Physics of Self-assembly in Complex Matter

by
Yossi Eliaz

A dissertation submitted to the Department of Physics,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Physics

Chair of Committee: Margaret S. Cheung

Committee Member: Donald Kouri

Committee Member: Gemunu Gunaratne

Committee Member: Krešimir Josić

Committee Member: Andre C. Barato

University of Houston
December 2020

DEDICATION

To my dearest family, my partners, my friends, my mentors, and HaShem–Yitbarach ($\beth_\infty$)

ACKNOWLEDGMENTS

I have been fortunate to receive support from and work with many special people over the course of my graduate school experience.

I would like to thank three amazing scholars and mentors that have supported me along the way. First and foremost, I want to thank my advisor and mentor Dr. Margaret Cheung who has been an inspiring role model since the first time I visited her lab a year before joining the graduate program. Her tremendous guidance, wise advice, scientific perspective, and outstanding leadership have reshaped me as an aspiring scholar and as a person. I am grateful for her constant support to pursue my passions along the years and for the fact that she was always open to new ideas.

Second, I want to thank Dr. Herbert Levine for teaching me how a scholar can be brilliant, influential, and humble at the same time, through his work as co-director of the Center for Theoretical Biological Physics (CTBP). I benefited many times during my work in CTBP from his sublime ability to break down complex physics systems into the simplest possible yet still useful scientific terms. Last, but not least, I thank Dr. Erez Lieberman Aiden. His enthusiasm for science, life, and innovation has motivated me frequently to stay as optimistic as he has always been. Thanks to Erez, I have acquired the ability to stay on track and prioritize my work better.

Drs. Cheung, Levine, and Aiden have been giving me the best guidance I could have asked for and for that I am incredibly grateful. At the same time, I would like to thank all my committee members: Dr. Cheung, Dr. Kouri, Dr. Gunaratne, Dr. Josić, and Dr. Barato for their great advice along the way and their firm support that put me on track and enabled me to accomplish this work. I am delighted to have your names listed on the front page of this work. Evidently, this would not have been accomplished without the help and collaboration with many fabulous people including Ms. Louise Miller, Dr. Aviva Presser Aiden, Andrei Gasic, Dr. Greg Gasic, Dr. Fabio Zaeggara, Carlos Bueno, James Liman, Muhammad Shamim, Dr. Cynthia Perez, Dr. Olga Dudchenko, David Weistz, Alyssa Blackburn, Dat Nguyen, Dr. Neva Durand, Suhas Rao, Dr. Huiya Gu, Prof. Jordan Rowley, Brian-Tyler St.Hilaire, Dr. Marianna Zazhytska, Dr. Nikos Sidiropoulos, ENCODE, Prof. Bill Nobel, Prof. Maria Cristina Gambetta, Dr. Arina Omer, Rebecca and Ivan Bochkov, Shibali Patra, Prof. Lena Simine, Dr. Tomer Markovich, Dr. Dana Krepel, Prof. Peter Wolynes, Prof. Neal Waxham, Dr. Nick Sheffer, Dr. Qian Wang, Ms. Naomi Haynes, Jacob Tinnin, Dr. Jake Ezerski, Jules Nde, Brett Velasquez, Atrayee Sarkar, Rauf Giwa, Hoa, Millad, Dirar Homouz, Chengxuan Li, Prof. Greg Morrison, Dr. Pengzhi Zhang, and the Late Prof. Kwong Hon Lau. I want to thank the entire Jewish community in Houston and the UOS families including: the Gilad, Pala, Yehudai, Tessler, Kazinnik, Amsalem, Ogorek, Wadler, Passi, Gelman, Yoshor, Morgan, Hoffman, Levin, Hafkin, Cohen, Tzur, Biran, Levy, Albi, Mitzner, Dinner, Gross (×2), Ouzan, Birnbaum, Lazaroff, Wolbe, Yanowitz, and Katz families.

Lastly, I would like to thank the incredible staff at the Physics Department at the University of Houston, at CTBP, at Rice University, and at the Baylor College of Medicine. Please accept my sincerest apologies to anyone I have overlooked here, surely you had a tremendous impact on me when our paths intersected throughout our journeys in this world. Thank you all! May the force be with you, amen.

ABSTRACT

This dissertation is based on my graduate research studying complex matter in multiple areas of biological physics. While the complex matter systems in each area differ vastly in scale and function, I use similar computational approaches on each to answer scientific queries about their structure and dynamics. In the first and second sections, I present how graph theory helps interpret actomyosin networks, which are complex biological active matter composed of filament, linker, and motor proteins. The results show how the network's dynamics and structure are reshaped by motor and multivalent actin-binding proteins ("multilinkers").

The third section presents my collaborative work with James Liman and Carlos Bueno on the effect of the actin-related protein (Arp2/3) complex on actomyosin dynamics. Generally, Arp2/3 forms a brancher between a mother and a daughter filament at an angle of 70°. I show that in percolating networks of actin monomers, Arp2/3 promotes avalanches (abrupt release of accumulated mechanical tension) in actomyosin.

In the fourth section, I use deep learning to identify substructures in the global three-dimensional (3D) folded structure of genomes inside the cell nucleus. Those 3D substructures were originally detected by the Aiden lab's Hi-C technology in 2014 when they reported the existence of approximately 10,000 long-range interactions in the human genome called loops. My deep learning model detected the most noticeable loops and alluded to the existence of many more loop-like interactions, which are not easily visible to the naked eye.

In the fifth and final section of this dissertation, I present a collaborative work led by Dr. Fabio Zegarra on the effect of hydrodynamic interactions on the folding of proteins in water. Here, I describe how our computational model of hydrodynamic interactions between proteins and live intracellular media resolved an open question in the literature about whether (a) the effect of hydrodynamics interactions is negligible; (b) hydrodynamics interactions accelerate the folding process; or (c) hydrodynamic interactions decelerate the folding process. I show how all three conclusions are correct under certain circumstances, with an intimate dependence on the system's temperature regime.

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGMENTS iv

ABSTRACT v

LIST OF FIGURES x

LIST OF TABLES xi

1 INTRODUCTION 1
- 1.1 Complex matter: when biology, computer science, and physics merge . . . 1
- 1.2 Introduction to graph and network theory . . . 5
- 1.3 Active matter and actomyosin . . . 9
- 1.4 Introduction to protein folding . . . 10
  - 1.4.1 The energy landscape theory (ELT) . . . 13
- 1.5 The human genome structure as a graph . . . 14

2 STUDYING BIOLOGICAL PHYSICS OF MATTER FROM THE VISTA OF GRAPH THEORY 18
- 2.1 Introduction and motivation . . . 18
- 2.2 Graph representation of complex biological matter . . . 19
  - 2.2.1 Phase separation . . . 19
  - 2.2.2 Conformation of proteins . . . 26
  - 2.2.3 Morphology in actomyosin networks . . . 28
  - 2.2.4 Inter and intra chromosomal subgraphs in the human genome multigraph . . . 28
- 2.3 Discussion and outlook . . . 33

3 MULTIVALENT ACTIN-BINDING PROTEINS AUGMENT THE MORPHOLOGIES OF ACTIN FILAMENT NETWORKS 34
- 3.1 Introduction . . . 34
- 3.2 Methods . . . 36
  - 3.2.1 Coarse-grained collective Langevin dynamics model of actomyosin in Cytosim 36
  - 3.2.2 Order parameters . . . 41
  - 3.2.3 Potential of mean force (PMF) of order parameters . . . 49
- 3.3 Results . . . 49
  - 3.3.1 Multilinkers enrich the morphologies of actin bundles . . . 49
  - 3.3.2 Correlation between order parameters from polymer physics, gelation theory, and network theory . . . 52
  - 3.3.3 Motor activity regulates temporal changes in the global structure and the shape of actomyosin networks . . . 52
  - 3.3.4 Gelation of filaments in actomyosin network by actin-binding proteins . . . 55
  - 3.3.5 Constructing graphs/networks using the spatial information of actin filaments 56
  - 3.3.6 Analysis of actin filament networks viewed by order parameters from graph theory . . . 57

3.3.7 Network theory order parameters reveal rich topologies in a dendritic actomyosin network . . . 59
3.3.8 State diagrams of actomyosin networks in the plane of order parameters from network theory . . . 63
3.4 Discussion . . . 63

4 AVALANCHES IN ACTOMYOSIN NETWORKS 69
4.1 Introduction . . . 69
4.2 Methods . . . 71
4.2.1 Coarse-grained mechanochemical model of actomyosin systems (MEDYAN) . . . 71
4.2.2 Mechanistic order parameters . . . 75
4.2.3 Gelation and percolation in the connectivity graph of actin monomers . . . 76
4.3 Results . . . 79
4.3.1 Arp2/3 branchers augment actomyosin expansion patterns . . . 79
4.3.2 Crosslinkers modulate contraction in unbranched actomyosin networks . . . 80
4.3.3 Branched actomyosin networks display convulsive movements . . . 83
4.3.4 Two modes of avalanches in branched networks . . . 85
4.3.5 Avalanches arise in branched networks when their corresponding graphs of actin monomers percolate . . . 88
4.4 Discussion . . . 89
4.4.1 Contractility in actomyosin depends on the ABPs-to-actin concentrations . . . 89
4.4.2 Avalanches are mechanochemical emergent phenomena in actomyosin . . . 90

5 QUANTITATIVE METHODS TO CHARACTERIZE STRUCTURES AND MUTATIONS IN GENOMES 91
5.1 Introduction . . . 91
5.2 Methods and results . . . 92
5.3 Discussion . . . 96

6 HYDRODYNAMIC INTERACTIONS ALTER THE FOLDING DYNAMICS OF PROTEINS BASED ON THE TEMPERATURE 98
6.1 Introduction . . . 98
6.2 Methods . . . 99
6.2.1 Coarse-grained protein model . . . 99
6.2.2 Brownian dynamics with or without HI . . . 101
6.2.3 Equilibrium thermodynamics simulations . . . 102
6.2.4 Non-equilibrium kinetic simulations . . . 103
6.2.5 The reaction coordinate Q measures the fraction of native contacts . . . 103
6.2.6 Data analysis . . . 104
6.3 Results . . . 106
6.3.1 The impact of BDHI on the folding time depends on temperature . . . 106
6.3.2 Hydrodynamic coupling of midrange and long-range contacts and their opposing impact on the general ordering of contact formation at $T < T_f$ and $T > T_f$ . . . 107
6.4 Discussion . . . 109

7 CONCLUSION AND OUTLOOK **111**

A APPENDIX **113**
A.1 Input file of the phase separation simulations 113
A.2 Input file of the Cytosim simulations 114
A.3 Chemical reaction input file for MEDYAN 116
A.4 System input file for MEDYAN 117

B APPENDIX Twitter university ranking table **121**

C APPENDIX COVID-19 Pool-testing **126**
C.1 Pooled Testing Strategy 126
C.2 Optimal Pooled Testing Strategy 126

D APPENDIX **129**

E APPENDIX **130**
E.1 Actomyosin and memory formation 130

BIBLIOGRAPHY **132**

LIST OF FIGURES

1.1 Simplified illustration of sample graphs and networks . . . 7
1.2 Mentioning graph of tweets by the Top 200 universities on Twitter . . . 8
1.3 Equilibrium free energy of Cahn-Hilliard system . . . 10
1.4 The four hierarchies in protein shapes . . . 12
1.5 Illustration of protein folding energy landscape . . . 14
1.6 The hierarchies of the human genome from DNA to chromosomes . . . 15
1.7 Hi-C map of the entire human genome . . . 17
2.1 Equilibrium free energy of Cahn-Hilliard system . . . 20
2.2 Finite element solution of Cahn-Hilliard equation . . . 21
2.3 Simulations of phase separation . . . 23
2.4 CI2 protein structure, contact map, and graph representation . . . 27
2.5 Actomyosin simulations and the corresponding graph representations . . . 29
2.6 Graph representation of intrachromosomal Hi-C interaction map . . . 31
2.7 Graph representation of interchromosomal interaction Hi-C map . . . 32
3.1 CaMKII structure and functional domains . . . 35
3.2 Early snapshots of simulations from Cytosim . . . 50
3.3 Late snapshots of simulations from Cytosim . . . 51
3.4 Heatmap of the Pearson correlation between all order parameters . . . 53
3.5 Power spectral analysis of mesoscopic polymer order parameters . . . 54
3.6 Gelation of filament clusters formed by actin-binding proteins . . . 56
3.7 Order parameters from network theory . . . 58
3.8 Node degree assortativity coefficient at low and high motor activity . . . 60
3.9 Graph density of actomyosin networks . . . 61
3.10 Potential of mean force (PMF) of two order parameters . . . 62
3.11 State diagram of the node degree assortativity coefficient . . . 64
3.12 State diagram of graph density . . . 65
4.1 Arp2/3 protein brancher in actomyosin networks . . . 69
4.2 Mechanochemical loop to simulate fine-grained actomyosin systems . . . 72
4.3 Illustration of the actin monomer graph construction . . . 77
4.4 Snapshots of MEDYAN simulations with and without Arp2/3 protein complexes . . . 81
4.5 Time signals of the ratio R_g/R_g^i in unbranched actomyosin networks . . . 82
4.6 Time signals of the ratio R_g/R_g^i in branched actomyosin networks . . . 84
4.7 Snapshots before and after an avalanche in actomyosin network . . . 85
4.8 Histogram of the mean displacement of actin monomers in branched and unbranched actomyosin . . . 86
4.9 Types of avalanches in branched actomyosin netwokrs . . . 87
4.10 Probability densities of the weighted mean cluster size in actomyosin networks . . . 88
5.1 LoopDetectNet's architecture and UMAP of the gold standard loop list . . . 93
5.2 Precision-recall plot after training only with the original annotations . . . 95
5.3 Precision-recall plot after retraining with the original annotations and false positive calls . . . 96
5.4 Calls of loops on chromosome 14 . . . 97
6.1 Representations of the studied CI2 protein . . . 105

6.2 Free energy at various conditions with and without HI 106
6.3 Temporal evolution of the fraction of native contacts $Q(t)$ 107
6.4 The effects of hydrodynamic interactions on the secondary structure 108
6.5 The effect of Brownian dynamics on the correlation between protein's beads 110
C.1 Illustration of nested pooling strategy 126
C.2 Simulation of the optimal sizes of the pools 128
D.1 Protein folding times show a non-Arrhenius profile 129

LIST OF TABLES

3.1 System configuration parameter space for multilinker simulations in Cytosim 40
4.1 Concentration ratios of actomyosin simulations with branchers 73
B.1 Top 200 US academic accounts 125

1 Introduction

The research reported herein focuses predominantly on complex biological systems, and in particular on investigations of actomyosin networks on the scale of dendritic spines, which are hallmark for memory formation and maintenance (Appendix E); the computational-based detection of loops in the human genome formed between its 3 billion deoxyribonucleic acid (DNA) base pairs; and the folding of a protein in the complex environment of the cell while considering solvent effects on the folding procedure. The common theme for all these problems is their complexity. They are complex systems that cannot easily be simplified ("reduced") using a straightforward and predictable physical model. For all, the knowledge of the inter-connectivity between the system's components facilitates the development of the main tools required to learn their complex nature. In particular, they all can be interpreted as complex networks of nodes and edges embedded in three-dimensional space that create and manage life. In this work, we start by defining the concepts of complexity and complex matter, and then discuss the concepts of active matter and actomyosin. Subsequently, we describe the basic definitions of graph theory and network theory, describe what is the human genome Hi-C map and briefly introduce the major concept of protein folding, all of which are relevant for this work.

1.1 Complex matter: when biology, computer science, and physics merge

As humans, we admire the skill required to forming a tangible piece of art out of an array of seemingly disparate elements. When we encounter a piece of art, we are amazed by the human ingenuity in crafting it. We appreciate both the craftsmanship process and the final artwork. By the same token, we are intrigued by complex systems in nature. The principle of building a complex entity from simple components occurs in many disciplines beyond art, including in natural sciences, mathematics, and social sciences. Since the days of the Greek philosopher Democritus, who was a younger contemporary of Socrates and lived in the 5th century BCE, scientists hypothesized that complex systems could emerge from the arrangement and rearrangement of atomic constituents

in space [1]. Even though we have an intuitive sense of what complexity and complex matter are, there is still a range of definitions of complexity and a broader range of how complexity can be measured [137]. While complexity theory and complex systems are widely accepted terms used by computer scientists and physicists, the concept of complex matter on a molecular level is relatively new. In fact, this concept was first introduced by the Nobel laureate Jean-Marie Lehn over the past two decades [125, 126]. After JM Lehn's major contribution to the field of supramolecular chemistry, he tried to spearhead the transformation of supramolecular chemistry of individual molecules to the higher complexity of self-organization of many supramolecular structures [125, 126]. In particular, Lehn wrote that matter evolves "under the pressure of information" [126], which naturally lends itself to the development of quantitative measures of complex matter based on the strong foundations of the theories of computational complexity, information, and statistical physics. However, complexity is a relative property, to some extent, and as such it is the assumption herein that all systems under investigation: actomyosin, the human genome, and the folding of a single protein are complex matter, defined both qualitatively and quantitatively. Qualitatively, these systems are complex because they exhibit the property of emergence, meaning the individual constituents and their local interactions evolve into self-assembled structures that are either: (a) chaotic in nature; (b) require significant energy consumption from the environment to be built; or (c) create a conformation which has "computational" capabilities to self-replicate and/or divide the space into functional compartments (e.g., the case of the human genome, which is organized into active and inactive gene regulation regions). Naturally, all the systems detailed above require computational power to be analyzed and understood. Quantitative metrics of complexity can be defined in many ways and shapes [137, 208]; herein, we suggest three possible metrics:

(i) The computational complexity $T(N)$ of the most efficient known algorithm that simulates a discrete representation of the system with N elements.

(ii) The minimal number of bits $I_b(N, t_0, t_1)$ required to represent a program that evolves the system for $t_1 - t_0$ timestamps from an initial configuration at $t = 0$. Of note, this metric is inspired by Kolmogorov complexity [131], and the program may be represented in any predetermined

programming language or a computational model.

(iii) If the dynamics of the system is nonlinear or chaotic.

The first complexity measure (i) is inspired by the computational complexity theory [202, 13] which has listed more than 500 classes of complexity, in the context of both classical and quantum computing [2]. For our purposes, we focus on three of the listed complexity classes: $\mathsf{P} \subseteq \mathsf{NP} \subseteq \mathsf{EXP}$. In order to define computational complexity classes, we first have to adapt a generic computational model. Because it is tedious and beyond the scope of this work to define the Turing machine (TM) computational model rigorously, we define and adhere to an equivalent and as powerful computational model called the random-access machine (RAM) [13]. The RAM model is a simple yet powerful computational model that implements the von Neumann architecture [173]. The von Neumann architecture resembles the abundant computer architecture we use today.

RAM has a finite set of registers that can store integer values, a program counter PC that points to the next instruction in the program to be executed, infinite tape of memory, and a program which is a definite sequence of primitive set of instructions: load, store, read, write, add, print, subtract, and conditional jump [51, 190].

Using the definition of the RAM model we now define the three complexity classes: $\mathsf{P} \subseteq \mathsf{NP} \subseteq \mathsf{EXP}$. The complexity class P is the equivalence class of all programs and/or problems whose execution times $T(N)$ are bounded from above by a polynomial in the input length N. Namely, P is the class of all programs to successfully complete their execution on a RAM machine within $T(N) = \mathcal{O}(N^k)$ steps, where $k \geq 1$ is a constant.

The complexity class NP is the equivalence class of programs and/or problems for which the execution time on a nondeterministic RAM is a polynomial in N. The term "nondeterministic RAM" [78, 3] defines a theoretical machine which always produces the same answer on a given input as RAM does, but instead of following a single branch of execution, all the possible branches can be invoked at the same time; as such, the nondeterministic computational machine explores all of the possibilities simultaneously. NP is the class of the problems that, when executed on a nondeterministic RAM, terminate after $T(N) = \mathcal{O}(N^k)$, where $k \geq 1$ is a constant, N is the input

size, and $T(N)$ is a polynomial in N. EXP is the class of all problems whose execution time scales as $\mathcal{O}(2^N)$ when running on a RAM.

The second complexity measure (ii) is inspired by Kolmogorov complexity [131] and the coding theory of lossless compression techniques [191], which are used in many communication and storage applications. Intuitively, the amount of information in a finite string is the size (number of binary digits, or bits) of the shortest program that without additional data, computes the exact string and terminates. A similar definition for both finite and infinite strings could be given through the program that generate the string. For example, if we consider the binary string:

$$\underbrace{101010\ldots101010}_{\text{2,000,000 digits}}, \tag{1.1}$$

and the Python program that generates the $2,000,000$ characters of the string in Eq. (1.1):

```
for i in range(0, 1_000_000):
    print('10', end='')
```

Strings and configurations of physical systems are equivalent in terms of evaluating their complexity. Using the second metric of complexity, we could ask how many bits are needed to encode a program that generates the final system configuration or a desired output string, given an initial configuration or an initial input string, respectively.

The third complexity measure (iii) stems from the notion that in some cases, despite the low Kolmogorov complexity of a system or the length of the generating program, the system is very complex and cannot be predicted without actually simulating the system. Two examples are the Conway's game of life [83, 107] and Wolfram's cellular automaton [228]. For both cases, the source code to create the simulate the game of life or the cellular automaton is fairly simple, but they are irreducible computational systems and the only way to predict the dynamics is to simulate the system. A similar example, although more predictable and reducible, is the program to plot the fractal geometry of the Mandelbrot set [147]. In all these cases, even though there is a simple program to simulate the system, the emergent dynamics may be very complicated and unexpected [228, 107, 83, 147].

Overall, we use the conceptual definitions of these three measures to define a physical system as complex if at least one or more of the aforementioned complexity measures (i complexity class, ii complex representation of states, or iii complex dynamics) are applicable to it.

According to this definition, the self-assembly of actomyosin, the process of protein folding, and the complexity of the genome structure are all examples of complex matter, determined by the computing power required to create the data, to represent it, and/or to analyze it.

Ultimately, complexity is hard to define and there are many ways to define what is complexity, only some of which are selected to emphasize that the physical system studied in this dissertation are complex. Nonetheless, maybe the simple duck test–"*When I see a bird that walks like a duck and swims like a duck and quacks like a duck, I call that bird a duck.*" [67]–should work here too: is it complicated to solve? If the answer to this question is yes, then it is a complex system. Alternatively, another definition of complexity was provided by Tamas Vicsek, who stated that a system is complex if *"the laws that describe the behavior of a complex system are qualitatively different from those that govern its units."* [218].

1.2 Introduction to graph and network theory

The term "graph" defines a mathematical structure of a pair of sets (V, E), where V is a set of vertices (nodes of the graph) and E is the set of edges (links between the nodes). Such graphs can be divided into "directed" and "undirected" subcategories. In a directed graph, the edges have directionality, e.g., $(v_1, v_2) \in E$ is an edge that goes from node v_1 to node v_2. In an undirected graph, by contrast, the edges do not have directionality and if nodes v_1 and v_2 are linked, then $\{v_1, v_2\} \in E$. The formalism of using graphs and networks provides a powerful tool for humans as well as computers to describe and model complex systems, including [15]: (a) social networks; (b) communications infrastructure involving billions of devices; (c) biological networks responsible for gene regulation and metabolites; (d) neural connectivity; and (e) public transportation. Using graph or network theory for problems from different domains may help in the development of universal laws [17], or in drawing parallels between the dynamics of a social media, the stock

market, or pandemic spread. In this dissertation, graphs are used to represent structural biological systems, with the goal of using such graphical representations to extract relatively simple insights. Beyond translating a snapshot of a biological system to a graph $\mathbf{G} = (V, E)$, we compute a few order parameters based on the network theory literature. Originally, those order parameters were developed to compute structural properties of networks, including: (a) the degree of connectivity between nodes in the graph; (b) the distance between nodes in the graph; (c) the communities that exist in that network; (d) the robustness of the network and its internal communication, and many more [50]. We compute such order parameters on the graph representations of physical systems to gain an insight into the underlying behavior of the systems, and compare them via their graph representations [160, 50].

Fig. 1.1 shows three different examples of graphs. The three graphs have the same number of nodes $V = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$ with differences among the graphs noted in the connections between the nodes. While the graphs in Fig. 1.1(a,b) are undirected, the graph in Fig. 1.1(c) is a directed one. In particular, node A in Fig. 1.1(a) has a node degree, defined as the number of the neighbors of the node, of 2 because it is directly connected to its two neighbors: node B and node C. In the graph in Fig. 1.1(b), the degree of node A is 1 since it is only linked to one node, C. In the directed graph shown in Fig. 1.1(c), node A has an output degree of 0 and an input degree of 1.

To illustrate how a network representation can expose complex pairwise interactions in a wide range of contexts, we have created a directed graph whose nodes are the top 200 Twitter accounts of universities (as listed in B), and we have added a directed edge origins from a node e.g., v_{UH} (University of Houston), to a second node e.g., v_{MIT} (Massachusetts Institute of Technology), if there was at least one tweet by the UH account that tagged MIT. Fig. 1.2 shows how the different universities can be divided into nine clusters that puts the universities from California (cluster 4) together, the universities from the east coast together in cluster 2, Rice University and the University of Houston in cluster 1, and Duke University and the University of North Carolina in cluster 5.

For convenience, we summarize herein the basic definitions and notations used in graph theory. Let $\mathbf{G} = (V, E)$ be a graph with a set of nodes V connected with set of edges E. The degree of a

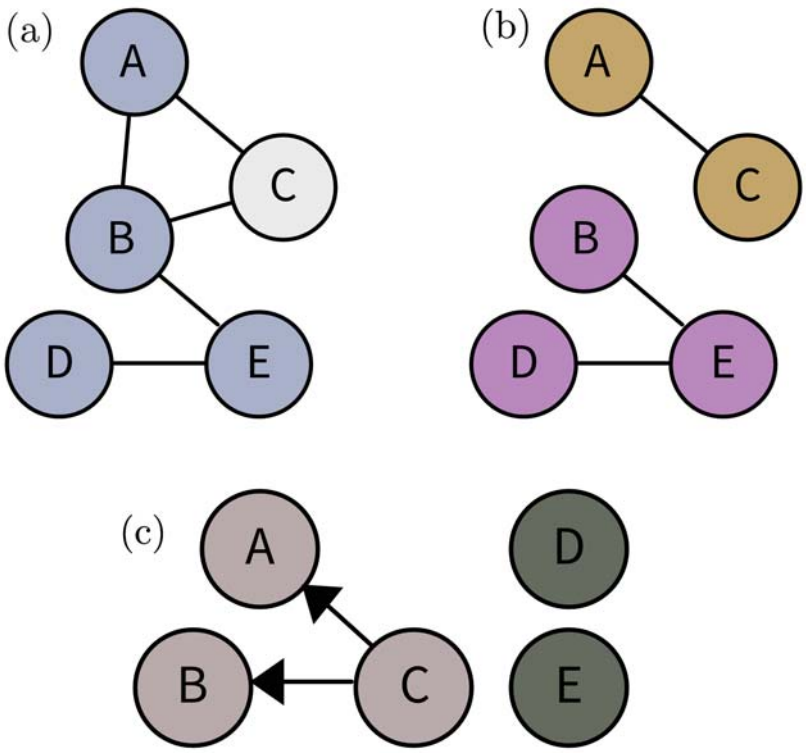

Figure 1.1 Simplified illustration of sample graphs and networks. There are two examples shown of undirected graphs in (a) and (b), and of a directed graph in (c). All three graphs have the same set of nodes $V = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$, as such, the differences between the graphs are found in the connectivity between the nodes. In graph (a), all nodes are connected; in graph (b), there are two connected components: $\{\mathrm{A}, \mathrm{C}\}$ and $\{\mathrm{B}, \mathrm{D}, \mathrm{E}\}$. In the directed graph (c), there are no strongly connected components besides the individual nodes. In graph (a), the clique number, which is the largest subset of nodes that are all directly connected, is 3, and the largest clique is A, B, C. In contrast, the largest clique in graph (b) is A, C. In graph (c), there are no cliques beyond the single nodes. The path length between node A and node E equals 2 (going from A via node C or node B towards node E). In graph (b) and graph (c), the path from node A to node E has length of infinity because they are disconnected.

node $v \in V$ is defined as the number of edges connected to the node. The "betweenness" of a node is defined as the number of shortest paths between two other nodes that passes through this node, i.e., paths in which this node is "between" two other nodes. Betweenness is a simple measure of how central the node is in connecting other nodes. For example, an airport which is a hub (e.g., IAH or JFK international airports in Houston and New York City, respectively) will have high degree and

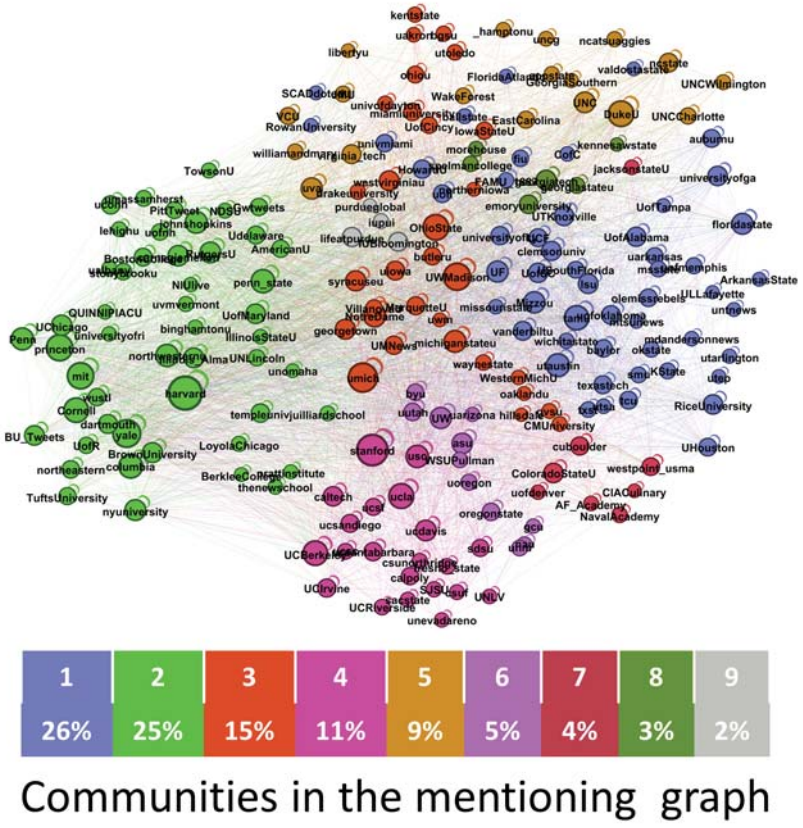

Figure 1.2 Mentioning graph of the Top 200 accounts of universities on Twitter(Appendix B). This graph is based on the analysis of all the tweets by the Top 200 Twitter accounts of universities. There is a directed edge between node $A \to B$ if the account of A tweeted and tagged node B. The universities are divided into nine modularity classes [27] based on tweeting patterns used Gephi [18]. The size of the node correspond to the input degree of the node in the mentioning graph.

high betweenness values. A "clique" in a graph is defined as a subset of nodes in V that are all connected with pairwise edges, much like a clique in real life defines a high degree of connectivity within the clique and exclusion with respect to outside of the clique. Later in Chapter 2, Chapter 3, and Chapter 4, I define order parameters that are computed on the graph representation of many physical systems.

1.3 Active matter and actomyosin

Living cells require a constant supply of energy to create and preserve biological order [196]. In order to achieve that energy, cells often extract energy by breaking the chemical bonds found in food [6].

The required intracellular processes are fairly complicated and rely on complex biochemistry in cells to control numerous biological feedback loops [6, 9]. This intracellular process is complex, and relies on complex biochemistry in cells to couple an energetically favorable reaction with an energetically unfavorable one. Overall, the total energetic cost in such coupled reaction systems will be favorable [6]. Reaction coupling enables cells to build complex assemblies of new morphologies from subunit constituent proteins, a highly energetically unfavorable process, by coupling these reactions with highly favorable ones [176].

