We performed three independent molecular dynamics simulations to produce three conformational ensembles of polymers. The monomers are connected to nearest neighbors with the finite extensible nonlinear elastic (FENE) potential

$$u_{\text{bond}}(r_{i,i+1}) = -\frac{1}{2}KR_0^2 \ln\left[1 - \left(\frac{r_{i,i+1}}{R_0}\right)^2\right], K_b = 30\varepsilon, R_0 = 1.5\sigma. \tag{B.2}$$

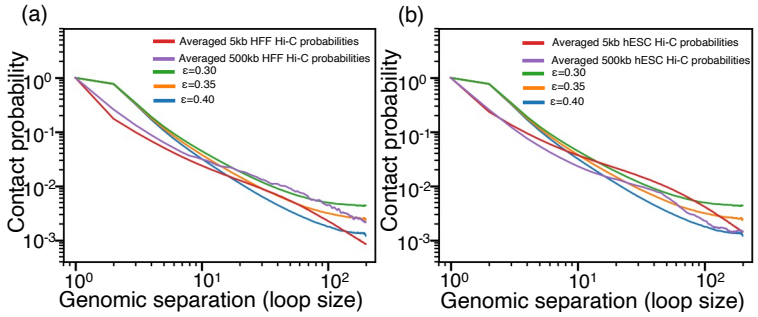

Figure B.1: For most loop sizes relevant to this study, the loop size-averaged contact probabilities associated with the three polymer models agree with Hi-C data representing the first chromosome of (a) human foreskin fibroblasts (HFF) and (b) human embryonic stem cells (hESC) cells at 5 and 500 kb resolution. Plotting the contact probabilities against their loop size, $|j-i|$, highlights the influence of topological constraints in each system.

The Lennard-Jones (LJ) potential was applied between all monomer pairs

$$U_{\text{wall}}(r) = \begin{cases} 4\varepsilon_{LJ}[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] + E_{\text{cut}}, & r < r_c \\ 0, & \text{otherwise,} \end{cases} \tag{B.3}$$

where E_{cut} is the energy of the LJ potential at the cutoff distance $r_c = 2.6\sigma$. We chose ε_{LJ} as 0.35 for the LJ potential because it creates a sequence separation dependence for contact probabilities similar to that expected in chromatin systems *in vivo* [233]. Fig. B.1 illustrates this, showing the relationship between contact probability and sequence separation for each of the homopolymer models and for Hi-C data from (a) human foreskin fibroblasts (HFF) and (b) human embryonic stem cells (hESC). Each Hi-C contact probability represents the average of all contact probabilities in the first chromosome with the same sequence separation, so the plot illustrates non-specific effects; similarly, homopolymer contact probabilities were averaged for monomer pairs separated by the same number of bonds. In addition, we simulated homopolymer models using $\varepsilon_{LJ} = 0.3$ and $\varepsilon_{LJ} = 0.4$ to obtain two additional sets of homopolymer conformations that remain biologically plausible while being qualitatively distinct; these sets improve the result of our pseudolikelihood maximization procedure discussed in the next subsection. The simulated polymer is 500 beads in length, and only the conformations of the central 40 beads were recorded. Simulating a longer polymer avoids potential edge effects that might produce different statistics

for polymer beads at the boundary.

The LAMMPS software package [347] was used to perform the simulations using reduced units and shrink-wrapping boundary conditions with a time step of 0.005. Langevin dynamics with a damping parameter of 10 were used to maintain the temperature. Temperature replica exchange was used with seven temperatures evenly spaced from 0.7 to 1.3, and data was collected from the replica with $T = 1.0$. Exchanges were performed every 100^{th} timestep, and configurations were collected once every 5000 timestep over one billion total timesteps. This yields $200,000$ configurations for each of the three homopolymers.

B.1.2 Parameter optimization with the pseudolikelihood approach

The model parameters h_{ij} and J_{ijkl} were learned using a pseudolikelihood maximization approach developed in previous work [235]. This approach adjusts the parameters to optimize the probability, or likelihood, of the conformations from a reference ensemble. From the 3D structures in Cartesian space obtained using molecular dynamics simulations, we obtained a list of contacts between monomer pairs using a distance cutoff of 1.707σ. This conversion produces three ensembles of polymer conformations in the contact space, denoted as $\mathbb{B}_1$, $\mathbb{B}_2$, and $\mathbb{B}_3$, which were used in pseudolikelihood optimization.

The function used for parameter optimization is defined as

$$\ell_{\text{pseudo},\mathbb{B}}\left(\boldsymbol{h},\boldsymbol{J},\boldsymbol{\Delta h}\right) = \tag{B.4}$$
$$\sum_{b\in\mathbb{B}_1}\sum_{(i,j)}\log\left[\frac{1}{1+\exp\left((2q_{ij}-1)(h_{ij}+\Delta h_1+\sum_{kl}J_{ijkl}q_{kl}^{(b)})\right)}\right]$$
$$+\sum_{b\in\mathbb{B}_2}\sum_{(i,j)}\log\left[\frac{1}{1+\exp\left((2q_{ij}-1)(h_{ij}+\Delta h_2+\sum_{kl}J_{ijkl}q_{kl}^{(b)})\right)}\right]$$
$$+\sum_{b\in\mathbb{B}_3}\sum_{(i,j)}\log\left[\frac{1}{1+\exp\left((2q_{ij}-1)(h_{ij}+\Delta h_3+\sum_{kl}J_{ijkl}q_{kl}^{(b)})\right)}\right]$$

Here, $\boldsymbol{h}$ ($\boldsymbol{J}$) is the model's full set of h_{ij} (J_{ijkl}) parameters. The three terms on the right side correspond to the total pseudolikelihood of the energy function defined in Eq. (B.1) over the

configurations from the three ensembles. Since the molecular dynamics simulations used to produce the three ensembles differ in the non-bonded interaction energy, we introduced $\boldsymbol{\Delta h} = \{\Delta h_1, \Delta h_2, \Delta h_3\}$ to account for the difference. Δh_1, corresponding to $\varepsilon_{\mathrm{LJ}} = 0.35$, was fixed at 0. Our use of three conformational ensembles with different degrees of polymer collapse provides a wide variety of conformations essential for probing the correlation between contact pairs.

As mentioned in the main text, the first-order parameter h_{ij} accounts for the entropic penalty associated with forming contact (i, j). In addition, it includes the contact potential associated with the interaction between i and j in the homopolymer, so shifting the well depth associated with each Lennard-Jones potential, ε_{LJ}, causes a constant shift in all h_{ij}. $\Delta h_\varepsilon \in \boldsymbol{\Delta h}$ accounts for these shifts, and we fixed $\Delta h_1 = 0$ so that the computed h parameters correspond to the simulation with $\varepsilon_{LJ} = 0.35$. We related h to the homopolymer parameterized by $\varepsilon_{LJ} = 0.35$ because the sequence-separation dependence of contact probabilities in that model agree with the probabilities observed in chromatin systems *in vivo* [233]. Meanwhile, the supporting simulations use ε_{LJ} that differ from this biologically accurate parameterization by the small $\Delta\varepsilon_{LJ} = \pm 0.05$ to ensure that the resulting sets of conformations remain biologically plausible while providing qualitatively distinct sets of conformations (Fig. B.2a).

On the other hand, J accounts for the correlation between contacts, which primarily depends on connectivity and excluded volume effects [231, 232]. The former is independent of ε_{LJ}, and the latter is negligibly affected by it. Consequently, J is unaffected by the choice of ε_{LJ}, and all models share the same J values. Therefore, using PLM to infer the Ising-like parameters of all three models simultaneously provides additional information regarding first- and higher-order processes, encouraging h and J to capture the intended entropic effects.

Additional regularization was introduced to ensure the robustness of parameter optimization, leading to the final objective function

$$\ell_{\mathbb{B}}\left(\boldsymbol{h}, \boldsymbol{J}, \boldsymbol{\Delta h}\right) = \ell_{pseudo,\mathbb{B}} + \gamma \sum_{ijkl} J_{ijkl}^2 \tag{B.5}$$

with γ set to 0.6.

The function was optimized with the limited-memory Broyden–Fletcher–Goldfarb–Shannon with bound variables (L-BFGS-B) algorithm using SciPy version 1.5.2. All parameters were initialized at 0.

B.1.3 Characterizing the parameterized model

In Fig. B.2(a), we plot the distribution of the homopolymer's radius of gyration (R_g) within each simulated set of homopolymer conformations, i.e., the set of conformations associated with each ε value and obtained via MD simulation. For each conformation, we computed the R_g of the central 40 monomers of the polymer alone, as this is the polymer region considered by our pseudolikelihood maximization procedure. The observed distributions indicate that the homopolymer by itself is expected to be relatively open, and we reiterate that the kinetic model incorporates Ising-like parameters representing a homopolymer with $\varepsilon_{LJ} = 0.35$. We note that each set of conformations was used only once, namely to parameterize the Ising-like representation of a homopolymer model. Afterwards, the kinetic simulations discussed in the main text utilized this parameterization, which remained constant throughout; the attractions in the system that arise from nucleosome marking are introduced during the kinetic simulation and are entirely independent of the pseudolikelihood maximization task and the specific h_{ij} and J_{ijkl} values. As to the specific ε_{LJ} values chosen in this work, the contact probabilities produced by the polymer model using $\varepsilon_{LJ} = 0.35$ agree with the sequence separation-averaged contact probabilities observed in *in vivo* chromatin systems. (See [233], which uses an identical parameterization for the homopolymer model described in its Fig. 5A.) Therefore, the $\varepsilon_{LJ} = 0.35$ model is biologically relevant, and the supporting simulations use ε_{LJ} that differ from this biologically accurate parameterization by the small $\Delta\varepsilon_{LJ} = \pm 0.05$ to ensure that the resulting sets of conformations remain biologically plausible while still providing qualitatively distinct sets of conformations. We note a reasonable agreement between the statistics obtained from explicit molecular dynamics simulations and those reproduced by our learned model (Fig. B.2). This is more than sufficient for our intended primary purpose of demonstrating that coupling between the stochastic epigenetic reaction network and structure pro-

duces interesting dynamical behavior when the underlying structure relaxes on disparate timescales. The subtleties of parameterizing Ising-like models of homopolymers are the subject of other works [233].

Fig. B.2 includes plots that compare the first- (b,c) and second-order (d) contact statistics of the Ising-like model, obtained via a TRE MCMC simulation, to the statistics computed from the homopolymer conformations obtained via MD simulation. These plots demonstrate the quality of the parameters composing the homopolymer model.

In addition, Fig. B.3 compares h and J to the MD-derived homopolymer contact statistics that illustrate the physical effects they capture. First, topological constraints cause the contact probability between beads i and j to decrease as their sequence separation, quantified by the loop size $j-i$ increases [230–232]. The left panel shows this effect, and the Ising-like model captures it by increasing h as loop size increases. Second, topological constraints couple the formation of different contacts [231, 232], which we illustrate with the covariance between each pair of interactions, i.e. $C_{ijkl} \equiv \langle q_{ij} q_{kl} \rangle - \langle q_{ij} \rangle \langle q_{kl} \rangle$. To visualize second-order data, the right panel considers the interaction between the contact $(16,26)$ and all other contacts (k,l), and the grid uses k and l (lower triangle) or l and k (upper triangle) to index the x- and y-axis, respectively. The covariance $C_{16,26,kl}$ is positive for interactions in which one loop contains the other, negative for partially overlapping loops, and near-zero for loops that don't overlap. (Chan and Dill provide a thorough description of the distinct topological conditions associated with each of these interaction types in [231, 232].) Supporting these correlations, the $J_{16,26,kl}$ values (lower triangle) are negative-valued for correlated contact pairs ($C_{16,26,kl} > 0$), positive-valued for anticorrelated contact pairs ($C_{16,26,kl} < 0$), and zero-valued for uncorrelated contact pairs ($C_{16,26,kl} \approx 0$). Green indicates values that are undefined in the Ising-like model.

B.2 Deriving The pseudo-potential and transition rates

Following Ref. [67], we derive the pseudo-potential $V(\overline{n}, \overline{q})$ that dictates the transition rates between marked and unmarked states. We consider $\overline{n} = \sum_i n_i / N$, the fraction of nu-

cleosomes in the modified state. The kinetic equation for $\mathrm{d}\overline{n} = 1/\mathbf{N}$ is

$$\frac{\mathrm{d}\overline{n}}{\mathrm{d}t} = (R_+(\overline{n}) - R_-(\overline{n}))\mathrm{d}\overline{n} \tag{B.6}$$

where,

$$R_+(\overline{n}) = c_r\overline{n}^2(1-\overline{n})\overline{q}^2 + c_n(1-\overline{n}), \tag{B.7a}$$

$$R_-(\overline{n}) = c_r\overline{n}(1-\overline{n})^2\overline{q}^2 + c_n\overline{n}. \tag{B.7b}$$

The $c_r\overline{n}^2(1-\overline{n})\overline{q}^2$ term in above equations represents the rate of recruited conversion from unmarked to marked nucleosomes. Similarly, $c_r\overline{n}(1-\overline{n})^2\overline{q}^2$ corresponds to recruited conversions from marked to unmarked nucleosomes. The terms proportional to c_n represent noisy conversions. c_r and c_n are the same rates introduced in the main text. For convenience, we introduce the feedback ratio $F = c_r/c_n$, the ratio of recruited to random conversions. We formulate the master equation for this system,

$$\begin{aligned}\partial_t P(\overline{n},t) = {} & R_-(\overline{n}+\mathrm{d}\overline{n})P(\overline{n}+\mathrm{d}\overline{n},t) + R_+(\overline{n}-\mathrm{d}\overline{n})P(\overline{n}-\mathrm{d}\overline{n},t) \\ & - [R_+(\overline{n}) + R_-(\overline{n})]P(\overline{n},t)\end{aligned} \tag{B.8}$$

We expand Eq. (B.8) to second-order to obtain the following Fokker-Planck equation

$$\begin{aligned}\partial_t P = -\frac{\partial J}{\partial \overline{n}} &= -\frac{\partial}{\partial \overline{n}}\left[-\mu(\overline{n})\frac{\mathrm{d}U(\overline{n})}{\mathrm{d}\overline{n}}P - \frac{\partial(D(\overline{n})P)}{\partial \overline{n}}\right] \\ &= -\frac{\partial}{\partial \overline{n}}\left[-\mu(\overline{n})\frac{\mathrm{d}V(\overline{n})}{\mathrm{d}\overline{n}}P - D(\overline{n})\frac{\partial P}{\partial \overline{n}}\right]\end{aligned} \tag{B.9}$$

J in the above equation describes the probability flux, $\mu(\overline{n})$ is the mobility and $D(\overline{n})$ is the diffusion coefficient. $V(\overline{n})$ is an effective potential that includes both drift and noise events, defined as $\frac{\mathrm{d}V}{\mathrm{d}\overline{n}} = \frac{\mathrm{d}U}{\mathrm{d}\overline{n}} + \frac{D}{\mu}\frac{\mathrm{d}\ln(D)}{\mathrm{d}\overline{n}}$. The pseudo-potential, $V(\overline{n})$ is obtained by expanding Eq. (B.8) to second-order and comparing it with with Eq. (B.9). We identify the drift

$\left(\left\langle \frac{d\overline{n}}{dt} \right\rangle = \mu(\overline{n}) \frac{dU}{d\overline{n}}\right)$, the pseudo-potential $(V(\overline{n}))$, diffusion $(D(\overline{n}))$ and mobility $(\mu(\overline{n}))$ as follows,

$$\left\langle \frac{d\overline{n}}{dt} \right\rangle = \frac{c_r q^2}{\mathbf{N}} (2\overline{n} - 1) \left(\overline{n}(1-\overline{n}) - \frac{1}{Fq^2} \right) \tag{B.10}$$

$$V(\overline{n}) = 2\mathbf{N}\overline{n}(1-\overline{n}) + \left(1 - \frac{4\mathbf{N}}{Fq^2}\right) \log[Fq^2 \overline{n}(1-\overline{n}) + 1] \tag{B.11}$$

$$\mu(\overline{n}) = D(\overline{n}) = \frac{c_r q^2}{2\mathbf{N}^2} \left(\frac{1}{Fq^2} + \overline{n}(1-\overline{n}) \right) \tag{B.12}$$

The transition from $s = 0$ to $s = 1$ in Eq. (3.4) occurs with rate $h(\overline{n}, \overline{q}) = k_m \exp(-(V(\overline{n} = 0.5, \overline{q}) - V(\overline{n} = 1, \overline{q}))$ while the transition from $s = 1$ to $s = 0$ occurs with rate $f(\overline{n}, \overline{q}) = k_m \exp(-(V(\overline{n} = 0.5, \overline{q}) - V(\overline{n} = 0, \overline{q})))$.