In one biological mechanism that uses reaction coupling, actomyosin networks, found in living cells, use the active process of adenosine triphosphate (ATP) hydrolysis (an energetically favorable reaction) to assemble and contract biological structures (an energetically unfavorable one) [154]. Actomyosin is complex material composed of filamentous actin polymers (F-actin) that are connected with actin-binding proteins (ABPs) to enable a range of mechanical tasks such as cell migration and adhesion [179, 172], mechanical elasticity sensing [66], cell division [138], and the process of memory formation [123].

There are two main active processes that occur in actomyosin networks, both of which require coupling with an energetically favorable reaction to proceed. The first is the polymerization of filament actin (F-actin) from globular actin (G-actin) monomers. The second is the fact that motor proteins require ATP to walk along the F-actin polymer towards its positive end. When motors walk on the network, they exert forces that can contract [154] the actomyosin networks. In Chapter 3, we model an actomyosin network with passive linkers that connect two or more filaments, motors that can walk on filaments, and filaments without the active processes of polymerization and depolymerization. In Chapter 4, we simulate actomyosin networks with filaments that are able

to both polymerize and depolymerize.

1.4 Introduction to protein folding

Proteins are a key player for many biological processes [200], and they account for more than half the dry mass of the cell [150]. While the complexity related to the physics of proteins has many facets and active frontiers in modern scientific research, in this dissertation, the main focus is how a single protein folds under the influence of hydrodynamic interactions with the aqueous solvent, and on self-assembly processes with many proteins that form complex physical structures, i.e., in the case of actomyosin networks and the human genome.

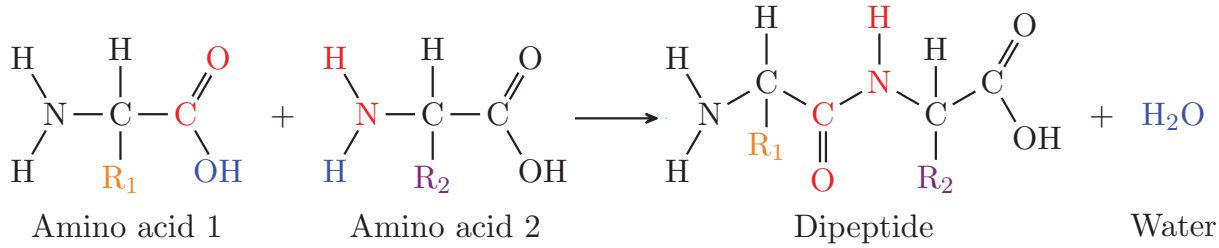

Figure 1.3 Condensation of two amino acids with the side chains R_1 and R_2 forms a peptide bond and releases a water molecule.

The building blocks of proteins are amino acids, mostly chiral monomers that contain both amino (NH_2) and carboxylic acid (COOH) functional groups. Amino acids can form long peptides (the average length of a protein is 350 amino acids [150]) by joining together amino acids via a condensation reaction to generate an amide (specifically a peptide) bond. In doing so, the two amino acids release a molecule of water and forms a dipeptide (Fig. 1.3). There are 20 different amino acids (which can be thought of as letters in an alphabet) and any string using this 20-character alphabet can plausibly be formed. Interestingly, similar sequences (strings) of amino acids may result in drastically different 3D conformations (Fig. 1.4). The amino acids, although they have very similar structures, differ in the structures of their side chains (represented as R_1 and R_2 in Fig. 1.3). The initial chain of the amino acids (residues) is formed inside the ribosome, by using the encoding properties of the transfer ribonucleic acid (tRNA) molecule that encodes for the desired protein

sequence. While being formed, the protein chain immediately begins to adopt a more complex 3D structure, which is defined as the "native structure", and represents the most stable (i.e., lowest energy) conformation. The process by which a chain of amino acid residues encodes the specific 3D native structure is called protein folding [55]. Of note, the precise sequences of proteins are encoded by the DNA, where three letters (i.e., bases) of DNA encode a particular letter in the amino acid alphabet.

The problem of understanding protein folding includes three major questions [55]: (1) How does the amino acid linear sequence govern the 3D conformation of the protein?; (2) How do proteins rapidly fold into their final configuration (referred to as their "native structures") in less than a few milliseconds [134]?; and (3) How can we use a computational model to predict the 3D protein structures from the corresponding linear sequences? Berger et al. [24] previously proved that the hydrophobic-hydrophilic (HP) lattice model is an NP-complete problem, meaning a problem that if we find an efficient solution for protein folding, we could solve efficiently any other NP problem. One of the prevalent hypotheses, which explains the rapid folding times of proteins ($\sim$ µs) in nature [97], is the hydrophobic collapse hypothesis. This hypothesis postulates that mutual attractions of hydrophobic residues reduce the number of degrees of freedom to be explored by the protein on its quest for its minimum energy native state on the energy landscape manifold. As such, water induces the enthalpic and entropic forces, such as hydrophobic collapse, that direct the protein on the global energy landscape through a funnel-like surface and towards its native structure [31, 129].

A competing hypothesis, that of prevalent hydrophobic collapse, postulates that mutual attractions of hydrophobic residues reduce the number of degrees of freedom to be explored by the protein on its quest for its minimum energy native state. Additionally, water induces the enthalpic and entropic forces such as hydrophobic collapse that direct the protein through a funnel-like surface and towards its native structure [28, 117]. There are three commonly used experiments to find the native structures of proteins: X-ray crystallography [61], nuclear magnetic resonance (NMR) spectroscopy [42], and cryo-electron microscopy (Cryo-EM) [38]. The retrieved 3D structures of thousands of proteins are available in a public database known as the Protein Data Bank (PBD) [21].

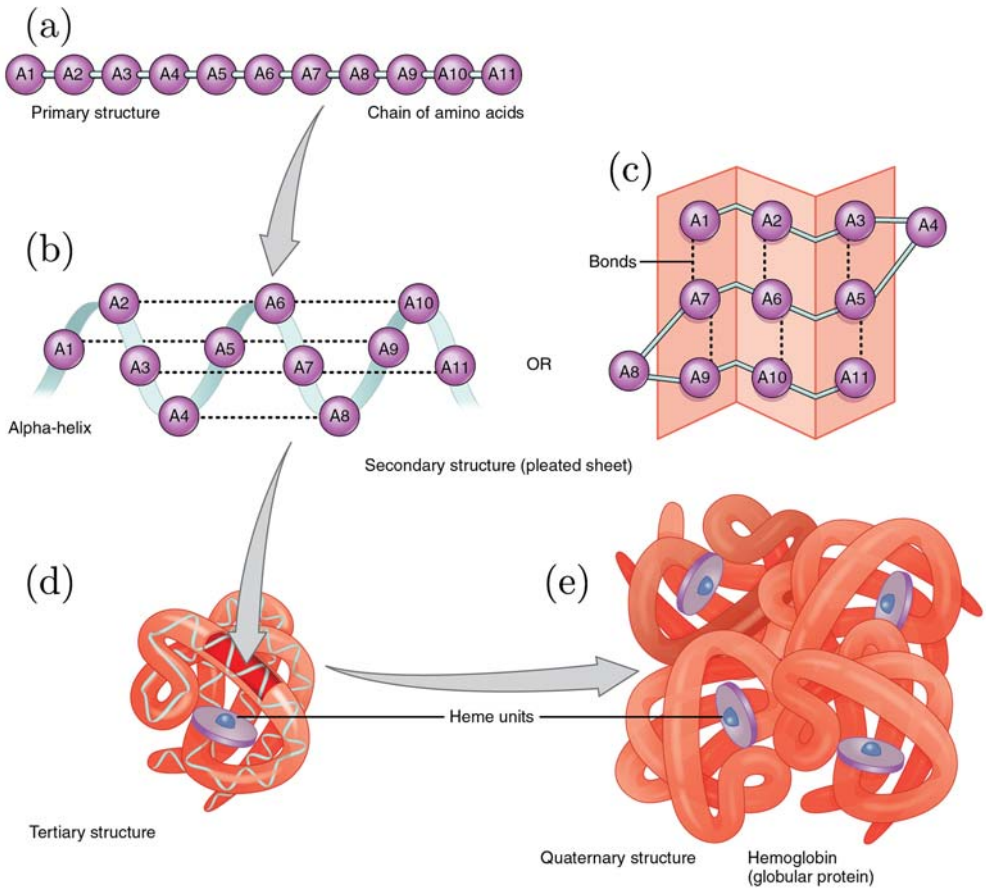

Figure 1.4 The four hierarchies in protein folding. (a) The primary structure is a linear polypeptide chain of amino acids. The two types of secondary structures are (b) α-helix and (c) β-sheets (pleated sheets). (d) The tertiary structure is formed as a result of folding and bonding within the secondary structure. (e) The quaternary structure is the result of interaction with and binding between tertiary subunits. Yet, not all proteins have quaternary structures. The example shown is of hemoglobin protein which has four tertiary subunits and is found inside red blood cells. Adapted from reference [25].

Proteins structures have typically four levels of structural organizations (i.e., hierarchies) [175]. The primary chain of amino acids Fig. 1.4(a) forms two main types of secondary structures: α-helices Fig. 1.4(b) and β-sheets Fig. 1.4(c). The tertiary structure Fig. 1.4(d) is the next hierarchy of organization between the different secondary structures formed from and within a single protein chain. While the stability of the individual secondary structures is maintained by intramolecular

hydrogen bonds, the stability of the tertiary structure is generally maintained by intramolecular electrostatic interactions between amino acid residues and analogous electrostatic interactions between residues and water molecules. The quaternary structure, the highest level of organization, occurs when multiple folded subunits of a protein form a multi-subunit protein complex, as shown in Fig. 1.4(e) for hemoglobin, a protein comprised of four distinct subunits.

1.4.1 The energy landscape theory (ELT)

ELT can be used to understand the complex nature of protein folding. In particular, ELT postulates that a protein in a random unfolded state is funneled on a multidimensional energy landscape towards the configuration with the lowest configurational entropy and minimum free energy. The minimum of the energy landscape is called the protein's native state. ELT assumes that the folding dynamics occurs on a multidimensional landscape and can be reduced to one or a few reaction coordinates without losing crucial information about the complex system.

When a protein navigates down the funnel, there is a mismatch between the configurational entropy and the free energy, which results in a free-energy barrier that separates the "unfolded" states from the "folded" states (usually a few k_BT, where k_B represents the Boltzmann constant and T represents the temperature). Since the activation barrier depends on the global system temperature, this leads to a range of temperatures at which the protein could fold in biological systems, with the rate of folding highly dependent on the system temperature. Using the collective fraction of native amino acid contacts, $Q \in [0, 1]$ as the reaction coordinate for a two-state protein (folded/unfolded) allows us to generate an illustration of the energy landscape, depicted for an unfolded protein in Fig. 1.5. Of note, along the pathway towards energy minimization, there are large numbers of local energetic minima that may trap the protein. Nonetheless, the native state of the protein is still accessible due to the principle of minimal frustration, which states that all configurations and energetic differences evolved in order to facilitate a first-order phase transition for the protein from an unfolded to a folded state [32].

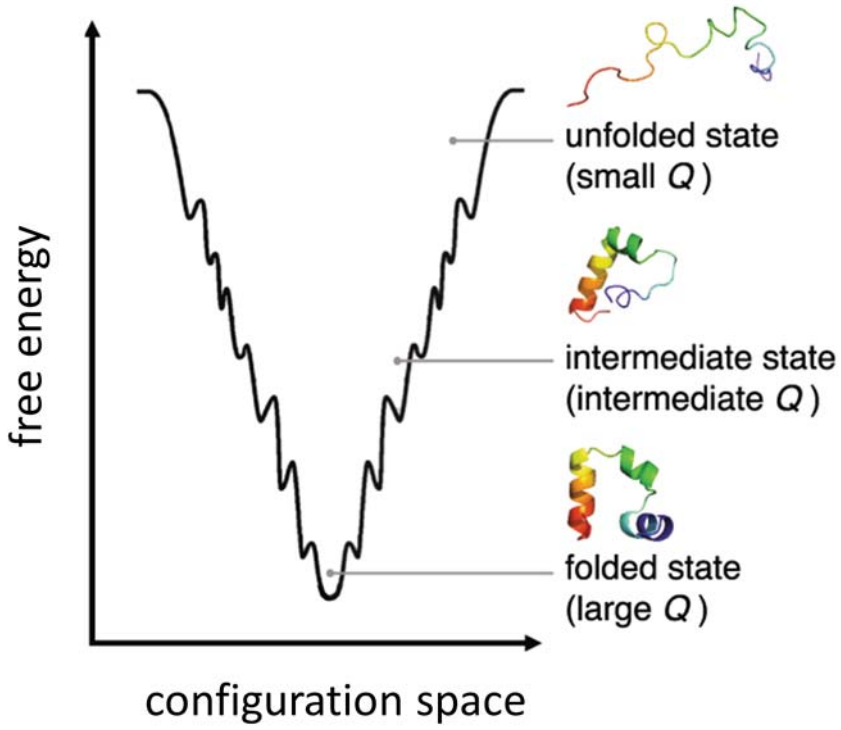

Figure 1.5 A funnel diagram of the energy landscape of a protein. The width represents the configurational entropy of the protein and the depth represents the energy of the protein. An unfolded structure navigates through the energy landscape towards the folded state (native state) located at the lowest global minimum of the funnel. The small peaks represent kinetic traps. The figure was adapted from [45].

1.5 The human genome structure as a graph

In the case of the human genome, which can be defined as a hierarchically organized complex polymer, there are two major groups of proteins that allow the compact folding and regulation of the three billion DNA base pairs found in the micron-sized nucleus of the cell. The genome structure is also referred to as chromatin (DNA bound to RNA and protein molecules) remodeling. Of note, this highly complex structure is responsible for a broad variety of functions, including messenger ribonucleic acid (mRNA) transcription, DNA replication, and DNA repair. Within the genome, the smallest subunit that packs 146 base pairs of DNA is the nucleosome [30, 166]. The nucleosome is shaped like a spool made of eight highly packed histone proteins [143] around which the 146

base pairs of DNA are wrapped. Fig.1.6 illustrates the different hierarchical ordering in the human genome: starting from DNA base pairs, through the formation of a double stranded DNA helix, to the packing of of the DNA around the nucleosomes, all the way to the formation of chromatin (packed DNA that is bound to protein) and the chromosome.

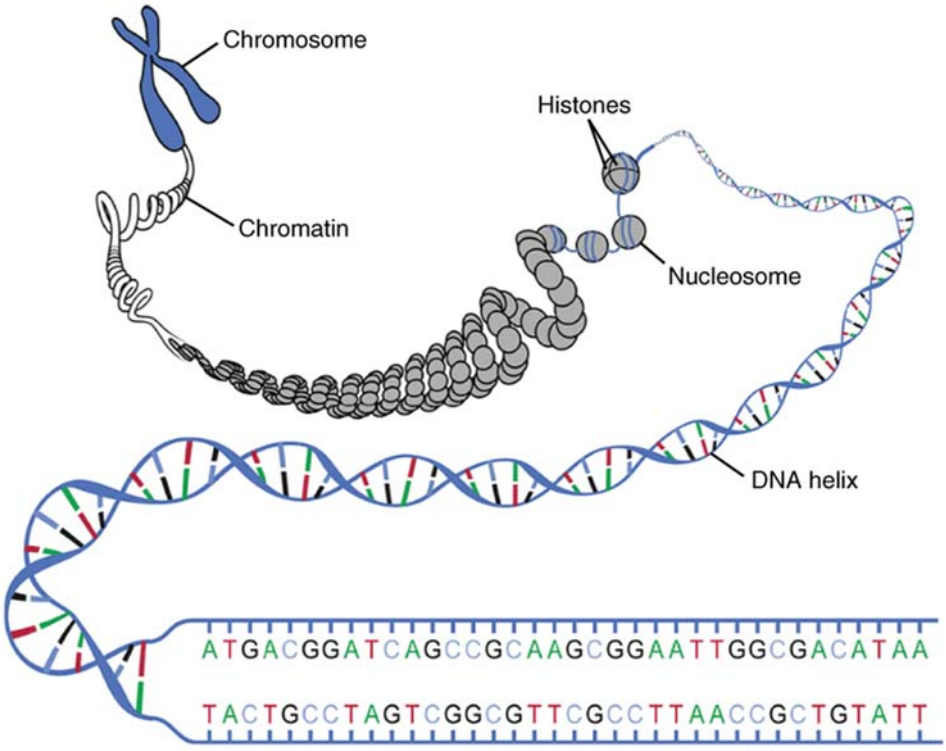

Figure 1.6 The hierarchy of DNA macrostructure, and how it is assembled and wrapped around histone proteins, which bundle together to create condensed chromatin, which in turn is tightly packed into a chromosome. Adapted from [25].

The higher order organization of how nucleosomes pack into chromatin and chromosomes is still under active debate in the field of genomics, with a few models invoked to explain this structure [164]. One such model, which is particularly widespread, is based on electron microscopy (EM) and X-ray results, which suggest that nucleosomes are packed into a solenoidal/helical fiber with a diameter of 30 nm [193, 192]. An *in vitro* reconstitution of chromatin fibers has suggested that nucleosome compaction varies under different epigenetic conditions, such as histone modifications, and the binding of chromatin-associating proteins to the genome [130]. In parallel with extensive research

of how nucleosomes organize along the chromatin fiber, there was an appreciable improvement in the understanding of chromatin conformation using next-generation DNA sequencing. One of the leading methods for this sequencing-based method is the high-throughput chromatin conformation capture (Hi-C) technique, which has revealed the higher order conformation of the entire genome, with particular information provided about loops, domains and subcompartments that are regulated by proteins and other epigenetic conditions [132, 181]. Of note, Fig. 1.7(a) shows a typical Hi-C map of a human blood cell, with each box in Fig. 1.7(a) representing interactions between two different chromosomes. For example, the intrachromosomal interactions between the loci of chromosome 2 are enlarged in Fig. 1.7(b), and the interaction between different chromosomes (chr1 and chr14) is shown in Fig. 1.7(c).

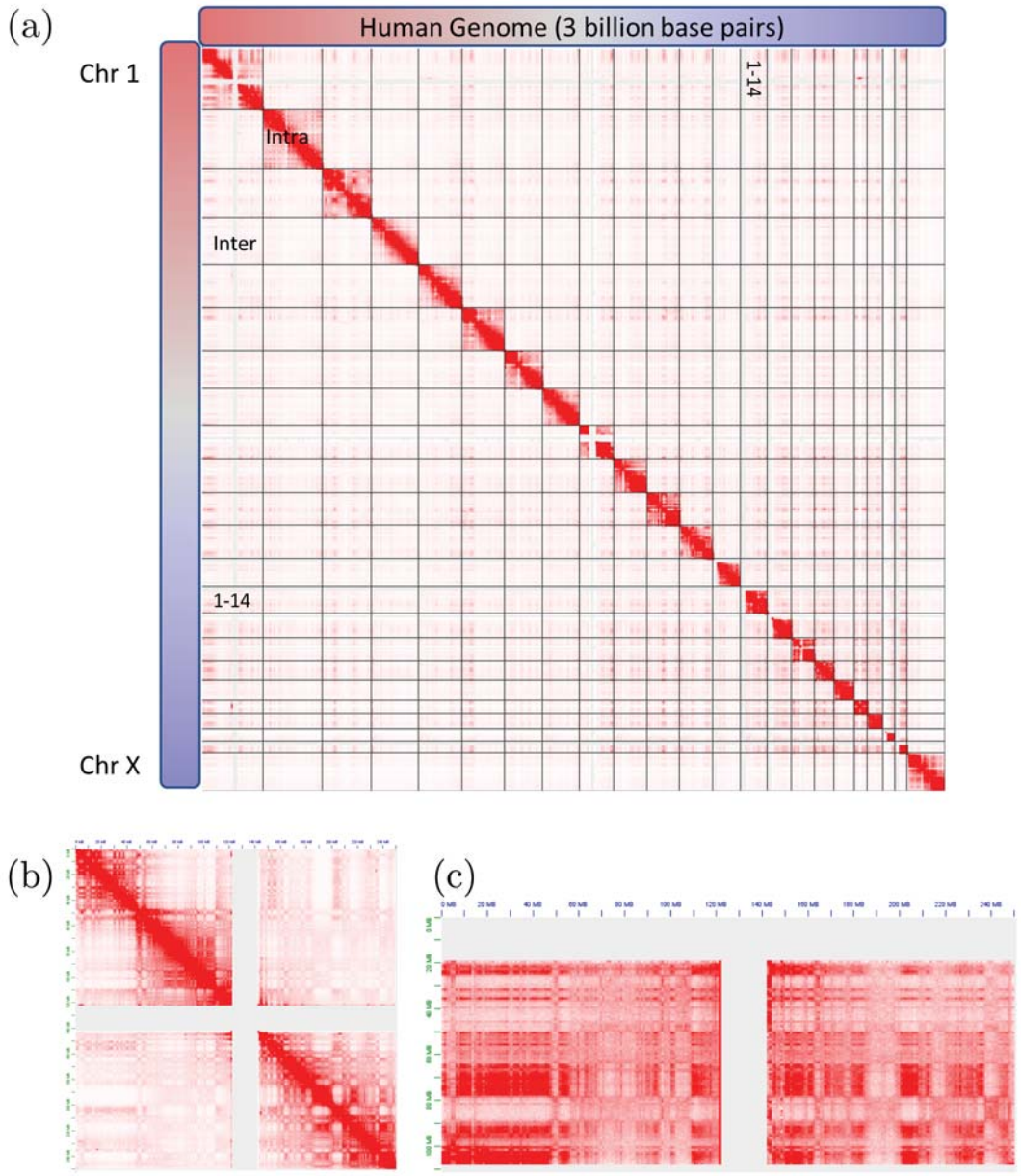

Figure 1.7 The Hi-C map of the GM12878 human cell line. (a) The entire Hi-C map of the human genome, where each rectangle corresponds to the interaction between two chromosomes. (b) Hi-C map of the intrachromosomal interactions of "chromosome 2" at 1 mbp resolution. (c) Interchromosomal interaction Hi-C map between chromosome 1 and chromosome 14 at 1 mbp resolution.

2 Studying biological physics of matter from the vista of graph theory

This chapter is based on an arXiv preprint [64] published in 2020, and a talk at the APS (American Physical Society) conference [63] in 2018.

2.1 Introduction and motivation

Graph theory is a power formalism that originated from discrete mathematics and is used to represent systems with countable elements (discrete systems). Since Euler's era, graph theory has been widely adapted by computer scientists, applied mathematicians, and abstract mathematicians. Although notable polymaths, such as Euler and Hamilton, developed graph theory in the 18th and 19th centuries, only in the late 1990s did physicists started to use this formalism to study complex systems. Since that time, physicists have contributed tremendously to graph theory and developed its synonym, network theory [4, 50]. Because much of the field of physics was founded on a backbone of differential equations and calculus, physics subfields, including general relativity, electrodynamics, quantum mechanics, and classical mechanics, rely on similar mathematical underpinnings.

Many of these theories rely on a set of differential equations that has advanced our understanding of our world, and our technology. One example of such technology is the development of the engine, which led to a need to understand macroscopic thermodynamic properties of engine systems. The demand to reduce the complexity of such a system, containing an immense number of particles (ca. on the order of Avogadro's constant $N_A = 6.0221408 \times 10^{23}\,\mathrm{mol}^{-1}$ [12]), led to the evolution of statistical physics, which treats the large number of degrees of freedom statistically.

The need for further dimensionality reduction for enhanced system understanding dramatically intensified with the development of computers, large man-made networks, as well as the advancement in biological experiments using optical devices and next-generation sequencing (NGS) techniques.

Although there are intrinsically irreducible complex systems and computationally intractable problems (Chapter 1), for almost any type of complex system or problem we can compute an

insightful quantity to assess the rough yet statistically valid state of the system. This quantity could be some statistically-driven aggregate signal or some local property of the complex system. Since graph theory formalism is a generic way to describe discrete systems, I describe herein four examples of physical systems with their graph representations. We start with a simple continuous version of the Cahn-Hilliard equation [35, 36, 162] that promotes phase separation [28] in continuous media, then we simulate a discrete system that undergoes phase separation using off-lattice dissipative particle dynamics (DPD) [100, 71]. Later, I demonstrate how graph theory representation lays the foundation that enables accurate predictions of the 3D structures of a folding protein, which can be represented as a contact map (adjacency matrix) between amino acid residues in the native structure. We also delineate briefly how an actomyosin network can be mapped as a graph representation, based on pairwise distances between filaments or based on the connectivity of the filaments using actin-binding proteins (ABP). Lastly, I explain how Hi-C maps can be represented as weighted graphs, i.e., graphs in which each edge has a real number weight associated with it.

2.2 Graph representation of complex biological matter

In this section, I list four examples of how to map a complex system into graphs. Once a representation graph is built for a physical system, an array of order parameters can be derived from the graph.

2.2.1 Phase separation

Cells organize their complex biochemical reactions into compartments to create distinct chemical regions. Each intracellular compartment then has boundaries within which the chemical compounds can diffuse freely [105]. Phase separation is the term used to describe the physical processes that form many of these compartments and divide the space of the cell into regional independent authorities that can interact with the environment at their boundary positions [5].

Phase field model (continuous) of phase separation

The Cahn-Hilliard equation [35, 36, 162] is a simple continuous model that simulates the process of

phase separation in a mixed liquid solution.

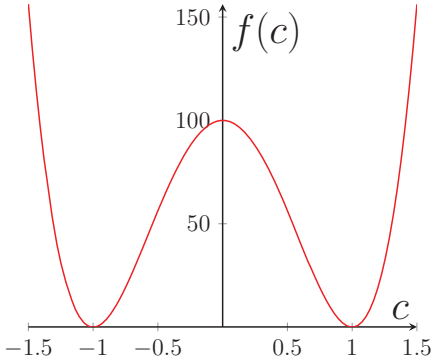

Figure 2.1 Equilibrium free energy profile of the Cahn-Hilliard which drives the system towards phase separation with two states: $c = -1$ and $c = 1$.

The underlying equation for this model can be represented as shown below:

$$\begin{aligned} \frac{\partial c}{\partial t} &= D\nabla^2\mu \\ \mu &= \frac{df(c)}{dc} - \nabla^2 c, \end{aligned} \tag{2.1}$$

where $c \in [-1, 1]$ is the concentration of the fluid also named the "phase field", D is the diffusion coefficient, μ is the chemical potential, $f(c) = 100(1-c)^2(1+c)^2$ plotted in Fig. 2.1, and $\sqrt{\gamma} = \sqrt{0.01}$ is defined as a characteristic length of the transition. We solved Eq. (2.1) on the unit square $\Omega \in (0, 1) \times (0, 1)$ using the finite element method in FEniCS [139, 8]. I ran the system in Eq. (2.1) for 125 seconds and visualized the system at $t \in [0, 25, 50, 75, 100, 125]$ seconds, as shown in Fig. 2.2. The results demonstrate how a sample mixed liquid is expected to undergo phase separation.

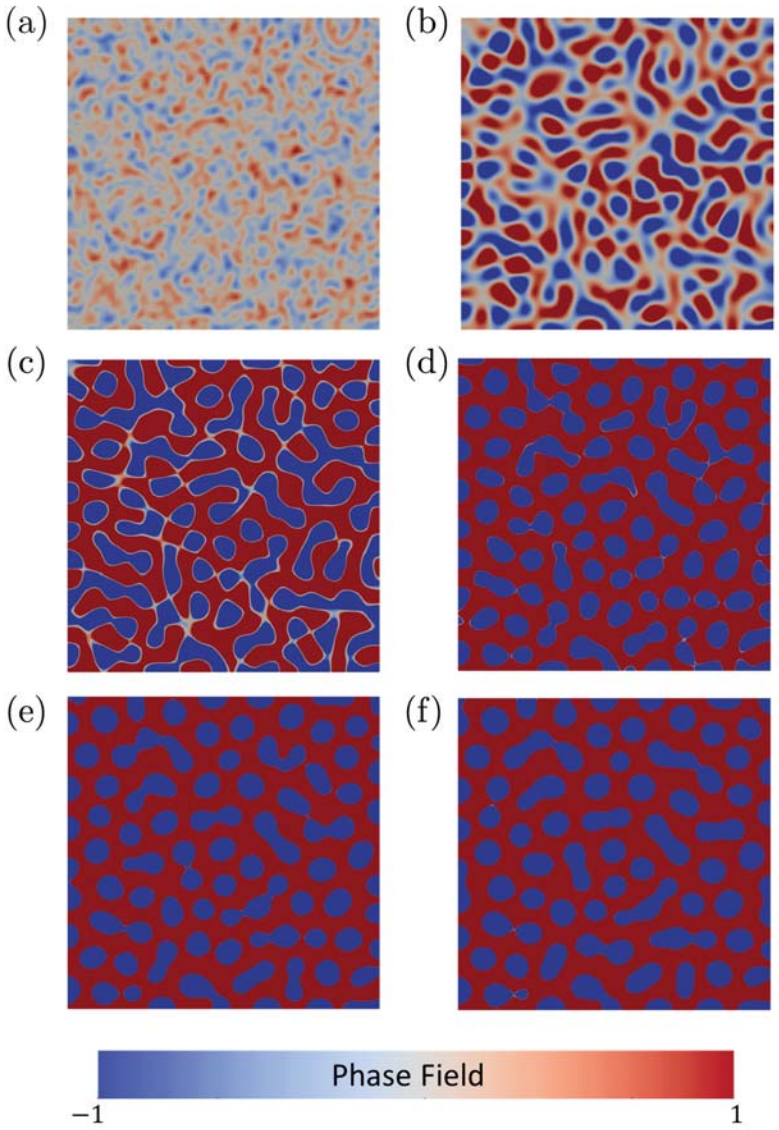

Figure 2.2 Finite element solution of Cahn-Hilliard equation using FEniCS [139, 8]. The phase field $c(x, y)$ is plotted versus $(x, y) \in \Omega$ at the times (a) $t = 0\,\mathrm{s}$, (b) $t = 25\,\mathrm{s}$, (c) $t = 50\,\mathrm{s}$, (d) $t = 75\,\mathrm{s}$, (e) $t = 100\,\mathrm{s}$, and (f) $t = 125\,\mathrm{s}$. The figure goes from (a) a mixed liquid to (f) a completely phase-separated system.

Discrete simulation of phase separation using dissipative particle dynamics (DPD)

In order to show how graph theory can assist in building a meaningful reaction coordinate or order parameter, we set up a binary mixture system using two types of DPD particles [72] in the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [177, 91]. The relevant

force field equations for this system are defined below:

$$\mathbf{F}_{ij}^{\text{tot}}(r) = \begin{cases} \left(F^C + F^D + F^R\right)\hat{\mathbf{r}}_{ij}, & r < r_c \\ 0, & \text{otherwise} \end{cases}$$
$$F^C = Aw(r)$$
$$F^D = -\gamma w^2(r)\left(\hat{\mathbf{r}}_{ij} \cdot \mathbf{v}_{ij}\right) \tag{2.2}$$
$$F^R = \sigma w(r)\alpha(\Delta t)^{-1/2}$$
$$w(r) = 1 - r/r_c,$$

where $\mathbf{F}_{ij}^{\text{tot}}$ is the total force between particles i and j, including a conservative force F^C, a dissipative force F^D, and a random force F^R. $\hat{\mathbf{r}}_{ij}$ is defined as the unit vector between two particles i and j in the direction $\mathbf{r}_i - \mathbf{r}_j$ and magnitude $r := |\mathbf{r}_i - \mathbf{r}_j|$. $\mathbf{v}_{ij} = \dot{\mathbf{r}}_i - \dot{\mathbf{r}}_j$ is the vector difference of the velocities. α is a Gaussian random number with zero mean and unit variance, dt is the timestep, and w(r) is defined as a weighting factor that varies between 0 and 1 . r_c is the cutoff distance. $\sigma := \sqrt{2k_BT\gamma}$, where k_B represents the Boltzmann constant and T is the temperature parameter.

As part of the simulation conditions, the temperature, the volume and the number of particles within the system are all held constant. There were $4,050$ "Type 1" particles and $20,000$ "Type 2" particles investigated. The simulated system considered three types of DPD pairwise interactions, enumerated below:

1. Intra-type DPD force field between two Type 1 particles
2. Intra-type DPD force field between two Type 2 particles
3. Inter-type DPD force field between a Type 1 particle and a Type 2 particle

The difference between the three types of DPD interactions was in their interaction cutoff lengths: the DPD cutoff distance for the inter-type $r_c^{(1-2)}$ was twice as large as the cutoff distance observed for intra-type particle interactions, i.e., $r_c^{(1-1)}$ and $r_c^{(2-2)}$ [53, 163]. I developed three graphical representations from which I derived three possible reaction coordinates that capture the phase separation phenomena in the discrete system of the DPD particles, as it evolves from a system with

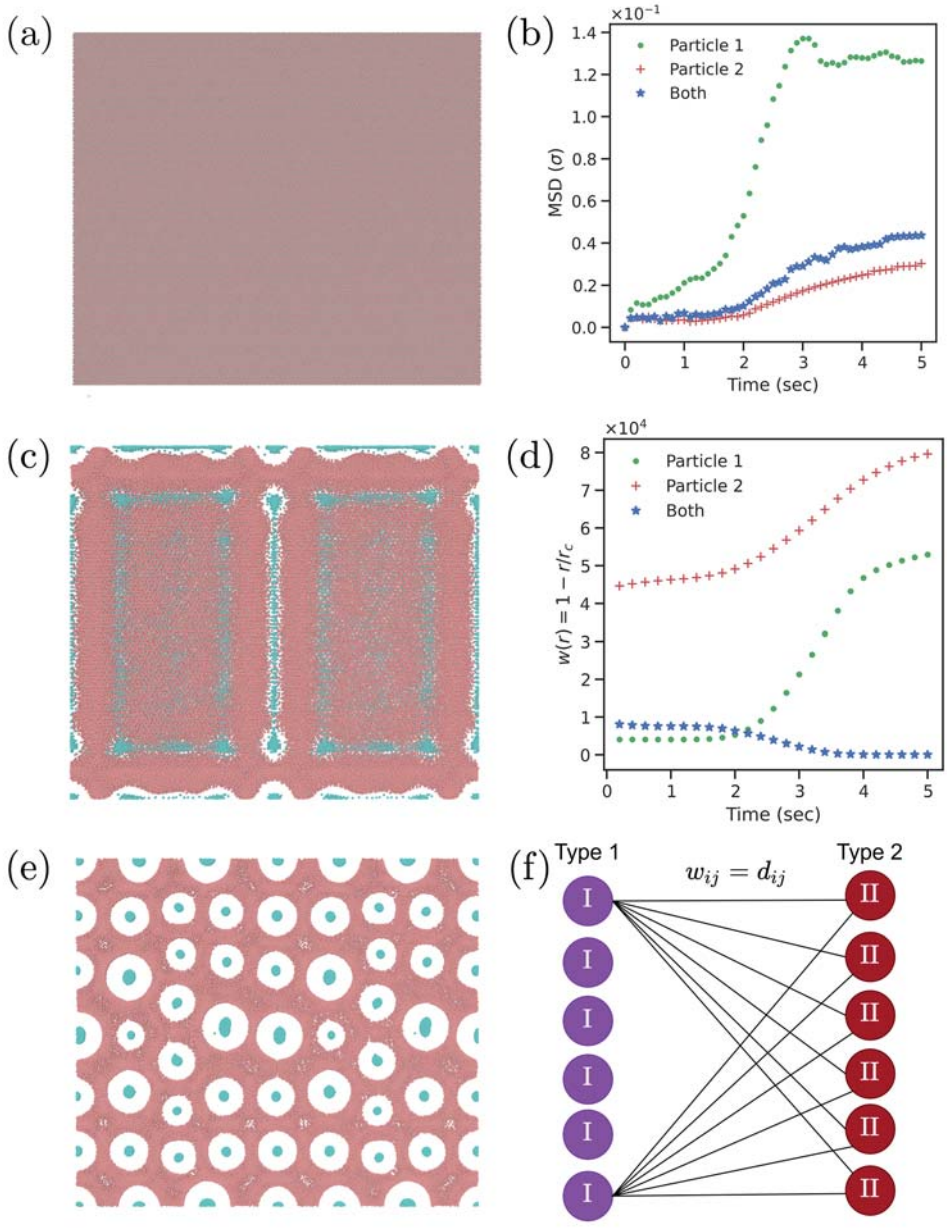

Figure 2.3 Simulations of phase separation. (a) A system with two types of DPD particles at $t = 0$, (c) $t = 2.5\,\mathrm{s}$, and (e) $t = 5\,\mathrm{s}$. (b) The time signal of the mean squared displacement (MSD), as defined in Eq. (2.4), plots, when the length units are normalized with the interaction length constant σ. (d) The time signal of the value of the cut of the bipartite graph representation of the system as function of time. (f) Illustration of a bipartite graph, where the cut is the sum of all the weights of the edges between the L (Type 1) and R (Type 2) nodes in the graph.

uniformly distributed two types of DPD particles in Fig. 2.3(a) into a phase-separated system in Fig. 2.3(e).

These bipartite graphs are defined as $\mathbf{G}_{\text{bipartite}} = (L, R, E)$, where the usual set of nodes V in $G = (V, E)$ is split into two sets of disjoint nodes L (represented as Type 1 (I) in Fig. 2.3) and R (Type 2 (II) in Fig. 2.3). In this bipartite graph representation, edges are allowed only between two nodes that are elements in the two different sets L and R. If there is a node between $i \in L$ and $j \in R$, then $\{i, j\} \in E$. The bipartite graphs are equipped with a weight function that assigns values to edges, for example, $d(i, j) : L \times R \mapsto \mathbb{R}$. In particular, given two nodes $i \in L$ and $j \in R$, I assign the a value to the undirected edge (i, j). In the particular case of DPD simulations, I refer to the three bipartite graphs with the weight function $d(i, j)$ taken to be $w(r) = 1 - r/r_c$ from Eq. (2.2) if $r \leq r_c$ and 0 otherwise. The weight function reads

$$d(i,j) = \begin{cases} w(r) = 1 - r/r_c, & r \leq r_c \\ 0, & \text{otherwise} \end{cases} . \tag{2.3}$$

Overall, the three types of bipartite graphical representations are:

1. $\mathbf{G_{1-1}} = (L_1, R_1, E)$ is the representation graph of the interactions between "Type 1" particles. For each "Type 1" particle in the system, I assign two nodes at both the left L_1 and the right R_1 sets of nodes, and the weight of edges between any two nodes $l \in L_1$ and $r \in R_1$ is set to be $d(l \in L_1, r \in R_1)$, as defined in Eq. (2.5).

2. $\mathbf{G_{2-2}} = (L_2, R_2, E)$ is the representation graph of the interactions between "Type 2" particles. For each "Type 2" particle in the system, I assign two nodes at both the left L_2 and the right R_2 sets of nodes, and the weight of edges between any two nodes $l \in L_2$ and $r \in R_2$ is set to be $d(l \in L_1, r \in R_2)$, as defined in Eq. (2.5).

3. $\mathbf{G_{1-2}} = (L_1, R_2, E)$ is the representation graph of the interactions between "Type 1" and "Type 2" particles. For each "Type 1" particle in the system, I assign a node in L_1 and for each "Type 2" particle, I assign a node in R_2. The weight of the any edge between two nodes $l \in L_1$ and

$r \in R_2$ is set to be $d(l \in L_1, r \in R_2)$, as defined in Eq. (2.5).

From each of the three graphs defined above, I use the cut of the graph order parameters that demonstrate how the order parameter from the graph representation of the original system can capture the process of phase separation.

To compare the diffusion of the particles that form the droplets with the particles that form the background in the phase-separated system (Fig. 2.3(e)), I plot the mean square displacement (MSD), defined in Eq. (2.4), for a Type 1 particle and a Type 2 particle in Fig. 2.3(b) to show how they diffuse in the system while undergoing phase separation. "Type 1" particle is either in the super-diffusion or the normal diffusion regime and the "Type 2" particle diffuses in a subdiffusive regime. This is due to the fact that the "Type 1" particles will form the droplets shown in Fig. 2.3(e).

$$\text{MSD} := \langle |\mathbf{r_i}(t) - \mathbf{r}_i(t=0)|^2 \rangle, \tag{2.4}$$

where the average $\langle \cdot \rangle$ is calculated using over $1,000$ randomly chosen particles, without replacement, in the three cases: Type 1, Type 2, or both Type 1 and Type 2 simultaneously. The particles of Type 1 diffuse more rapidly than the Type 2 particles. Additionally, I computed the cut of the graph, which is defined as the sum of all the graph's edges between the left L and right R sets of nodes. Therefore, a cut $C(\mathbf{G})$ of a bipartite graph $\mathbf{G} = L, R, E$ is defined as:

$$C(\mathbf{G}) := \sum_{i \in L, j \in R} d(i,j), \tag{2.5}$$

where $d(i,j)$ is the weight of the edge between i and j. In Fig. 2.3(b) are the plots of the graph cut $C(\mathbf{G})$ for the three cases: $\mathbf{G_{1-1}}$, $\mathbf{G_{2-2}}$, and $\mathbf{G_{1-2}}$. In the three curves, there is a sigmoidal shape to the curve starting to emerge after $t = 2\,\text{s}$, which capture the phase separation using a graph-based order parameter. This simple example of a system of two types of DPD particles that undergoes phase separation is an example of how a graph representation of a physical system may assist in developing straightforward order parameters to capture nontrivial dynamics and to translate them into a time signal of a scalar.The ideas of coarse-graining and dimensionality reduction are crucial

in physics since they help to develop a simplified understanding of the physical system with many particles, while preserving and capturing the essence of the phase transition or the dynamics of the complex system.

2.2.2 Conformation of proteins

Since the start of *in-silico* protein modeling in 1975, scientists have been using a graph-like representation named "contact maps" [128]. The idea of using contact maps for folded structures has been widely adapted in the field of protein folding [168, 200], and allowed for the easy adaptation of deep learning methods for protein folding and the evolutionary study of how proteins coevolve across biological species, through analysis of their contact maps [151]. There are many ways to represent the conformation of an amino acid chain through dimensionality reduction on a graph, most of which utilize pairwise distances between the α-carbon and or the β-carbon atoms of each amino-acid [205].

Although the distance-based contact maps provide concise structural information, they do not consider specific secondary interactions within the protein. These secondary interactions include hydrophobic interactions, van der Waals forces, electrostatic attraction and repulsion, intermolecular hydrogen bonds and intramolecular covalent bonds. Whereas most secondary interactions between residue i and j are symmetric with respect to the replacement of chain j and chain i (i.e., non directional), both hydrogen bonds and covalent bonds are highly directional. For example, a hydrogen atom in chain i is attracted to an electronegative atom (usually nitrogen or oxygen) in chain j to form a directional hydrogen bond; the direction of this bond cannot easily be reversed. Moreover, in contrast to covalent bonds that are usually established between adjacent residues to form the initial peptide chain, hydrogen bonds are established between distant amino acid residues and considered a major driving force for 3D structure stabilization and flexibility. Since hydrogen bonds reflect non-trivial structural information, the idea of describing the secondary structure (Fig. 2.4(a)) of a protein by visualizing the hydrogen bonding interactions (HB) plots [26] between the protein's chain of residues representation (Fig. 2.4(b)) was also adapted (Fig. 2.4(c)),

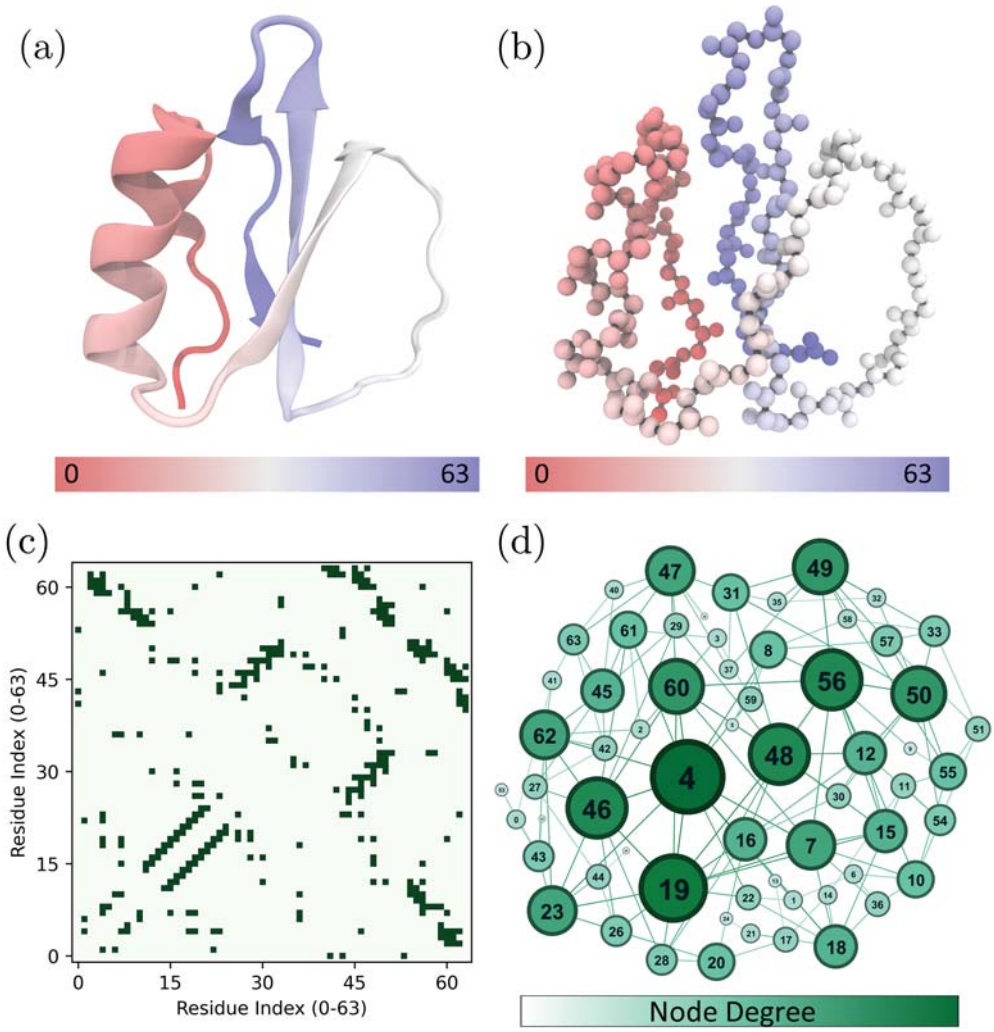

Figure 2.4 Representations of the chymotrypsin inhibitor 2 (CI2) protein. (a) Cartoon representation of the CI2 protein secondary structure using the Visual Molecular Dynamics (VMD) program [102]. (b) The chain of beads (residues) coarse-grained representation of the CI2 protein using VMD [102]. (c) The contact map of the CI2 protein. (d) The corresponding undirected graph representation of the CI2 contact map, where the node color and size corresponds to the degree of the protein, as visualized in Gephi [18].

where the undirected graph is used to indicate the presence of a hydrogen bond between chain i and chain j (Fig. 2.4(d)).

2.2.3 Morphology in actomyosin networks

We use two general graph-based approaches to describe the morphology of a generic actomyosin network: (a) a connectivity-based graph that represents how the actin-binding proteins (ABPs) connect the different actin filaments (F-actins) in the network; and (b) a proximity-based graph that delineates the pairwise distances among the different filaments in space. Moreover, our framework is highly relevant for comparisons between biophysical theories and experiments in which usually only F-actin is visible via fluorescence microscopy. For the ABP connectivity-based graph approach, we consider each filament as a single entity (i.e., a filament represents one node in the graph) or its monomers as single entities (i.e., one monomer represents a node in the graph), depending on the coarse graining level of the simulator.

In one example, as seen in Chapter 3, I assign a node in a graph $\mathbf{G}$ for each filament, while in a finer coarse graining simulations, as presented in Chapter 4, where the monomers of the filament can polymerize and depolymerize and the level of details of the actomyosin species is in the monomer resolution, we build a graph where each node in the graph corresponds to a single actin monomer. Fig. 2.5(a)-(d) show how the actomyosin system evolves when passive ABP linkers are bivalent from time $t = t_0 = 1\,\mathrm{s}$ (Fig. 2.5(a)) to $t = 600\,\mathrm{s}$ (Fig. 2.5(b)) along the reduced graph of filaments, based on proximity with $d_{\mathrm{cutoff}} = 200\,\mathrm{nm}$ in (Fig. 2.5(b),(d). The effect of a passive ABP linker with a multivalency of six is depicted in Fig. 2.5(e)-(h) and compared with the bivalent morphologies in Fig. 2.5(a)-(d).

2.2.4 Inter and intra chromosomal subgraphs in the human genome multigraph

The human genome representation as a Hi-C map is a multigraph (i.e., it is permitted to have multiple edges between nodes) that accounts for both intrachromosomal and interchromosomal interactions. In particular, the Hi-C map measures how frequently two genomic loci were in closed proximity in the folded genome. For simplicity, I converted the original adjacency matrix, which contains integer values depending on the frequency of interactions between loci, to a binary adjacency

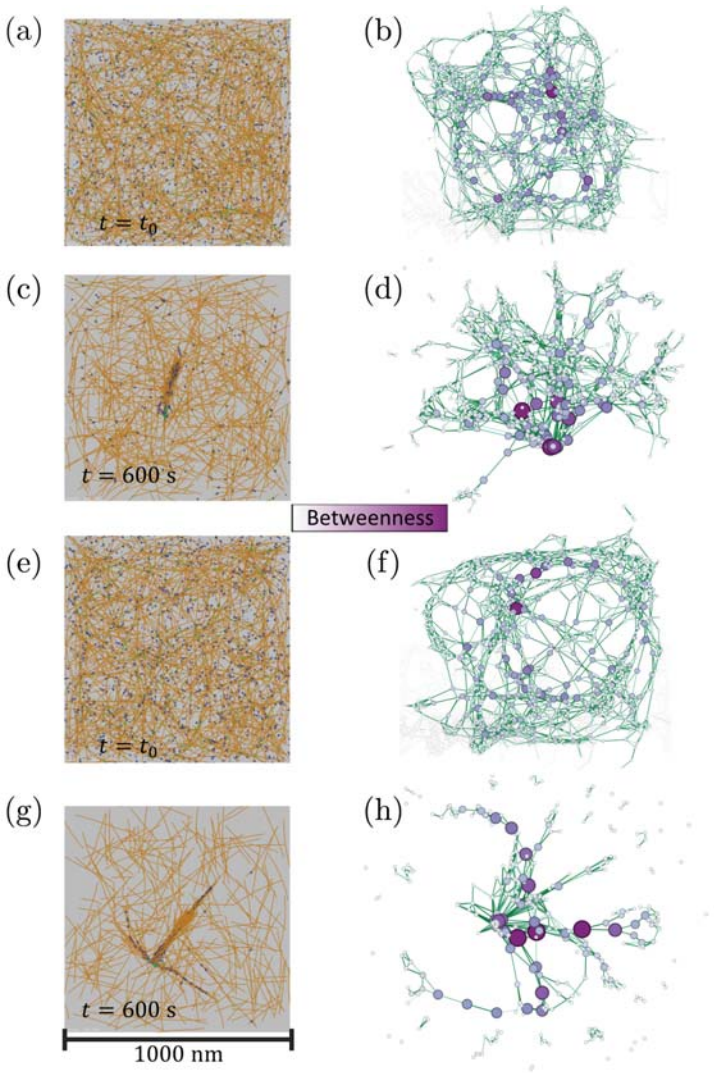

Figure 2.5 Snapshots of actomyosin networks and their graph representations. In (a) and (c), there are linkers that connect two filaments at most, and (b) and (d) are their graph representations, respectively. Paths (e) and (g) are simulations at $t = 1\,\mathrm{s}$ and at $t = 600\,\mathrm{s}$ of actomyosin with linkers that can bind up to six filaments at the same time, and (f) and (g) are their graph representation, respectively. All the graph layouts in (b), (d), (f), and (h) were visualized following the same procedure: First, I produced a random network layout with OpenOrd, and then ran Yifan Hu's algorithm in Gephi [18]. The betweenness of a node is defined as the number of times that the node lies on the shortest path between any two other nodes in Eq. (24).

matrix. For example, to build a simplified subgraph of a subregion in the genome, I converted the frequency-based adjacency matrix into a binary adjacency matrix by setting to 1 entries in the matrix if their value is above some threshold and to 0 if the their value equal to or less than

the threshold. For this section, I worked at a resolution of 10^6 nucleotide base pairs (1 mbp), so each node on the graph corresponds to the aggregated signal from a locus of 10^6 contiguous DNA bases along the chromosome in question. I built a graph for the intrachromosomal region of the largest chromosome in the human genome, represented by the abbreviation "chr1", which has a measured length 249.2 mbp. The raw signal from the Hi-C experiment (represented as the frequency adjacency matrix in Fig. 2.6(a)) is converted into a binary matrix (Fig. 2.6(b)) after applying the threshold θ_{intra}, which leads to the graph G shown in Fig. 2.6(c).

For the case in which inter-chromosomal interactions were investigated, I used a bipartite graph with each side representing one of the two chromosomes, e.g., chromosome 1 and chromosome 14. Of note, in this particular case, the length of chromosome 14 (chr14) is less than half the length of chromosome 1 (chr1). Specifically, I first visualized the bipartite graph as a regular graph using Fruchterman Reingold algorithm [81, 18], and the results of this graph visualization indicated four different subcompartments at which chr14 (represented by green nodes in Fig. 2.5(a) and on the left in Fig. 2.5(b)) forms with chr1 (represented by pink nodes in Fig. 2.5(a) on the right side of the bipartite graph in Fig. 2.5(b)). These graph representations of interchromosomal interactions demonstrate how chr1 and chr14 interact with each other. There are compartments that highly interact (i.e., green stripes in Fig. 2.5(d) and with close proximity in Fig. 2.5(a)) and there are compartments that evade each other (i.e., white stripes in Fig. 2.5(d)).

To construct the intrachromosomal graph $\mathbf{G} = (V, E)$ in Fig. 2.7, I used the cutoff value, according to the relation below:

$$\vartheta_{\text{intra}}^{\text{cutoff}} = \frac{\sum_{i=0}^{|V|-1} \mathbf{A}_{ii}}{|V|^2}, \tag{2.6}$$

where $\mathbf{A}$ is adjacency matrix of the intrachromosomal graph $\mathbf{G} = (V, E)$.

To construct the bipartite interchromosomal graph $\mathbf{G} = (R, L, E)$ in Fig. 2.5, I used the cutoff value defined according to the relation below:

$$\vartheta_{\text{inter}}^{\text{cutoff}} = \frac{\sum_{i=0}^{|L|-1} \sum_{j}^{|R|-1} \mathbf{A}_{ij}}{|L| \cdot |R|}, \tag{2.7}$$

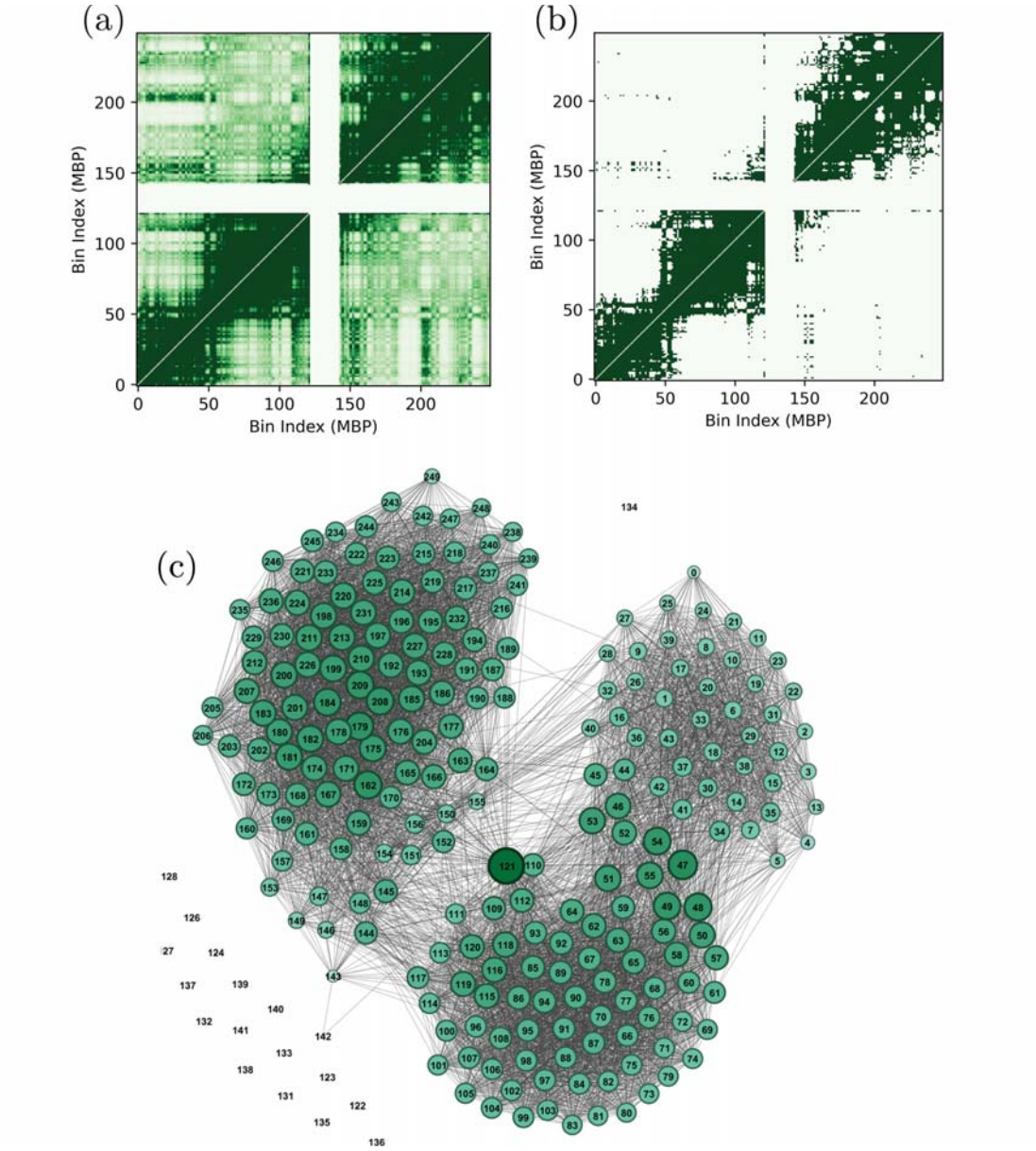

Figure 2.6 Depiction of chromosome 1 of the human genome as a graph. (a) The intrachromosomal contact map of chromosome 1 from Rao et al. [181] at 1 mbp resolution. (b) The binary adjacency matrix of chromosome 1 after applying a threshold based on the average signal along the diagonal of (a). (c) The graph of chromosome 1 is shown based on the values of the binary adjacency matrix in (b). Note that the sizes of the nodes in graph (c) are scaled according to the betweenness values of each node.

where $\mathbf{A}$ is adjacency matrix of the interchromosomal graph $\mathbf{G} = (R, L, E)$.

Using the Clauset-Newman-Moore greedy modularity maximization algorithm [48], I found that

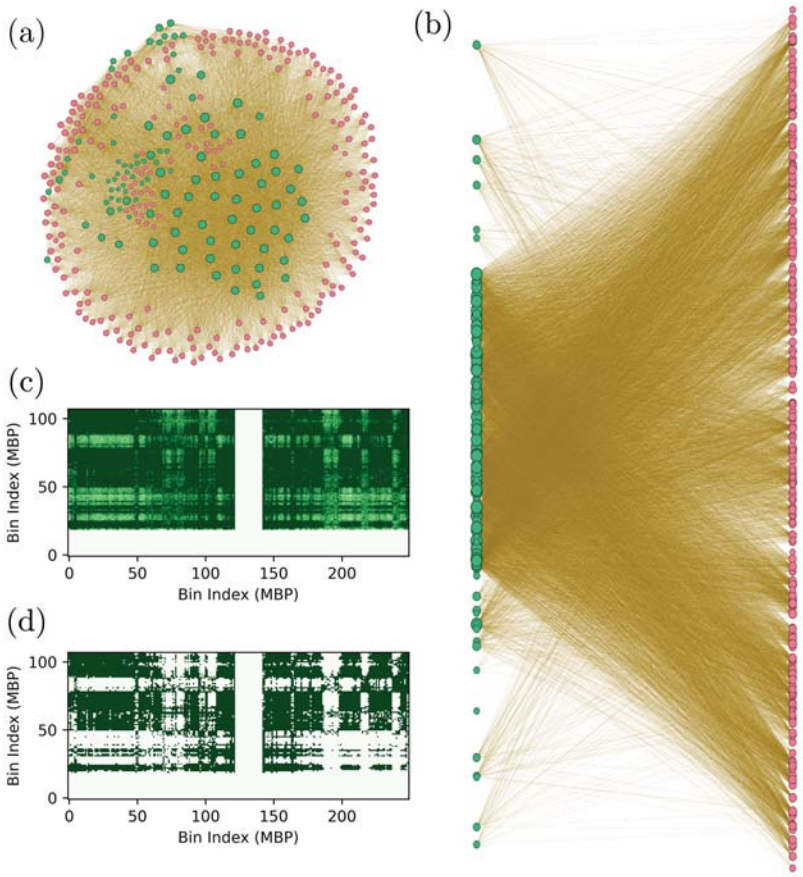

Figure 2.7 The interchromosomal interaction graph between chr1 and chr14 in the human genome. (a) The interchromosomal bipartite graph after running the Force Atlas algorithm in Gephi [18]. (b) The same graph where the nodes on the left in green are the 1 mbp bins in chromosome 14 and the nodes on the right are the nodes of the chromosome 1. (c) The Hi-C matrix of chr1 and chr14 at 1 mbp from Rao et al. [181]. (d) The binary adjacency matrix of the interchromosomal graph after applying a threshold based on the average signal along the the matrix in (c).

there are 3 communities in the bipartite interchromosomal graph of chr1 and chr14, which correspond to four subcompartments (Fig. 2.5(d)) in the Hi-C map at a resolution of 1 mbp [181]. In contrast to the results obtained with the inter-chromosomal graph, the results of the intra-chromosomal graph indicate that there are three communities according to the Clauset-Newman-Moore greedy modularity algorithm [48], which correspond to the most striking and largest domains in chr1 (Fig. 2.3(a)-(c)). In addition, I note that the graph representation shown in Fig. 2.3(c) indicates

that the node 121, which is near the centromere of chr1, has the highest betweenness value. The centromere region is depicted in Fig. 2.3(a)-(b) as a white cross, because the Hi-C technology cannot map DNA sequencing reads to this region. However, node 121 is found on the boundary of the centromere and it directly connects the three large genomic domains in chr1. As a reminder, the centromere is the region of the chromosome that links the the two copies of the chromosomes in the human genome. During mitosis (cell division), fibers (e.g., actin filaments) are attach to the centromere and work with motor proteins to divide the chromosome into its two constituent copies, allowing replication to occur.

2.3 Discussion and outlook

In summary, in this chapter, I have briefly outlined a graph-based paradigm to look at complex biological systems. The ability to look at those systems as networks allows us to develop quantitative order parameters to assess the phase of the system, to gain insight into the system conformation, and to learn about the structure of a single protein, complex structures built from proteins, or the entire human genome. Although contact maps are widely used in biological physics, network theory tools have been used only rarely to study the dynamics and structure of the conformation of biological systems. I expect this work to motivate future research directions in this field, with a particular renewed focus on network theory as a tool to analyze experiments as well as to establish a quantitative understanding of morphologies and phases in complex biological matter. The realization that biological systems could be thought of as graphs and networks will allow researchers to automate the process of coarse-graining using graph neural networks [234]. Presumably, the integration of nonequilibrium physics theories [108, 197, 198] and statistical physics on probabilistic graphical models [120] could develop a theory of nonequilibrium ensembles and landscape reconstruction [140, 73] for living complex matter.