B.3 Deriving an imaginary-time Schrödinger equation

In Section 3.3.3 we rewrite the master equation as an imaginary time Schrödinger equation (Eq. (3.6)), where the stochastic Hamiltonian is given by Eq. (3.7). Since we model the contacts as a birth-death process, we use the following Poisson *ansatz* [69, 205, 224],

$$|\Psi\rangle = \begin{pmatrix} c_1 \exp\left(\overline{q}_1 (a^\dagger - 1)\right) |\mathbf{0}\rangle \\ c_0 \exp\left(\overline{q}_0 (a^\dagger - 1)\right) |\mathbf{0}\rangle \end{pmatrix}, \tag{B.13}$$

$$\langle \Phi | = \begin{pmatrix} \langle \mathbf{0} | e^a \exp(\alpha_1 + \lambda_1 a) & \langle \mathbf{0} | e^a \exp(\alpha_0 + \lambda_0 a) \end{pmatrix}. \tag{B.14}$$

Furthermore, we impose $\langle \Phi(\alpha_L = 0) | \Psi(\alpha_R) \rangle = 1$. Plugging (B.14) and (B.13) into (3.8), we obtain the following set of variational equations given in Eq. (3.9).

B.4 Details of Gillespie Stochastic Simulations

Stochastic simulations were carried out for the reaction network using an implementation of the Gillespie stochastic simulation algorithm [240] in Python using standard libraries. We set the parameters as follows: $c_n = 1.0\ \tau^{-1}$, $c_r/c_n = 100.0$. We simulate a

system of size $\mathbf{N} = 40$ sites, and $\mathbf{M} = 741$ mutable, non-backbone contacts.

We ran simulations of length 3×10^5 τ using specified $k_c/c_n, \varepsilon, \lambda$ values, approximately corresponding to $\sim 10^8 - 10^{10}$ Gillespie moves. We discarded the first half of each trajectory to remove the influence of initial conditions on the simulation results. Steady-state probability distributions and contact maps were obtained by averaging over the remaining half of each simulated trajectory.

To better understand the origin of the asymmetry in the epigenetic landscape presented in Fig. 3.2 of the main text, we computed the steady-state lifetime as follows. We identified the steady states using the fraction of modified nucleosomes as $\overline{n} \in (0.8, 1.0]$ and $\overline{n} \in [0.0, 0.2)$. To compute the average lifetimes of each, we sub-divided the second half of each of the simulated trajectories into ~ 10 sub-trajectories each of length $10^4\tau$; this allows one to observe at least $\sim 10^2$ transitions between marked and unmarked states in each sub-trajectory. A successful transition from the unmarked to marked state was defined as the event wherein a system starting in the unmarked state arrives in the set defining the marked state (and vice versa). The lifetime of the unmarked (marked) state was defined to be the time between a transition to the unmarked (marked) state and transition to the marked (unmarked) state. The average lifetime was then obtained by computing the average time spent in each state in each sub-trajectory, and the error bars in Fig. 3.3c reflect the standard error of the mean across the 10 sub trajectories. A similar procedure was carried out to prepare Figs. 3.4c, 3.5c, 3.6d, 3.7a-c. We also tested the sensitivity of the results to the definition of marked (and unmarked) states used. We performed similar analysis by defining a marked (unmarked) state to be $\overline{n} \in (0.9, 1.0]$ ($\overline{n} \in [0.0, 0.1)$) and $\overline{n} \in (0.7, 1.0]$ ($\overline{n} \in [0.0, 0.3)$). Fig. B.4 demonstrates qualitatively similar trends for the average lifetimes with varying k_c using these thresholds to the ones seen in Fig. 3.3c of the main text.

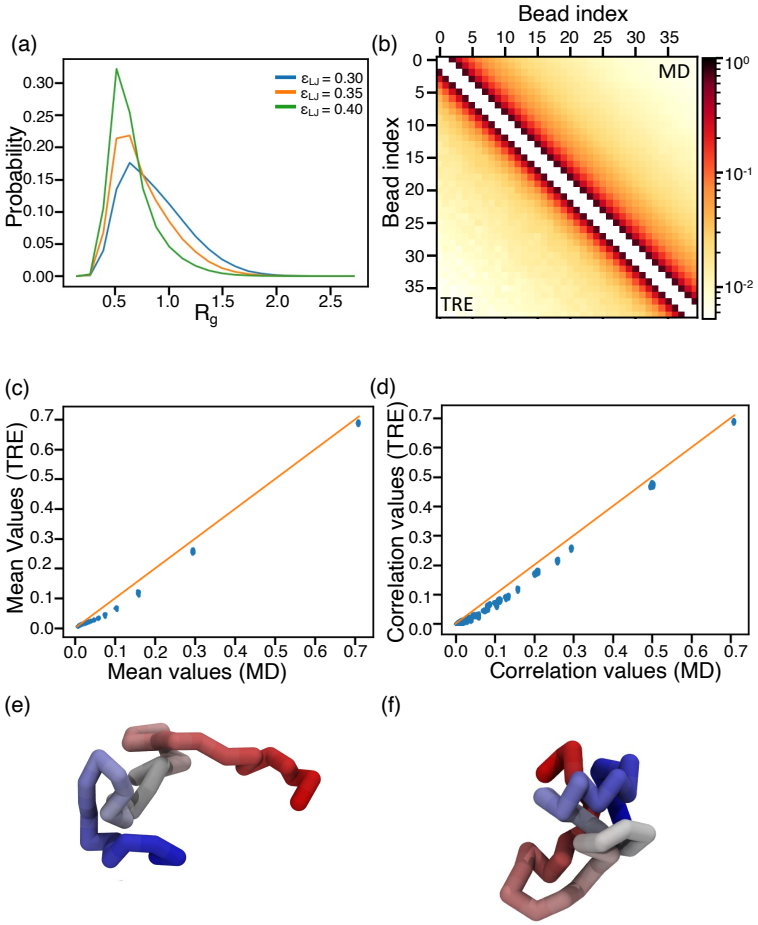

Figure B.2: Characterizing and comparing the homopolymer conformations determined via an MD simulation of the $\lambda = 0.01$ polymer model and via a temperature replica exchange (TRE) Markov chain Monte Carlo (MCMC) simulation of the Ising-like model. (a) Compactness, measured by the distribution of the homopolymer's radius of gyration (R_g), differs between models parameterized with different ε_{LJ} values in each of three MD simulations. (b,c) The mean contact probabilities associated with the Ising-like model and computed via TRE MCMC (lower triangle, y-axis) agree reasonably well with the mean contact probabilities computed from MD simulation-derived polymer conformations (upper triangle, x-axis). (d) The correlation between each contact pair, defined as $\langle q_i q_j \rangle$, agrees well in both the TRE MCMC-simulated Ising-like model (y-axis) and MD-simulated polymer model (x-axis). Open (e) and collapsed (f) polymer conformations are shown, having R_g values of ≈ 2.01 and 0.22, respectively. Each conformation was computed via MD simulation of the homopolymer model that uses Lennard-Jones interaction potentials with magnitude $\varepsilon_{LJ} = 0.35$.

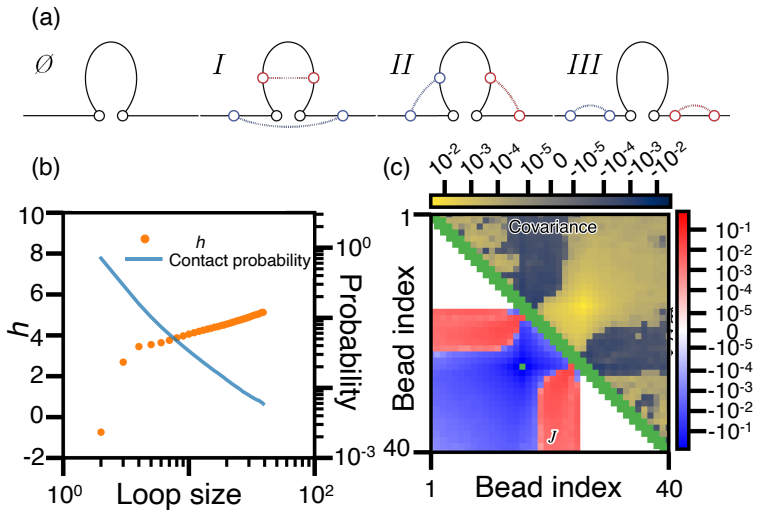

Figure B.3: The Ising-like model parameters are compared to the physical effects they capture. The MD simulation of the homopolymer using $\varepsilon = 0.35$ provided the included contact statistics. (a) Illustration of various topological relationships between loops. $\emptyset$ illustrates the loop associated with one contact. This loop resides within or fully contains loops in region I, partially overlaps with loops in region II, and is independent of loops in region III. (b) The contact probabilities, $\langle q_{ij} \rangle$, and entropic penalties, h_{ij}, are plotted against their associated loop size, $j - i$.For visual clarity, the plot displays the average value of all $\langle q_{ij} \rangle$ corresponding to each loop size. However, as shown in the upper triangle of Fig. B.2 (b), they are similar to the plotted probability whose contact has the same loop size. Meanwhile, $h_{ij}\ \forall\ (i,j)$ are plotted against their associated loop size. (c) The coupling (lower triangle) and correlation (upper triangle) between contact $(16,26)$ and all other contacts (k,l) are plotted on a two-dimensional grid. $J_{16,26,kl}$ quantifies coupling, where k and l index the x- and y-axis, respectively. Mirroring each interaction across the diagonal, the covariance $\langle q_{16,26} q_{kl} \rangle - \langle q_{16,26} \rangle \langle q_{kl} \rangle$ quantifies the correlation between the formation of the relevant contacts, where l and k index the x- and y-axis, respectively. Green indicates values that are undefined in the Ising-like model.

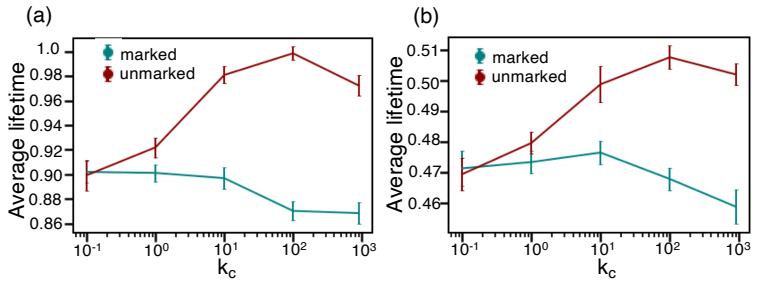

Figure B.4: Steady state lifetime estimation is robust with respect to the state definition. Qualitatively similar trends to Fig. 3.3c of the main text are observed for marked (unmarked) state defined as (a) $\langle n \rangle \in (0.9, 1.0]$ ($\langle n \rangle \in [0.0, 0.1)$) and (b) $\langle n \rangle \in (0.7, 1.0]$ ($\langle n \rangle \in [0.0, 0.3)$).

Appendix C

Supporting Information for Chapter 4

C.1 Derivation of trial function for the epigenetic switch

The connection between multinomial probability distributions and SU(n) algebras has been studied from an algebraic perspective, and we refer interested readers to the literature [281, 282]. Here we present an alternate derivation that is perhaps simpler and more intuitive. Keeping with standard literature conventions, we start with the trial function in the Poisson ansatz [291] for a two particle system,

$$|\Psi(t)\rangle = \sum_{\{n_x,n_y\}} P(\{n_x,n_y\};t)\left|n_x,n_y\right\rangle. \tag{C.1}$$

with $\left|n_x,n_y\right\rangle = (a_x^\dagger)^{n_x}(a_y^\dagger)^{n_y}|0,0\rangle$, and $P(\{n_x,n_y\};t)$ being the product of two independent Poisson distributions in the Poisson ansatz,

$$P(\{n_x,n_y\};t) = \frac{e^{-x}x^{n_x}}{n_x!}\frac{e^{-y}y^{n_y}}{n_y!}. \tag{C.2}$$

Here x,y are parameters of the Poisson distribution, and we can treat them as time dependent variational parameters. One observes that in the Poisson ansatz, $|\Psi\rangle = |\psi_x\rangle \otimes \left|\psi_y\right\rangle$, where one frequently encounters the abbreviated form for $|\psi_i\rangle = \exp\Big(\mu_i(a_i^\dagger - 1)\Big)|0_i\rangle$. In order to proceed, we impose $P(\{n_x,n_y\};t) = \frac{e^{-x}x^{nx}}{n_x!}\frac{e^{-y}y^{ny}}{n_y!}\delta_{n_x+n_y,N}$, and normalized by $\sum \frac{e^{-x}x^{nx}}{n_x!}\frac{e^{-y}y^{ny}}{n_y!}\delta_{n_x+n_y,N} = 1$. We then note $x+y=N$, $n_x \equiv n$, and $n_y = N-n_x$. Introducing

a new parameter θ such that $x = N\theta$, and $y = N(1-\theta)$, one arrives at, $P(\{\theta, N\}; t) = \binom{N}{n_x}\theta^{n_x}(1-\theta)^{N-n_x}$, which is the Binomial probability distribution. Thus, the variational ansatz for two-particle system with a constraint $N = n_x + n_y$ is reduced to,

$$|\Psi(t)\rangle = \sum_n \mathrm{Bin}(\theta, N; t)\,|n\rangle\,, \tag{C.3}$$

where θ is the appropriate time-dependent variational parameter. Any arbitrary state is now, $|n_x, N-n_x\rangle = (a_x^\dagger)^{n_x}(a_y^\dagger)^{N-n_x}|0,0\rangle \equiv \frac{J_+^n}{N^{\underline{n}}}|0\rangle$, where $x^{\underline{n}} = x(x-1)(x-2)\ldots(x-n+1)$ denotes the falling factorial. This allows us to write the binomial varitional ansatz more succinctly as,

$$|\Psi\rangle = (1-\theta)^N \exp\left(\frac{\theta}{1-\theta}J_+\right)|0\rangle\,. \tag{C.4}$$

The action of the operator J_+ on a ket, and its algebraic properties are discussed in greater detail in the literature [69].