3 Multilinkers augment actomyosin morphology

This chapter is based on an arXiv preprint [64] published in 2020.

3.1 Introduction

Living cells actively regulate the morphologies of actomyosin networks to control the force produced by these networks during various cellular processes [207, 145]. They do so, in part, by controlling the activity of specific actin-binding proteins (ABPs) that bind to actin filaments.

Some ABPs, such as Myosin II (a molecular motor) [186], require the hydrolysis of ATP to actively transmit mechanical changes in actomyosin networks, whereas passive ABPs such as α-actinin (a bivalent ABP) coordinate contractions and control the density of actomyosin networks by crosslinking two filaments [155]. A third distinct category of ABPs are those that can bind to more than two filaments at the same time, such as the multivalent linker CaMKII [114, 220].

Overall, both motors and linkers collectively regulate the morphology of a dendritic spine in neuronal cells as well as the movement of a cell [215, 172]. ABPs do so by altering the connectivity between actin filaments, which enables various biological functionalities due to the robust, yet flexible, nature of actomyosin networks [172, 114]. Most ABPs and actin-related proteins (ARPs), such as Arp2/3 (discussed in more detail in Chapter 4) [133, 204], operate on two filaments as "bivalent linkers". Interestingly, in addition to these bivalent linkers, ABPs with higher valencies (greater than 2) have also been reported. One example of such a multivalent ABP is the Tau protein, which acts as both a multivalent linker [76] and as a mediator of neurodegeneration [62]. The Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) protein is another multivalent actin-binding protein [114, 220], which plays a mechanobiological role in tuning and controlling the intricate and functional shape of dendritic spines in neurons. In this work, I explore how multivalent ABPs change the morphology of actomyosin networks.

For these simulations, I used Cytosim [156], a software developed to simulate mesoscopic cytoskeletal biological systems [22, 141]. In this system, polymeric actin is represented as filaments

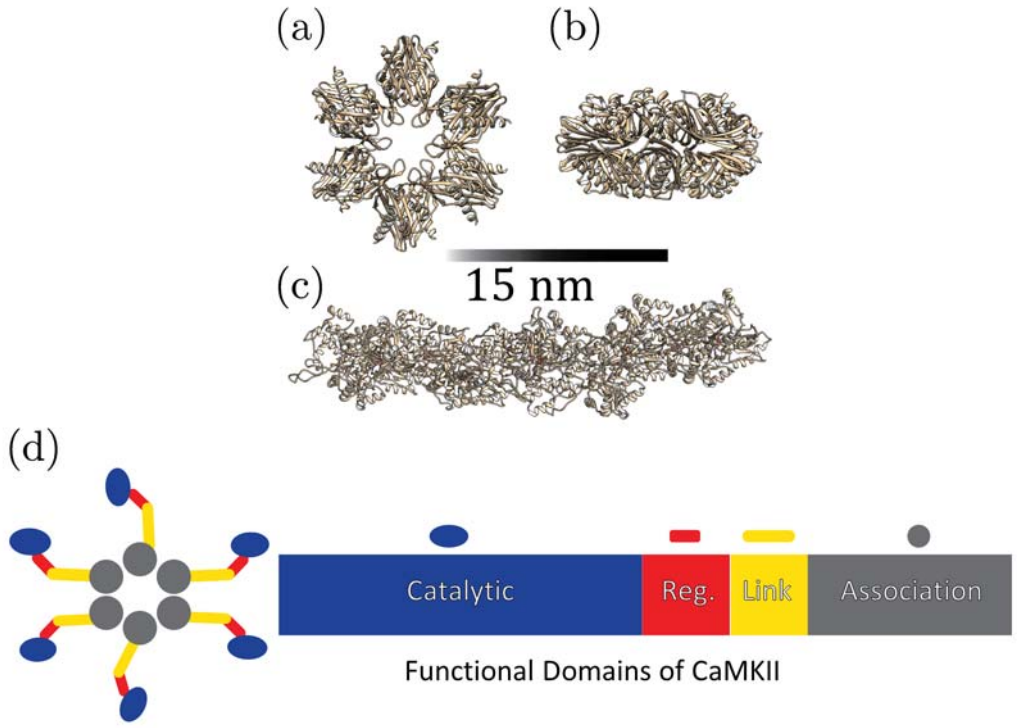

Figure 3.1 CaMKII structure and functional domains. (a) Side and (b) front views of CaMKII crystal structure (PDB ID: 3SOA), with only the association domains of CaMKII shown [220]. (c) The crystal structure of an actin filament made of 10 globular actin (G-actin) monomers assembled into a filamentous actin (F-actin) in (PDB ID: 3J8I). (d) CaMKII's functional domains, including the association domain that binds F-actin [220] and the catalytic domain that binds to ATP and is regulated by the regulatory (Reg.) domain, which becomes active via the process of calcium/calmodulin signaling [90].

with bending elasticity and linkers are represented as stochastic entities that can bind and connect these filaments into a network. Extending Cytosim's modeling of motors, actin filaments and bivalent ABPs, I have worked with Prof. Francois Nedelec to develop a model for multivalent ABPs. With this addition, Cytosim allowed me to investigate the structural and dynamic principles that govern the organization of random actomyosin networks, with the final simulated structure highly dependent on both the motor concentration and on the valency of the multivalent ABPs.

Through this work, I have discovered that multilinkers enrich the variety in the dynamics and structures of actomyosin networks as they evolve heterogeneously in time and space. In order to

analyze and characterize the emergent networks, I used order parameters from different domains of science in terms of polymer physics, gelation theory, and graph theory to compare the simulation results with different multivalent ABPs. I found that the order parameters from network theory capture local and non-local features in actomyosin networks. I established the state diagrams of selected network theory order parameters to identify and quantify the effects of motor activity and multivalent ABP concentration on the actomyosin architecture. Since graph theory considers the extent of nodes' inter-connectivity, the graph theory order parameters have facilitated the possibility of distinguishing between gelated and bundled actin networks.

3.2 Methods

3.2.1 Coarse-grained collective Langevin dynamics model of actomyosin in Cytosim

We used a coarse-grained model to study the morphology and structure of actomyosin networks. In our 3D model, actin filaments are represented as incompressible bendable polar fibers (i.e., fibers have a "plus" end and a "minus" end) with a rigidity of $0.075\,\mathrm{pN}\,\mu\mathrm{m}^2$ [88]. Each of the motors has two walking heads that operate independently and walk towards a filament's plus end. Additionally, there are two types of linker species present in the model: (a) the crosslinker (bivalent linker) that resembles the α-actinin bivalent crosslinker and (b) the multivalent linker inspired by CaMKII with a variety of permissible valencies, referred to as the "multilinker" throughout this work. The CaMKII-inspired multilinker is depicted in Fig. 3.1. In summary, the model contains the following components: filaments, passive bivalent crosslinkers, passive multivalent linkers (multilinkers), and active bivalent motors.

Collective constrained Langevin dynamics

The vector $\mathbf{x}(t)$ stores the coordinates of the N three-dimensional vertices describing the physical objects in the system at time t. For a fixed temperature T, the physics of an actomyosin system in Cytosim [156] is described by a Langevin equation [127] that encapsulates the Brownian motion of the filaments and their bending elasticity, the filament-filament interactions, and the external force

fields [156]:

$$d\mathbf{x}(t) = \boldsymbol{\mu}\mathbf{f}_{\text{tot}}(\mathbf{x}, t)dt + d\mathbf{B}(t), \tag{3.1}$$

where $\boldsymbol{\mu}$ is a $3N \times 3N$ diagonal matrix consisting the mobility coefficients of all objects, and the term $\mathbf{f}_{\text{tot}}(\mathbf{x}, t)$ represents a vector that contains all the forces acting on the points $\mathbf{x}(t)$ at time t. $\mathbf{B}(t)$ represents the random Brownian noise vector due to molecular collisions and its ith coordinate, $\mathbf{B}_i(t)$, is a temperature-dependent value drawn from a normal distribution with a mean of zero and a standard deviation that equals $\sqrt{2dt\mathbf{D}_i}$, where dt represents an infinitesimal time interval. Per Einstein's relation [194], we set the diffusion coefficient of the ith random molecular degree of freedom to be $\mathbf{D}_i := \boldsymbol{\mu}_{ii}k_BT$, where k_B is the Boltzmann constant.

Motor activity and binding and unbinding events

We used the software Cytosim [156] to propagate the equations of motion. After the collective Brownian mechanics are calculated, Cytosim executes two sub-routines to account for chemical processes such as binding, unbinding, and motor walking. The first subroutine simulates binding and unbinding events of ABPs according to the k_{on} and k_{off} rates. The second subroutine simulates motor activity, computes the motor's exerted forces on the filaments, and recalculates the locations of motors on the filaments [156].

Geometry and mobility of fibers and solid objects

Filaments and multilinkers are described by geometrical vertices, each of which has three specific spatial coordinates. A filament is represented as an elastic, polar, and non-extendable rod, which can be quantified as a set of p equidistant vertices $\{m_i\}_{i=1}^{p}$, where m_1 and m_p are the minus and the plus ends of the filaments, respectively. A solid object in Cytosim is a nondeformable set of vertices with a fixed size and shape. For example, the multilinker is made of a solid object in Cytosim. A solid object is a set of k vertices $\{s_i\}_{i=1}^{k}$ with hard constraints on their pairwise distances ($|s_i - s_j| = d_{\text{ij}} = d_{\text{ji}}$); e.g., the multilinker's binding entities that can possibly bind to filaments lay on the surface of a solid sphere, which due to the hard constraints have effectively only

translational and rotational degrees of freedom available.

Interactions between objects

In Cytosim [156], any interaction between objects is linearized $\mathbf{f}(\mathbf{r},t) = \mathbf{A}(t)\,\mathbf{r} + \mathbf{g}(t)$. The matrix $\mathbf{A}(t)$ and the vector $\mathbf{g}(t)$ contain the contributions from all the elementary interactions that are represented by the constant and linear terms of the Taylor expansion. Within this model, two points $\mathbf{r}_a$ and $\mathbf{r}_b$ from different objects a and b can be connected by a link, with Hookean stiffness denoted as k. The forces between the points can be represented according to the following equation:

$$\mathbf{f}_a \;=\; -\mathbf{f}_b \;=\; k\,\left(1 - \frac{r_0}{|\mathbf{r}_b - \mathbf{r}_a|}\right)(\mathbf{r}_b - \mathbf{r}_a), \tag{3.2}$$

where $r_0 \geq 0$ is the resting length of the link. When a motor is attached to a fiber, the motor's position is given by a curvilinear abscissa measured from a fixed reference point on the fiber $\mathbf{x}_0$. The abscissa is increased by $\delta = \tau v_{\max}^{\text{motor}}\left(1 - \frac{f}{f_{\text{stall}}}\right)$, where $v_{\max}^{\text{motor}}$ is a constant real value representing the maximum speed of a motor walking along a fiber, and the sign of this number indicates the motor's tendency to walk toward the filament's minus or plus terminus. The value of f is defined as the load that the motor experiences projected along the direction of the filament on which the motor walks. The stall force f_{stall} is defined as the amount of force that is sufficient to stop the motor from moving. In general, Cytosim uses a force-dependent unbinding rate $k_{\text{off}} = k_0 \exp\left(\frac{|f|}{f_0}\right)$ to model the dissociation from the fiber. Motors have constant binding $k_{\text{on}}^{\text{motor}}$ and unbinding rates $k_{\text{off}}^{\text{motor}}$. Namely, they work in the limit of a constant unbinding rate ($f_0 \to \infty$. In addition, we set $v_{textmax}^{\text{motor}} = 0.2\,\mu\text{m}\,\text{s}^{-1}$ and $f_{\text{stall}} = 6\,\text{pN}$). Bivalent crosslinkers and multilinkers have the unbinding and binding rates denoted as: $k_{\text{on}}^{\text{linker}}$ and $k_{\text{off}}^{\text{linker}}$. The interaction of a linker and filaments is assumed to be Hookean with a resting length of zero and a stiffness of $50\,\text{pN}\,\mu\text{m}^{-1}$.

Excluded volume interactions between multilinkers

Cytosim [156] incorporates excluded volume and steric effects that exist between objects in the system. In general, the simulations account for both attractive and repulsive forces as the interaction

is taken as piecewise linear radial force, according to the equation shown below:

$$f(d) = \begin{cases} k_{\text{repulsion}}\,(d - d_0)\,, & d \leq d_0 \\ 0, & d > d_0 \end{cases}, \tag{3.3}$$

where d is the distance between two interacting elements, and d_0 is the distance at which the elements are at equilibrium. For the particular system being investigated, $d_0 = 6\,\text{nm}$ (the radius of CaMKII [165]) and $k_{\text{repulsion}} = 500\,\text{pN}\,\mu\text{m}^{-1}$.

Parameter space of the simulations

Our system is a $1\,\mu\,\text{m}^3$ cube with a viscosity of $0.5\,\text{pN}\,\mu\text{m}^{-2}$. The filaments have a length of $0.5\,\mu\text{m}$ and are uniformly distributed within the cubic system at $t = 0$. The actin filaments have a flexural rigidity of $0.0752\,\text{pN}\,\mu\text{m}^{-2}$ (corresponding to a persistence length of $\sim 17\,\mu\text{m}$ at the configuration working temperature, for which $k_BT = 4.2\,\text{pN}\,\text{nm}$) [88]. An α-actinin-like bivalent crosslinker was modeled as a Hookean spring of zero resting length between two binding entities, with the spring stiffness constant equals to $k_{\text{crosslinker}} = 250\,\text{pN}\,\mu\text{m}^{-1}$.

Each crosslinker's binding entity had a binding range of 17.5 nm, a binding rate $k_{\text{on}}^{\text{linker}} = 5\,\text{s}^{-1}$, and an unbinding rate $k_{\text{off}}^{\text{linker}} = 0.1\,\text{s}^{-1}$.

A myosin-like motor is represented as an inextensible object with two independent motor heads. Each head of a motor can walk on a different filament, with the two heads separated by a resting length of 100 nm. When attached, the connection between a motor and two filaments is modeled as a Hookean spring, with a stiffness of $k_{\text{motor}} = 250\,\text{pN}\,\mu\text{m}^{-1}$. Motors unbind at the same rate as linkers $k_{\text{off}}^{\text{motor}} = 0.1\,\text{s}^{-1}$, and bind twice as fast as linkers, $k_{\text{on}}^{\text{motor}} = 10\,\text{s}^{-1}$, and each motor site has a binding range of 50 nm. A multilinker of valency ν has ν hands residing on the surface of a sphere of radius 6 nm and are linked to the sphere with a stiffness of $k_{\text{multilinker}} = 200\,\text{pN}\,\mu\text{m}^{-1}$. A multilinker's binding entity has a binding range of 5 nm and binding and unbinding rates equal to the rates of a crosslinker:$k_{\text{on}}^{\text{linker}}$ and $k_{\text{off}}^{\text{linker}}$. The parameters that are systematically varied in the simulations include the total number of multilinkers in the system ($N_{\text{multilinkers}} \in \{250, 500, 1000\}$), the valencies

of the multilinkers ($\nu \in \{2, 3, 4, 5, 6, 7\}$), the number of motors ($N_{\text{motors}} \in \{250, 500, 1000\}$), the number of filaments ($N_{\text{filaments}} \in \{250, 500, 1000\}$), and the number of α-actin-like crosslinkers ($N_{\text{crosslinkers}} \in \{10, 250, 500, 1000\}$). The simulations explored the Cartesian product of all possible values of the above five parameters. For example, one possible system configuration included 250 pentavalent ($\nu = 5$) multilinkers, 500 filaments, 1000 motors, and 10 crosslinkers. For readability, Table 3.1 summarizes the possible values of the parameters and the binding properties of objects in the simulations:

Parameter	**Description**	**Values**
ν	Valency of a multilinker	$\{2, 3, 4, 5, 6, 7\}$
$k_{\text{multilinker}}$	Hookean stiffness of a multilinker binding entity	$200\,\text{pN}\,\mu\text{m}^{-1}$
$k_{\text{crosslinker}}$	Hookean stiffness of a crosslinker	$250\,\text{pN}\,\mu\text{m}^{-1}$
$k_{\text{off}}^{\text{linker}}$	Unbinding rate of a multilinker or a crosslinker binding entity	$0.1\,\text{s}^{-1}$
$k_{\text{on}}^{\text{linker}}$	Binding rate of a multilinker or a crosslinker binding entity	$5\,\text{s}^{-1}$
$k_{\text{off}}^{\text{motor}}$	Unbinding rate of a motor binding entity	$0.1\,\text{s}^{-1}$
$k_{\text{on}}^{\text{motor}}$	Binding rate of a motor binding entity	$10\,\text{s}^{-1}$
N_{motors}	Number of motors	$\{10, 250, 500, 1000\}$
$N_{\text{filaments}}$	Number of filaments	$\{250, 500, 1000\}$
$N_{\text{crosslinker}}$	Number of α-actinin crosslinkers	$\{10, 250, 500, 1000\}$
V_{box}	Volume of the simulation box	$1\,\mu\text{m}^3$
$N_{\text{multilinkers}}$	Number of multilinkers	$\{250, 500, 1000\}$

Table 3.1 System configuration parameter values for the computational model, with, when applicable, the set of values that have been explored. In all simulations, these parameters are held constant from start to end of the simulation, and are varied only between simulations.

3.2.2 Order parameters

I explored the use of various order parameters to effectively analyze actomyosin network simulations. In general, an order parameter should reflect an interesting feature of the system, ideally reducing the configuration of a system characterized by a myriad of degrees of freedom into a single, mesoscopic, physically meaningful value. To achieve this goal, I first examined geometrical quantities such as the radius of gyration (R_g), the shape (S), and the asphericity (Δ) parameters [104, 57] among all the filament vertices. Next, I used gel-sol theory order parameters, which considered the molecular connectivity of APBs, i.e., motors and linkers, to filaments. Each filament was defined as an elementary unit for this analysis, and if two filaments are connected by at least one ABP, then the filaments are considered to be in the same cluster. These gel-sol order parameters measure the number of clusters, the number of filaments in any cluster, the largest cluster size, the smallest cluster size, the average cluster size, and the standard deviation in cluster sizes. Finally, I created a network/graph by considering only the spatial information of the filaments and used order parameters on the undirected graph representation of the actomyosin network: $\mathbf{G} = (V, E)$. This graph allows a rich representation of the system, based only on spatial information of the filaments. The set of nodes V in this graph corresponds to filaments, and E is defined as the set of edges between the nodes. An edge is considered to exist between node i and node j if the minimal Euclidean distance between the filaments is less than the cutoff distance, $d_{\text{cutoff}} = 200\,\text{nm}$. The distance d between two filaments is taken to be the minimal distance among all possible pairwise distances of their segments' center of masses. Thus, the graph $\mathbf{G}$ is constructed only from the information of distances between filaments and without any consideration or knowledge of the microscopic connectivity between ABPs and filaments. The network-theory-based list of order parameters includes the following order parameters: the number of communities [48], average clustering [189], clique number of the graph (the maximal clique size), average closeness centrality [221, 54], average eigenvector centrality [29], mean average neighbor degree [16], degree assortativity coefficient [159], and graph density. The Python package NetworkX 2.2 [95] was used to construct the graph $\mathbf{G}$ and to compute the order

parameters.

Geometrical quantities derived from the moment of inertia

We used three order parameters derived from the general field of protein folding [57]: the radius of gyration R_g (the variance of all the positions of geometrical filament vertices), which is a proxy for the macroscopic structure, the shape parameter $-\frac{1}{4} \leq S \leq 2$, and the asphericity measure $0 \leq \Delta \leq 1$. The shape order parameter [57] specifies how prolate ($S > 0$) or oblate ($S < 0$) is the conformation of an actomyosin network, while asphericity measures how an actomyosin network conformation differs from a perfect sphere ($\Delta = 0$). The fluctuations in these order parameters reveal the stress relaxation the system undergoes, and its response at different ABP-filaments concentration ratios. Using spectral and fluctuation analysis, we evaluate the differences [214] at high versus low motor content. The three order parameters R_g, Δ, and S are readily derived from the moment of inertia tensor of filament vertices at time t:

$$\mathbf{T}_{\alpha\beta}(t) = \frac{1}{2N^2} \sum_{i,j=1}^{N} \left(\mathbf{r}_{i\alpha}(t) - \mathbf{r}_{j\alpha}(t)\right)\left(\mathbf{r}_{i\beta}(t) - \mathbf{r}_{j\beta}(t)\right), \tag{3.4}$$

where $\mathbf{r}_{i\alpha}(t)$ is the α-component of a filament vertex, N is the number of filament vertices, and $\alpha, \beta \in \{x, y, z\}$ are the indices of the Cartesian elements. Following this general approach, the order parameters are given by the following expressions as functions of the eigenvalues of the moment of inertia tensor $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The radius of gyration is defined as:

$$R_g(t) = \sqrt{\operatorname{tr}\mathbf{T}(t)} = \sqrt{\sum_{i=1}^{3} \lambda_i(t)}. \tag{3.5}$$

The shape parameter is defined as:

$$S(t) = \frac{\prod_{i=1}^{3}\left(\lambda_i(t) - \overline{\lambda}(t)\right)}{\left(\frac{1}{3}\operatorname{tr}\mathbf{T}(t)\right)^3}, \tag{3.6}$$

where $\overline{\lambda}(t) = \frac{\operatorname{tr}\mathbf{T}(t)}{3}$.

Finally, the asphericity measure is defined as:

$$\Delta\left(t\right)=\frac{3}{2}\frac{\left(\sum_{i=1}^{3}\quad\left(\lambda_{i}\left(t\right)-\overline{\lambda}\left(t\right)\right)^{2}\right)}{\left(\mathrm{tr}\,\mathbf{T}\left(t\right)\right)^{2}}. \tag{3.7}$$

To calculate the power spectral density (PSD) of the three order parameter detailed above, I used Welch's method for spectral density approximation [224] with a flat top window size of 10 samples and a window overlap of 50%. Then we compared the PSD of $R_g(t), S(t)$, and $\Delta(t)$ under conditions of low motor concentration versus conditions of high motor concentrations. The ratio between any two power measurements P_1 and P_2 was given in decibels, following the convention $10\log_{10}(\frac{P_1}{P_2})$.

Gelation of actin filaments into clusters formed by actin-binding proteins

A complementary approach to study actomyosin networks is to explore the sol-gel phase transition with respect to the molecular connectivity of ABPs and filaments. The ith cluster at time t is defines as a gelated group of filaments connected by either motors, linkers, or both; we denote as $N_{c_i}(t)$ the number of filaments in the ith cluster. By convention, in our framework, a pair of filaments forms the smallest cluster of size two, meaning that $N_{c_i}(t)\geq 2$ always holds. The term $N_c\left(t\right)$ is denoted as the total number of clusters that exist in the system at time t. The total number of gelated filaments in the system is given by the following equation:

$$N_{\mathrm{gel}}\left(t\right)=\sum_{i=1}^{N_c(t)}N_{c_i}\left(t\right), \tag{3.8}$$

and the gelation ratio order parameter is defined by the following equation:

$$\overline{N}_{\mathrm{gel}}\left(t\right)=\frac{1}{N_{\mathrm{filaments}}}N_{\mathrm{gel}}\left(t\right). \tag{3.9}$$

The largest cluster size is given by

$$N_{\max}(t) = \max_{1 \leq i \leq N_c(t)} \{N_{c_i}(t)\}, \tag{3.10}$$

then the normalized largest cluster size order parameter reads

$$\overline{N}_{\max}(t) = \frac{1}{N_{\text{filaments}}} N_{\max}(t). \tag{3.11}$$

The average cluster size is given by

$$\mu_c(t) = \frac{1}{N_c(t)} \sum_{i=1}^{N_c(t)} N_{c_i}(t), \tag{3.12}$$

then the normalized cluster size order parameter reads

$$\overline{\mu}_c(t) = \frac{1}{N_{\text{filaments}}} \mu_c(t). \tag{3.13}$$

The standard deviation cluster size is defined according to the relation

$$\sigma_c(t) = \sqrt{\frac{1}{N_c(t)} \sum_{i=1}^{N_c(t)} [N_{c_i}(t) - \mu_c(t)]^2}, \tag{3.14}$$

then the normalized standard deviation cluster size order parameter reads

$$\overline{\sigma}_c(t) = \frac{1}{N_{\text{filaments}}} \sigma_c(t). \tag{3.15}$$

The reasoning behind dividing gelation-related quantities by the total number of filament $N_{\text{filaments}}$ is to map the order parameter onto the segment $[0, 1]$.

Network theory order parameters

The third approach to quantify the morphology of an actomyosin network is to consider the graph

generated from the positions of the filaments. This framework is highly relevant in scenarios of experiments in which only F-actin is visible via fluorescence microscopy. For this purpose, we harness ideas and order parameters from graph theory or network theory [4, 50, 68]. As mentioned above, the actin graph at time t is denoted by $\mathbf{G}(t) = (V(t), E(t))$, after applying a cutoff distance $d_{\text{cutoff}} = 200\,\text{nm}$, which corresponds to the length of a motor (i.e., length of the motor is 100 nm and binding range of each motor head is 50 nm). For higher readability, we occasionally omit the time argument, t. Because of the applicability of graph theory to a wide range of scientific disciplines, there are an enormous number of topological order parameters available in the scientific literature. As a result, once $\mathbf{G}(t)$ is constructed for the system, one can compute the global order parameters on properties such as the community structure [48, 77] and the clustering coefficient [189], as well as order parameters defined on the level of individual nodes, such as centrality [29, 157].

In general, community structure in a network indicates the presence of heterogeneity in the density of links in a network: if some nodes are more likely to be connected with each other than with the rest of the network, they may be said to form a community within the network. This heuristic definition of community has spawned an enormous literature on the subject of community structure in networks [77], although in this work, I focus solely on a commonly used method for community detection: the Clauset-Newman-Moore greedy modularity maximization algorithm [48]. The algorithm first assign a unique community C_v to each node $v \in V$; then the greedy modularity maximization procedure of the algorithm joins pairs of communities to maximize the modularity, defined as

$$Q = \frac{1}{|E|} \sum_{(v,w)\ \in\ E} \left[\mathbf{A}_{\text{vw}} - \frac{k_v k_w}{|E|}\right] \delta_{s_v, s_w}, \tag{3.16}$$

where $|E|$ is the number of edges in the graph between any two nodes v and w where $(w, v), (v, w) \in E$.

$\mathbf{A}$ is a symmetric adjacency matrix of $\mathbf{G}$, defined as:

$$\mathbf{A}_{vw} = \begin{cases} 1, & (v,w) \in E \wedge (w,v) \in E \\ 0, & \text{otherwise} \end{cases}, \tag{3.17}$$

where k_v is the degree of node v and set as equal to the number of edges connected to v. $\delta_{a,b}$ is the Kronecker delta, and equals 1 if $a = b$ and 0 otherwise. Using these definitions and the Clauset-Newman-Moore greedy algorithm [48], we define the order parameter $N_{\text{communities}}(t)$ to be the number of communities that are represented in graph $\mathbf{G}$.

In many real-world networks, it has been found that triplets of nodes share edges more often than would be expected by random chance [189]. In particular, the average clustering coefficient of a network $\mathbf{G} = (V, E)$ is defined as

$$\Gamma = \frac{1}{|V|} \sum_{v \in V} \gamma_v, \tag{3.18}$$

where γ_v is the clustering coefficient of node v, and is defined as

$$\gamma_v = \frac{\frac{1}{2} \sum_{i,j} \mathbf{A}_{\text{vi}} \mathbf{A}_{\text{ij}} \mathbf{A}_{\text{vj}}}{k_v \left(k_v - 1\right)}. \tag{3.19}$$

Of note, the clustering coefficient referred to in network theory is related to the density of triangles of edges, and is conceptually distinct from the clustering discussed in the context of gelation literature.

In addition to the aforementioned concepts, the concept of node centrality has been useful in better understanding the structure of complex networks on the level of individual nodes [29, 157]. There are many possible measures of centrality found in the literature, each measuring the 'importance' of a node in a slightly different way. Herein, I used a few of these centrality measures, discussed in more detail below [50].

The eigenvector centrality of a node i is the ith element of the eigenvector $\mathbf{v}$ of the adjacency matrix $\mathbf{A}$:

$$\mathbf{A}\mathbf{v} = \lambda_{\max} \mathbf{v}, \tag{3.20}$$

where $\lambda_{\max} \geq \lambda_1 \geq \ldots \geq \lambda_{|V|}$ is the maximal eigenvalue of the symmetric adjacency matrix $\mathbf{A}$ of the graph $\mathbf{G}$. In general, eigenvector centrality tends to be higher for nodes that are connected to other nodes with high centrality. Then, the average eigenvector centrality is defined as:

$$E_{\lambda_{\max}}(t) = \frac{1}{|V(t)|} \sum_{1 \leq i \leq |V(t)|} \lambda_i. \tag{3.21}$$

A second measure of the centrality of a node v is its average neighbor degree, as evaluated in v's neighborhood, with the set of the nearest neighbors of v, $\mathcal{N}(v) = \{u \,|\, (u,v) \in E\}$:

$$\kappa_{nn,i} = \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} k_u, \tag{3.22}$$

where k_u is the degree of the node u. Like eigenvector centrality, neighbor degree centrality is higher for nodes that are connected to other central nodes, but unlike eigenvector centrality, longer paths are not explicitly considered in this definition. Within the concept of neighbor degree centrality, I use the mean average neighbor degree as an order parameter, according to the equation shown below:

$$k_{\mathrm{nn}}(t) = \frac{1}{|V|} \sum_{v \in V} \kappa_{nn,v}. \tag{3.23}$$

In addition to the above centrality parameters, the parameter of "betweenness centrality" of a node (referred to herein as "betweenness") measures how central a node is by how often it acts as a 'bridge' between other pairs of nodes (sketched in Fig. 2). This parameter is quantified by counting the number of shortest paths in which the node participates:

$$\beta(v \in V) = \sum_{s,s' \in V^*} \frac{g(s,s'|v)}{g(s,s')}, \tag{3.24}$$

where $V^* = V \setminus \{v\}$ is defined as the set of nodes excluding node v, the function $g(s,s')$ counts the number of geodesics (shortest paths) between node s and node s' , and $g(s,s'|v)$ is the number of shortest paths between node s and s' such that v lies along those shortest paths.

Beyond evaluating the centrality of nodes in a graph, one can measure the similarity of connections within the network. One such connection similarity measure is the "degree assortativity coefficient" [159]. The degree assortativity, previously described by Newman et al. [37, 158], measures whether nodes in the graph with similar degrees are also directly connected.

Mathematically, the assortativity is derived from the nodes' joint probability matrix, a symmetric matrix in which the $\mathbf{E}_{lm}$ element equals the number of edges in the graph that link a vertex of degree l with another vertex of degree m:

$$\mathbf{E}_{lm} = \left|\{(u,v) \ : \ v,u \in E \ \wedge k_v = l \ \wedge k_u = m\}\right|. \tag{3.25}$$

The joint probability matrix is achieved by normalizing $\mathbf{E}_{lm}$ with the sum of its elements $\sum_{l,m} \mathbf{E}_{lm}$. Using this definition, the degree assortativity coefficient order parameter is defined as:

$$\rho\left(t\right) = \frac{\sum_{l,m} l \cdot m \cdot (\mathbf{E}_{lm}(t) - p_l(t)p_m(t))}{\sigma_p^2(t)}, \tag{3.26}$$

where $p_i := \sum_j \mathbf{E}_{ij} = \sum_j \mathbf{E}_{ji}$ for any possible node's degree i, and σ_p is the standard variation of the set of the p_i values, $\{p_i\}_{i=1}^{|V|}$.

Of note, for brevity, I refer to the degree assortativity coefficient as assortativity.

In addition to the detailed order parameters discussed above, the global details of the network can also be characterized using coarse network measures. One example of such an order parameter is the graph density, defined as:

$$d_{\mathbf{G}}\left(t\right) = \frac{2|E|}{|V| \cdot |V-1|}, \tag{3.27}$$

which measures the number of edges observed in the network compared to the maximum number that could be observed. Besides the graph density, I also evaluated, as an order parameter, the maximal clique size, which is the largest subset of vertices from V such that every pair of nodes in the subset has an edge between them.

3.2.3 Potential of mean force (PMF) of order parameters

The potential of mean force (PMF) is the natural logarithm of any probability density function (PDF), defined on a specific degree of freedom of the system [217]. The PDF of an order parameter $\xi \in [0,\ 1]$, computed along a trajectory of the system, is given by the following relations [217]:

$$P\left(\xi\right) = \left\langle \delta\left(\xi - \xi\left(\mathbf{x}\right)\right)\right\rangle_{\text{trajectory}} := \frac{Q\left(\xi\right)}{\sum_{\xi} Q\left(\xi\right)}, \tag{3.28}$$

where $Q\left(\xi\right)$ is the partition function used to find the system in a state at which the order parameter of interest has the value ξ. The entropy $\mathcal{S}\left(\{\xi\}\right)$ [109, 199] of the PDF of ξ (i.e., $P(\xi)$) is given by:

$$\mathcal{S}\left(\{\xi\}\right) = -k_B \sum_{\xi'} P\left(\xi'\right) \ln P\left(\xi'\right) = k_B \sum_{\xi'} P\left(\xi'\right) I\left(\xi'\right), \tag{3.29}$$

where $I\left(\xi\right) := -k_B \ln P\left(\xi\right)$ is the Shannon information (self-information) as a function of ξ. If k_B is defined as one, both the entropy and the information measures have the "nat" (natural unit of information) units [201].

In order to compute the PMF, I used the kernel density estimator function, implemented in the Scikit Python library [174], to estimate $I\left(\xi\right)$ via the negative logarithm of the PDF $P\left(\xi\right)$. Since the self-information is related to the potential of mean force (PMF), denoted with $\mathcal{U}\left(T,\ \xi\right)$, the following relations between the entropy, information, the PMF, and the PDF of xi hold [217, 43]:

$$\mathcal{U}\left(T,\ \xi\right) \ :=\ TI\left(\xi\right) := -k_B T \ln P\left(\xi\right) := k_B T \ln\left(\frac{\sum_{\xi'} Q\left(\xi'\right)}{Q\left(\xi\right)}\right). \tag{3.30}$$

3.3 Results

3.3.1 Multilinkers enrich the morphologies of actin bundles

I have conducted mesoscopic simulations using Cytosim, over a time course of 600 seconds, to investigate the role of multivalent linkers on the structural changes that occur in actomyosin networks.

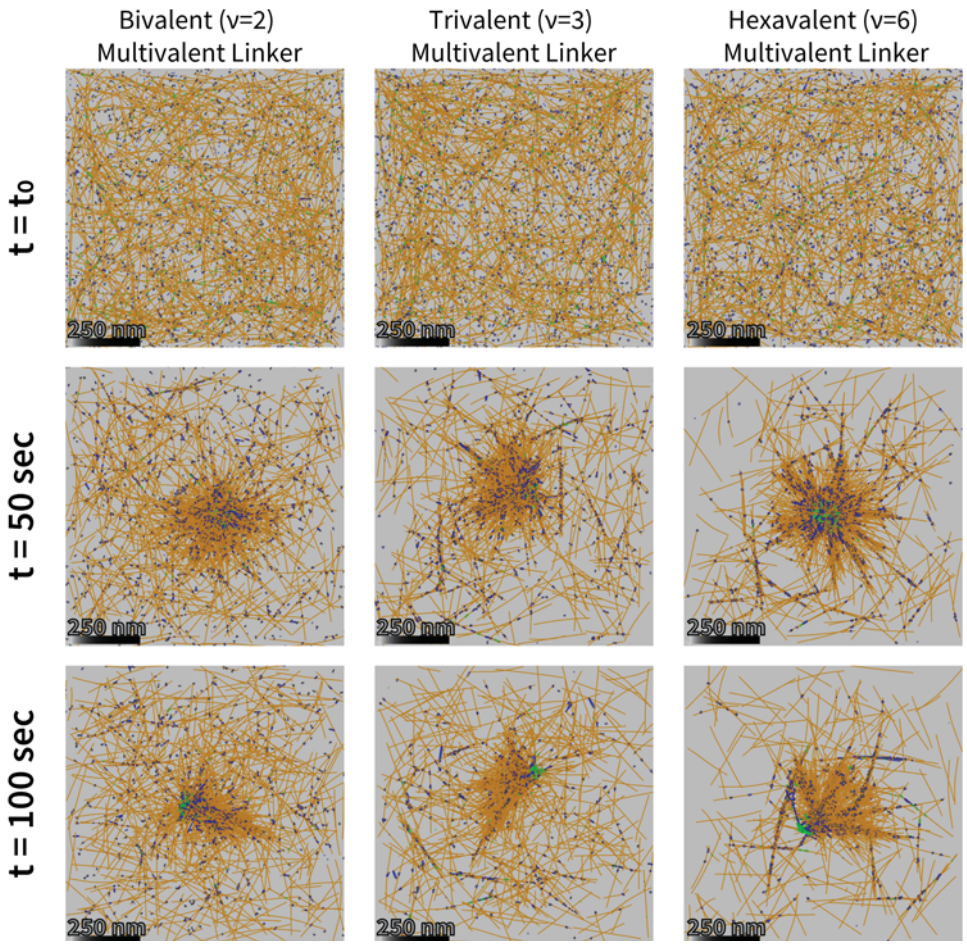

Figure 3.2 Snapshots of simulation results from Cytosim of three $1\,\mu m^3$ box systems with a $1:1:1:1$ concentration ratio of filaments, motors, multilinkers, and α-actinin like crosslinkers. The three systems differ only in the valency of the available multilinkers. In the leftmost column, the multilinkers are bivalent ($\nu = 2$). In the middle column, the multilinkers are trivalent ($\nu = 3$). In the rightmost column, the multilinkers are hexavalent ($\nu = 6$). The time increases from the top panels to the bottom panels, with snapshots occurring at $t = t_0 = 1\,s$, $t = 50\,s$, and $t = 100\,s$.

In the simulations, the size of the simulation box is set at $1\,\mu m^3$ with 1:1:1:1 concentration ratio among filaments, motors, multilinkers, and α-actinin crosslinkers. The snapshots from simulations with three types of multilinkers, differing in their valency numbers (2, 3, and 6), are provided in Fig 3.2 at $t = 1\,s$, $t = 50\,s$, $t = 150\,s$ and Fig 3.3 at $t = 200\,s$, $t = 400\,s$, and $t = 600\,s$.

In the systems with both crosslinkers and bivalent multilinkers, actin bundles are formed [216, 226], with a smaller variation in the cluster sizes as time evolves compared to systems with

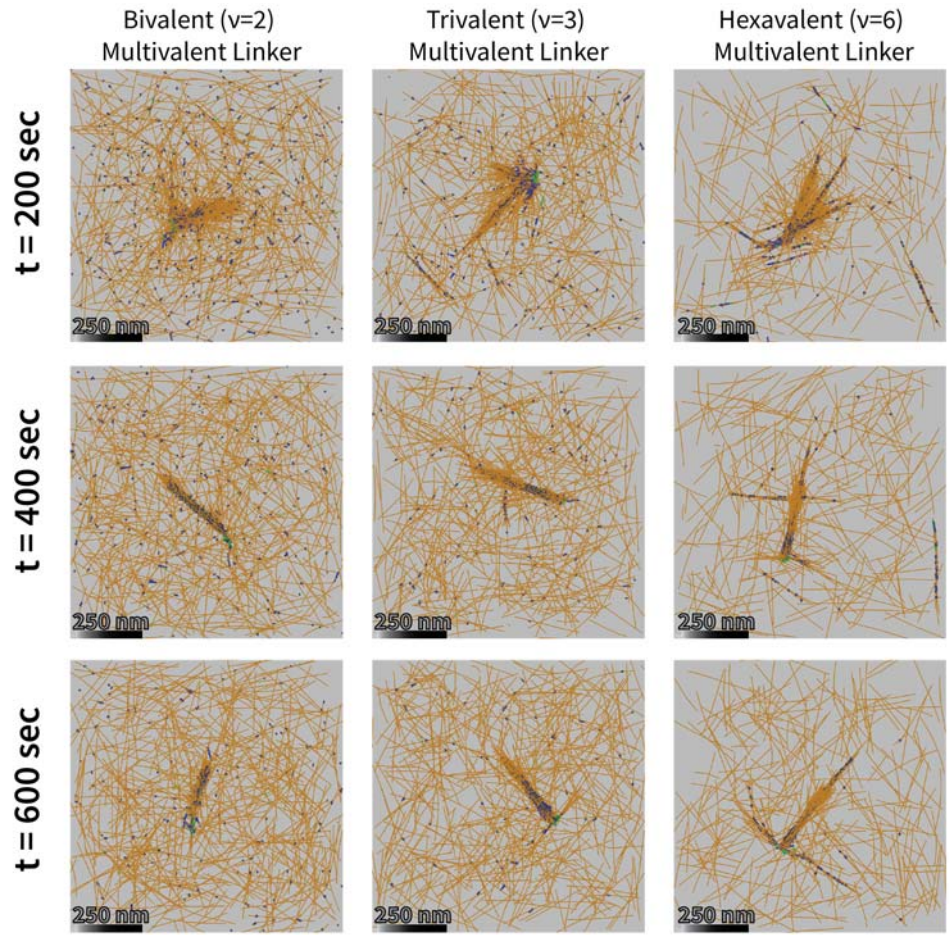

Figure 3.3 Snapshots of simulation results from Cytosim of three $1\,\mu m^3$ box systems with a $1:1:1:1$ concentration ratio of filaments, motors, multilinkers, and α-actinin like crosslinkers. The three systems differ only in the valency of the available multilinkers. In the leftmost column, the multilinkers are bivalent ($\nu = 2$). In the middle column, the multilinkers are trivalent ($\nu = 3$). In the rightmost column, the multilinkers are hexavalent ($\nu = 6$). The time increases from the top panels to the bottom panels, with snapshots occurring at $t = 200\,s$, $t = 400\,s$, and $t = 600\,s$.

multilinkers of valency ($\nu \geq 3$).

In the systems with trivalent multilinkers, the nascent "arborization" of bundles (the appearance of branched actin bundles) starts to appear at $t = 150\,s$ in Fig. 3.3. This phenomenon becomes more prominent in systems with multilinkers of higher valencies, as shown in Fig. 3.3 at time $t = 600\,s$ for a system with hexavalent ($\nu = 6$) multilinkers.

3.3.2 Correlation between order parameters from polymer physics, gelation theory, and network theory

The rich actomyosin structures shown in Fig. 3.2 and Fig. 3.3 have prompted the search for appropriate order parameters that characterize the complexity and heterogeneity of actomyosin dynamics in space over time. These order parameters include common measures from polymer physics, gelation theory, and network theory, as described in the Methods section (*vide supra*). Fig. 3.4 shows the Pearson correlation of all order parameters, using the data collected from the Cytosim simulations. In particular, each group of order parameters presents a different insight into the actomyosin morphology. The three polymer physics order parameters: radius of gyration (R_g), the shape parameter S , and the asphericity Δ, are highly correlated; these three parameters are nearly uncorrelated with the order parameters obtained from gelation theory or with those that are obtained from graph theory.

To some extent, the order parameters from network theory correlate with the gelation order parameters (e.g., the gelation ratio order parameter positively correlates with network's average clustering order parameter). However, the correlation is below 0.5, because the network order parameters have more complex time evolution patterns, as shown in detail below. Both the number of gelated clusters and the node degree assortativity order parameters have negative correlation with the other order parameters. This is because the number of gelated cluster decays with time, as the gelation ratio arises. On the other hand, the degree assortativity has a negative correlation with the gelation and network order parameters, due to the arborization of bundles, as discussed later in detail.

3.3.3 Motor activity regulates temporal changes in the global structure and the shape of actomyosin networks

The fluctuations in the distribution of filament vertices in space can be used to provide insight into the temporal dynamics of the actomyosin networks, since the filaments are the scaffold of the

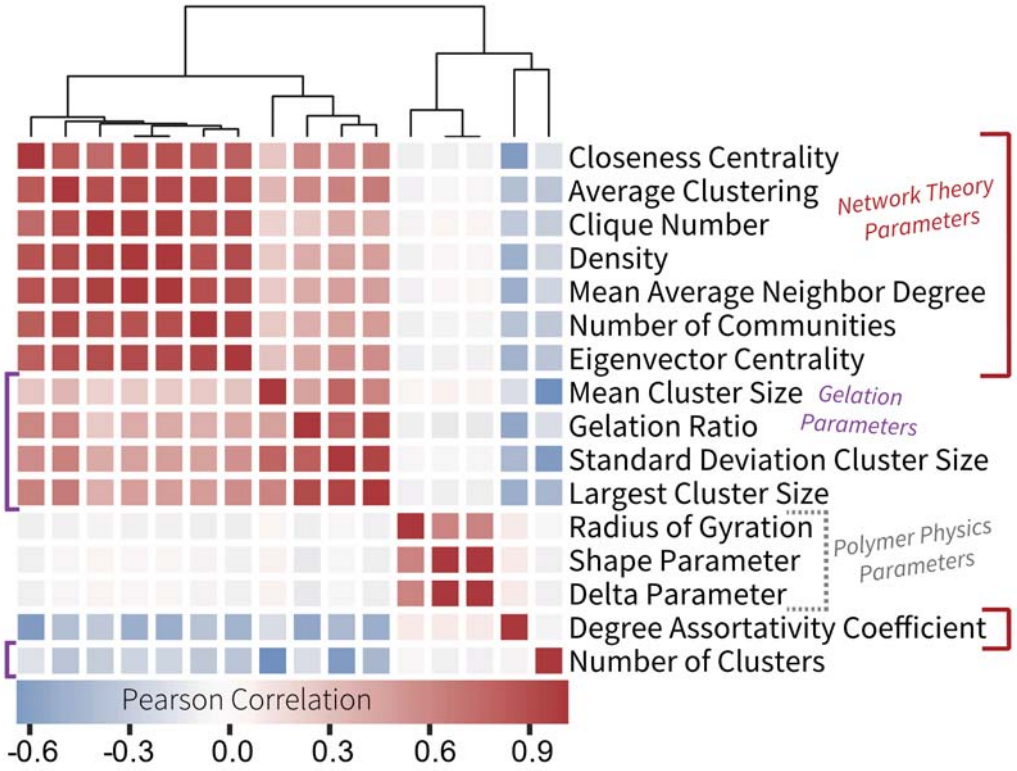

Figure 3.4 Heatmap of the Pearson correlation between the order parameters used for this study. The order parameters are grouped according to the pairwise correlation distance and annotated by their physical order parameter group: polymer physics (labeled in gray dashed brackets on the right), gelation theory (labeled in purple brackets on the left), or network theory (labeled in red brackets on the right).

network. I used the power spectral density functions (PSD) [224] of the three polymer physics-derived order parameters (R_g, S, and Δ from Eq. (3.5-3.7) to compare the fluctuations at high and low motor-to-filament ratio values, as shown in Fig. 3.5. Of note, each of these three order parameters measures a distinctive mesoscopic reshaping mode of the actomyosin network, derived from the moment of inertia of the filament vertices (Eq. (3.4)). In particular, the PSD of the shape parameter, S, is a mesoscopic assessment of filament sliding modes (Fig. 3.5(a-b)). The PSD of the asphericity Δ is a proxy for the rotational modes of filaments (Fig. 3.5(c-d)). The PSD of the radius of gyration R_g order parameter provides information about the mesoscopic expansion and contraction modes (Fig. 3.5(e-f)). One observation is that the filament sliding modes in Fig. 3.5(a-b) are energetically

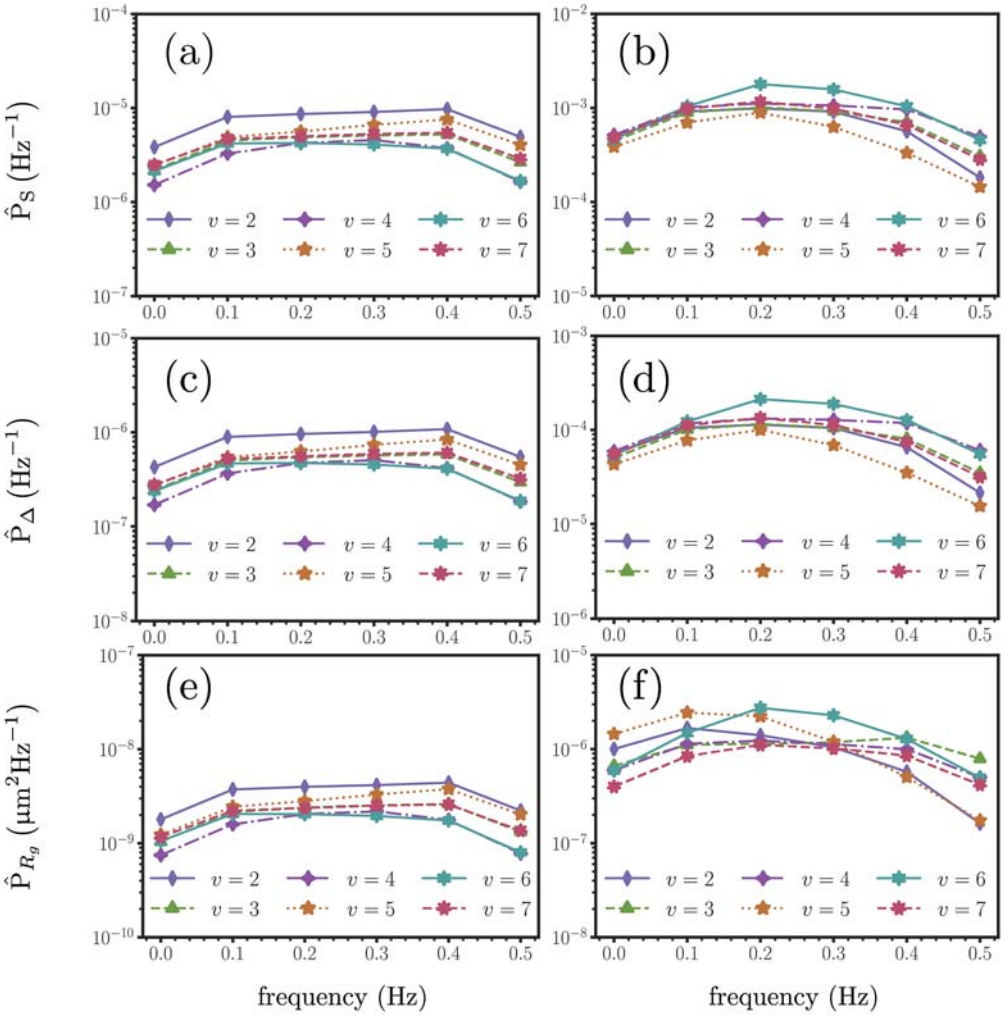

Figure 3.5 Power spectral analysis of mesoscopic polymer order parameters. Low and high motor-filament-crosslinker-multilinker ratios, $1:100:100:100$ and $1:1:1:1$, are examined in systems with multilinkers of various valencies $\nu \in \{2,3,4,5,6,7\}$. (a), (c) and (e) show the power spectral density (PSD) from the systems with low motor-filament ratios, while (b), (d), and (f) show the PSD from the systems with high motor-filament ratios. (a)-(b) The topmost row shows the PDF of the shape parameter S. (c)-(d) The middle row shows the PDF of the asphercity parameter Δ. (e)-(f) The bottom row shows the PDF of the radius of gyration R_g.

more dominant than the rotational modes in (Fig. 3.5(c-d), which is reflected in the fact that the PSD of the former is one order of magnitude (10 dB) higher than the PSD of the latter. When comparing the PSD at the two low and high conditions of motor-filament-crosslinker-multilinker concentrations, $(1:100:100:100)$ vs. $(1:1:1:1)$, all three of the PSD values in the case of high

motor-filament ratio are two order of magnitudes (20 dB) greater than those obtained under low motor-filament ratio conditions.

3.3.4 Gelation of filaments in actomyosin network by actin-binding proteins

Although the shape and structure parameters R_g, S, and Δ from polymer physics provide information about the fluctuations derived from the various configurations of actin filaments within a network, they cannot directly capture the individual characteristics of a gelated, or percolated cluster, in the actomyosin networks [82]. By contrast, the order parameters derived from gelation theory, which consider the connectivity between actin filaments through actin-binding proteins, can be used precisely for such detailed analysis, i.e., to monitor the development of a percolated cluster.

In one example, the data shown in Fig. 3.6 demonstrates how motors can regulate the total number of clusters in a way that is independent of the valency of multilinkers, even at very high multilinker valency values. Fig. 3.6(a-b) show the normalized largest cluster size over time, defined as $\overline{N}_{\max}(t)$ in Eq. (3.11). This data demonstrates that multilinkers with higher valency promote larger cluster formation, regardless of the motor concentration. Another way of viewing this data, shown in Fig. 3.6(c), is by examining the cumulative distribution functions (CDFs) of the average cluster size, $\overline{\mu}_c(t)$ at high $(1:1:1:1)$ and low $(1:100:100:100)$ motor-filament-crosslinker-multilinker concentration ratios, in the presence of multilinkers with varying valency values. Interestingly, multilinkers with valencies of $\nu = 2$ and $\nu = 3$ create smaller clusters on average at low motor concentration. However, for a system with multilinkers of $\nu = 4$, systems with both high or low motor content display a similar distribution of $\overline{\mu}_c(t)$. Finally, at valency values $\nu = 5$, $\nu = 6$, and $\nu = 7$, higher motor activity promotes bigger "largest clusters" by connecting smaller clusters together, as the average cluster size histograms (Fig. 3.6(c)) are shifted to the right compared to the system with low motor content.

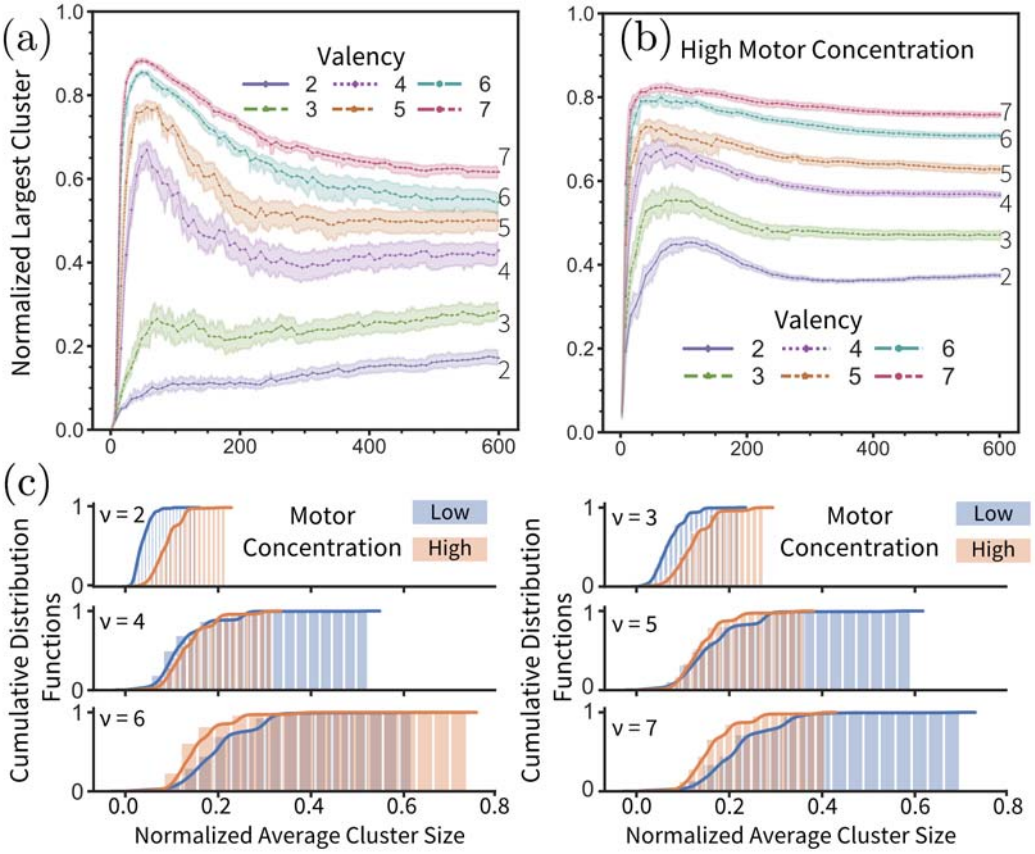

Figure 3.6 Gelation of filament clusters formed by crosslinkers, multilinkers, or motors. The condition of "low motor concentration" corresponds to motor-filament-multilinker-crosslinker ratios of $1 : 100 : 100 : 100$, shown in (a). The condition of "high motor concentration" corresponds to motor-filament-multilinker-crosslinker ratios of $1 : 1 : 1 : 1$, shown in (b). The time evolution of the normalized largest cluster, as defined in Eq. (3.11), is shown in (a) and (b). (c) The cumulative distribution functions (CDF) of the normalized average cluster size order parameter $\overline{\mu}_c(t)$, defined in Eq. (3.13), are plotted in both motor concentration conditions and at the six different multilinker's valency conditions. The plots shown in (a) and (b) are averaged over 30 independent simulation trajectories using the same species parameters, with a confidence interval CI = 90%.

3.3.5 Constructing graphs/networks using the spatial information of actin filaments

In real-world systems, the local microscopic-level information of actin-binding connectivity to filaments is not always known nor it is easy to determine. In addition, treating actomyosin as a percolating network or as a gel limits the availability of the information about the hierarchy of

connectivity within the network scaffold. To conduct further analysis about the morphology of actomyosin networks, I calculated the order parameters derived from network theory on the graph of filaments, which is built by accounting only for the distances between filaments, without considering the types of ABPs responsible for such connectivity, as described also in Section 2.2.3 (vide supra).

In this section, I use this approach to develop an alternative, big-picture representation of actomyosin networks. In this graph representation of the actomyosin networks, the filaments are defined as nodes in the graph, and an edge between two nodes exists if they are within a cutoff distance $d_{\text{cutoff}} = 200\,\text{nm}$.

3.3.6 Analysis of actin filament networks viewed by order parameters from graph theory

The topology of an actomyosin network becomes increasingly complex over time, as shown in Fig. 3.2 and Fig. 3.3. Network theory provides a simplified method to describe the hierarchical ranks of a network built by these filaments. Within such a description, clique numbers, number of communities, and clustering numbers are used to describe the status of a node, or filament, within a growing actomyosin network. Graph theory allows a variety of perspectives, from a fine-grained (zoomed-in) to a coarse-grained (zoomed-out) view. Moreover, the particular strength of network theory lies primarily in its ability to describe scales beyond the connectivity of ABPs and to provide the social status of a node (or filament) in a hierarchical actomyosin network. I explore the effect and interplay of multilinkers with a variety of valencies and motor proteins on the morphology of actomyosin networks, by studying the time signal of several order parameters such as clique number, node degree assortativity, and graph density in Fig. 3.7, 3.8, and 3.9, respectively. Lastly, I compare the PMF of a graph theory order parameter with the PMF of gelation theory; this demonstrates that network theory may distinguish between bundles and gels in actomyosin networks, as shown in Fig. 3.10.

We first zoomed in on an actomyosin network at the fine-grained level where multivalency differentiates topologies at either high or low motor content. At low motor content, filaments

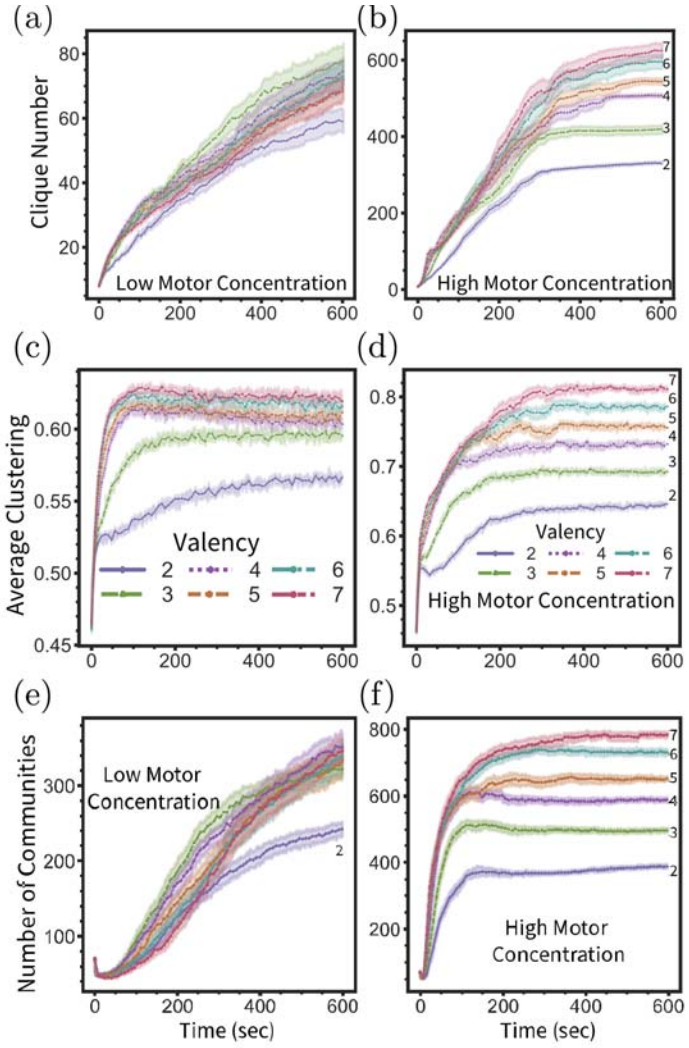

Figure 3.7 Order parameters derived from network theory, measuring proximity within the network in the presence of multilinkers for two systems with 1000 nodes (filaments). The left-side column (a, c, e) represents the system with "low motor concentration", i.e., $1:100:100:100$ concentration ratios among the motors, filaments, multilinkers, and crosslinkers. The right-side column (b, d, f) represents the systems with "high motor concentration", i.e., $1:1:1:1$ concentration ratios among the motors, filaments, multilinkers, and crosslinkers. The clique number of the network as a function of time is plotted in (a, b). The average clustering, defined in Eq. (3.18), is plotted as a function of time in (c, d). The number of communities [48] of the network is plotted as a function of time in (e, f). Of note, each system profile represented herein is averaged over 30 independent trajectories with a confidence interval of CI = 90%.

produce a similar kind of bundles at all valency values, from $\nu = 2$ to $\nu = 7$, according to the results depicted in Fig. 3.7(a, c, e). In contrast, at high motor content, the differences in network

topology obtained with multilinkers of various valencies are more prominent, as the clique number (Fig. 3.7(b)) profiles, the average clustering profiles, (Fig. 3.7(d)), and the number of communities (Fig. 3.7(f)) deviate significantly over time. Notably, the number of communities [48] that maximize the modularity in Eq. (3.16) is the highest for systems with multilinkers of valency $\nu = 7$, as shown in Fig. 3.7(f). This result provides important insight into systems with multilinkers of higher valencies, which is due to the effect of high valency on the average degree of the nodes in the graph (as shown in Fig. 3.7(c-d)).

When zooming out to view the network at a coarse-grained level, it is possible to examine the node degree relations and their global contributions to hierarchical connectivity in the network. Higher valency multilinkers reduce the node degree assortativity coefficient defined in Eq. (3.26) both at systems with low and systems with high motor content, as shown in Fig. 3.8(a) and Fig. 3.8(c), respectively.