Appendix D

Supporting Information for Chapter 5

D.1 Interrogating the Effect of Parameters u_a, l on the Single Component System

We demonstrate that the qualitative trends discussed in the main text remain robust despite variations in sticker strength (u_a) and patterning ($l = N/f$). In Fig. D.1 (a) and (b), we qualitatively replicate the increase in critical temperature and concentration highlighted in the main text for $u_a = 2.5$ and $u_a = 10$, respectively. Similarly, in Fig. D.1 (c) and (d), we observe a similar qualitative dependence of the degree of conversion on $\Delta\varepsilon$ for $u_a = 2.5$ and $u_a = 10$, respectively. Finally, Fig. D.1 (e) and (f) demonstrate, akin to the observations in the main text, an increase in the concentration of chains with all stickers free, $\mathscr{C}_{\text{free}}$, up to a maximum value in the pre-gel regime but a monotonic decrease in the post-gel regime for $\Delta\varepsilon = 0$. $\mathscr{C}_{\text{free}}$ exhibits non-monotonic behavior in the post-gel regime for $\Delta\varepsilon \neq 0$ for both $u_a = 2.5$ and $u_a = 10$, respectively.

The parameter l can also modulate the amount of specific interactions in the system, and our qualitative observations and conclusions in the main text remain independent of the specific choice of l. In Fig. D.2 (a) and (b), we qualitatively reproduce the increase in critical temperature and concentration outlined in the main text for $l = 20$ and $l = 5$, respectively. Similarly, in Fig. D.2 (c) and (d), we note a qualitatively similar dependence of the degree of conversion on $\Delta\varepsilon$ for $l = 20$ and $l = 5$, respectively. Finally, Fig. D.2 (e)

and (f) demonstrate, similar to the main text, an increase in the concentration of chains with all stickers free up to a maximum value in the pre-gel regime. However, this concentration monotonically decreases in the post-gel regime for $\Delta\varepsilon = 0$, while displaying non-monotonic behavior in the post-gel regime for $\Delta\varepsilon \neq 0$ for both $l = 20$ and $l = 5$, respectively.

At fixed T, u_a, an initial increase $\Delta\varepsilon$ promotes macro-phase separation initially and sticker-sticker cross-links can form more readily in the dense phase, leading to the initial increase in degree of conversion (Fig. 3, Fig.S1c-d and Fig. D.3c). However, as we further increase $\Delta\varepsilon$ the non-specific interaction compete with and overwhelm the specific interaction, explaining the decrease in degree of conversion. For convenience, this cross over point is denoted $(p^*, \Delta\varepsilon^*)$. The strength of sticker-sticker cross-links (u_a) plays an important role in determining the behavior of the degree of conversion (p) as a function of $\Delta\varepsilon$.

First, at small u_a ($u_a = 1.25$ and $u_a = 2.5$) and low $\Delta\varepsilon$ ($\Delta\varepsilon \lesssim 0.7$), $T = 1$ is below the critical temperature and therefore the degree of conversion is not defined. This explains the lack of data near the origin in the blue and orange curves in Fig. D.3c. At larger values of u_a ($u_a = 5.0$ and $u_a = 10.0$), the initial macro-phase separation is easier, and therefore the cross-over point $\Delta\varepsilon^*$ occurs at very slightly lower values (Fig. D.3a). Second, p^* increases with increasing u_a as expected (Fig. D.3b) Third, indeed at higher u_a it is harder for non-specific interactions to overwhelm the specific ones. This behavior can be understood not by looking at the cross-over point $(p^*, \Delta\varepsilon^*)$ but rather at how these curves decay to zero. In Fig. D.3c we see that at low u_a, the curves decay to zero more readily than at high u_a.

D.2 Interrogating the Effect of $\Delta\varepsilon_{ab}, \Delta\varepsilon_{bb}$ on the Phase Behavior of the Two Component System

In this section, we discuss the changes in the phase behavior of the two-component system as we modulate $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb}$ in both the low-sticker strength ($u_{ab} = 2$) and high-sticker strength ($u_{ab} = 5$) regimes. Since computing binodals is often numerically challenging, we plot the spinodals to efficiently scan the parameter space and gain qualitative insights (see

Fig. D.4 and Fig. D.5). In Fig. D.4a-e, we observe the system's transition from the demixing regime to the co-condensation regimes, depicting phase behavior at intermediate values of $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb}$ in the regime where sticker-sticker interaction strengths are relatively weak ($u_{ab} = 2$). Conversely, in Fig. D.5a-e, we observe the system's transition from the demixing regime to the co-condensation regimes, exhibiting phase behavior at intermediate values of $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb}$ in the regime where sticker-sticker interaction strengths are relatively strong ($u_{ab} = 5$).

We observe the U-shaped region at the center of the phase diagram in Fig. D.5a, primarily stabilized by $A-B$ cross-links, which also emerges in the weak sticker regime, albeit at $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb} \neq 0$ (Fig. D.4b). Notably, this U-shaped region, characterized by a narrow composition range for the A-B mixtures, appears only at intermediate values of $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb}$. As concluded in the main text, higher values of $\Delta\varepsilon_{ab} = \Delta\varepsilon_{bb}$–indicating stronger spacer interactions–eliminate this region, resulting in condensates with poorly defined compositions.

D.3 Interrogating the effects of N, l on the Phase Behavior of the Two Component System

In Fig. D.6, we qualitatively reproduce phase diagrams similar to those discussed in Fig. 4 of the main text, but now for longer polymer chains, specifically $N_a = N_b = 100$. All other parameters are held constant: $u_{ab} = 5, l = 2, z = 6, \bar{\varepsilon_{aa}} = 1$.

Additionally, we observe that the parameter $l = N_x/f_x$ can significantly impact the phase behavior as it relates to the relative abundance of stickers. Increasing l from $l = 2$ to $l = 10$, for a bead system with $N_a = N_b = 100$, is equivalent to reducing the number of stickers by a fifth. Consequently, maintaining the interaction strength at $u_{ab} = 5$ while setting $l = 10$ places us within the weak sticker regime previously discussed in the main text (see Fig. D.7a-b). However, we can still qualitatively reproduce the phase behavior at strong sticker interactions by simultaneously increasing the sticker-sticker interaction strength to $u_{ab} = 10$ (Fig. D.7c-d). This adjustment compensates for the reduction in sticker abundance

caused by the increase in l, allowing us to maintain similar qualitative phase characteristics.

D.4 Details of Numerical Computations

D.4.1 Computing spinodal, binodal and critical points for the single component system

The spinodal is given by the nullcline $\partial_\varphi \mu = 0$. The critical point is then the intersection of the nullclines $\partial_\varphi \mu = 0$ and $\partial^2_\varphi \mu = 0$ [298]. φ_{dense} was estimated $\Pi \approx 0$ [301]. This was used as an initial guess in the root finding algorithm. Initialisation for φ_{dilute} values for the root-finding algorithm were obtained using a line search in the interval $0 \leq \varphi_{\text{dilute}} < \varphi_{\text{dense}}$. Using these initialisations, the binodal was computed by equating the chemical potentials in the dilute and dense phases using a root-finding algorithm. We used *Mathematica* (Version 13.2) and the `FindRoot` function therein to execute these computations.

D.4.2 Computing spinodal and binodal for the two component system

For the two component system, the spinodal can be obtained from the nullcline of the determinant of the hessian matrix of $\mathscr{F}$, namely, $\det(\mathscr{H}_{\mathscr{F}}) = 0$. We determine the binodal by solving the following system of equations

$$\begin{aligned}
\mu_a^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \mu_a^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right), \\
\mu_b^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \mu_b^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right), \\
\Pi^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) &= \Pi^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right).
\end{aligned} \tag{D.1}$$

where, $\mu_k = \partial_{\varphi_k}\mathscr{F}$ and $\Pi = \sum_k \varphi_k(\partial_{\varphi_k}\mathscr{F}) - \mathscr{F}$. These equations ensure that the chemical potentials and the osmotic pressure are identical between coexisting phases. We solve the above equations by numerically finding roots to the the function $\mathscr{B}(\{\varphi_x^k\}) \equiv (\mu_a^{(0)} - \mu_a^{(1)}, \mu_b^{(0)} - \mu_b^{(1)}, \Pi^{(0)} - \Pi^{(1)}, 0) = 0$ using the *Scipy* library (Version 1.10.0) `scipy.optimize.root` function using the default implementation of Powell's hybrid method [348–350]. In addition, analytic derivatives and Jacobian's computed using the python *Sympy* library (Version

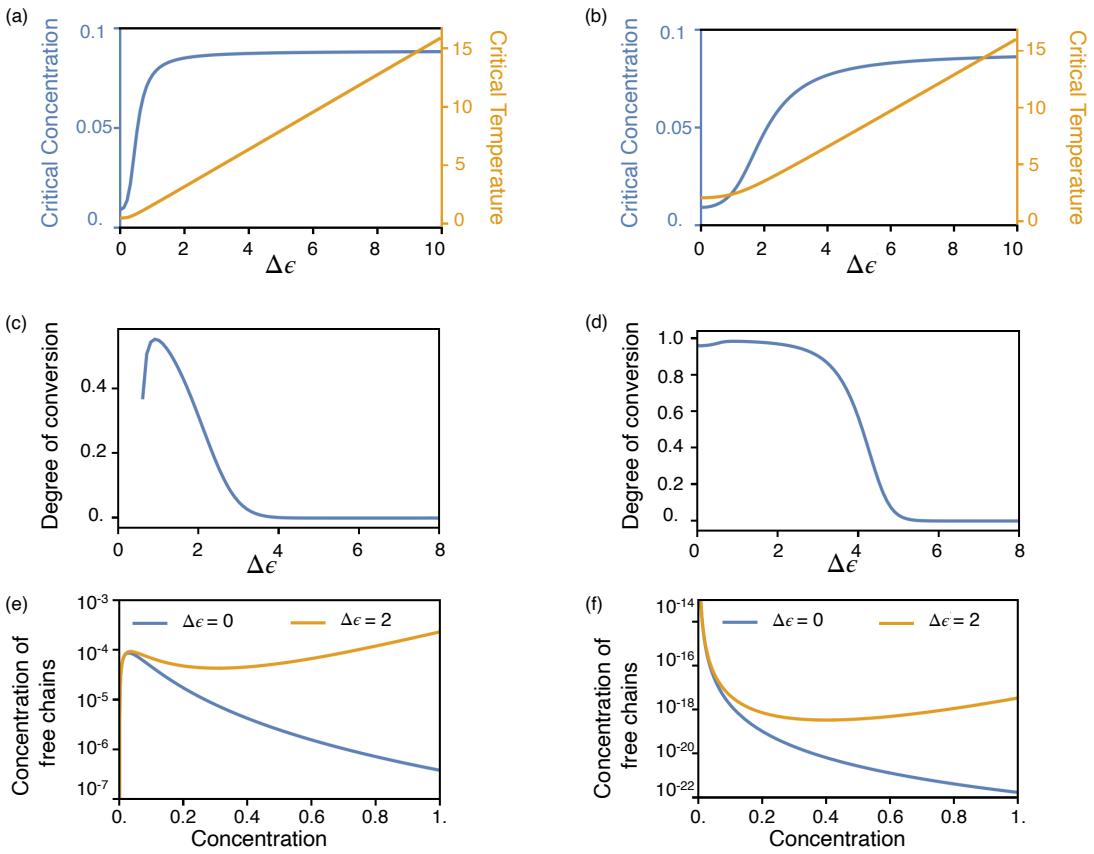

Figure D.1: The results discussed in the main text for the one component system can be recapitulated for $u_a = 2.5$ and $u_a = 10$. We plot how the critical point shifts as we vary $\Delta\varepsilon$ for (a) $u_a = 2.5$ and (b) $u_a = 10$. We also plot the degree of conversion evaluated at φ_{dense} (concentration of polymers in the dense phase in the gel regime) as a function $\Delta\varepsilon$ for (c) $u_a = 2.5$ and (d) $u_a = 10$ respectively. The concentration of free chains (with all stickers free) increases with concentration in the pre-gel regime and admits a maximum at the gel-point. The concentration of free chains is monotonically decreasing in the post-gel regime when $\Delta\varepsilon = 0$ but exhibits non-monotonic behaviour for $\Delta\varepsilon \neq 0$ for both (e) $u_a = 2.5$ and (f) $u_a = 10$ respectively. We set $\bar{\varepsilon} = 0$, $l = 10$, $N = 100$, and $z = 6$. Additionaly, for (c)-(f) we held fixed $T = 1$. See text *Section: Interrogating the effect of parameters u_a, l on the single component system* for more details.

1.11.1) and *Numpy* (Version 1.24.1) library were utilised for other incidental numerical manipulations. Initial values for $\varphi_x^{(k)}$ for $x \in a, b$ provided to the root-finding algorithm, were chosen on a fine-grid in the triangle bounded $\varphi_a + \varphi_b \leq 1$ and $\varphi_a \geq 0$ and $\varphi_b \geq 0$. In general, precisely computing binodals can be numerically challenging, and we direct interested readers to sophisticated numerical approaches in the literature [351–357].

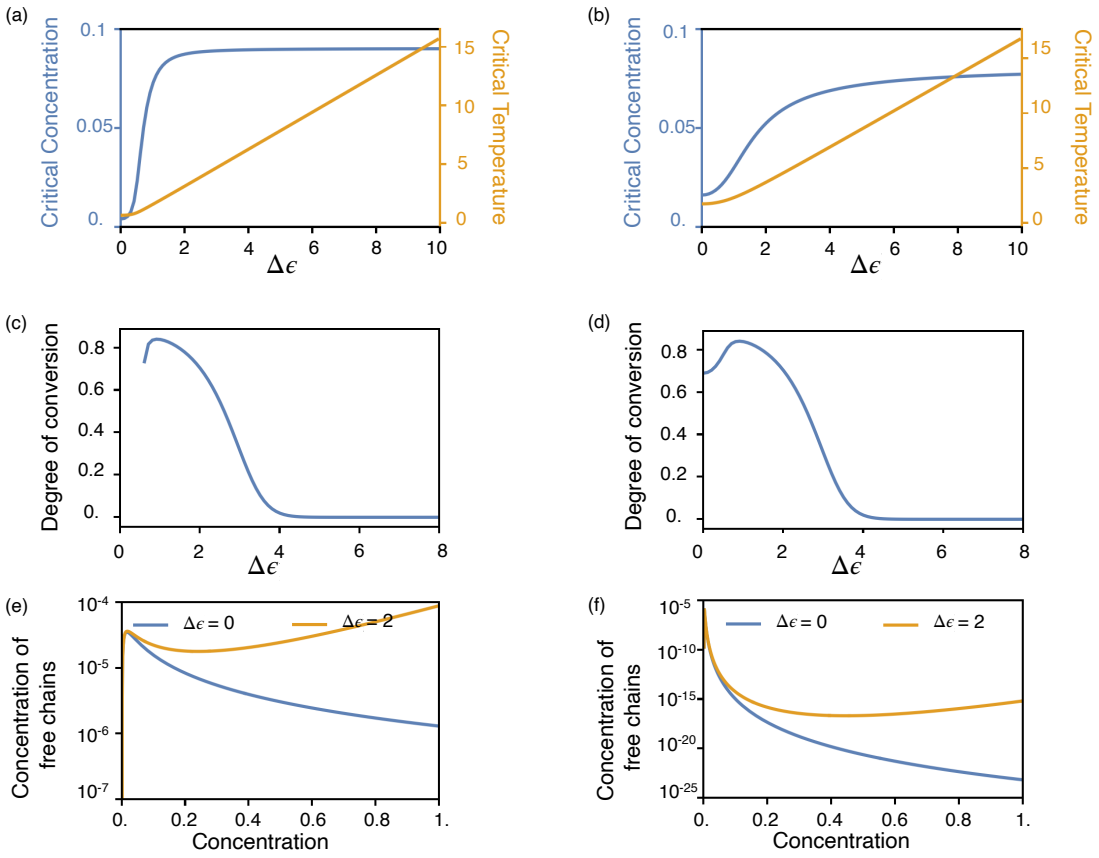

Figure D.2: The results discussed in the main text for the one component system can be recapitulated for $l = 20$ and $l = 5$. We plot how the critical point shifts as we vary $\Delta\varepsilon$ for (a) $l = 20$ and (b) $l = 5$. We also plot the degree of conversion evaluated at φ_{dense} (concentration of polymers in the dense phase in the gel regime) as a function of $\Delta\varepsilon$ for (c) $l = 20$ and (d) $l = 5$ respectively. The concentration of free chains (with all stickers free) increases with concentration in the pre-gel regime and admits a maximum at the gel-point. The concentration of free chains is monotonically decreasing in the post-gel regime when $\Delta\varepsilon = 0$ but exhibits non-monotonic behaviour for $\Delta\varepsilon \neq 0$ for both (e) $l = 20$ and (f) $l = 5$ respectively. We set $\bar{\varepsilon} = 0$, $u_a = 5$, $N = 100$, and $z = 6$. Additionally, for (c)-(f) we held fixed $T = 1$. See text *Section: Interrogating the effect of parameters u_a, l on the single component system* for more details.