The observation that in systems with high multivalency values (e.g., $\nu \geq 4$), that can interconnect more filaments with a single multilinker, the degree assortativity is reduced suggests that motors may break the large bundles into dendritic network by the arborization of the large bundle, as observed in Fig. 3.8(d). Indeed, at high motor content, motors reduce the assortativity of the network from 0.7 to 0.5 when the multilinkers are hexavalent ($\nu = 6$) or heptavalent ($\nu = 7$). However, the tendency of the network at low motor content is to create a meshgrid of bundles (can be thought of as a gel of actin bundles), where the assortativity is higher, as shown in Fig. 3.8(a,b). Notably, at high motor content, when a large bundle is emerged in the system the motors arborize the large bundle into a thick bundles with actin branches, as shown in Fig. 3.8(c,d).

3.3.7 Network theory order parameters reveal rich topologies in a dendritic actomyosin network

Beyond the order parameters that are used to quantify the connectivity of both of the nodes (filaments), the non-local order parameters in graph theory, such as the graph density, reveal the global heterogeneity of actomyosin networks. In Fig. 3.9(a) in particular, the observed graph density

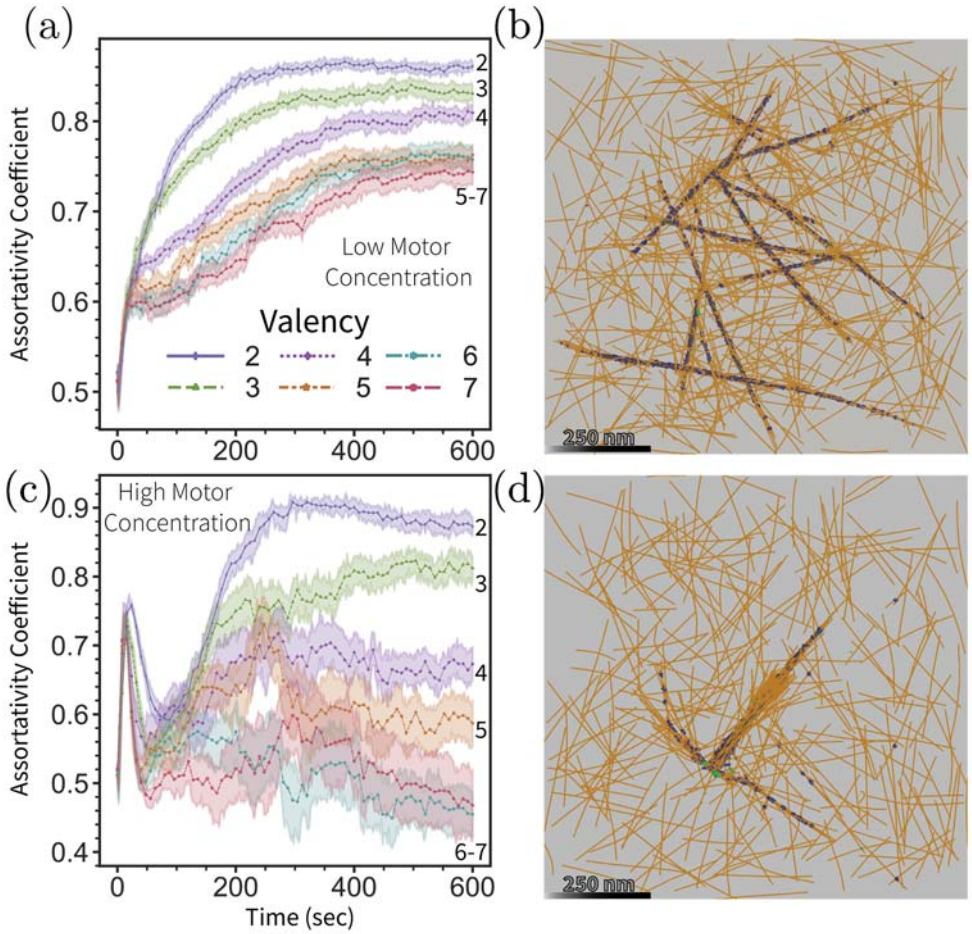

Figure 3.8 Node degree assortativity coefficient of actomyosin networks observed in the presence of multilinkers with varying valencies. The node degree assortativity, as defined in Eq. (3.26), is plotted as a function of time at "low motor concentration" (i.e., $1:100:100:100$ ratios between motors, filaments, multilinkers, and crosslinkers) in (a) and at "high motor concentration" (i.e., $1:1:1:1$ ratios) in (c). Each profile in panel (a) and panel (c) was averaged over 30 independent trajectories with a confidence interval of CI = 90%. The snapshots depicted in (b) and (d) are of $1\,\mu\text{m}^3$ cubic systems with multilinkers of valency $\nu = 6$ at low and high motor content, respectively, at time $t = 600\,\text{s}$.

signal shows that even at low motor activity, the presence of multilinkers with valencies $\nu > 2$ increases the density by three-fold. At high motor content, such as in the case depicted in Fig. 3.9(b), results strongly suggest that the dynamics of the network has more than two phases as in the case of gel-sol theory, since the graph density can distinguish between a gelated network and a bundled network. For the case of a multilinker with $\nu = 3$, the graph density shows that there are two stages

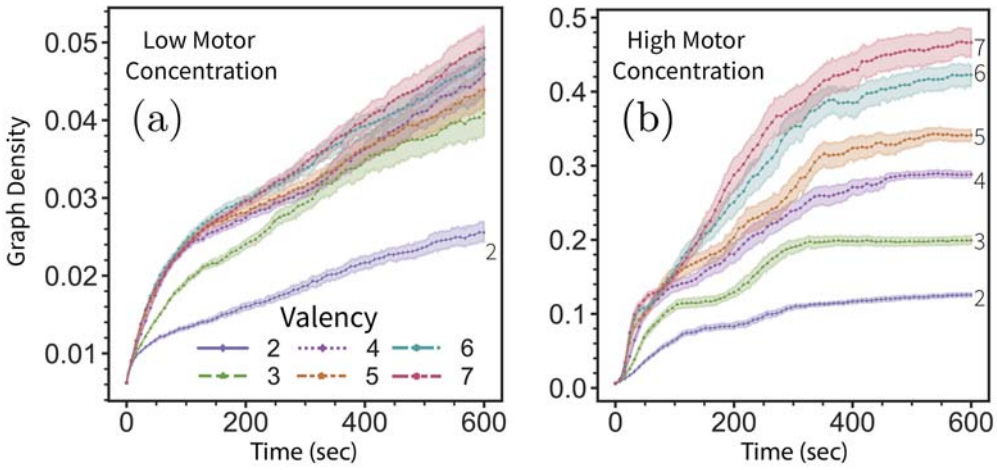

Figure 3.9 Graph density of actomyosin networks in the presence of multilinkers with varying valencies. The graph density, defined in Eq. 3.27, is plotted as a function of time at 1:100:100:100 concentration ratios between motors, filaments, multilinkers, and crosslinkers in (a) and at 1:1:1:1 ratios in (b). Each profile depicted in (a) and (b) is an average of 30 independent trajectories, with the same species parameters, with an overall confidence interval of CI=90%.

of evolution characterized by a double sigmoidal curve: the graph density is less than 0.1 before $t = 200\,\mathrm{s}$ and abruptly grows to 0.2 after $t > 300\,\mathrm{s}$. This double sigmoidal curve suggests that the disordered network becomes gelated and then bundled. Therefore, tracking only percolation using the order parameters in gelation theory may be inadequate.

Because network theory order parameters capture the hierarchical order of actin filaments in an actomyosin network, they are useful in revealing the states of matter in actomyosin. It is desirable to develop them into the order parameters that describe the complexity of an energy landscape. To show such a possibility, I show the potential of the mean force (PMF), defined in Eq. (3.30), as a function of an order parameter from graph theory and as a function of an order parameter from gelation theory in Fig. 3.10. This figure includes information for a variety of the sizes of gel clusters and compares it with the PMF of the average node neighbor degree of each node along a trajectory of 600 s that was sampled every second. The results depicted in Fig. 3.10 support the observation that with network theory information, there are more than two phases (noted by the number of local minima) of the actomyosin networks with multilinkers (Fig. 3.10(a)), compared

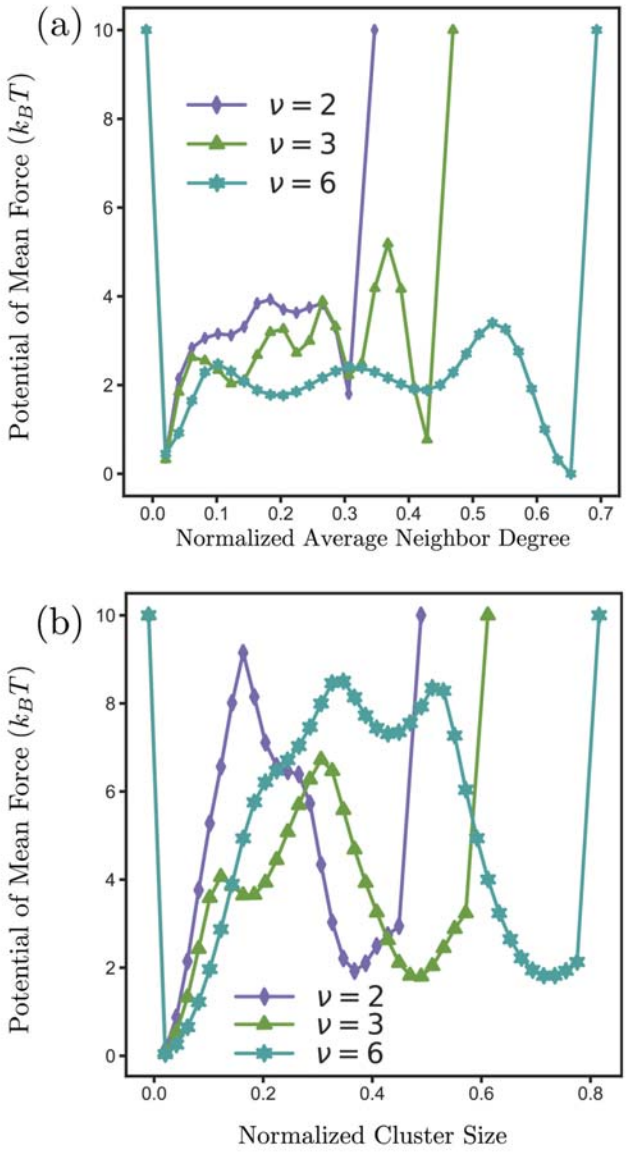

Figure 3.10 Potential of Mean Force (PMF) of (a) network and (b) gelation theory parameters. The potential of mean force is defined in Eq. 3.30 of the probability density function over the cluster sizes in (a) of systems with multilinkers of different valencies. The PMF of the probability density function of the average neighbor degree of each node is plotted in (b). The results plotted in panels (a) and (b) were computed at 1:100:100:100 concentration ratios between motors, filaments, multilinkers, and crosslinkers.

with the information about the states revealed from sol-gel theory, as there are only two global minima in Fig. 3.10(b).

3.3.8 State diagrams of actomyosin networks in the plane of order parameters from network theory

Overall, through the use of the proper order parameters connecting local and non-local features in a complex network established from network theory, I have outlined the state diagrams in Fig. 3.11 and Fig. 3.12 that allow further experimental validation of our hypothesis that both motor activity and the valency of the multilinkers in the system regulate the dendritic state of the actomyosin network. The state diagrams in Fig. 3.11 and Fig. 3.12 are computed by averaging over the last 20 seconds of simulations, for which a specific ratio of motor-filament and multilinker-filament holds.

First, I show the state diagram of the node degree assortativity coefficient. At a $2:1$ (horizontal dashed line in Fig. 3.11) motor-filament concentration ratio or above, the degree assortativity is reduced as shown in Fig. 3.11. At a multilinker-filament concentration ratio of $1:1$ (vertical dashed line in Fig. 3.11) the network behaves in a non-linear and complex manner, as the degree assortativity first rises and then drops. Both a high concentration of multilinkers and a high activity of motors are necessary to regulate the correlation between the nodes in the network. At a multilinker-filament concentration ratio above $4:1$, as shown in Fig. 3.11, the motor activity does not attenuate the assortativity, as in the case where the motor-filament concentration ratio is between $1:2$ and $2:1$.

The horizontal dashed line in Fig. 3.12 shows that the graph density of the network does not necessarily increase by adding more multilinkers, nor does it increase through the addition of more motors, as suggested by the vertical dashed line in Fig. 3.12. Rather, the data suggests that by increasing the concentration of multilinkers in the actomyosin system, the graph density was reduced overall in the filament network.

3.4 Discussion

I simulated in Cytosim [156] the evolution of actomyosin networks that occurs through mediation of filaments by ABPs. The utilization of several order parameters, adapted from three scientific subdisciplines, helped to characterize the morphologies of the networks. Order parameters that are

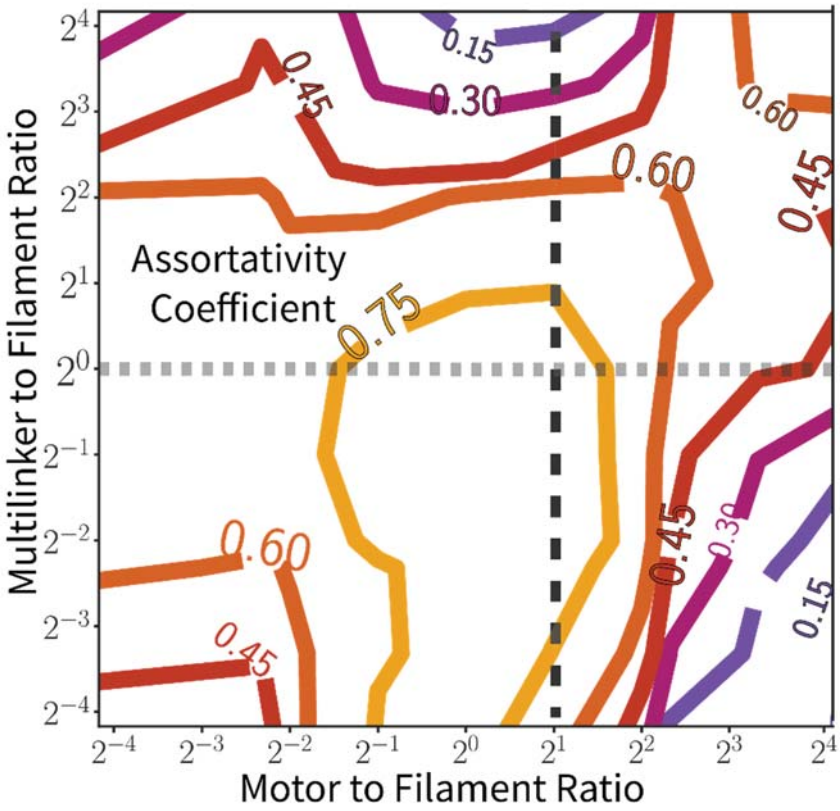

Figure 3.11 State diagram of the node degree assortativity coefficient. The average over the steady state values of the node degree assortativity coefficient, defined in Eq. (3.26).

derived from polymeric physics and gelation theory are both valuable, but they lack the ability to characterize the overall complexity of the topology within the clusters of filaments. Polymeric physics is inadequate because it was originally developed to classify the conformation of a single complex polymer; gelation theory on the other hand lacks the formalism to distinguish between a weakly connected gel and a fully connected bundle, as both conformations have similar gelation order parameter values. In contrast, using network theory-derived parameters allowed me to distinguish between three types of gels: a meshgrid of bundles, a large bundled gel, and an arborized bundled gel. In particular, the clique number derived from network theory shows the organization of filaments in systems with different types of multilinkers and motor concentrations. At the coarse-grained level,

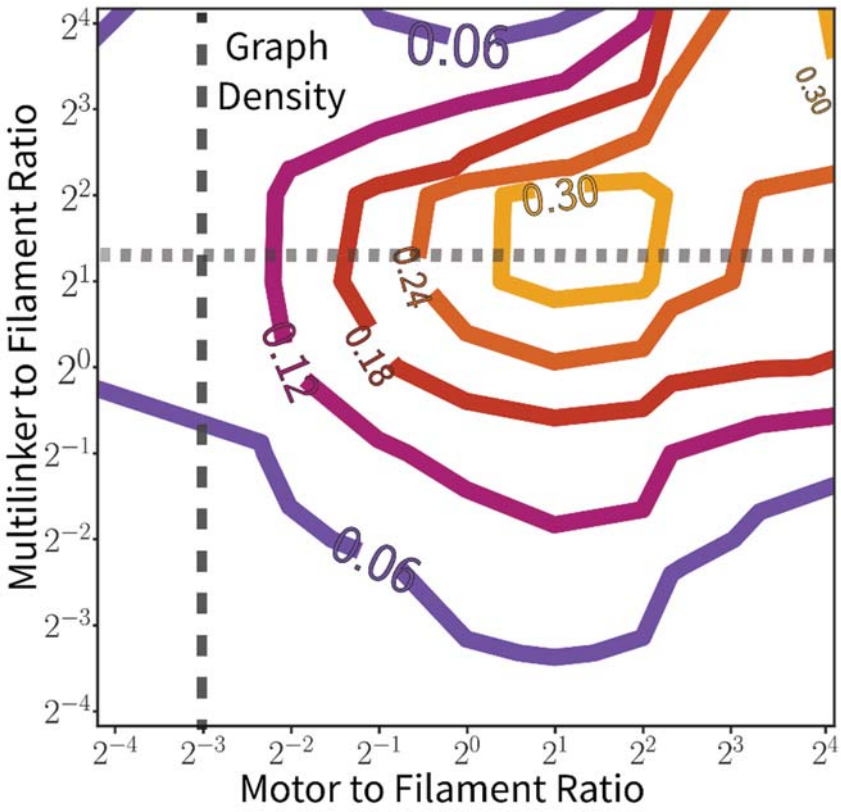

Figure 3.12 State diagram of graph density. The average values are defined over the steady state values of the graph density, defined in Eq. (3.27).

the graph density from network theory allowed us to predict the effects of the valency multilinkers versus motor concentration on the topology of the networks, through the use of a state diagram. Such descriptions function effectively across multiple length and time scales, and are particularly useful in connecting microscopic properties to macroscopic phenomena that can be validated experimentally. For example, when the concentration of motors or of multilinkers increases, at a one-grained level, this leads to a reduction in the order parameters, such as the clique number of the network as well as the degree correlation between the nodes. At a coarse-grained level, similar concentration changes results in a decrease in the overall system density, as a result of the formation of hierarchical network structures. This result parallels the known effects, observed experimentally, that adding

more crosslinkers into systems prevents overall system contraction [23].

This work shows that many interesting features of actomyosin networks are captured by focusing on the nodes (or filaments) from the system level, starting from the level of connectivity through actin-binding proteins to a larger scale. This movement of scales can occur by knowing only the distances between actin filaments, and without specific knowledge of the actin-binding proteins that are present. Such a reduced graphical representation can markedly ease the analysis of the dynamics and structures of the actomyosin networks, through which they will accurately reflect the network reorganization caused by motor and actin-binding protein variations. Such a framework to evaluate actomyosin networks is valuable also in the context of analysis of live high-resolution experiments, in which real-time location tracking for the filament components is observed.

Actomyosin structures mediated by multivalent actin-binding proteins shape the synaptic plasticity in a dendritic spine

CaMKII is a multivalent ABP that serves a dual role: as a structural protein that organizes actin filaments and as a signaling protein that mediates calcium signaling. Regarding its structural role, Khan's as well as Waxham's group have shown that CaMKII is a multivalent linker that forms actin bundles [98] as well as junctions [114]. CaMKII binds actin with high affinity that maintains the stability of actin bundles [98] over a long period of time, i.e., days or months. My work agrees with the experimental findings that high valency promotes actin-bundling, as in the case of CaMKII. In addition, my work suggests that motor activity promotes larger bundles and, at high multivalencies, I observed arborization of bundles that can explain the dendritic shapes of the spine in neurons [170, 136].

Possible experimental validation of the effect of multilinkers and motors

Although this study offers only qualitative insights into the role of multilinkers, such as CaMKII, in regulating the morphologies of actomyosin networks, it should be possible to validate those by *in vitro* experiments. For example, I will describe how one could design an experimental setup to validate the predicted state diagrams of the degree assortativity (Fig. 3.11) and the graph density (Fig. 3.12).

My suggestion is based on the protocol by Bendix et al. [23], who studied the role of the α-actinin concentration and the motors concentration on actomyosin architectures. Building on their protocol, one may add another multivalent actin-binding protein, such as a purified CaMKII [165], and study a few (e.g., 16) combinations of motor-to-actin and multilinker-to-actin concentration ratios. For each combination of motor-to-actin and multilinker-to-actin concentration ratios, one would reconstitute an actomyosin system with a fixed amount of purified actin and α-actinin crosslinker proteins. The systems may be then visualized using adaptive rheology, as demonstrated by Gupta et al. [93], and by recording the dynamics of fluorescent labeled F-actin using the "RFP-Ftractin" actin labeling protocol [232]. In order to compute from experiments the state diagram of any order parameter from graph theory (e.g., graph density), one could use high-resolution images of the steady states of the different systems and translate them into graphs using an edge detection algorithm such as Canny's algorithm [41]. Once the graphical representations of the experimental images are available, one can compute the order parameters using NetworkX [95], similar to what I did with the results from the Cytosim [156] simulations. Overall, I expect the results from the experiments to be qualitatively similar to the predictions obtained by the multilinker simulations in Cytosim, namely that there should be a "Goldilocks Zone" (i.e., an optimal amount of multilinkers and motors) at which the assortativity and graph density achieve their maximum values.

In summary, I introduced a computational mesoscopic model for multilinkers built into Cytosim and ran the simulations of actomyosin dynamics in the presence of multilinkers with varying binding valencies to actin filaments. I then analyzed the effects of motor activity and the valency of the multilinkers on the morphologies of actomyosin, using order parameters from polymer physics, gelation theory and network theory. Overall, the results elucidate the relevance of these order parameters for the study of the structure and dynamics of actomyosin networks. I suggest a possible reduced space, using network theory methods, to understand and characterize complex actomyosin networks. Although this is only a start, hopefully this will motivate future directions to embrace network theory as a tool to analyze actomyosin experiments, as well as to establish a quantitative understanding of cytoskeletal systems. I believe that the representation of actomyosin as graphs

would be useful beyond the scope of our analysis. The realization that actomyosin could be thought of as graphs will allow one to apply graph neural network [234] models to analyze and predict actomyosin dynamics and presumably to integrate nonequilibrium physics theories [197, 198, 108] with probabilistic graphical models [140, 73] to describe complex active matter, such as actomyosin networks.

4 Avalanches in Actomyosin Networks

This chapter is based on a paper published in PNAS [133] in 2020.

4.1 Introduction

Actomyosin networks are vital for proper functioning of living matter [154]. In particular, two major biological processes that rely on the non-equilibrium behavior of actomyosin are memory formation [94] and cell motility [39, 179, 152]. Relatedly, a key actin-binding (actin-related) protein Arp2/3 complex [52] is also implicated in cell motility [39, 153] and memory formation [94, 117, 118]. The Arp2/3 complex (Fig. 4.1(b)) is a protein complex with seven subunits that can bind to the side of an existing actin filament and initiate the nucleation of a daughter filament at a 70° angle from the mother filament [184], as shown schematically in Fig. 4.1(a).

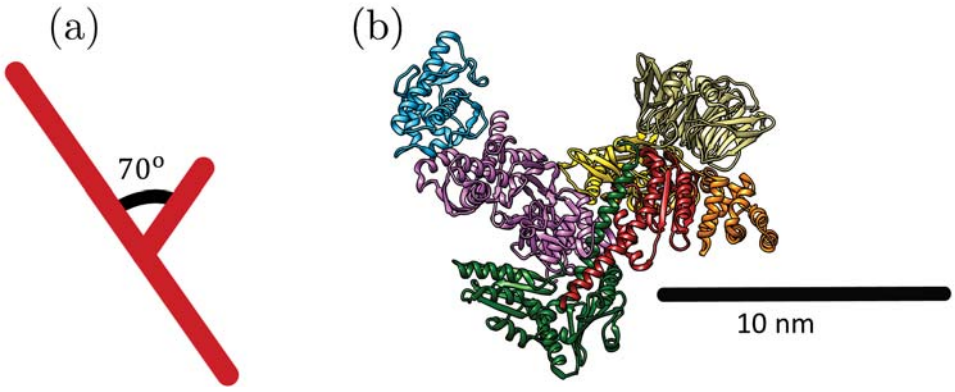

Figure 4.1 Arp2/3, defined as the brancher protein in actomyosin networks. (a) Conceptual illustration of an Arp2/3 brancher that creates a 70° angle between mother and daughter filaments. (b) The crystal structure of Arp2/3 [182], as visualized using VMD [102].

Actomyosin networks may also contract, and this motion has been replicated *in vitro* using purified actin filaments, crosslinkers, and active motors [23]. Experiments have shown that contractility emerges in the presence of ATP and above a critical concentration of myosin motors. However, the dependence of contractility on the concentration of bivalent crosslinkers has shown that such contractility only occurs at intermediate crosslinker concentrations, when found in a solution of

actin filaments, motors, and crosslinkers [23]. In addition, it is unclear how the formation of branches in actomyosin influence system contractility, overall system structure, and the force generation of the network. Herein, I present the simulations of a $1\,\mu m^3$ system to study how branchers, crosslinkers, and motors reshape the actin filaments morphologies and dynamics in the dendritic spine whose morphological plasticity is strongly implicated in learning and memory processes [99, 47]. Contractility in actomyosin networks resembles the Feynman-Smoluchowski thermal ratchet [75, 144, 110, 84], which is a classical thought experiment describing how directional behavior can stem from randomness. In a similar manner to how ratchets rectify applied force to allow motion in one direction, actomyosin networks may undergo contraction events in the presence of an energy-consuming reaction and a broken symmetry to facilitate motion in the target direction [144]. Contraction events in actomyosin networks originate mostly from forces generated by motors (across actin filaments), but may also stem from the compression and expansion of actin filaments due to their polymerization [74, 119, 229, 223, 227]. When forces deform the network and build up strain, this in turn causes notable changes in the rates of various biochemical processes. The goal of this research is to explore the contrast between the dynamics of branched and unbranched networks, via computational simulations of systems with and without the Arp2/3 brancher complex. The simulations were done on the Mechanochemical Dynamics of Active Networks (MEDYAN) framework, which is characterized as a mechanochemical computational software that incorporates a stochastic description of individual chemical reaction events and a deterministic force-minimization scheme [180] while accounting for the mechanochemical feedback loops. The simulations show that branched networks (i.e. those that contain Arp2/3) can become marginally stable and abruptly release stress in the network via a sudden mechanical avalanche. In contrast, in unbranched networks (i.e. those without Arp2/3), no such event has been observed in our simulations. Finally, I discuss the finding of avalanche events that occur in actomyosin in light of a recent discovery of earthquakes in the cytoskeleton ("cytoquakes") and the implication of such cytoquakes to the micron-length scale of dendritic spines.

The modeling of Arp2/3 in MEDYAN was a collective effort by James Liman, Carlos Bueno, and

I. The simulations were mainly carried out by James Liman. Most of the analysis was done by James Liman with extensive help from Carlos Bueno. My main contribution was the characterization of contractility and developing the order parameters for the analysis of avalanches.

4.2 Methods

4.2.1 Coarse-grained mechanochemical model of actomyosin systems (MEDYAN)

For this project, the mechanochemical model of actomyosin networks MEDYAN (vide infra) [180] was used. MEDYAN models the mechanical and the chemical kinetics coarse-grained actomyosin networks in rounds of mechanochemical steps. In the chemical step of the simulation, MEDYAN uses a stochastic scheme based on the spatially resolved Gillespie algorithm [86, 87, 85]. In the mechanical step of the simulation, MEDYAN computes the total mechanical energy in the system and relaxes the system by minimizing the energy of the actomyosin network [180]. MEDYAN's mechanochemical feedback loop is depicted schematically in Fig. 4.2.

The coarse-grained actomyosin system is initialized with filaments that are made of segments consisting of 10 G-actin monomers each, non-muscle myosin IIA heavy chain (NMIIA) motor proteins (i.e. motors), α-actinin cross-linking proteins (i.e. linkers), and actin-related protein 2/3 (Arp2/3) branching protein complexes (i.e. branchers). James Liman ran two types of simulations: unbranched simulations and branched simulations. In the unbranched simulations, the chemical species included actin monomers, crosslinkers, and motors at various concentration ratios. In the branched simulations, branchers were also included in the simulations while keeping the same concentration ratios of actin monomers, crosslinkers, and motors as in the unbranched simulations.

Chemical reactions in MEDYAN

The simulations included chemical events of actin polymerization and depolymerization, motor walking, and binding and unbinding events of actin-binding proteins (ABPs) such as: motors, crosslinkers, and branchers. The system was confined to a $11\,\mu m^3$ box with $15,000$ actin monomers, corresponding to $25\,\mu M$ [23], with initial 50 filaments at length of one actin segment (10 monomers),

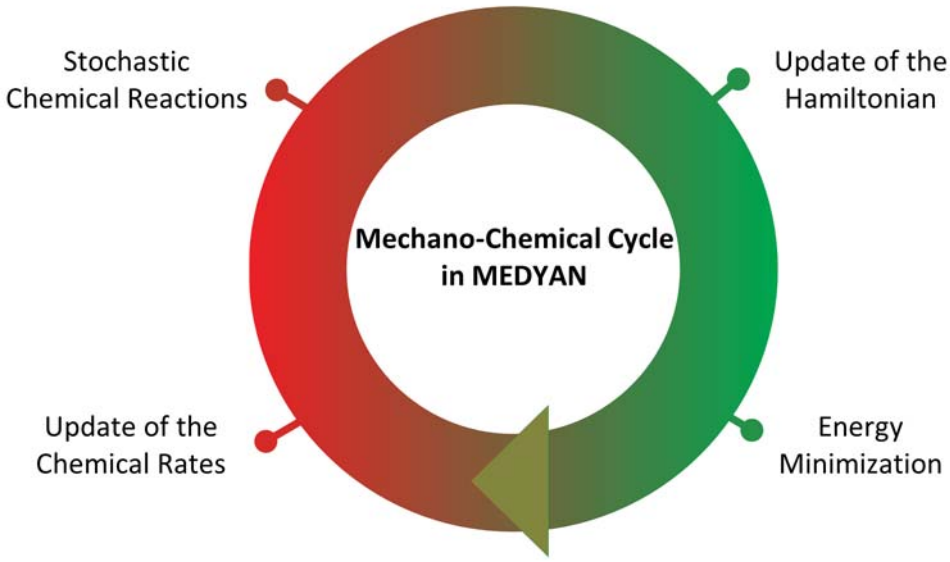

Figure 4.2 MEDYAN mechanochemical loop to simulate actomyosin networks. Upon stochastic simulation of chemical reactions, the mechanical Hamiltonian is updated and the total energy of the system is minimized. Then the mechanical forces update the chemical reactions before executing the chemical simulation [180].

meaning that the total number of monomers in the pool at time = 0 was 14,500. Of note, the presence of branchers in the system allows the nucleation of a new daughter filaments after binding to a mother filament. James Liman considered nine different concentration ratios of motor proteins to actin monomers $x_{m:a}$ and eight concentration ratios of α-actinin crosslinker proteins to actin monomers $x_{\alpha:a}$ in the simulations, and one concentration ratio of Arp2/3 brancher proteins to actin monomers $x_{b:a}$ which corresponds to 300 branchers (0.5 µM). The on and off rates for the crosslinkers, motors, polymerization and depolymerization chemical events follow the values provided in the original MEDYAN publication [180], which is based on values derived from experimental data [23, 121, 153, 178, 219]. Due to the known high stability of Arp2/3-actin protein complexes, the unbinding rate of branchers was set to zero and the binding rate to $10^{-4}\mathrm{s}^{-1}$, which allowed gradual brancher nucleation of binding sites along the simulated system trajectories [178].