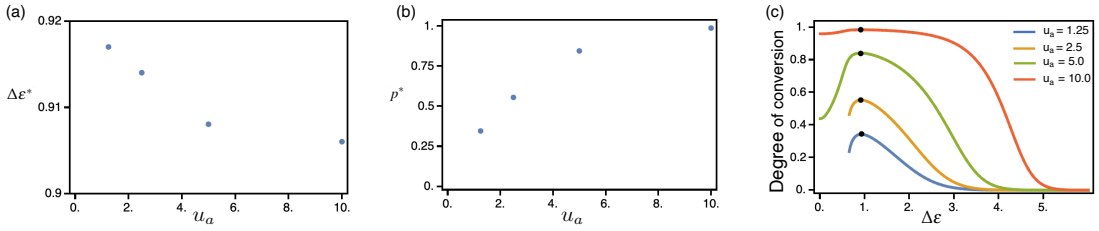

Figure D.3: The effect of sticker interaction strength (u_a) on degree of conversion (p) vs $\Delta\varepsilon$ curves. (a) The cross-over point ($\Delta\varepsilon^*$) decreases very slightly as with an increase in sticker-sticker interaction strength. (b) The maximum degree of cross-linking (p^*) achieved at the cross-over point increases with a corresponding increase in sticker-sticker interaction strength. (c) The degree of conversion as a function of $\Delta\varepsilon$ decays to zero faster when sticker-sticker interactions strength is low than when it is high. The cross-over point $(p^*,\ \Delta\varepsilon^*)$ is indicated with a black dot. We set $\bar{\varepsilon} = 0, l = 10, N = 100,\ T = 1$ and $z = 6$. See text *Section: Interrogating the effect of parameters u_a, l on the single component system* for more details.

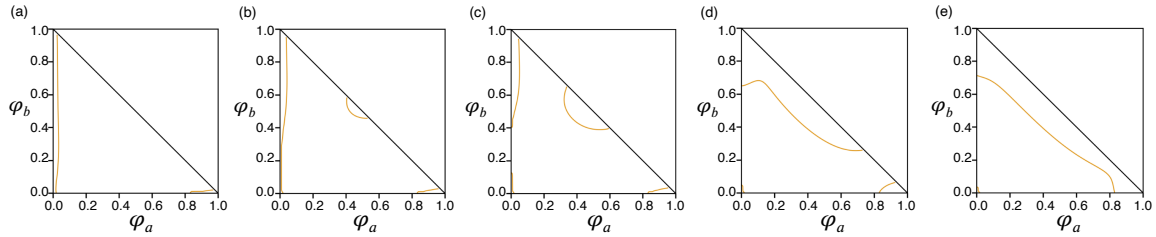

Figure D.4: Phase diagrams showing the spinodal line (orange, solid) for a two component system with with $u_{ab} = 2$ (a) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, (b) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0.7$, (c) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0.8$, (d) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.0$ and (e) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.1$. We set $N_a = N_b = 10, l = 2, z = 6, \bar{\varepsilon}_{aa} = 1, \Delta\varepsilon_{aa} = 0$, and $\bar{\varepsilon}_{bb} = \bar{\varepsilon}_{ab} = 0$ in all systems. See text *Section: Interrogating the effects of $\Delta\varepsilon_{ab}, \Delta\varepsilon_{bb}$ on the Phase Behavior of the Two Component System* for more details.

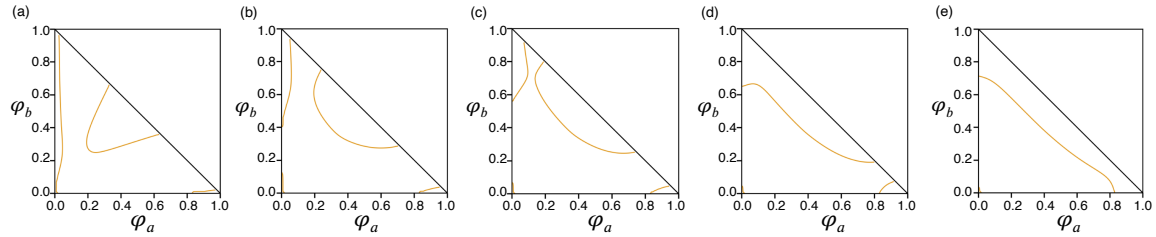

Figure D.5: Phase diagrams showing the spinodal line (orange, solid) for a two component system with $u_{ab} = 5$ (a) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, (b) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0.8$, (c) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0.9$, (d) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.0$ and (e) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.1$. We set $N_a = N_b = 10, l = 2, z = 6, \bar{\varepsilon}_{aa} = 1, \Delta\varepsilon_{aa} = 0$, and $\bar{\varepsilon}_{bb} = \bar{\varepsilon}_{ab} = 0$ in all systems. See text *Section: Interrogating the effects of $\Delta\varepsilon_{ab}, \Delta\varepsilon_{bb}$ on the Phase Behavior of the Two Component System* for more details.

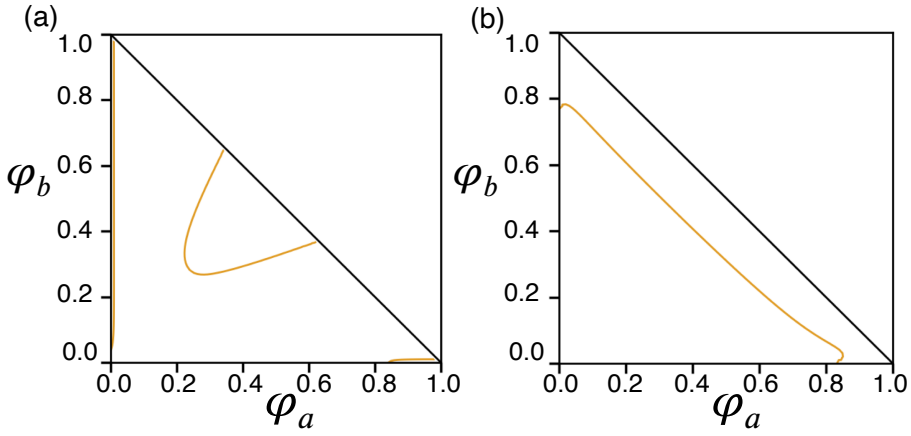

Figure D.6: Phase diagrams showing the spinodal line (orange, solid) for a two component system with $u_{ab} = 5$ for a larger $N_a = N_b = 100$ system for (a) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, (b) $\Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.1$. Here we qualitatively recover the same phase behavior seen in the main text prior. We set $l = 2, z = 6, \bar{\varepsilon}_{aa} = 1, \Delta\varepsilon_{aa} = 0$, and $\bar{\varepsilon}_{bb} = \bar{\varepsilon}_{ab} = 0$ in all systems. See text *Section: Interrogating the effects of N, l on the Phase Behavior of the Two Component System* for more details.

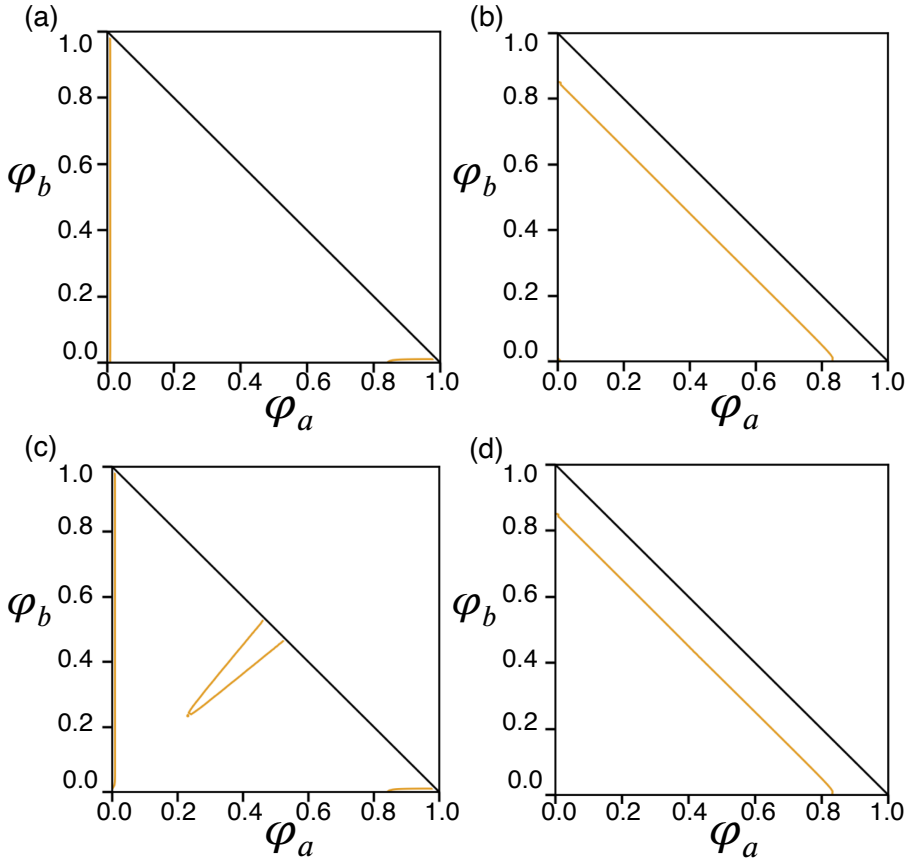

Figure D.7: Phase diagrams showing the spinodal line (orange, solid) for a two component system with (a) $u_{ab} = 5, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, (b) $u_{ab} = 5, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.5$, (c) $u_{ab} = 10, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 0$, and (d) $u_{ab} = 10, \Delta\varepsilon_{bb} = \Delta\varepsilon_{ab} = 1.5$. We set $N_a = N_b = 100, l = 10, z = 6, \bar{\varepsilon}_{aa} = 1, \Delta\varepsilon_{aa} = 0$, and $\bar{\varepsilon}_{bb} = \bar{\varepsilon}_{ab} = 0$ in all systems. See text *Section: Interrogating the effects of N, l on the Phase Behavior of the Two Component System* for more details.

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell* (W.W. Norton & Company, Aug. 2017).

[2] B. R. Sabari, A. Dall'Agnese, and R. A. Young, "Biomolecular condensates in the nucleus", Trends Biochem. Sci. **45**, 961 (2020).

[3] R. V. Pappu, S. R. Cohen, F. Dar, M. Farag, and M. Kar, "Phase transitions of associative biomacromolecules", Chem. Rev. **123**, 8945 (2023).

[4] J.-M. Choi, A. S. Holehouse, and R. V. Pappu, "Physical principles underlying the complex biology of intracellular phase transitions", Annu. Rev. Biophys. **49**, 107 (2020).

[5] K. L. Saar, D. Qian, L. L. Good, A. S. Morgunov, R. Collepardo-Guevara, R. B. Best, and T. P. J. Knowles, "Theoretical and data-driven approaches for biomolecular condensates", Chem. Rev. **123**, 8988 (2023).

[6] H. R. Kilgore and R. A. Young, "Learning the chemical grammar of biomolecular condensates", Nat. Chem. Biol. **18**, 1298 (2022).

[7] K. L. Saar, A. S. Morgunov, R. Qi, W. E. Arter, G. Krainer, A. A. Lee, and T. P. Knowles, "Learning the molecular grammar of protein condensates from sequence determinants and embeddings", Proc. Natl. Acad. Sci. U. S. A. **118**, e2019053118 (2021).

[8] B. S. Schuster, R. M. Regy, E. M. Dolan, A. Kanchi Ranganath, N. Jovic, S. D. Khare, Z. Shi, and J. Mittal, "Biomolecular Condensates: Sequence Determinants of Phase Separation, Microstructural Organization, Enzymatic Activity, and Material Properties", J. Phys. Chem. B **125**, 3441 (2021).

[9] J. Wang, J. M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, and A. A. Hyman, "A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins", Cell **174**, 688 (2018).

[10] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, and M. K. Rosen, "Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein", Mol. Cell **63**, 72 (2016).

[11] D. W. Sanders, N. Kedersha, D. S. Lee, A. R. Strom, V. Drake, J. A. Riback, D. Bracha, J. M. Eeftens, A. Iwanicki, A. Wang, M. T. Wei, G. Whitney, S. M. Lyons, P. Anderson, W. M. Jacobs, P. Ivanov, and C. P. Brangwynne, "Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization", Cell **181**, 306 (2020).

[12] T. Kaur, M. Raju, I. Alshareedah, R. B. Davis, D. A. Potoyan, and P. R. Banerjee, "Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies", Nat. Commun. **12**, 872 (2021).

[13] S. Liu, A. Athreya, Z. Lao, and B. Zhang, "From nucleosomes to compartments: physicochemical interactions underlying chromatin organization", Annu. Rev. Biophys. **53**, 10.1146/annurev-biophys-030822-032650 (2024).

[14] L. D. Gallego, M. Schneider, C. Mittal, A. Romanauska, R. M. Gudino Carrillo, T. Schubert, B. F. Pugh, and A. Köhler, "Phase separation directs ubiquitination of gene-body nucleosomes", Nature **579**, 592 (2020).

[15] S. Redding, "Dynamic asymmetry and why chromatin defies simple physical definitions", Curr. Opin. Cell Biol. **70**, 116 (2021).

[16] R. Cortini, M. Barbi, B. R. Care, C. Lavelle, A. Lesne, J. Mozziconacci, and J.-M. Victor, "The physics of epigenetics", Rev. Mod. Phys. **88**, 025002 (2016).

[17] A. Z. Abdulla, H. Salari, M. M. C. Tortora, C. Vaillant, and D. Jost, "4D epigenomics: deciphering the coupling between genome folding and epigenomic regulation with biophysical modeling", Curr. Opin. Genet. Dev. **79**, 102033 (2023).

[18] M.-H. Han, D. Issagulova, and M. Park, "Interplay between epigenome and 3D chromatin structure", BMB Reports **56**, 633 (2023).

[19] L. Wang, Y. Gao, X. Zheng, C. Liu, S. Dong, R. Li, G. Zhang, Y. Wei, H. Qu, Y. Li, C. D. Allis, G. Li, H. Li, and P. Li, "Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism", Mol. Cell **76**, 646 (2019).

[20] A. Klosin and A. A. Hyman, "A liquid reservoir for silent chromatin", Nature **547**, 168 (2017).

[21] T. Kouzarides, "Chromatin modifications and their function", Cell **128**, 693 (2007).

[22] A. D. Goldberg, C. D. Allis, and E. Bernstein, "Epigenetics: a landscape takes shape", Cell **128**, 635 (2007).

[23] C. A. Musselman, M.-E. Lalonde, J. Côté, and T. G. Kutateladze, "Perceiving the epigenetic landscape through histone readers", Nature Structural & Molecular Biology **19**, 1218 (2012).

[24] B. Xhemalce, M. A. Dawson, and A. J. Bannister, "Histone modifications", in *Reviews in cell biology and molecular medicine* (John Wiley & Sons, Ltd, 2011).

[25] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications", Cell Res. **21**, 381 (2011).

[26] X. Ding, X. Lin, and B. Zhang, "Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface", Nat. Commun. **12**, 1091 (2021).

[27] A. P. Latham and B. Zhang, "Consistent Force Field Captures Homologue-Resolved HP1 Phase Separation", J. Chem. Theory Comput. **17**, 3134 (2021).

[28] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin", Nature **547**, 236 (2017).

[29] M. M. Keenen, D. Brown, L. D. Brennan, R. Renger, H. Khoo, C. R. Carlson, B. Huang, S. W. Grill, G. J. Narlikar, and S. Redding, "HP1 proteins compact DNA into mechanically and positionally stable phase separated domains", eLife **10**, e64563 (2021).

[30] R. Tatavosian, S. Kent, K. Brown, T. Yao, H. N. Duc, T. N. Huynh, C. Y. Zhen, B. Ma, H. Wang, and X. Ren, "Nuclear condensates of the polycomb protein chromobox 2 (cbx2) assemble through phase separation", J. Biol. Chem. **294**, 1451 (2019).

[31] N. J. Francis, R. E. Kingston, and C. L. Woodcock, "Chromatin compaction by a polycomb group protein complex", Science **306**, 1574 (2004).

[32] S. Mujtaba, L. Zeng, and M.-M. Zhou, "Structure and acetyl-lysine recognition of the bromodomain", Oncogene **26**, 5521 (2007).

[33] "Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes", Cell **111**, 369 (2002).

[34] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, "Phase separation drives heterochromatin domain formation", Nature **547**, 241 (2017).

[35] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin", Nature **547**, 236 (2017).

[36] B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, et al., "Coactivator condensation at super-enhancers links phase separation and gene control", Science **361**, eaar3958 (2018).

[37] A. J. Plys, C. P. Davis, J. Kim, G. Rizki, M. M. Keenen, S. K. Marr, and R. E. Kingston, "Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2", Genes Dev. **33**, 799 (2019).

[38] J. M. Eeftens, M. Kapoor, D. Michieletto, and C. P. Brangwynne, "Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction", Nat. Commun. **12**, 5888 (2021).

[39] S. I. S. Grewal and D. Moazed, "Heterochromatin and epigenetic control of gene expression", Science **301**, 798 (2003).

[40] R. H. Jacobson, A. G. Ladurner, D. S. King, and R. Tjian, "Structure and Function of a Human TAFII250 Double Bromodomain Module", Science **288**, 1422 (2000).

[41] D. J. Owen, P. Ornaghi, J.-C. Yang, N. Lowe, P. R. Evans, P. Ballario, D. Neuhaus, P. Filetici, and A. A. Travers, "The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p", EMBO J. **19**, 6141 (2000).

[42] G. Schotta, A. Ebert, V. Krauss, A. Fischer, J. Hoffmann, S. Rea, T. Jenuwein, R. Dorn, and G. Reuter, "Central role of Drosophila SU(VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing", EMBO J. **21**, 1121 (2002).

[43] K. Newar, A. Z. Abdulla, H. Salari, E. Fanchon, and D. Jost, "Dynamical modeling of the H3K27 epigenetic landscape in mouse embryonic stem cells", PLOS Comput. Biol. **18**, 1 (2022).

[44] T. Schlick, J. Hayes, and S. Grigoryev, "Toward Convergence of Experimental Studies and Theoretical Modeling of the Chromatin Fiber", J. Biol. Chem. **287**, 5183 (2012).

[45] T. Parsons and B. Zhang, "Critical role of histone tail entropy in nucleosome unwinding", J. Chem. Phys. **150**, 185103 (2019).

[46] Y. Qi and B. Zhang, "Predicting three-dimensional genome organization with chromatin states", PLOS Comput. Biol. **15**, e1007024 (2019).

[47] Z. Jiang and B. Zhang, "Theory of active chromatin remodeling", Phys. Rev. Lett. **123**, 208102 (2019).

[48] W. J. Xie and B. Zhang, "Learning the formation mechanism of domain-level chromatin states with epigenomics data", Biophys. J. **116**, 2047 (2019).

[49] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker, "Human dna methylomes at base resolution show widespread epigenomic differences", Nature **462**, 315 (2009).

[50] R. Lu, F. Markowetz, R. D. Unwin, J. T. Leek, E. M. Airoldi, B. D. MacArthur, A. Lachmann, R. Rozov, A. Ma'ayan, L. A. Boyer, O. G. Troyanskaya, A. D. Whetton, and I. R. Lemischka, "Systems-level dynamic analyses of fate change in murine embryonic stem cells", Nature **462**, 358 (2009).

[51] R. Margueron and D. Reinberg, "Chromatin structure and the inheritance of epigenetic information", Nat. Rev. Genet. **11**, 285 (2010).

[52] D. Moazed, "Mechanisms for the inheritance of chromatin states", Cell **146**, 510 (2011).

[53] R. Krishnakumar and W. L. Kraus, "Parp-1 regulates chromatin structure and transcription through a kdm5b-dependent pathway", Mol. Cell **39**, 736 (2010).

[54] T. Jenuwein and C. D. Allis, "Translating the histone code", Science **293**, 1074 (2001).

[55] C. D. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control", Nat. Rev. Genet. **17**, 487 (2016).

[56] F. Nicol-Benoît, P. Le-Goff, Y. Le-Dréan, F. Demay, F. Pakdel, G. Flouriot, and D. Michel, "Epigenetic memories: structural marks or active circuits?", Cell. Mol. Life Sci. **69**, 2189 (2012).

[57] C. Pieterman, E. Conemans, K. Dreijerink, J. Laat, M. Timmers, M. Vriens, and G. Valk, "Thoracic and duodenopancreatic neuroendocrine tumors in multiple endocrine neoplasia type 1: natural history and function of menin in tumorigenesis", Endocr.-Relat. Cancer **21**, 121 (2014).

[58] J. Padeken, S. P. Methot, and S. M. Gasser, "Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance", Nat. Rev. Mol. Cell Biol **23**, 623 (2022).

[59] R. Villaseñor and T. Baubec, "Regulatory mechanisms governing chromatin organization and function", Curr. Opin. Cell Biol. **70**, 10 (2021).

[60] I. B. Dodd, M. A. Micheelsen, K. Sneppen, and G. Thon, "Theoretical analysis of epigenetic cell memory by nucleosome modification", Cell **129**, 813 (2007).

[61] H. Zhang, X.-J. Tian, A. Mukhopadhyay, K. S. Kim, and J. Xing, "Statistical mechanics model for the dynamics of collective epigenetic histone modification", Phys. Rev. Lett. **112**, 068101 (2014).

[62] M. Ancona, D. Michieletto, and D. Marenduzzo, "Competition between local erasure and long-range spreading of a single biochemical mark leads to epigenetic bistability", Phys. Rev. E **101**, 042408 (2020).

[63] D. Jost, "Bifurcation in epigenetics: implications in development, proliferation, and diseases", Phys. Rev. E **89**, 010701 (2014).

[64] M. Sedighi and A. M. Sengupta, "Epigenetic chromatin silencing: bistability and front propagation", Phys. Biol. **4**, 246 (2007).

[65] A. Dayarian and A. M. Sengupta, "Titration and hysteresis in epigenetic chromatin silencing", Phys. Biol. **10**, 036005 (2013).

[66] K. Sneppen and N. Mitarai, "Multistability with a metastable mixed state", Phys. Rev. Lett. **109**, 100602 (2012).

[67] M. A. Micheelsen, N. Mitarai, K. Sneppen, and I. B. Dodd, "Theory for the stability and regulation of epigenetic landscapes", Phys. Biol. **7**, 026010 (2010).

[68] C. Alabert, C. Loos, M. Voelker-Albert, S. Graziano, I. Forné, N. Reveron-Gomez, L. Schuh, J. Hasenauer, C. Marr, A. Imhof, and A. Groth, "Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes", Cell Rep. **30**, 1223 (2020).

[69] A. Sood and B. Zhang, "Quantifying epigenetic stability with minimum action paths", Phys. Rev. E **101**, 062409 (2020).

[70] W. Zhao, L. Qiao, S. Yan, Q. Nie, and L. Zhang, "Mathematical modeling of histone modifications reveals the formation mechanism and function of bivalent chromatin", iScience **24**, 102732 (2021).

[71] A. Z. Abdulla, C. Vaillant, and D. Jost, "Painters in chromatin: a unified quantitative framework to systematically characterize epigenome regulation and memory", Nucleic Acids Res. **50**, 9083 (2022).

[72] S. Mukhopadhyay and A. M. Sengupta, "The role of multiple marks in epigenetic silencing and the emergence of a stable bivalent chromatin state", PLOS Comput. Biol. **9**, edited by N. Maheshri, e1003121 (2013).

[73] I. B. Dodd and K. Sneppen, "Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states", J. Mol. Biol. **414**, 624 (2011).

[74] J. F. Nickels, A. K. Edwards, S. J. Charlton, A. M. Mortensen, S. C. L. Hougaard, A. Trusina, K. Sneppen, and G. Thon, "Establishment of heterochromatin in domain-size-dependent bursts", Proc. Natl. Acad. Sci. U.S.A. **118**, e2022887118 (2021).

[75] I. O. Torres and D. G. Fujimori, "Functional coupling between writers, erasers and readers of histone and dna methylation", Curr. Opin. Struct. Biol. **35**, 68 (2015).

[76] R. Leicher, E. J. Ge, X. Lin, M. J. Reynolds, W. Xie, T. Walz, B. Zhang, T. W. Muir, and S. Liu, "Single-molecule and in silico dissection of the interaction between Polycomb repressive complex 2 and chromatin", Proc. Natl. Acad. Sci. **117**, 30465 (2020).

[77] F. Erdel and E. C. Greene, "Generalized nucleation and looping model for epigenetic memory of histone modifications", Proc. Natl. Acad. Sci. **113**, E4180 (2016).

[78] X. Lin, R. Leicher, S. Liu, and B. Zhang, "Cooperative DNA looping by PRC2 complexes", Nucleic Acids Res. **49**, 6238 (2021).

[79] L. Zhao, S. Wang, Z. Cao, W. Ouyang, Q. Zhang, L. Xie, R. Zheng, M. Guo, M. Ma, Z. Hu, W.-K. Sung, Q. Zhang, G. Li, and X. Li, "Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation", Nat. Commun. **10**, 3640 (2019).

[80] G. Fourel, F. Magdinier, and É. Gilson, "Insulator dynamics and the setting of chromatin domains", BioEssays **26**, 523 (2004).

[81] D. Michieletto, E. Orlandini, and D. Marenduzzo, "Polymer model with Epigenetic Recoloring Reveals a Pathway for the de novo Establishment and 3D Organization of Chromatin Domains", Phys. Rev. X **6**, 041047 (2016).

[82] D. Michieletto, D. Coli, D. Marenduzzo, and E. Orlandini, "Nonequilibrium theory of epigenomic microphase separation in the cell nucleus", Phys. Rev. Lett. **123**, 228101 (2019).

[83] D. Michieletto, M. Chiang, D. Coli, A. Papantonis, E. Orlandini, P. R. Cook, and D. Marenduzzo, "Shaping epigenetic memory via genomic bookmarking", Nucleic Acids Res. **46**, 83 (2017).

[84] J. A. Owen, D. Osmanović, and L. Mirny, "Design principles of 3D epigenetic memory systems", Science **382**, 10.1126/science.adg3053 (2023).

[85] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, "Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains", Nucleic Acids Res. **42**, 9553 (2014).

[86] D. Jost, C. Vaillant, and P. Meister, "Coupling 1D modifications and 3D nuclear organization: data, models and function", Curr. Opin. Cell Biol. **44**, 20 (2017).

[87] D. Jost and C. Vaillant, "Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance", Nucleic Acids Res. **46**, 2252 (2018).

[88] J. D. Olarte-Plata, N. Haddad, C. Vaillant, and D. Jost, "The folding landscape of the epigenome", Phys. Biol. **13**, 026001 (2016).

[89] M. Katava, G. Shi, and D. Thirumalai, "Chromatin dynamics controls epigenetic domain formation", Biophys. J. **121**, 2895 (2022).

[90] S. H. Sandholtz, Q. MacPherson, and A. J. Spakowitz, "Physical modeling of the heritability and maintenance of epigenetic modifications", Proc. Natl. Acad. Sci. U.S.A. **117**, 20423 (2020).

[91] J. G. Wakim, S. H. Sandholtz, and A. J. Spakowitz, "Impact of chromosomal organization on epigenetic drift and domain stability revealed by physics-based simulations", Biophys. J. **120**, 4932 (2021).

[92] D. Coli, E. Orlandini, D. Michieletto, and D. Marenduzzo, "Magnetic polymer models for epigenetics-driven chromosome folding", Phys. Rev. E **100**, 052410 (2019).

[93] K. Adachi and K. Kawaguchi, "Chromatin state switching in a polymer model with mark-conformation coupling", Phys. Rev. E **100**, 060401 (2019).

[94] M. Aertsens and C. Vanderzande, "The ising model on a SAW", J. Phys. A: Math. Gen. **25**, 735 (1992).

[95] D. P. Foster and D. Majumdar, "Critical behavior of magnetic polymers in two and three dimensions", Phys. Rev. E **104**, 024122 (2021).

[96] K. Faizullina, I. Pchelintsev, and E. Burovski, "Critical and geometric properties of magnetic polymers across the globule-coil transition", Phys. Rev. E **104**, 054501 (2021).

[97] T. Garel, H. Orland, and E. Orlandini, "Phase diagram of magnetic polymers", Eur. Phys. J. B **12**, 261 (1999).

[98] K. Faizullina and E. Burovski, "Globule-coil transition in the dynamic HP model", J. Phys. Conf. Ser. **1740**, 012014 (2021).

[99] A. Zidovska, "The self-stirred genome: large-scale chromatin dynamics, its biophysical origins and implications", Curr. Opin. Genet. Dev. **61**, 83 (2020).

[100] I. Eshghi, J. A. Eaton, and A. Zidovska, "Interphase chromatin undergoes a local sol-gel transition upon cell differentiation", Phys. Rev. Lett. **126**, 228101 (2021).

[101] J. Lerner, P. A. Gomez-Garcia, R. L. McCarthy, Z. Liu, M. Lakadamyali, and K. S. Zaret, "Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin", Mol. Cell **79**, 677 (2020).

[102] S. S. Ashwin, T. Nozaki, K. Maeshima, and M. Sasai, "Organization of fast and slow chromatin revealed by single-nucleosome dynamics", Proc. Natl. Acad. Sci. U.S.A. **116**, 19939 (2019).

[103] S. Fujishiro and M. Sasai, "Generation of dynamic three-dimensional genome structure through phase separation of chromatin", Proc. Natl. Acad. Sci. U.S.A. **119**, e2109838119 (2022).