Parameter	Description	Values
$x_{m:a}$	Motor to actin monomer ratio	$\left\{0, \frac{5}{10^3}, \frac{1}{10^2}, \frac{2}{10^2}, \frac{5}{10^2}, \frac{1}{10}, \frac{2}{10}, \frac{5}{10}, 1\right\}$
$x_{\alpha:a}$	Crosslinker to actin monomer ratio	$\left\{0, \frac{1}{10^2}, \frac{2}{10^2}, \frac{5}{10^2}, \frac{1}{10}, \frac{2}{10}, \frac{5}{10}, 1\right\}$
$x_{b:a}$	Brancher to actin monomer ratio	$\frac{1}{50}$

Table 4.1 The concentration ratios between ABPs and actin monomer used in the simulations.

Mechanical forces in MEDYAN

After the changes of the chemical composition in the system are simulated, MEDYAN then computes the total forces in the system. Included within the system design are actin filament bending, stretching, branching, and exclusion volume potentials; motor and linker stretching harmonic potentials; and filament-boundary interaction potentials, with the default constant values and configurations set according to reference from [180]. The bending potential of the semiflexible rod is defined by the equation

$$U_i^{\text{bend}} = k_{\text{bend}} \left[1 - cos(\theta_{i,i+1})\right], \tag{4.1}$$

where $k_{\text{bend}} = 2690\,\text{pN nm}$ [180] is the bending energy constant and $\theta_{i,i+1}$ is the angle between neighboring filament segments at index i and $i+1$. The stretching potential of the semiflexible rod is given by the equation

$$U_i^{\text{stretch}} = \frac{1}{2} k_{\text{stretch}} \left(|\mathbf{l}_i| - l_0\right)^2, \tag{4.2}$$

where k_{stretch} is the stretching energy constant, $\mathbf{l}_i$ is the vector connecting the endpoints of the ith filament segment, and l_0 is the equilibrium length of a filament segment.

The polymer branching potential is given by the equation

$$U_{i,j}^{\text{branch}} = U_{i,j}^{\text{branch,stretch}} + U_{i,j}^{\text{branch,angular}}, \tag{4.3}$$

where i and j are filament segments in the mother and the daughter filaments, respectively. The

branched polymer stretching potential is given by the equation

$$U_{i,j}^{\text{stretch}} = k_{\text{stretch}}^{\text{branch}} \left(|\mathbf{d}_{i,j}| - d_0\right)^2, \tag{4.4}$$

where $k_{\text{stretch}}^{\text{branch}} = 100\,\text{pN nm}$ is the stretching constant of the branched filament, $\mathbf{d}_{i,j}$ is the distance between the binding site in segment i in the mother filament and segment j in the daughter filament, and $d_0 = 6\,\text{nm}$ is the equilibrium length of $\mathbf{d}_{i,j}$. The branched filament angular potential is given by the equation

$$U_{i,j}^{\text{angular}} = k_{\text{angular}}^{\text{branch}} \left[1 - cos(\theta_{i,j} - \theta_0)\right], \tag{4.5}$$

where $k_{\text{angular}}^{\text{branch}} = 100\,\text{pN nm}$ is the angular energy constant for the branched filament, $\theta_{i,j}$ is the angle between the binding site in segment i in the mother filament and segment j in the daughter filament, and $\theta_0 = 70^\circ$ is the equilibrium angle between branched polymers [180, 153].

The motor stretching harmonic potential is given by the equation

$$U_{i,j}^{\text{motor}} = \frac{1}{2} k_{\text{motor}} \left(\left|\mathbf{l}_{i,j}^m - l_0^m\right|\right)^2, \tag{4.6}$$

where $k_{\text{motor}} = 2.5\,\text{pN nm}^{-1}$ is the stretching energy constant of motors, $\mathbf{l}_{i,j}^m$ is the instantaneous length of the motor, as measured by the distance between its binding sites, and $l_0^m = 200\,\text{nm}$ is the equilibrium length of the motor. Crosslinkers are also modeled with a stretching harmonic potential analogous to those defined in Eq. (4.6), but with $k_{\text{crosslinker}} = 8.0\,\text{pN nm}^{-1}$ and $l_0^l = 35\,\text{nm}$.

The exclusion volume potential between two filament segments on two adjacent polymers is given by

$$U_{i,j}^{\text{vol}} = k_{\text{vol}} \int_0^1 \int_0^1 \frac{ds ds'}{\left|\mathbf{r}_i(s) - \mathbf{r}_j(s')\right|^4}\ , \tag{4.7}$$

where $k_{\text{vol}} = 10^5\,\text{pN nm}^3$ is the repulsion constant between two filament segments [180], $\mathbf{r}_i(s)$ and $\mathbf{r}_j(s')$ are parametric representation of segment i and segment j with $s, s' \in [0, 1]$.

The interaction potential between a filament and the boundary of the cubic box is given by

$$U_i^{\text{boundary}} = \begin{cases} k_{\text{boundary}}\, e^{-\frac{d_i}{\lambda}} & d_i \leq d_{\text{cutoff}} \\ 0 & d_i > d_{\text{cutoff}} \end{cases}, \tag{4.8}$$

where $k_{\text{boundary}} = 41\,\text{pN}\,\text{nm}$ is the repulsive energy constant from the boundary, d_i is the distance to the closest boundary and the ith filament segment, and $\lambda = 2.7\,\text{nm}$ is the screening length. After accounting for all the mechanical potentials listed above, MEDYAN minimizes the energy of the system using the well-precedented Polak-Ribière conjugate gradient method [180]. After the energy minimization is completed, a new mechanical configuration of the system (including new locations and stress profiles of filaments, bound motors and linkers, and branchers) is determined.

Mechanochemical coupling in MEDYAN

MEDYAN runs the chemical and mechanical steps in tandem. This means that once a new mechanical configuration is determined by the mechanical step, after the energy minimization step, MEDYAN then updates the rates of the simulated chemical reactions. The reaction rates change in this context due to the mechanical force that is acting on all system components. In one example of such force-dependent reaction kinetics, a polymerization reaction is less likely to occur on a filament that is located closer to the boundary, due to greater repulsion boundary forces that slow the reaction kinetics. Moreover, the unbinding reaction of linkers is modeled using a slip bond and the unbinding reaction of motors is modeled using a catch bond. As a result, a linker is more likely to unbind when pulling forces are applied to the linker, whereas a motor is less likely to unbind under the same force conditions.

4.2.2 Mechanistic order parameters

We used three order parameters to characterize and analyze contractions and avalanches in the networks. Two of them are derived from the inertia tensor of filament beads, $\mathbf{T}_{\alpha\beta}$, defined as

$$\mathbf{T}_{\alpha\beta} = \frac{1}{2N^2} \sum_{i,j=1}^{N} \left(\mathbf{r}_{i\alpha} - \mathbf{r}_{j\alpha}\right)\left(\mathbf{r}_{i\beta} - \mathbf{r}_{j\beta}\right), \tag{4.9}$$

where $\mathbf{r}_{\alpha\beta}(t)$ represents the α-component of a filament bead, $N = 15,000$ is the number of filament beads, and $\alpha, \beta \in \{x, y, z\}$ are the indices of the Cartesian elements. The radius of gyration is given by:

$$R_g = \sqrt{\operatorname{tr} \mathbf{T}} = \sqrt{\sum_i \lambda_i}, \tag{4.10}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues of $\mathbf{T}$. The shape parameter S is also derived from $\mathbf{T}$ [56]

$$S = \frac{\Pi_{i=1}^{3}(\lambda_i - \bar{\lambda})}{(\frac{1}{3}\operatorname{tr} \mathbf{T})^3} \sqrt{\operatorname{tr} \mathbf{T}}, \tag{4.11}$$

where $\bar{\lambda} = \frac{1}{3}\sum_i \lambda_i$.

To accurately identify an event as an avalanche, the mean displacement of the center of masses of actin filaments was used, with the understanding that avalanche events often move a few filaments simultaneously. The mean filament center of mass displacement $\delta x_F(t)$ was carried out by computing the following equation

$$\delta x_F(t) = \frac{1}{N_F} \sum_{i}^{N_F} \left|\mathbf{x}_i^{cm}(t) - \mathbf{x}_i^{cm}(t-1)\right|, \tag{4.12}$$

where N_F represents the number of filaments in the system, $\mathbf{x}_i^{cm}$ is the center of mass of the ith filament, $|\cdot|$ represents the norm of a vector in 3D. An event was classified as being an avalanche when the mean filament displacement $\delta x_F(t)$ between a successive pair of snapshots was "sufficiently large", defined by requiring its z-score value relative to the displacements for the other intervals to exceed five standard deviations (corresponds to a p-value of $3 \cdot 10^{-7}$).

4.2.3 Gelation and percolation in the connectivity graph of actin monomers

In order to determine how avalanche events are related to the connectivity between actin monomers in the network, the theories of gelation and percolation were used. These theories have been studied

extensively over decades [34]: since the work in the early 1940s by Flory and Stockmayer studying the mechanics, kinetics, and thermodynamics of the gelation process. The name "percolation" was coined in 1957 by Broadbent and Hammersley to describe the properites of fluids that spread within a random media. Erdős–Rényi then studied percolation in a random graph, followed by Mandelbrot's work that added the idea of a fractal into the theory of percolation. In order to study the percolation in the finite graph of actin monomers, I first constructed an undirected graph $\mathbf{G}(t) = (V(t), E(t))$ at time t, where each actin monomer represents a node in V and denote the degree of a node v with $k(v)$. There is an edge between two nodes v_1 and v_2 if their corresponding actin monomers are linked by an ABP or are consecutive monomers along the same filament polymer.

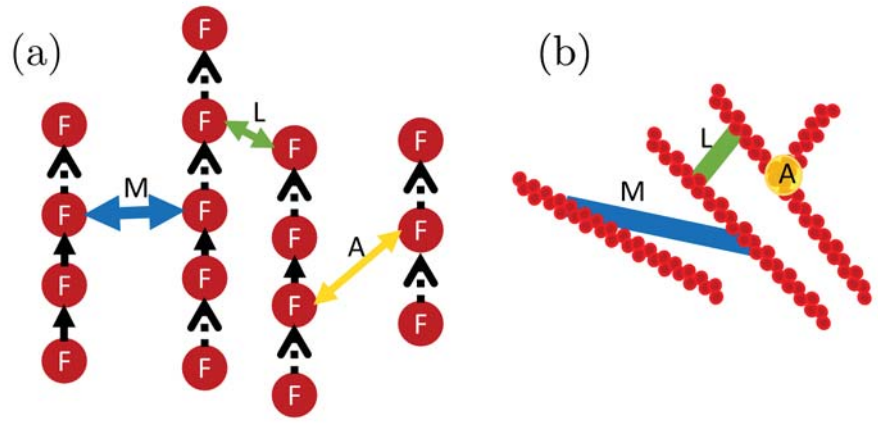

Figure 4.3 Illustration of the actin monomer graph construction. (a) The graph representation of a simple configuration of actin monomers nodes connected with edges according to the connectivity of the linkers, motors, and branchers. (b) A toy physical configuration of four filaments connected with one motor, one linker and a brancher.

Of note, in MEDYAN, an actin monomer can bind to up to one ABP. Therefore, within this graph representation of actin monomers, there are four possible values of node degree $k(v) \in \{0, 1, 2, 3\}$:

$$k(v) = \begin{cases} 0 & \text{node } v \text{ is a floating monomer} \\ 1 & \text{node } v \text{ is at the end of a filament} \\ 2 & \text{node } v \text{ is in the middle of a filament} \\ 3 & \text{node } v \text{ is in a filament and connected by an ABP to a second filament} \end{cases}, \quad (4.13)$$

where an ABP can be a brancher, a motor, or a crosslinker that is represented as an edge between two beads (i.e. actin monomers). For simplicity, from the representation $\mathbf{G}$, I considered only the subgraph $\mathbf{G}_{\text{sub}}(t) = (V_{\text{sub}}, E_{\text{sub}})$, where $V_{\text{sub}} \subsetneq V$ and $E_{\text{sub}} \subsetneq E$. The subgraph $\mathbf{G}_{\text{sub}}$ contains only the monomers that are connected to a filament, i.e., only nodes with the degree greater than 0: for any $v \in V_{\text{sub}}$ the degree $k_v \in \{1, 2, 3\}$. A sample illustration of such a graphical representation is shown schematically in Fig. 4.8(a), which corresponds to the sample actomyosin network shown schematically in Fig. 4.8(b). Since any possible degree distribution of $k_{\text{sub}}(v)$ defined on the subgraph $\mathbf{G}_{\text{sub}}$ has only 3 possible values, it is straightforward to predict if a network will be percolated or gelated, given the number of actin monomers (beads), number of branchers, number of crosslinkers, and the number of motors. I defined the probability of an edge emanating from a node v_{sub} of degree $k(v_{\text{sub}})$ to lead to a node v'_{sub} of degree $k(v'_{\text{sub}})$ as $P\left(k(v_{\text{sub}}) \mid k(v'_{\text{sub}})\right)$, which can be simplified to the notation $P\left(k' \mid k\right)$. Starting from a random node in V_{sub}, one could consider the following graph traversal scheme, using the assumption that the graph has local tree-like properties [50]: at each step l, one is allowed to go to a neighbor node only if it has not been visited at step $l-1$. Now, when one has arrived at node of degree k there are $k-1$ possible steps one could take at the $l+1$ step.

This traversal procedure can be written using a transition matrix for any possible subgraph G_{sub}. The transition matrix from a node of degree k at step l to a node of degree k' at step $l+1$ is defined as

$$\mathbf{M}_{k',k} = (k-1)P(k'|k). \quad (4.14)$$

A subgraph is percolated when $\lim_{l\to\infty} \mathbf{M}^l\mathbf{P}(k)$ diverges, where the vector $\mathbf{P}(k)$ is defined for a graph $G_{\text{sub}} = (V_{\text{sub}}, E_{\text{sub}})$ as [50] as below:

$$\mathbf{P}(k) = \frac{\left|\{v \in V_{\text{sub}} \wedge k(v) = k\}\right|}{\left|V_{\text{sub}}\right|}. \tag{4.15}$$

The matrix formalism predicts whether at specific conditions of concentration ratios between actin monomers, branchers, crosslinkers, and motors the network will percolate. The formalism above is only a theoretical tool to predict if an actin monomer network will percolate. In MEDYAN, however, all the information of the connectivity between ABPs and filaments is available. Therefore, it is possible to evaluate the finite component size distribution $p(s)$ on the actin monomer graph [50], within the distribution most readily defined as the distribution of cluster sizes in the graph. In other words, $p(s)$ is the probability of having a a finite connected component (cluster) of size s. For a percolating network, the probability to see a finite connected component that includes all the nodes goes to one: $p(s = |V|) \to 1$. Following that, it is possible to define the weighted cluster size of a graph $\mathbf{G}$ order parameter as the ratio between the second and the first moments of $p(s)$, according to the equation shown below:

$$N_w = \frac{\sum_{s=1}^{|V|} s^2 p(s)}{\sum_{s=1}^{|V|} sp(s)}. \tag{4.16}$$

4.3 Results

4.3.1 Arp2/3 branchers augment actomyosin expansion patterns

Overall, the results of the MEDYAN simulations show that Arp2/3 branchers change the morphologies of actomyosin networks by promoting the creation of a daughter filament at 70° from the mother filament, leading to a situation in which the geometrical 10 nm brancher protein has a mesoscopic effect on the global structure of actomyosin. Fig. 4.4 shows two snapshots, taken at $t = 1000$ s, from a simulation of an unbranched network (Fig. 4.4(a)) and from a simulation of a branched network (Fig. 4.4(b)). The filament monomers (G-actin), the filaments (F-actin), the motors,

the linkers, and the branchers (Arp2/3) were initially distributed randomly within a $1\,\mu m^3$ cubic container. Over the time course of the simulation, it became clear that branchers rigidify actomyosin networks and mitigate their filamentous layout, as seen in Fig. 4.4(b), especially when compared with the unbranched network shown in Fig. 4.4(a). Overall, after 1000 seconds of the simulation, the unbranched networks consist of long and parallel filaments displaying liquid crystalline order [59], while the branched networks consist of short and bifurcated filaments and appear more nearly isotropic.

4.3.2 Crosslinkers modulate contraction in unbranched actomyosin networks

To study the contribution of the crosslinker concentration on the contraction of actomyosin in unbranched networks, James Liman focused on four out the nine different concentration ratios of motor to actin monomers ($x_{m:a}$) and eight concentration ratios of α-actinin crosslinkers to actin monomers ($x_{\alpha:a}$) in the simulations, as detailed below:

(a) Low motor $x_{m:a} = 0.01$ and low linker $x_{\alpha:a} = 0.01$ concentrations

(b) Medium motor $x_{m:a} = 0.05$ and medium linker $x_{\alpha:a} = 0.1$ concentrations

(c) High motor $x_{m:a} = 0.5$ and low linker $x_{\alpha:a} = 0.01$ concentrations

(d) High motor $x_{m:a} = 0.5$ and high linker $x_{\alpha:a} = 0.5$ concentrations

For every condition of motor and crosslinker to actin ratios, 16 simulation replicates were performed, differing only in their initial configuration of chemical components, which were uniformly and randomly distributed throughout the simulation box. The average length of a filament in the simulations for unbranched networks is approximately $0.85\,\mu m$. Throughout the simulations, the structure of the networks were recorded at ten second time intervals. Fig. 4.5 shows how the radius of gyration of the actin monomers R_g, normalized by the "initial" R_g^i, changes with time. R_g^i is in fact the radius of gyration measured after the initial configuration ran for 10 seconds. When an actomyosin network contracts due to motor action, the radius ratio R_g/R_g^i decays below 1; in contrast, when R_g/R_g^i is above 1, this indicates that the system has expanded during the self-assembly process.

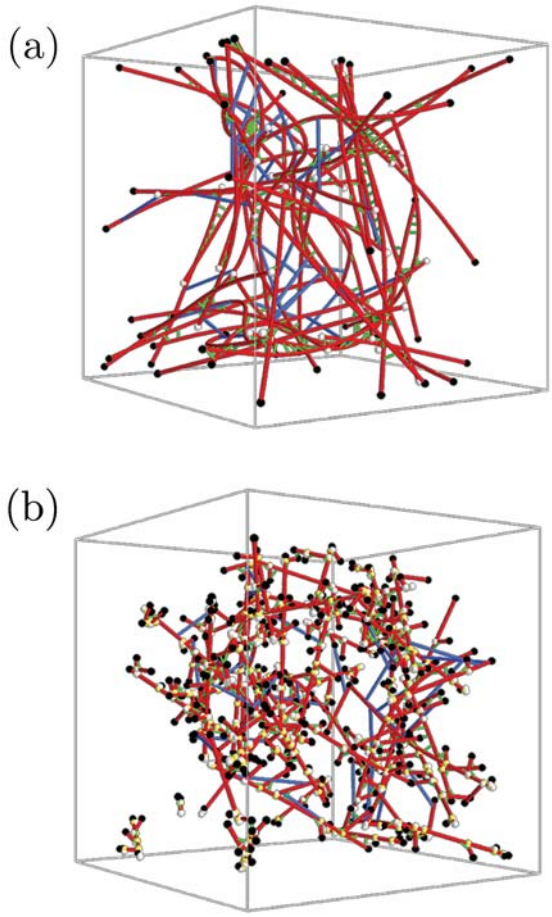

Figure 4.4 Typical snapshots of MEDYAN simulations without Arp2/3 protein complexes (a) or with Arp2/3 protein complexes (b). The red cylinders represent actin filaments. A black bead represents a positive (barbed) end of an actin filament. A white bead represents a negative (pointed) end of an actin filament. A green cylinder represents an α-actinin crosslinker protein. The blue cylinders represent NMIIA motor proteins. A yellow bead represents an Arp2/3 protein brancher complex between two filaments, where the angle between the mother and its daughter filaments through the brancher is 70° [212, 153].

Unbranched actomyosin systems with a low motor concentration $x_{m:a} = 0.01$ and a low linker concentration $x_{\alpha:a} = 0.01$ expand at a relatively slow rate (Fig. 4.5(a)) in comparison with the other three conditions investigated (Fig. 4.5(b-d)). Such expansions are caused by actin polymerization, which occurs without any significant opposing forces exerted by motors that would pull the actin filaments together.

Increasing the ratio $x_{\alpha:a}$ leads to a competition between actin polymerization and motor pulling,

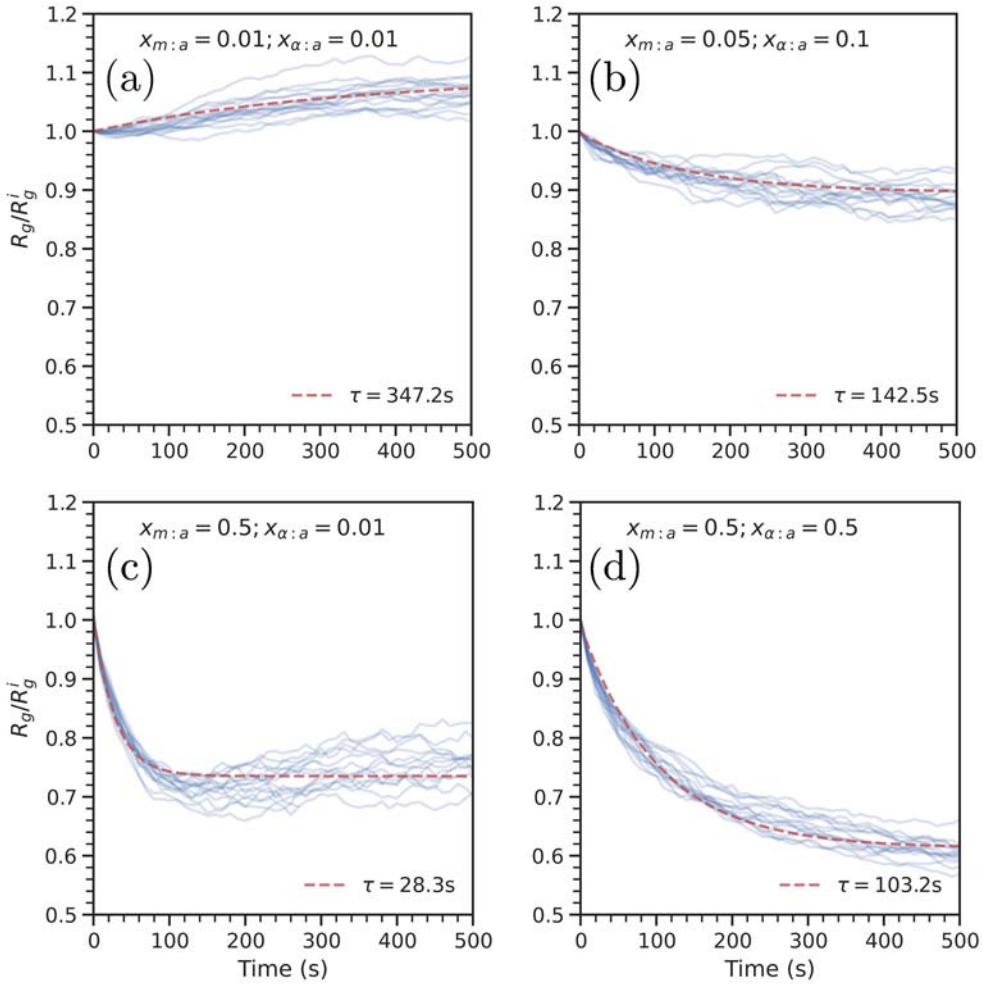

Figure 4.5 Time signals of the ratio R_g/R_g^i in unbranched actomyosin networks at four motor-to-actin ($x_{m:a}$) and linker-to-actin ($x_{\alpha:a}$) concentration ratios. (a) Systems with low motor and low linker concentrations ($x_{m:a} = 0.01$ and $x_{\alpha:a} = 0.01$). (b) Systems with intermediate motor and intermediate linker concentrations ($x_{m:a} = 0.05$ and $x_{\alpha:a} = 0.1$). (c) Systems with high motor and low linker concentrations ($x_{m:a} = 0.5$ and $x_{\alpha:a} = 0.01$). (d) Systems with high motor and high linker concentrations ($x_{m:a} = 0.5$ and $x_{\alpha:a} = 0.5$). The red and dashed lines represent the fit of $R_g/R_g^i(t)$ to the exponential form $R_g(t)/R_g^i \approx (1 - e^{-t/\tau})$.

which in turn accounts for the differing extent of the contraction observed. The average of the normalized radius of gyration ratio $\langle R_g/R_g^i \rangle$ among the replicates for a given condition is then fitted to an exponential form $A + B^{-t/\tau}$, where both A and B are constant values. The radius ratios $\langle R_g/R_g^i \rangle$ achieved at steady state values in the high motor concentration and high linker concentration systems $x_{m:a} = 0.5$, $x_{\alpha:a} = 0.5$, depicted in Fig. 4.5(d), were approximately 16%

lower than the steady state $\langle R_g/R_g^i \rangle$ values in systems with high motor concentration and low linker concentration systems $x_{m:a} = 0.5$, $x_{\alpha:a} = 0.01$ in Fig. 4.5(c). However, in the case of high motor and low linker concentrations, the systems contract 5-fold faster compared with the rate of contraction under conditions of high motor and high linker concentrations. Notably, the contractility in the unbranched systems occurs only in cases where the motor concentration exceeds a threshold of $x_{m:a} = 0.01$ in Fig. 4.5(b-d). These simulation results are in qualitative agreement with the experimental data obtained by Bendix et al. [23] for analogous unbranched systems.

4.3.3 Branched actomyosin networks display convulsive movements

When the Arp2/3 protein in included in the branched network simulations, the R_g/R_g^i time courses in Fig. 4.6 behave differently during the actomyosin assembly process, compared with the behavior of the unbranched networks in Fig. 4.5. The branched networks with low concentrations of motors and low concentrations of linkers $x_{m:A} = 0.01$ and $x_{\alpha:A} = 0.01$ Fig. 4.6(a) do not expand like the unbranched networks expanded in Fig. 4.5(a), due to the fact that the presence of the branches inhibits actin polymerization and depolymerization. Such inhibition has multiple possible causes: first, branchers prevent the actin depolymerization reaction from occurring at the branch junction between the mother and daughter filaments. Second, because the brancher junctions are stable and their disassociation is unlikely, this inhibits the overall depolymerization reactions at the minus end of the mother filaments. This in turn increases the favorability of the daughter filament growth and attenuates significantly the turnover of actin monomers (G-actin). As a result, the network grows in such a way that at low motor activity the network is approximately static Fig. 4.6(a). Of note, the concentration of filamentous actin (F-actin) in the branched simulations is slightly larger than the concentration of F-actin in the unbranched simulations, which is due to the capping of the minus ends of both the mother and the daughter filaments. Nonetheless, the average length of a filament is in fact reduced when Arp2/3 is added, due to the increased number of total filaments. The average length of a filament in branched networks is $\approx 0.15\,\mu\text{m}$, compared with an average length of $0.85\,\mu\text{m}$ in the unbranched networks.

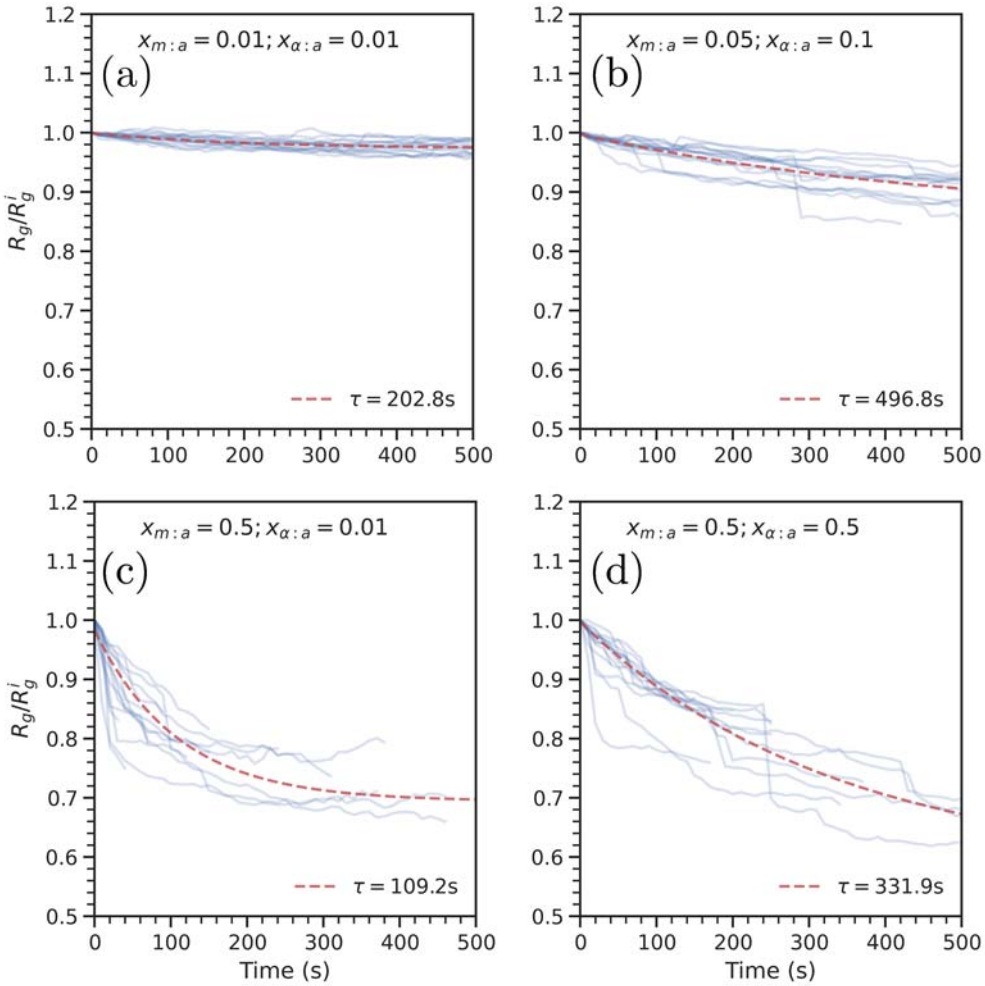

Figure 4.6 Time signals of the ratio R_g/R_g^i in branched actomyosin networks at four motor-to-actin ($x_{m:a}$) and linker-to-actin ($x_{\alpha:a}$) concentration ratios. (a) Systems with low motor and low linker concentrations ($x_{m:a} = 0.01$ and $x_{\alpha:a} = 0.01$). (b) Systems with intermediate motor and intermediate linker concentrations ($x_{m:a} = 0.05$ and $x_{\alpha:a} = 0.1$). (c) Systems with high motor and low linker concentrations ($x_{m:a} = 0.5$ and $x_{\alpha:a} = 0.01$). (d) Systems with high motor and high linker concentrations ($x_{m:a} = 0.5$ and $x_{\alpha:a} = 0.5$). The red and dashed lines represent the fit of $R_g/R_g^i(t)$ to the exponential form $R_g/R_g^i(t) \approx (1 - e^{-t/\tau})$.

In addition to the aforementioned differences between branched and unbranched networks, an additional difference is the fact that the contraction of branched networks occurs in a more irregular fashion than does the contraction of unbranched assemblies. In particular, in branched assemblies, contraction occurs intermittently as the network develops, and then abruptly the radius of gyration ratio signal R_g/R_g^i drops. These convulsive events are inherently stochastic and hence do not occur

at the same time in different replicates of the simulation. Moreover, these abrupt drops in R_g/R_g^i occur very quickly. Individually, these events have a duration of less than 10 seconds, and the sharpest drops reduce the signal by as much as 20%. These drops, referred to as avalanches, are observed more frequently at higher concentrations of motors and linkers (as seen in Fig. 4.6(d)) than at lower concentrations of these species (as seen in Fig. 4.6(a)). For the range of parameters studied in the simulations, James Liman, Carlos Bueno, and I have never observed such discrete and large-scale avalanches in the unbranched actomyosin networks that were created in the absence of Arp2/3. Fig. 4.7 shows how the accumulated tension circled in the top left area of the network dissipates and a large movement in the entire network occurs in less than 10 seconds.