[104] H. Salari, M. D. Stefano, and D. Jost, "Spatial organization of chromosomes leads to heterogeneous chromatin motion and drives the liquid- or gel-like dynamical behavior of chromatin", Genome Res. **32**, 28 (2021).

[105] K. Kamat, Z. Lao, Y. Qi, Y. Wang, J. Ma, and B. Zhang, "Compartmentalization with nuclear landmarks yields random, yet precise, genome organization", Biophys. J. **122**, 1376 (2023).

[106] N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, and C. Murre, "Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions", Nat. Commun. **10**, 10.1038/s41467-019-10628-9 (2019).

[107] C. P. Johnstone, N. B. Wang, S. A. Sevier, and K. E. Galloway, "Understanding and engineering chromatin as a dynamical system across length and timescales", Cell Systems **11**, 424 (2020).

[108] B. Fierz and M. G. Poirier, "Biophysics of chromatin dynamics", Annu. Rev. Biophys. **48**, 321 (2019).

[109] H. Strickfaden, T. O. Tolsma, A. Sharma, D. A. Underhill, J. C. Hansen, and M. J. Hendzel, "Condensed chromatin behaves like a solid on the mesoscale in vitro and in living cells", Cell **183**, 1772 (2020).

[110] Y. Zhang, N. Liu, W. Lin, and C. Li, "Quantifying the interplay between genetic and epigenetic regulations in stem cell development", New J. Phys. **21**, 103042 (2019).

[111] N. A. Hathaway, O. Bell, C. Hodges, E. L. Miller, D. S. Neel, and G. R. Crabtree, "Dynamics and memory of heterochromatin in living cells", Cell **149**, 1447 (2012).

[112] M. J. Obersriebnig, E. M. H. Pallesen, K. Sneppen, A. Trusina, and G. Thon, "Nucleation and spreading of a heterochromatic domain in fission yeast", Nat. Commun. **7**, 11518 (2016).

[113] A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C. T. Wu, and X. Zhuang, "Super-resolution imaging reveals distinct chromatin folding for different epigenetic states", Nature **529**, 418 (2016).

[114] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping", Cell **159**, 1665 (2014).

[115] A. P. Latham and B. Zhang, "Molecular determinants for the layering and coarsening of biological condensates", Aggregate **3**, e306 (2022).

[116] G. M. Ginell and A. S. Holehouse, "An introduction to the stickers-and-spacers framework as applied to biomolecular condensates", in *Phase-separated biomolecular condensates* (Springer US, Oct. 2022), pp. 95–116.

[117] J. Berry, C. P. Brangwynne, and M. Haataja, "Physical principles of intracellular organization via active and passive phase transitions", Rep. Prog. Phys. **81**, 046601 (2018).

[118] V. N. Uversky, "Intrinsically disordered proteins in overcrowded milieu: Membraneless organelles, phase separation, and intrinsic disorder", Curr. Opin. Struct. Biol. **44**, 18 (2017).

[119] J. B. Woodruff, A. A. Hyman, and E. Boke, "Organization and Function of Non-dynamic Biomolecular Condensates", Trends Biochem. Sci. **43**, 81 (2018).

[120] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, "A phase separation model for transcriptional control", Cell **169**, 13 (2017).

[121] S. Sanulli and G. J. Narlikar, "Liquid-like interactions in heterochromatin: implications for mechanism and regulation", Curr. Opin. Cell Biol. **64**, 90 (2020).

[122] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: organizers of cellular biochemistry", Nat. Rev. Mol. Cell Biol **18**, 285 (2017).

[123] P. Dogra, A. Joshi, A. Majumdar, and S. Mukhopadhyay, "Intermolecular charge-transfer modulates liquid–liquid phase separation and liquid-to-solid maturation of an intrinsically disordered ph-responsive domain", J. Am. Chem. Soc **141**, 20380 (2019).

[124] S. Ye, A. P. Latham, Y. Tang, C.-H. Hsiung, J. Chen, F. Luo, Y. Liu, B. Zhang, and X. Zhang, "Micropolarity governs the structural organization of biomolecular condensates", Nat. Chem. Biol., 443 (2023).

[125] A. P. Latham, L. Zhu, D. A. Sharon, S. Ye, A. P. Willard, X. Zhang, and B. Zhang, "Microphase separation produces interfacial environment within diblock biomolecular condensates", Elife, 10.1101/2023.03.30.534967 (2024).

[126] S. Boeynaems, A. S. Holehouse, V. Weinhardt, D. Kovacs, J. Van Lindt, C. Larabell, L. Van Den Bosch, R. Das, P. S. Tompa, R. V. Pappu, and A. D. Gitler, "Spontaneous driving forces give rise to protein-rna condensates with coexisting phases and complex material properties", Proc. Natl. Acad. Sci. U.S.A. **116**, 7889 (2019).

[127] J. Ahlers, E. M. Adams, V. Bader, S. Pezzotti, K. F. Winklhofer, J. Tatzelt, and M. Havenith, "The key role of solvent in condensation: mapping water in liquid-liquid phase-separated fus", Biophys. J. **120**, 1266 (2021).

[128] I. A. Klein, A. Boija, L. K. Afeyan, S. W. Hawken, M. Fan, A. Dall'Agnese, O. Oksuz, J. E. Henninger, K. Shrinivas, B. R. Sabari, et al., "Partitioning of cancer therapeutics in nuclear condensates", Science **368**, 1386 (2020).

[129] A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner, "Phase separation provides a mechanism to reduce noise in cells", Science **367**, 464 (2020).

[130] A. S. Lyon, W. B. Peeples, and M. K. Rosen, "A framework for understanding the functions of biomolecular condensates across scales", Nat. Rev. Mol. Cell Biol **22**, 215 (2020).

[131] D. Deviri and S. A. Safran, "Physical theory of biological noise buffering by multicomponent phase separation", Proc. Natl. Acad. Sci. U.S.A. **118**, e2100099118 (2021).

[132] W. Borcherds, A. Bremer, M. B. Borgia, and T. Mittag, "How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation?", Curr. Opin. Struct. Biol. **67**, 41 (2021).

[133] A. A. Hyman, C. A. Weber, and F. Jülicher, "Liquid-liquid phase separation in biology", Annu. Rev. Cell Dev. Biol **30**, 39 (2014).

[134] T. Mittag and R. V. Pappu, "A conceptual framework for understanding phase separation and addressing open questions and challenges", Mol. Cell **82**, 2201 (2022).

[135] G. L. Dignon, R. B. Best, and J. Mittal, "Biomolecular phase separation: from molecular driving forces to macroscopic properties", Annu. Rev. Phys. Chem. **71**, 53 (2020).

[136] B. A. Gibson, L. K. Doolittle, M. W. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen, "Organization of chromatin by intrinsic and regulated phase separation", Cell **179**, 470 (2019).

[137] A. G. Larson and G. J. Narlikar, "The role of phase separation in heterochromatin formation, function, and regulation", Biochemistry **57**, 2540 (2018).

[138] S. Sanulli, M. J. Trnka, V. Dharmarajan, R. W. Tibble, B. D. Pascal, A. L. Burlingame, P. R. Griffin, J. D. Gross, and G. J. Narlikar, "HP1 reshapes nucleosome core to promote phase separation of heterochromatin", Nature **575**, 390 (2019).

[139] A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V. Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, and R. A. Young, "Transcription factors activate genes through the phase-separation capacity of their activation domains", Cell **175**, 1842 (2018).

[140] S. Chong, C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq, and R. Tjian, "Imaging dynamic and selective low-complexity domain interactions that control gene transcription", Science **361**, eaar2555 (2018).

[141] W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and I. I. Cisse, "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates", Science **361**, 412 (2018).

[142] K. Shrinivas, B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, et al., "Enhancer features that drive formation of transcriptional condensates", Mol. Cell **75**, 549 (2019).

[143] Y. E. Guo, J. C. Manteiga, J. E. Henninger, B. R. Sabari, A. Dall'Agnese, N. M. Hannett, J.-H. Spille, L. K. Afeyan, A. V. Zamudio, K. Shrinivas, B. J. Abraham, A. Boija, T.-M. Decker, J. K. Rimel, C. B. Fant, T. I. Lee, I. I. Cisse, P. A. Sharp, D. J. Taatjes, and R. A. Young, "Pol ii phosphorylation regulates a switch between transcriptional and splicing condensates", Nature **572**, 543 (2019).

[144] Y. Chen and A. S. Belmont, "Genome organization around nuclear speckles", Curr. Opin. Genet. Dev. **55**, 91 (2019).

[145] M. W. Parker, M. Bell, M. Mir, J. A. Kao, X. Darzacq, M. R. Botchan, and J. M. Berger, "A new class of disordered elements controls dna replication through initiator self-assembly", eLife **8**, e48562 (2019).

[146] M. Altmeyer, K. J. Neelsen, F. Teloni, I. Pozdnyakova, S. Pellegrino, M. Grøfte, M.-B. D. Rask, W. Streicher, S. Jungmichel, M. L. Nielsen, and J. Lukas, "Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose)", Nat. Commun. **6**, 8088 (2015).

[147] S. Kilic, A. Lezaja, M. Gatti, E. Bianco, J. Michelena, R. Imhof, and M. Altmeyer, "Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments", EMBO J. **38**, e101379 (2019).

[148] A. J. Ruthenburg, H. Li, D. J. Patel, and C. David Allis, "Multivalent engagement of chromatin modifications by linked binding modules", Nat. Rev. Mol. Cell Biol **8**, 983 (2007).

[149] S. Chong, T. G. W. Graham, C. Dugast-Darzacq, G. M. Dailey, X. Darzacq, and R. Tjian, "Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription", Mol. Cell **82**, 2084 (2022).

[150] T. Jana, S. Brodsky, and N. Barkai, "Speed–specificity trade-offs in the transcription factors search for their genomic binding sites", Trends in Genetics **37**, 421 (2021).

[151] Z. Liu and R. Tjian, "Visualizing transcription factor dynamics in living cells", Journal of Cell Biology **217**, 1181 (2018).

[152] S. Kent, K. Brown, C.-h. Yang, N. Alsaihati, C. Tian, H. Wang, and X. Ren, "Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging", Cell Reports **33**, 108248 (2020).

[153] I. Izeddin, V. Récamier, L. Bosanac, I. I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, and X. Darzacq, "Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus", eLife **3**, 10.7554/elife.02230 (2014).

[154] X. Chu, T. Sun, Q. Li, Y. Xu, Z. Zhang, L. Lai, and J. Pei, "Prediction of liquid–liquid phase separating proteins using machine learning", BMC Bioinformatics **23**, 72 (2022).

[155] R. M. Vernon and J. D. Forman-Kay, "First-generation predictors of biological protein phase separation", Curr. Opin. Struct. Biol. **58**, 88 (2019).

[156] A. P. Latham and B. Zhang, "Unifying coarse-grained force fields for folded and disordered proteins", Curr. Opin. Struct. Biol. **72**, 63 (2022).

[157] A. P. Latham and B. Zhang, "On the stability and layered organization of protein-DNA condensates", Biophys. J. **121**, 1727 (2022).

[158] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, "Relation between single-molecule properties and phase behavior of intrinsically disordered proteins", Proc. Natl. Acad. Sci. U.S.A. **115**, 9929 (2018).

[159] S. Liu, C. Wang, A. P. Latham, X. Ding, and B. Zhang, "OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates", PLOS Comput. Biol. **19**, e1011442 (2023).

[160] W. Zheng, G. L. Dignon, N. Jovic, X. Xu, R. M. Regy, N. L. Fawzi, Y. C. Kim, R. B. Best, and J. Mittal, "Molecular details of protein condensates probed by microsecond long atomistic simulations", J. Phys. Chem. B **124**, 11671 (2020).

[161] M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne, "Coexisting liquid phases underlie nucleolar subcompartments", Cell **165**, 1686 (2016).

[162] J.-M. Choi, F. Dar, and R. V. Pappu, "LASSI: A lattice model for simulating phase transitions of multivalent proteins", PLOS Comput. Biol. **15**, e1007028 (2019).

[163] A. N. Semenov and M. Rubinstein, "Thermoreversible gelation in solutions of associative polymers. 1. statics", Macromolecules **31**, 1373 (1998).

[164] A. V. Dobrynin, "Phase diagram of solutions of associative polymers", Macromolecules **37**, 3881 (2004).

[165] F. Tanaka, "Theory of Thermoreversible Gelation", Macromolecules **22**, 1988 (1989).

[166] K. Bhandari, M. A. Cotten, J. Kim, M. K. Rosen, and J. D. Schmit, "Structure-Function Properties in Disordered Condensates", J. Phys. Chem. B **125**, 467 (2021).

[167] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, "Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components", Proc. Natl. Acad. Sci. U. S. A. **117**, 13238 (2020).

[168] W. M. Jacobs and D. Frenkel, "Phase transitions in biological systems with many components", Biophys. J. **112**, 683 (2017).

[169] K. Shrinivas and M. P. Brenner, "Phase separation in fluids with many interacting components", Proc. Natl. Acad. Sci. U.S.A. **118**, e2108551118 (2021).

[170] A. V. Ermoshkin and M. Olvera de la Cruz, "A modified random phase approximation of polyelectrolyte solutions", Macromolecules **36**, 7824 (2003).

[171] A. V. Ermoshkin and M. Olvera De La Cruz, "Gelation in strongly charged polyelectrolytes", J. Polym. Sci. B: Polym. Phys. **42**, 766 (2004).

[172] Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, "Sequence-specific polyampholyte phase separation in membraneless organelles", Phys. Rev. Lett. **117**, 178101 (2016).

[173] Y.-H. Lin, J. Song, J. D. Forman-Kay, and H. S. Chan, "Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins", J. Mol. Liq. **228**, 176 (2017).

[174] Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, "Theories for sequence-dependent phase behaviors of biomolecular condensates", Biochemistry **57**, 2499 (2018).

[175] J.-M. Choi, A. A. Hyman, and R. V. Pappu, "Generalized models for bond percolation transitions of associative polymers", Phys. Rev. E **102**, 042403 (2020).

[176] D. Prusty, V. Pryamitsyn, and M. Olvera de la Cruz, "Thermodynamics of associative polymer blends", Macromolecules **51**, 5918 (2018).

[177] J.-M. Choi, F. Dar, and R. V. Pappu, "Lassi: a lattice model for simulating phase transitions of multivalent proteins", PLOS Comput. Biol. **15**, e1007028 (2019).

[178] Y. Zhang, B. Xu, B. G. Weiner, Y. Meir, and N. S. Wingreen, "Decoding the physical principles of two-component biomolecular phase separation", elife **10**, 1 (2021).

[179] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, "Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins", eLife **6**, e30294 (2017).

[180] A. Bremer, M. Farag, W. M. Borcherds, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, "Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains", Nat. Chem. **14**, 196 (2022).

[181] M. Farag, S. R. Cohen, W. M. Borcherds, A. Bremer, T. Mittag, and R. V. Pappu, "Condensates formed by prion-like low-complexity domains have small-world network structures and interfaces defined by expanded conformations", Nat. Commun. **13**, 7722 (2022).

[182] M. Farag, W. M. Borcherds, A. Bremer, T. Mittag, and R. V. Pappu, "Phase separation of protein mixtures is driven by the interplay of homotypic and heterotypic interactions", Nat. Commun. **14**, 5527 (2023).

[183] T. Mittag and R. V. Pappu, "A conceptual framework for understanding phase separation and addressing open questions and challenges", Mol. Cell **82**, 2201 (2022).

[184] S. Ranganathan and E. I. Shakhnovich, "Dynamic metastable long-living droplets formed by sticker-spacer proteins", elife **9**, 1 (2020).

[185] A. Chattaraj and E. I. Shakhnovich, "Separation of sticker-spacer energetics governs the coalescence of metastable biomolecular condensates", bioRxiv, 10.1101/2023.10.03.560747 (2023).

[186] G. Orlando, D. Raimondi, F. Tabaro, F. Codicè, Y. Moreau, and W. F. Vranken, "Computational identification of prion-like rna-binding proteins that form liquid phase-separated condensates", Bioinformatics **35**, 4617 (2019).

[187] R. M. Vernon, P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, and J. D. Forman-Kay, "Pi-pi contacts are an overlooked protein feature relevant to phase separation", eLife **7**, e31486 (2018).

[188] J. Wang, K. Zhang, L. Xu, and E. Wang, "Quantifying the Waddington landscape and biological paths for development and differentiation", Proc. Natl. Acad. Sci. U.S.A. **108**, 8257 (2011).

[189] S. S. Ashwin and M. Sasai, "Effects of Collective Histone State Dynamics on Epigenetic Landscape and Kinetics of Cell Reprogramming", Sci. Rep. **5**, 1 (2015).

[190] J. Lewis, J. Slack, and L. Wolpert, "Thresholds in development", J. Theor. Biol. **65**, 579 (1977).

[191] R. Thomas and M. Kaufman, "Multistationarity, the basis of cell differentiation and memory. i. structural conditions of multistationarity and other nontrivial behavior", Chaos **11**, 170 (2001).

[192] J. E. Ferrell, "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability", Curr. Opin. Cell Biol. **14**, 140 (2002).

[193] Z. Cao and R. Grima, "Linear mapping approximation of gene regulatory networks with stochastic dynamics", Nat. Commun. **9**, 3305 (2018).

[194] C. Jia and R. Grima, "Small protein number effects in stochastic models of autoregulated bursty gene expression", J. Chem. Phys **152**, 084115 (2020).

[195] A. J. Ruthenburg, C. D. Allis, and J. Wysocka, "Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark", Mol. Cell **25**, 15 (2007).

[196] A. M. Naar, B. D. Lemon, and R. Tjian, "Transcriptional coactivator complexes", Annu. Rev. Biochem. **70**, 475 (2001).

[197] L. N. Rusche, A. L. Kirchmaier, and J. Rine, "The Establishment, Inheritance, and Function of Silenced Chromatin in Saccharomyces cerevisiae", Annu. Rev. Biochem. **72**, 481 (2003).

[198] D. David-Rus, S. Mukhopadhyay, J. L. Lebowitz, and A. M. Sengupta, "Inheritance of epigenetic chromatin silencing", J. Theor. Biol. **258**, 112 (2009).

[199] M. Doi, "Second quantization representation for classical many-particle system", J. Phys. A: Math. Gen. **9**, 1465 (1976).

[200] M. Doi, "Stochastic theory of diffusion-controlled reaction", J. Phys. A: Math. Gen. **9**, 1479 (1976).

[201] P. Grassberger and M. Scheunert, "Fock-space methods for identical classical objects", Fortschr. Phys. **28**, 547 (1980).

[202] L. Peliti, "Path integral approach to birth-death processes on a lattice", J. Phys. France **46**, 1469 (1985).

[203] P. Thomas, C. Fleck, R. Grima, and N. Popović, "System size expansion using feynman rules and diagrams", J. Phys. A: Math. Theor. **47**, 455007 (2014).

[204] B. P. Lee and J. Cardy, "Renormalization group study of the A+B diffusion-limited reaction", J. Stat. Phys. **80**, 971 (1995).

[205] M. Sasai and P. G. Wolynes, "Stochastic gene expression as a many-body problem", Proc. Natl. Acad. Sci. U.S.A **100**, 2374 (2003).

[206] B. Zhang and P. G. Wolynes, "Stem cell differentiation as a many-body problem", Proc. Natl. Acad. Sci. U.S.A. **111**, 10185 (2014).

[207] U. C. Tauber, *Critical dynamics a field theory approach to equilibrium and non-equilibrium scaling behavior* (Cambridge University Press, 2014).

[208] L. Biedenharn and J. D. Louck, *Angular momentum in quantum physics* (Cambridge University Press, 1981).

[209] J. Tailleur, J. Kurchan, and V. Lecomte, "Mapping out-of-equilibrium into equilibrium in one-dimensional transport models", J. Phys. A: Math. Theor. **41**, 505001 (2008).

[210] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, "Large fluctuations and optimal paths in chemical kinetics", J. Chem. Phys. **100**, 5735 (1994).

[211] D. M. Roma, R. A. O'Flanagan, A. E. Ruckenstein, A. M. Sengupta, and R. Mukhopadhyay, "Optimal path to epigenetic switching", Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. **71**, 1 (2005).

[212] D. Liu, "Optimal transition paths of stochastic chemical kinetic systems", J. Chem. Phys. **124**, 6 (2006).

[213] J. Wang, K. Zhang, and E. Wang, "Kinetic paths, time scale, and underlying landscapes: A path integral framework to study global natures of nonequilibrium systems and networks", J. Chem. Phys. **133**, 10.1063/1.3478547 (2010).

[214] P. Zhou and T. Li, "Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, A-type integral and beyond", J. Chem. Phys. **144**, 094109 (2016).

[215] V. Elgart and A. Kamenev, "Rare event statistics in reaction-diffusion systems", Phys. Rev. E **70**, 041106 (2004).

[216] V. Elgart and A. Kamenev, "Classification of phase transitions in reaction-diffusion models", Phys. Rev. E **74**, 041101 (2006).

[217] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems* (Springer, New York, 1998).

[218] M. Heymann and E. Vanden-Eijnden, "The geometric minimum action method: A least action principle on the space of curves", Commun. Pure Appl. Math. **61**, 1052 (2008).

[219] E. Autieri, P. Faccioli, M. Sega, F. Pederiva, and H. Orland, "Dominant reaction pathways in high-dimensional systems", J. Chem. Phys. **130**, 1 (2009).

[220] A. Sood, G. Schuette, and B. Zhang, "Dynamical phase transition in models that couple chromatin folding with histone modifications", Phys. Rev. E **109**, 054411 (2024).

[221] X. Lin, Y. Qi, A. P. Latham, and B. Zhang, "Multiscale modeling of genome organization with maximum entropy optimization", J. Chem. Phys. **155**, 010901 (2021).

[222] K. Zhang, M. Sasai, and J. Wang, "Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics", Proc. Natl. Acad. Sci. U.S.A. **110**, 14930 (2013).

[223] B. Bhattacharyya, J. Wang, and M. Sasai, "Stochastic epigenetic dynamics of gene switching", Phys. Rev. E **102**, 042408 (2020).

[224] A. Sood and B. Zhang, "Quantifying the stability of coupled genetic and epigenetic switches with variational methods", Front. genet. **11**, 10.3389/fgene.2020.636724 (2021).

[225] S. H. Sandholtz, D. Kannan, B. G. Beltran, and A. J. Spakowitz, "Chromosome structural mechanics dictates the local spreading of epigenetic marks", Biophys. J. **119**, 1630 (2020).

[226] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding", Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).

[227] S. S. Plotkin and J. N. Onuchic, "Structural and energetic heterogeneity in protein folding. I. Theory", J. Chem. Phys. **116**, 5263 (2002).

[228] B. A. Shoemaker, J. Wang, and P. G. Wolynes, "Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble", J. Mol. Biol **287**, 675 (1999).

[229] B. A. Shoemaker and P. G. Wolynes, "Exploring structures in protein folding funnels with free energy functionals: the denatured ensemble", J. Mol. Biol **287**, 657 (1999).

[230] H. Jacobson and W. H. Stockmayer, "Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems", J. Chem. Phys. **18**, 1600 (1950).

[231] H. S. Chan and K. A. Dill, "The effects of internal constraints on the configurations of chain molecules", J. Chem. Phys. **92**, 3118 (1990).

[232] H. S. Chan and K. A. Dill, "Intrachain loops in polymers: effects of excluded volume", J. Chem. Phys. **90**, 492 (1989).

[233] G. Schuette, X. Ding, and B. Zhang, "Efficient Hi-C inversion facilitates chromatin folding mechanism discovery and structure prediction", Biophys. J. **122**, 3425 (2023).

[234] T. Nozaki, R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, K. S. Wendt, Y. Okada, T. Nagai, and K. Maeshima, "Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging", Mol. Cell **67**, 282 (2017).

[235] M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, and E. Aurell, "Improved contact prediction in proteins: using pseudolikelihoods to infer potts models", Phys. Rev. E **87**, 012707 (2013).

[236] D. E. Makarov and H. Metiu, "A model for the kinetics of protein folding: Kinetic Monte Carlo simulations and analytic results", J. Chem. Phys. **116**, 5205 (2002).

[237] J. D. Bryngelson and P. G. Wolynes, "Intermediates and barrier crossing in a random energy model (with applications to protein folding)", J. Phys. Chem. **93**, 6902 (1989).

[238] L. Kaustov, H. Ouyang, M. Amaya, A. Lemak, N. Nady, S. Duan, G. A. Wasney, Z. Li, M. Vedadi, M. Schapira, J. Min, and C. H. Arrowsmith, "Recognition and specificity determinants of the human Cbx chromodomains", J. Biol. Chem. **286**, 521 (2011).

[239] J. J. Funke, P. Ketterer, C. Lieleg, S. Schunter, P. Korber, H. Dietz, C. Lieleg, J. J. Funke, P. Ketterer, H. Dietz, and S. Schunter, "Uncovering the forces between nucleosomes using DNA origami", Sci. Adv. **2**, e1600974 (2016).

[240] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", J. Phys. Chem. **81**, 2340 (1977).

[241] C. Waddington and H. Kacser, *The strategy of the genes: a discussion of some aspects of theoretical biology* (Allen & Unwin, 1957).

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

[243] K.-Y. Kim and J. Wang, "Potential energy landscape and robustness of a gene regulatory network: toggle switch", PLOS Comput. Biol. **3**, 1 (2007).

[244] J. Venegas-Ortiz and M. R. Evans, "Analytical study of an exclusive genetic switch", J. Phys. A Math. Theor. **44**, 355001 (2011).

[245] B.-L. Xu and Y. Tao, "External noise and feedback regulation: steady-state statistics of auto-regulatory genetic network", J. Theor. Biol. **243**, 214 (2006).

[246] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression", Proc. Natl. Acad. Sci. U.S.A. **105**, 17256 (2008).

[247] C. Lv, X. Li, F. Li, and T. Li, "Energy Landscape Reveals That the Budding Yeast Cell Cycle Is a Robust and Adaptive Multi-stage Process", PLoS Comput. Biol. **11**, 1 (2015).

[248] P. Wang, C. Song, H. Zhang, Z. Wu, X.-J. Tian, and J. Xing, "Epigenetic state network approach for describing cell phenotypic transitions", Interface focus **4**, 20130068 (2014).

[249] A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, "Absolute rate theories of epigenetic stability", Proc. Natl. Acad. Sci. U.S.A. **102**, 18926 (2005).

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

[251] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: From theories to phenotypes", Nat. Rev. Genet. **6**, 451 (2005).

[252] Y. Cao, H.-M. Lu, and J. Liang, "Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda", Proc. Natl. Acad. Sci. U.S.A. **107**, 18445 (2010).

[253] B. Qiu, T. Zhou, and J. Zhang, "Molecular-memory-driven phenotypic switching in a genetic toggle switch without cooperative binding", Phys. Rev. E **101**, 022409 (2020).

[254] H. Chen, P. Thill, and J. Cao, "Transitions in genetic toggle switches driven by dynamic disorder in rate coefficients", J. Chem. Phys. **144**, 175104 (2016).

[255] C. D. Andrew Angel Jie Song and M. Howard, "A polycomb-based switch underlying quantitative epigenetic memory", Nature **476**, 105 (2011).

[256] M. N. Artyomov, A. Meissner, and A. K. Chakraborty, "A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency", PLOS Comput. Biol. **6**, 1 (2010).

[257] K. Miller-Jensen, S. S. Dey, D. V. Schaffer, and A. P. Arkin, "Varying virulence: epigenetic control of expression noise and disease processes", Trends Biotechnol. **29**, 517 (2011).

[258] L. Mariani, E. G. Schulz, M. H. Lexberg, C. Helmstetter, A. Radbruch, M. Löhning, and T. Höfer, "Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression", Mol. Syst. Biol. **6**, 359 (2010).

[259] T. S. Furey and P. Sethupathy, "Genetics driving epigenetics", Science **342**, 705 (2013).

[260] Y. Qi, A. Reyes, S. E. Johnstone, M. J. Aryee, B. E. Bernstein, and B. Zhang, "Data-Driven Polymer Model for Mechanistic Exploration of Diploid Genome Organization", Biophys. J. **119**, 1905 (2020).

[261] W. J. Xie, Y. Qi, and B. Zhang, "Characterizing chromatin folding coordinate and landscape with deep learning", PLOS Comput. Biol. **16**, 1 (2020).

[262] V. W. Zhou, A. Goren, and B. E. Bernstein, "Charting histone modifications and the functional organization of mammalian genomes", Nat. Rev. Genet. **12**, 7 (2011).

[263] M. J. Rowley and V. G. Corces, "Organizational principles of 3D genome architecture", Nat. Rev. Genet. **19**, 789 (2018).

[264] W. A. Flavahan, Y. Drier, S. E. Johnstone, M. L. Hemming, D. R. Tarjan, E. Hegazi, S. J. Shareef, N. M. Javed, C. P. Raut, B. K. Eschle, et al., "Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs", Nature **575**, 229 (2019).

[265] P. H. Tate and A. P. Bird, "Effects of DNA methylation on DNA-binding proteins and gene expression", Curr. Opin. Genet. Dev. **3**, 226 (1993).

[266] H. Zhu, G. Wang, and J. Qian, "Transcription factors as readers and effectors of DNA methylation", Nat. Rev. Genet. **17**, 551 (2016).

[267] M. Sasai, Y. Kawabata, K. Makishi, K. Itoh, and T. P. Terada, "Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells", PLOS Comput. Biol. **9**, 1 (2013).

[268] N. Folguera-Blasco, R. Pérez-Carrasco, E. Cuyàs, J. A. Menendez, and T. Alarcón, "A multiscale model of epigenetic heterogeneity-driven cell fate decision-making", PLOS Comput. Biol. **15**, 1 (2019).

[269] R. Huang and J. Lei, "Dynamics of gene expression with positive feedback to histone modifications at bivalent domains", Int. J. Mod. Phys. B **32**, 1850075 (2018).

[270] H. Binder, L. Steiner, J. Przybilla, T. Rohlf, S. Prohaska, and J. Galle, "Transcriptional regulation by histone modifications: towards a theory of chromatin reorganization during stem cell differentiation", Phys. Biol. **10**, 026006 (2013).

[271] Y. Matsushita and K. Kaneko, "Homeorhesis in waddington's landscape by epigenetic feedback regulation", Phys. Rev. Research **2**, 023083 (2020).

[272] G. L. Eyink, "Action principle in nonequilibrium statistical dynamics", Phys. Rev. E **54**, 3419 (1996).

[273] N. Rosenfeld, M. B. Elowitz, and U. Alon, "Negative autoregulation speeds the response times of transcription networks", J. Mol. Biol **323**, 785 (2002).

[274] M. Ptashne, A. Jeffrey, A. Johnson, R. Maurer, B. Meyer, C. Pabo, T. Roberts, and R. Sauer, "How the λ repressor and cro work", Cell **19**, 1 (1980).

[275] A. D. Johnson, A. R. Poteete, G. Lauer, R. T. Sauer, G. K. Ackers, and M. Ptashne, "λ Repressor and cro—components of an efficient molecular switch", Nature **294**, 217 (1981).

[276] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, "Noise-based switches and amplifiers for gene expression", Proc. Natl. Acad. Sci. U.S.A. **97**, 2075 (2000).

[277] A. Ralston and J. Rossant, "Genetic regulation of stem cell origins in the mouse embryo", Clin. Genet. **68**, 106 (2005).

[278] Y.-H. Loh, J.-H. Ng, and H.-H. Ng, "Molecular framework underlying pluripotency", Cell Cycle **7**, 885 (2008).

[279] S. H. Orkin and L. I. Zon, "Hematopoiesis: an evolving paradigm for stem cell biology", Cell **132**, 631 (2008).

[280] F. J. Alexander and G. L. Eyink, "Rayleigh-ritz calculation of effective potential far from equilibrium", Phys. Rev. Lett. **78**, 1 (1997).

[281] H.-C. Fu and R. Sasaki, "Negative binomial and multinomial states: probability distributions and coherent states", J. Math. Phys. **38**, 3968 (1997).

[282] H.-C. Fu and R. Sasaki, "Probability distributions and coherent states of B_r, C_r, and D_r algebras", J. Phys. A: Math. Gen. **31**, 901 (1998).

[283] A. Gaspar-Maia, A. Alajem, E. Meshorer, and M. Ramalho-Santos, "Open chromatin in pluripotency and reprogramming", Nat. Rev. Mol. Cell Biol **12**, 36 (2011).

[284] G. Mas, E. Blanco, C. Ballaré, M. Sansó, Y. G. Spill, D. Hu, Y. Aoi, F. Le Dily, A. Shilatifard, M. A. Marti-Renom, et al., "Promoter bivalency favors an open chromatin architecture in embryonic stem cells", Nat. Genet. **50**, 1452 (2018).

[285] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, et al., "A bivalent chromatin structure marks key developmental genes in embryonic stem cells", Cell **125**, 315 (2006).

[286] N. L. Vastenhouw and A. F. Schier, "Bivalent histone modifications in early embryogenesis", Curr. Opin. Cell Biol. **24**, 374 (2012).

[287] G. Kar, J. K. Kim, A. A. Kolodziejczyk, K. N. Natarajan, E. T. Triglia, B. Mifsud, S. Elderkin, J. C. Marioni, A. Pombo, and S. A. Teichmann, "Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression", Nat. Commun. **8**, 1 (2017).

[288] K. Sneppen and L. Ringrose, "Theoretical analysis of polycomb-trithorax systems predicts that poised chromatin is bistable and not bivalent", Nat. Commun. **10**, 2133 (2019).

[289] P. S. L. Lim and E. Meshorer, "Dppa2 and Dppa4 safeguard bivalent chromatin in order to establish a pluripotent epigenome", Nat. Struct. Biol. **27**, 685 (2020).

[290] V. Azuara, P. Perry, S. Sauer, M. Spivakov, H. F. Jørgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, and A. G. Fisher, "Chromatin signatures of pluripotent cell lines", Nat. Cell Biol. **8**, 532 (2006).

[291] J. Ohkubo, "Approximation scheme for master equations: variational approach to multivariate case", J. Chem. Phys. **129**, 044108 (2008).

[292] A. F. Ramos, G. C. P. Innocentini, and J. E. M. Hornos, "Exact time-dependent solutions for a self-regulating gene", Phys. Rev. E **83**, 062902 (2011).

[293] A. Sood and B. Zhang, "Preserving condensate structure and composition by lowering sequence complexity", bioRxiv, 10.1101/2023.11.29.569249 (2023).

[294] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding", en, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).

[295] V. Pande, A. Grosberg, and T. Tanaka, "Statistical mechanics of simple models of protein folding and design", Biophys. J. **73**, 3192 (1997).

[296] B. Derrida, "Random-Energy Model: Limit of a Family of Disordered Models", en, Physical Review Letters **45**, 79 (1980).

[297] S. S. Plotkin, J. Wang, and P. G. Wolynes, "Statistical mechanics of a correlated energy landscape model for protein folding funnels", J. Chem. Phys. **106**, 2932 (1997).

[298] M. Doi, *Introduction to polymer physics*, Paperback (Oxford University Press, Sept. 20, 1990), p. 136.

[299] A. P. Latham and B. Zhang, "Maximum Entropy Optimized Force Field for Intrinsically Disordered Proteins", J. Chem. Theory Comput. **16**, 773 (2020).

[300] R. M. Regy, J. Thompson, Y. C. Kim, and J. Mittal, "Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins", Protein Sci. **50**, pro.4094 (2021).

[301] M. Rubinstein and A. V. Dobrynin, "Solutions of associative polymers", Trends in Polymer Science **5**, 181 (1997).

[302] F. Tanaka, *Polymer physics* (Cambridge University Press, Apr. 2011).

[303] R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins", Chem. Rev. **114**, 6589 (2014).

[304] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence complexity of disordered protein", Proteins Struct. Funct. Genet. **42**, 38 (2001).

[305] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions", Nat. Rev. Mol. Cell Biol. **6**, 197 (2005).

[306] V. N. Uversky, "Intrinsically disordered proteins and their "Mysterious" (meta)physics", Front. Phys. **7**, 8 (2019).

[307] V. Pareek, Z. Sha, J. He, N. S. Wingreen, and S. J. Benkovic, "Metabolic channeling: predictions, deductions, and evidence", Mol. Cell **81**, 3775 (2021).

[308] D. M. Mitrea, J. A. Cika, C. B. Stanley, A. Nourse, P. L. Onuchic, P. R. Banerjee, A. H. Phillips, C.-G. Park, A. A. Deniz, and R. W. Kriwacki, "Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation", Nat. Commun. **9**, 842 (2018).

[309] N. Rostam, S. Ghosh, C. F. W. Chow, A. Hadarovich, C. Landerer, R. Ghosh, H. Moon, L. Hersemann, D. M. Mitrea, I. A. Klein, A. A. Hyman, and A. Toth-Petroczy, "CD-CODE: crowdsourcing condensate database and encyclopedia.", Nat. Methods, 673 (2023).

[310] R. Leicher, A. Osunsade, G. N. L. Chua, S. C. Faulkner, A. P. Latham, J. W. Watters, T. Nguyen, E. C. Beckwitt, S. Christodoulou-Rubalcava, P. G. Young, B. Zhang, Y. David, and S. Liu, "Single-stranded nucleic acid binding and coacervation by linker histone H1", Nat. Struct. Mol. Biol. **29**, 463 (2022).

[311] D. L. J. Lafontaine, J. A. Riback, R. Bascetin, and C. P. Brangwynne, "The nucleolus as a multiphase liquid condensate", Nat. Rev. Mol. Cell Biol. **22**, 165 (2021).

[312] J. A. Riback, L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M.-T. Wei, R. W. Kriwacki, and C. P. Brangwynne, "Composition-dependent thermodynamics of intracellular phase separation", Nature **581**, 209 (2020).

[313] H. Lyons, R. T. Veettil, P. Pradhan, C. Fornero, N. De La Cruz, K. Ito, M. Eppert, R. G. Roeder, and B. R. Sabari, "Functional partitioning of transcriptional regulators by patterned charge blocks", Cell **186**, 327 (2023).

[314] S. A. Thody, H. D. Clements, H. Baniasadi, A. S. Lyon, M. S. Sigman, and M. K. Rosen, "Small molecule properties define partitioning into biomolecular condensates", bioRxiv, 10.1101/2022.12.19.521099 (2022).

[315] H. R. Kilgore, P. G. Mikhael, K. J. Overholt, A. Boija, N. M. Hannett, C. Van Dongen, T. I. Lee, Y.-T. Chang, R. Barzilay, and R. A. Young, "Distinct chemical environments in biomolecular condensates", Nat. Chem. Biol. **20**, 291 (2024).

[316] W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and I. I. Cisse, "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates", Science **361**, 412 (2018).

[317] L. Wang, Y. Gao, X. Zheng, C. Liu, S. Dong, R. Li, G. Zhang, Y. Wei, H. Qu, Y. Li, et al., "Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism", Mol. Cell **76**, 646 (2019).

[318] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, et al., "Phase transitions in the assembly of multivalent signalling proteins", Nature **483**, 336 (2012).

[319] M. Vendruscolo and M. Fuxreiter, "Protein condensation diseases: therapeutic opportunities", Nat. Commun. **13**, 5550 (2022).

[320] S. Alberti and A. A. Hyman, "Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing", Nat. Rev. Mol. Cell Biol **22**, 196 (2021).

[321] S. Boeynaems, S. Chong, J. Gsponer, L. Holt, D. Milovanovic, D. M. Mitrea, O. Mueller-Cajar, B. Portz, J. F. Reilly, C. D. Reinkemeier, B. R. Sabari, S. Sanulli, J. Shorter, E. Sontag, L. Strader, J. Stachowiak, S. C. Weber, M. White, H. Zhang, M. Zweckstetter, S. Elbaum-Garfinkle, and R. Kriwacki, "Phase separation in biology and disease: current perspectives and open questions", J. Mol. Biol **435**, 167971 (2023).

[322] Y. Shin, Y.-C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, and C. P. Brangwynne, "Liquid nuclear condensates mechanically sense and restructure the genome", Cell **175**, 1481 (2018).

[323] T. Haaf, E. I. Golub, G. Reddy, C. M. Radding, and D. C. Ward, "Nuclear foci of mammalian Rad51 recombination protein in somatic cells after DNA damage and its localization in synaptonemal complexes", Proc. Natl. Acad. Sci. U.S.A. **92**, 2298 (1995).

[324] R. S. Maser, K. J. Monsen, B. E. Nelms, and J. Petrini, "hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks", Mol. Cell. Biol., 6087 (1997).

[325] S. C. Ng and D. Görlich, "A simple thermodynamic description of phase separation of nup98 fg domains", Nat. Commun. **13**, 6172 (2022).

[326] C. Mathieu, R. V. Pappu, and J. P. Taylor, "Beyond aggregation: pathological phase transitions in neurodegenerative disease", Science **370**, 56 (2020).

[327] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, "Valence and patterning of aromatic residues determine the phase behavior of prion-like domains", Science **367**, 694 (2020).

[328] O. M. Morris, J. H. Torpey, and R. L. Isaacson, "Intrinsically disordered proteins: Modes of binding with emphasis on disordered domains", Open Biol. **11**, 10.1098/rsob.210222 (2021).

[329] V. Vacic, C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, V. N. Uversky, and A. K. Dunker, "Characterization of molecular recognition features, MoRFs, and their binding partners", J. Proteome Res. **6**, 2351 (2007).

[330] K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, and N. E. Davey, "Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation", Chem. Rev. **114**, 6733 (2014).

[331] S. Basu and R. P. Bahadur, "Conservation and coevolution determine evolvability of different classes of disordered residues in human intrinsically disordered proteins", Proteins: Struct., Funct., Bioinf. **90**, 632 (2021).

[332] C. J. Brown, A. K. Johnson, A. K. Dunker, and G. W. Daughdrill, "Evolution and disorder", Curr. Opin. Struct. Biol. **21**, 441 (2011).

[333] T. Zarin, B. Strome, A. N. Nguyen Ba, S. Alberti, J. D. Forman-Kay, and A. M. Moses, "Proteome-wide signatures of function in highly diverged intrinsically disordered regions", eLife **8**, e46883 (2019).

[334] T. Zarin, C. N. Tsai, A. N. Nguyen Ba, and A. M. Moses, "Selection maintains signaling function of a highly diverged intrinsically disordered region", Proc. Natl. Acad. Sci. U.S.A. **114**, E1450 (2017).

[335] A. X. Lu, A. X. Lu, I. Pritišanac, T. Zarin, J. D. Forman-Kay, and A. M. Moses, "Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning", PLOS Comput. Biol. **18**, e1010238 (2022).

[336] A. S. Holehouse, G. M. Ginell, D. Griffith, and E. Böke, "Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates: published as part of the biochemistry virtual special issue "protein condensates"", Biochemistry **60**, 3566 (2021).

[337] M. C. Cohan, M. K. Shinn, J. M. Lalmansingh, and R. V. Pappu, "Uncovering non-random binary patterns within sequences of intrinsically disordered proteins", J. Mol. Biol **434**, 167373 (2022).

[338] S. M. Cascarina and E. D. Ross, "Proteome-scale relationships between local amino acid composition and protein fates and functions", PLOS Comput. Biol. **14**, e1006256 (2018).

[339] M. Kato, X. Zhou, and S. L. McKnight, "How do protein domains of low sequence complexity work?", RNA **28**, 3 (2021).

[340] D. B. Lukatsky, A. Afek, and E. I. Shakhnovich, "Sequence correlations shape protein promiscuity", J. Chem. Phys. **135**, 10.1063/1.3624332 (2011).

[341] J.-P. Gazeau, *The spin coherent states* (John Wiley & Sons, Ltd, 2010).

[342] A. Perelomov, *Generalized coherent states and their applications* (Springer, Berlin, Heidelberg, 1986).

[343] M. Stone, K.-S. Park, and A. Garg, "The semiclassical propagator for spin coherent states", J. Math. Phys. **41**, 8025 (2000).

[344] E. A. Kochetov, "SU(2) coherent-state path integral", J. Math. Phys. **36**, 4667 (1995).

[345] V. Vieira and P. Sacramento, "Path integrals of spin-j systems in the holomorphic representation", Nucl. Phys. B. **448**, 331 (1995).

[346] H. G. Solari, "Semiclassical treatment of spin system by means of coherent states", J. Math. Phys. **28**, 1097 (1987).

[347] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics", J. Comput. Phys. **117**, 1 (1995).

[348] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *User guide for minpack-1*, Report (Argonne National Laboratory, Argonne, Illinois, Aug. 1980).

[349] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", Nat. Methods **17**, 261 (2020).

[350] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: open source scientific tools for Python*, 2001.

[351] A. C. Sun and W. D. Seider, "Homotopy-continuation method for stability analysis in the global minimization of the gibbs free energy", Fluid Ph. Equilib. **103**, 213 (1995).

[352] O. Cots, J. Gergaud, and N. Shcherbakova, in *31st european symposium on computer aided process engineering* (Elsevier, 2021), pp. 1081–1086.

[353] N. Shcherbakova, I. Rodriguez-Donis, J. Abildskov, and V. Gerbaud, "A novel method for detecting and computing univolatility curves in ternary mixtures", Chem. Eng. Sci. **173**, 21 (2017).

[354] N. Shcherbakova, K. Roger, J. Gergaud, and O. Cots, in *Computer aided chemical engineering* (Elsevier, 2023), pp. 709–714.

[355] J. Bausa and W. Marquardt, "Quick and reliable phase stability test in VLLE flash calculations by homotopy continuation", Comput. Chem. Eng. **24**, 2447 (2000).

[356] A. Bot, B. P. C. Dewi, and P. Venema, "Phase-separating binary polymer mixtures: the degeneracy of the virial coefficients and their extraction from phase diagrams", ACS Omega **6**, 7862 (2021).

[357] E. L. Allgower and K. Georg, *Introduction to numerical continuation methods* (Society for Industrial and Applied Mathematics, Jan. 2003).