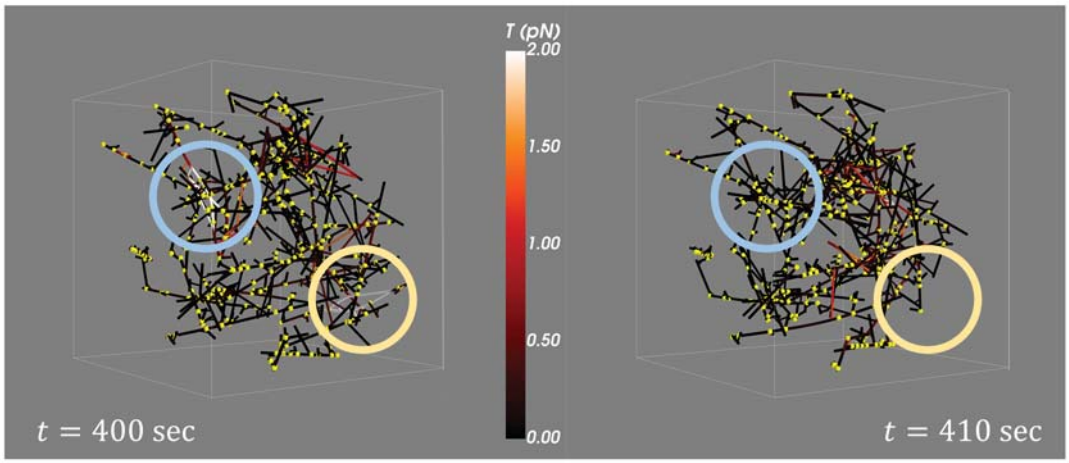

Figure 4.7 Two consecutive snapshots before and after an avalanche occurs, indicating the tension relaxation and long-range movement observed in a branched network. Actin filaments, motors, and linkers are shown and the tension they experience is indicated by a color bar. These show the morphology of the network before and after an abrupt drop in $R_g(t)/R_g^i$. The circled light blue region in the top-left corner highlights the concentrated epicenter of the avalanches where high-tension develops before the circled light-yellow region in the bottom-right corner of the figure occurs.

4.3.4 Two modes of avalanches in branched networks

Branched network are particularly susceptible to avalanches, which can be characterized as an emerged spontaneous phenomena. For example, when plotting the histogram of the frequencies

of the mean filament monomer displacement $\delta x_F(t)$ in Fig. 4.8 between all the simulations of branched and unbranched, it is evident that although polymerization and depolymerization events are inhibited in branched networks, nonetheless, abrupt movements of major parts of the network still occur.

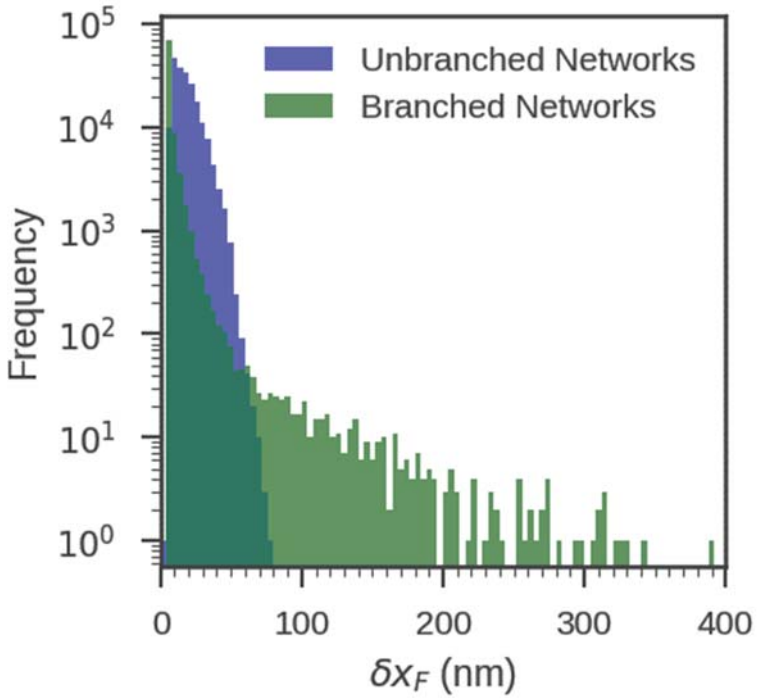

Figure 4.8 Graphical representation of the distributions of the mean displacement of filament monomers δx_F defined in Eq. (4.12) for both unbranched networks and branched networks.

To obtain a more comprehensive understanding of the mechanism underlying avalanches, James Liman examined changes in the radius of gyration ($\Delta R_g/R_g^i$) and the changes in the shape order parameter ($\Delta(S)$) that occur in one branched trajectory with two avalanches.

Results indicate the existence of two types of avalanches, both of which display differences in their shape and in the size of the reconfigured regions. In one group, which corresponds to shear events, the temporal changes are manifested by large changes of the shape parameter defined in Eq. (4.11) between successive pairs of snapshots, changes that occur through filament sliding (e.g.,

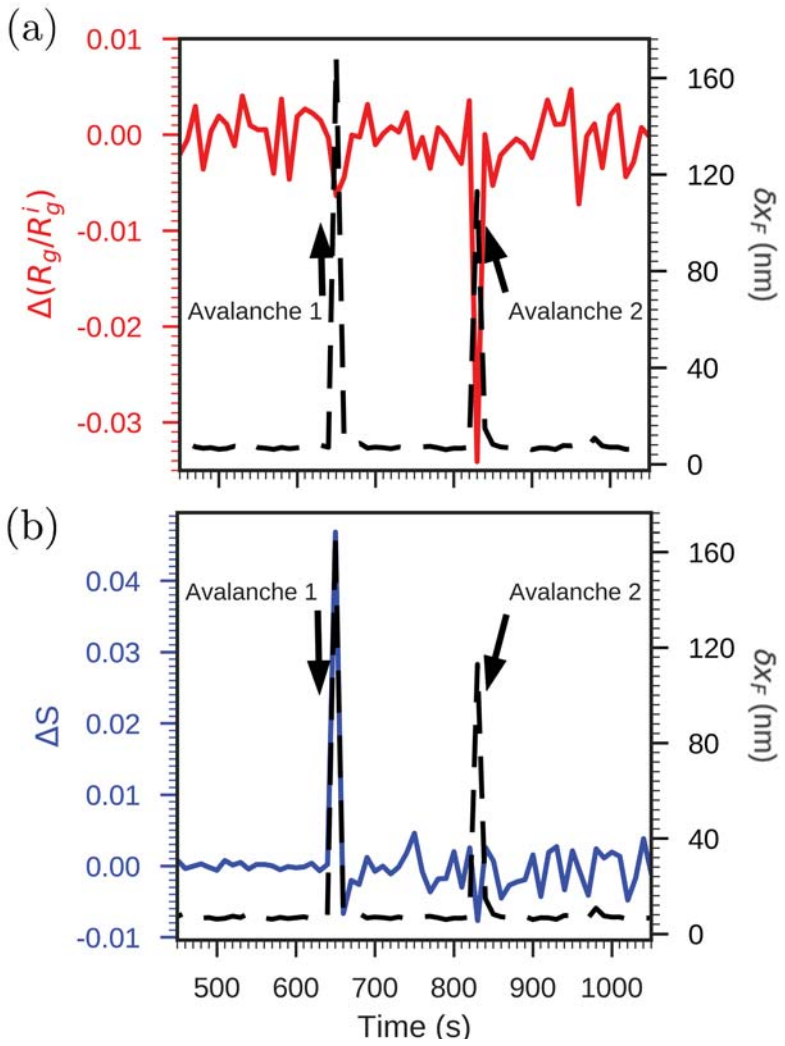

Figure 4.9 Examples of two avalanches in a branched network containing 16 motors and 3,000 linkers. In (a), the time course of the changes in R_g/R_g^i between successive pairs of snapshots that are separated by 10 s is plotted as: $\Delta(Rg/R_g^i)$ (red line). In (b), the time course of the changes in the shape parameter between successive pairs of snapshots that are separated by 10 s is plotted as: . In both (a) and (b), the time course of the mean filament displacement δx_F is plotted as a dashed black line.

labeled as "Avalanche 1" in Fig. 4.9). A second group of avalanches involve collapses of local regions and are characterized by significant changes in the normalized radius of gyration ratio between successive pairs of snapshots (e.g., labeled as "Avalanche 2" in Fig. 4.9). Presumably, the types of avalanches described herein are examples of two extreme cases that represent a broad distribution of avalanches.

4.3.5 Avalanches arise in branched networks when their corresponding graphs of actin monomers percolate

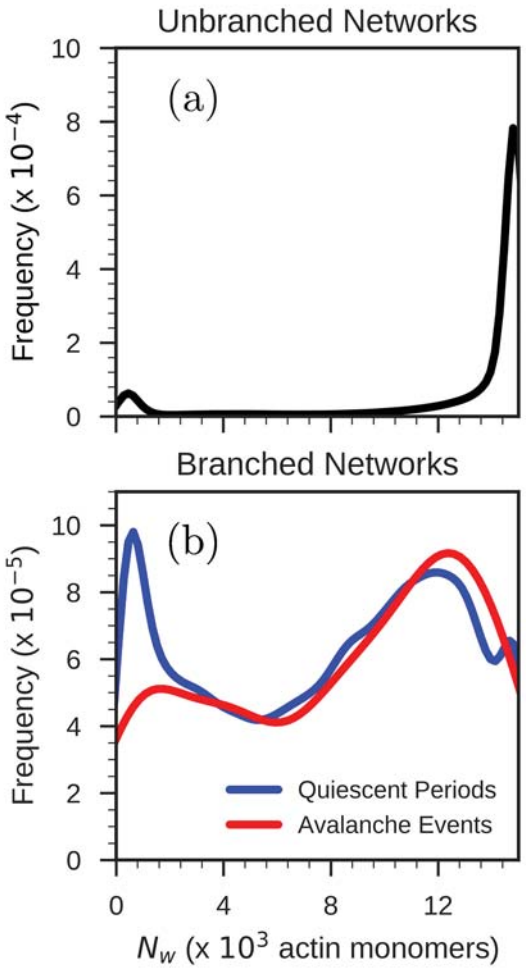

Figure 4.10 Probability density plots of the weighted mean cluster size N_w of the unbranched networks (a) and branched networks (b). The branched networks were further analyzed in the event of no avalanche occurrence (blue curve) and avalanches occurrence (red curve). A total of 1,152 simulations were carried out and analyzed both for the unbranched networks and for the branched networks.

Interestingly, Arp2/3 branchers promote avalanches, which occur in cases where the graph of the actin monomers percolates. Namely, in order for actomyosin network to kick, it must contain Arp2/3 and also must be connected by actin-binding proteins. Fig. 4.10(a) shows that percolation

is not a sufficient condition to have avalanches, as no avalanches were observed in the unbranched simulations. However, in the case of a branched network (Fig. 4.10(b)), when the network percolates (has a finite connected component or a cluster with most actin monomers in the network) avalanches occur, in agreement with the probability density function (PDF) of N_w defined in Eq.(4.16). What is more, even in cases when unbranched networks percolate, as determined by checking the graph of actin monomers or just by the theoretical mean field approximation of percolation as estimated by $\lim_{l\to\infty} \mathbf{M}^l\mathbf{P}(k)$ in Eq. (4.14), no avalanches or abrupt drops have been observed in any of the simulations of unbranched networks. Namely, even when both the mean field approximation and the actual actin monomer graphs predict and indicate that an unbranched system is percolating, no avalanches occur. In contrast, under similar conditions, a system with branches is likely to experience such avalanches. In one example of a branched system (depicted in Fig. 4.7), which has $G = 12,000$ actin monomers, $L = 3,000$ crosslinkers, $M = 16$ motors, $B = 300$ branchers, compared to an unbranched system with $G = 12,000$, $L = 3,000$, $M = 16$, $B = 0$, I report that even though both systems are likely to percolate, as determined by the mean field approximation and the observed actin monomer connectivity in the simulations (Fig. 4.10), only the system with branchers kicks and undergoes avalanches.

4.4 Discussion

4.4.1 Contractility in actomyosin depends on the ABPs-to-actin concentrations

Consistent with earlier studies, this study found that the contraction of unbranched actomyosin networks strongly depends on the concentrations of motor protein and linkers present in the system. In particular, systems with an intermediate value of linker concentration contract faster than systems with either a high or a low concentration of linkers [23]. This study suggests that the addition of Arp2/3 complexes to the system stabilizes the scaffolding of the actomyosin networks, which in turn can explain how actomyosin can drive cell motility and growth [117, 211] when the amount of actin monomers is indefinite e.g., due to protein synthesis.

4.4.2 Avalanches are mechanochemical emergent phenomena in actomyosin

Our results suggest that avalanches occur in actomyosin networks in the presence of Arp2/3 and when the network of actin monomer percolates. Arp2/3 brancher proteins overall act to reduce the polymerization rate of actin, and to change the actin mesh grid on which the motor protein can walk. This fascinating interdependent relationship between the mechanical, topological, and chemical aspects of complex actomyosin networks, which is remarkably modeled and captured by MEDYAN [180], may explain the abundant phenomena of avalanches in broader physical contexts such as nano-scale networks [146], earthquakes [188], sand piles [103, 80] and neural avalanches [21, 14], which occurs on a centimeter-length scale. From this work, I conclude that in order to observe avalanches in actomyosin networks, the network needs to percolate and ABP such as the Arp2/3 branchers must be present. From this, I suggest therefore the conditions under which a general active physical system could experience avalanche-like behavior: (1) the system needs to have a broken spatial symmetry that is being exploited by an energy-consuming process; (2) the graph representation of the elements of the system should be above or near the percolation threshold; and (c) there should be some macroscopic broken detailed balance or inhibition of one or more backward chemical reactions due to mechanical constraints or negative chemical feedback from other reactions. One of the striking findings is that having a percolated network and substantial motor activity are not by themselves sufficient to bring the network into the critical regime of avalanches. This unique feature of Arp2/3 to enable avalanches and cytoquakes [7] is very intriguing in the context of biological systems, which usually prefer to function near a critical regime to allow maximum flexibility while maintaining robustness [14].

5 Quantitative methods to Characterize Structures and mutations in genomes

This chapter is based on a poster presented at the ENCODE (Encyclopedia of DNA Elements) conference in 2018, and at 2019 Rice Data Science Conference.

5.1 Introduction

Hi-C experiments use high throughput next generation sequencing (NGS) of DNA to create two-dimensional (2D) matrix representations of the spatial proximity between pairs of genome loci [181]. These 2D matrices, called contact maps, share common global structural features across genomes of different species and cell types [181, 185]. The three major structural features observed to date in Hi-C of folded genomes are loops, domains, and subcompartments, each of which is defined herein: (a) loops are defined as genome loci that are far apart (up to millions of base pairs (bp) away on the linear (1D) DNA strand), which come into close proximity in the 3D folded structure of the genome inside the cell nucleus; (b) domains are defined as contiguous stretches of DNA that form self-associating contact domains; and (c) subcompartments are defined as areas of the contact domains, which segregate into two types of gene regulation areas, or subcompartments, characterized as "active" or "inactive". Reported herein are results obtained using both supervised and unsupervised deep learning approaches to detect loops in Hi-C contact maps, based on the public data that was published in 2014 by Rao et al. [181]. In their original work, Rao et al. [181] reported only 10,000 loops, which is between one to two orders of magnitude less than the anticipated number of loops one would expect to see in a functioning cell, which have hundreds of thousands of enhancer and promoter pairs that constantly interact. According to a large survey of cell lines and tissues, there exist hundreds of thousands of enhancers along the human genome, which vastly exceed the number of human $\sim 20,000$ protein-encoding genes.

In genetics, a promoter is defined as a sequence of DNA to which proteins bind and initiate

transcription of a single RNA downstream from the DNA. Promoters are located near the transcription start sites of genes and their length can be between 100 and 1,000 bp long [138]. Enhancers are defined as regions on the DNA strand (50-1,500 bp) that can increase the transcription of a particular gene. Enhancers may be located up to 1,000,000 bp away from the gene. Because of the prevalent hypothesis that gene regulation in cells is controlled by the looping of enhancers and promoters [65], there is a discrepancy between the number of reported loops in Hi-C maps and the expected enhancer-promoter (E-P) loops. Thus there is no solid evidence on whether Hi-C is a viable technology to find E-P loops nor if the hypothesis of E-P looping mechanism is valid [195]. In this work, I used a convolutional neural network (CNN) to detect loops in the deepest Hi-C maps to date from the work by Rao et al. [181], and compared the findings to their reported 10,000 loops. Although the results are inconclusive and it is still unclear whether E-P exists or if Hi-C is a valid technology to find them, the deep-learning results suggest that some hidden structures that were labeled with high confidence as loops by the CNN exist in regions within 1 Mbp from the Hi-C diagonal.

5.2 Methods and results

I used public Hi-C data and the the official list of loops listed by HiCCUPS, a Hi-C peak finder software [181], on the human lymphoblastoid cell line GM12878. This cell line is an immortalized cell lines that was manipulated to proliferate indefinitely. GM12878 is the extraction from the blood of a female donor with northern and western European ancestry. The loop list contains almost $10,000$ loops distributed throughout the genome, of which chr1 (the largest chromosome in the human genome) has 931 loops listed and chr14 has 315 listed loops. I used the data from chr1 as the training set and the data from chr14 as the validation set to train a binary classifier that predicts the existence of loops in a small 15×15 submatrix/subgraph of the intrachromosomal Hi-C map at $5\,\text{kbp}$ resolution. I trained a convolutional neural network (CNN) [124] to take a $N \times N$ Hi-C matrix and output a probability $p \in [0, 1]$, depending on whether loops are detected in the input 15×15 submatrix. A general illustration of this "LoopDetectNet" CNN is depicted in Fig. 5.1(a),

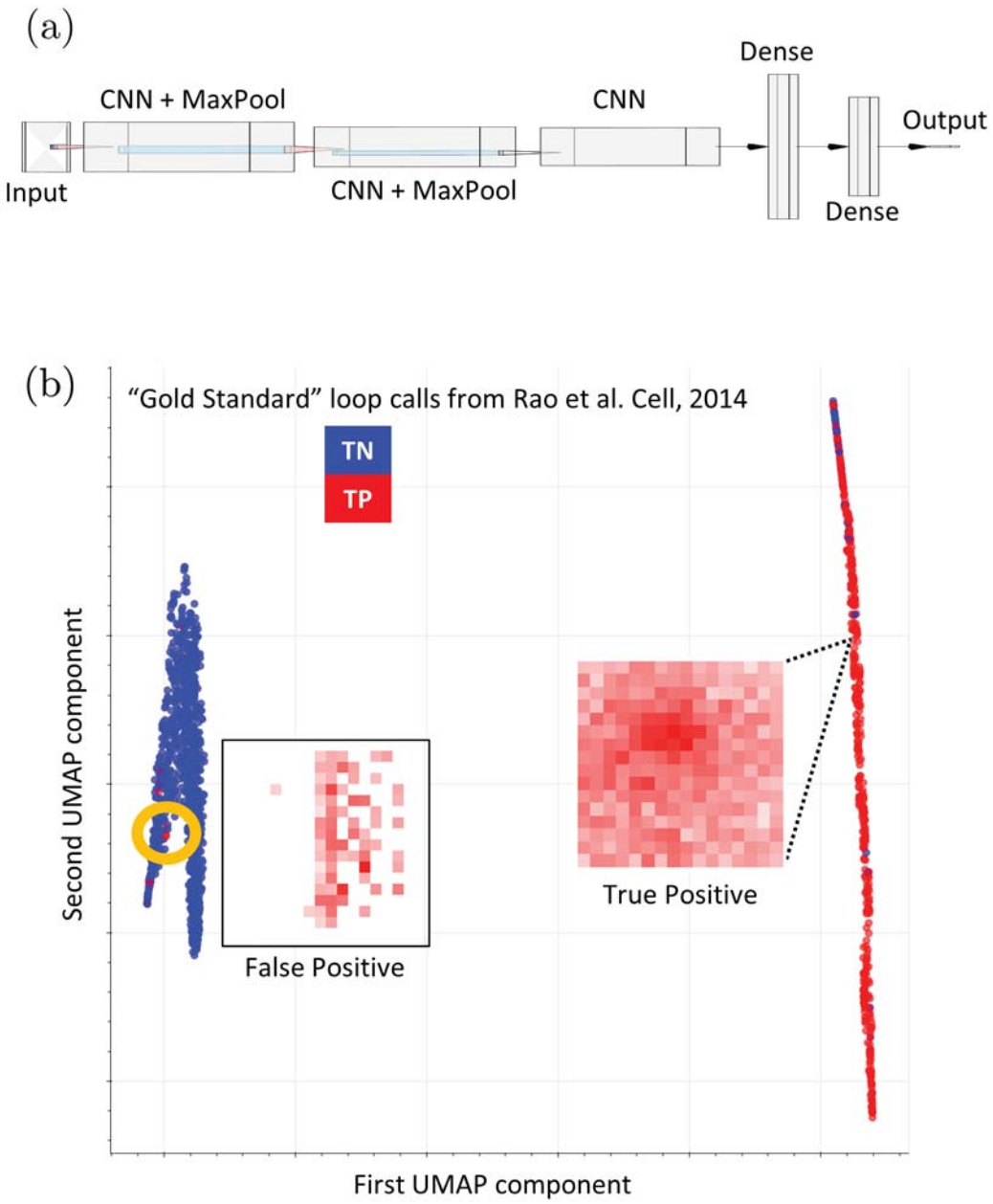

Figure 5.1 LoopDetectNet and UMAP of the gold standard data. (a) LoopDetectNet's illustration of the neural network architecture used as a binary classifier of loops. (b) The UMAP [148] projection of the training data that shows that Rao et al. [181] is mostly correct, with a few false annotations observed.

where I used the sigmoid $\sigma(x) = 1/(1 + e^{-x})$ function as an activation function in the last layer to retrieve the probability as an output.

CNN is a widely used deep learning technique inspired by the discovery in 1968 of how the visual cortex in living creatures interprets light hitting the retina [101]. After this discovery, it took 30 years until LeCun et al. [124] published in 1998 the seminal paper that established the foundation for the contemporary framework of CNN. In 2012, Krizhevsky et al. [122] used graphics accelerator units (GPUs) to speed up and outperform all prior competitors on image classifications, which started the new era of deep learning [92]. The main idea of a binary classifier CNN is to use the backpropagation algorithm to find the optimal convolution kernels and weights in an artificial neural network such that the cross entropy loss function is minimized on the training data. The convolutional layer in CNN is a set of K kernels. For this study, I used $K = 512$ kernels of size 3×3, where the operation of each kernel is given by the equation

$$\mathbf{O}_{i,j} = \sum_{m=i-1,l=j-1}^{i+1,j+1} \mathbf{W}_{k,l}\mathbf{I}_{m,l}, \tag{5.1}$$

where $\mathbf{W}$ is a 3×3 kernel matrix and $\mathbf{O}$ and $\mathbf{I}$ are defined as the output and input matrices before a 2D convolution is applied, both of which have the same dimension $N \times N$.

Another key layer in CNN models is the max-pooling layer, which can reduce the dimensions of the input matrix or preserve it, depending on the striding size used. The max-pooling layer is a non-linear operation that selects the maximal pixel within a contagious submatrix of size $m \times m$, where $m < N$.

Since 2014, the data from Rao et al. [181] has become a gold standard in the field of structural genomics. In Fig. 5.1(a), I used UMAP [148] (Uniform Manifold Approximation and Projection for Dimension Reduction) to verify that submatrices with loops, as annotated by the gold standard, are topologically different compared to submatrices with loops. Indeed, most positive calls are clustered together as shown as red dots on the right side of the plot in Fig. 5.1(b), while the submatrices without loops are clustered on the left side of the scatter plot. Noticeably, the UMAP of the training data suggests that the loop calls by Rao et al. [181] have a few false positive calls (red calls that UMAP classify near the negative cluster) as well as regions which Rao et al. [181] labels as regions

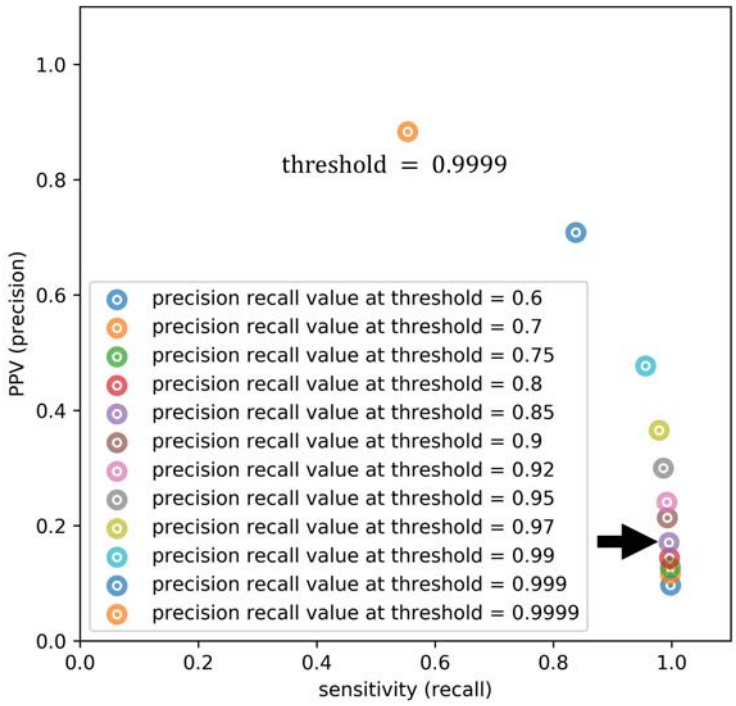

Figure 5.2 Precision-recall plot at various thresholds using the annotations previously reported by Rao et al. [181].

without loops, but UMAP projects them within the loop cluster on the right. Overall, UMAP suggests that the gold standard is mostly reliable and can be used as training data, however, it also implies that there are a few false positive results in the gold standard (See Fig. 5.1(b)) and may also overlook regions that contain loops and are not in the gold standard dataset.

After training a model on chr1 using the gold standard, I predicted loops in the testing dataset from chr14 and compared the results with the gold standard. Of note, I report that on the first iteration there are many "false positives" (Fig. 5.2), which occur under conditions of high sensitivity and low precision. Following that, I took the false positive submatrices and added them to the negative training set, to do another round of CNN retraining. Overall, therefore, I used, during the retraining, 25% of the time submatrices from the "false positive" regions, 25% of the time a random tile from the intrachromosomal region, and 50% of the time I took as a positive example a

submatrix which contains a loop (i.e., “true positive”).

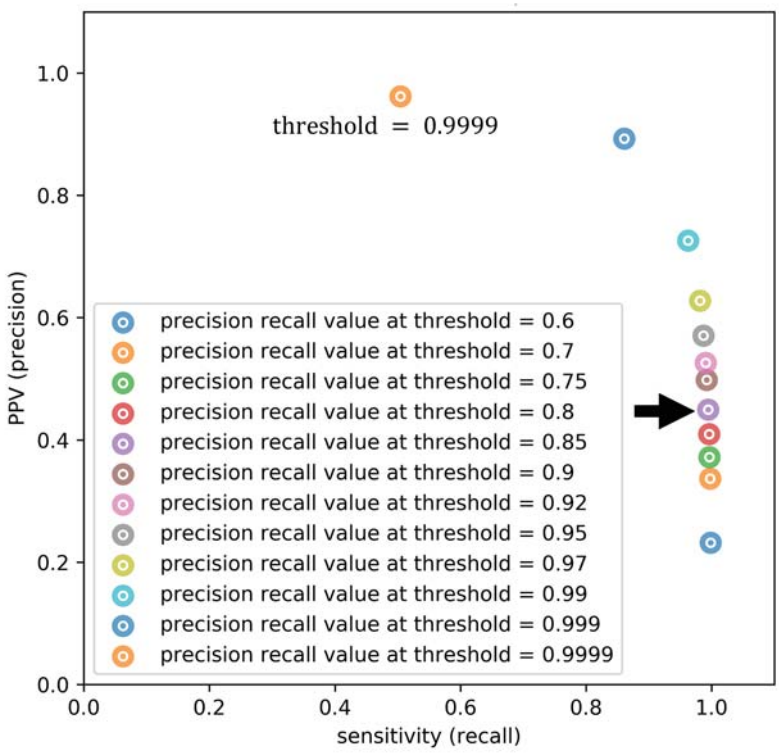

Figure 5.3 Precision-recall plot reported at various thresholds, after retraining the CNN model by adding the false positive data to the negative data set.

After the augmentation of the training data with the false positive regions, precision was improved (Fig. 5.3), but there were still many false positives. Further detailed analysis indicated that the deep learning network detects new regions that possibly have loops which are hard to see by human eyes and by using the current state-of-the-art Hi-C experimental data. The results reported in Fig. 5.4 indicate the degree of true positive and false positive loop identifications that appear for chromosome 14 (chr14).

5.3 Discussion

The discrepancy between the observed number of loops and the expected number of loops is one of the biggest hierarchy problems in structural biology. According our understanding of gene regulation by

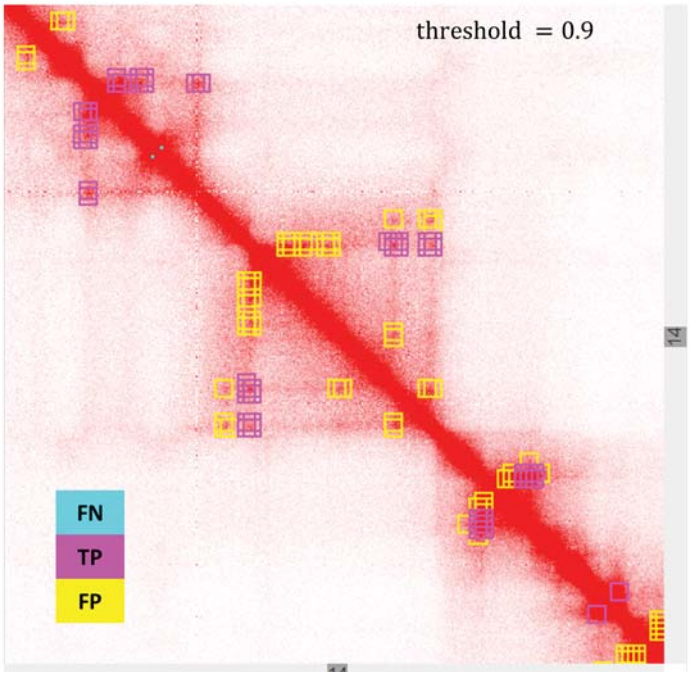

Figure 5.4 An example of all the the submatrices that LoopDetectNet predicts as loops with a $> 90\%$ confidence. The submatrices are then colored with three colors: FN (false negative), TP (true positive), and FP (false positive), based on the gold standard.

enhancer-promoter (E-P) interactions, one would expect to see hundreds of thousands of loops, but even in the deepest map generated to date, less than 10,000 of such loops have been reported [181]. Using deep learning and UMAP, I suggest that the gold standard of calls on GM12878 is indeed impressive, but significant discrepancies still remain. In addition, deep learning exposes new loops, mostly around areas that were previously classified as domain boundaries or stripes [181]. I believe that deeper Hi-C maps, especially those maps that are at extremely high (i.e., nucleosome-level) resolution will lead to improvements in the gold standard set of loops used in this field. Recently, a research group has reported that by using a computational enhancement algorithm they detect loop-like structures in subdomain and domain boundaries regions [142], which are the same regions where LoopDetectNet's "false positive" loops reside.

6 Hydrodynamic interactions alter the folding dynamics of proteins based on the temperature

This chapter is based on a paper published in PRE [233] in 2018.

6.1 Introduction

The three-dimensional (3D) conformation of a protein is crucial in order for the protein to function properly. In nature, proteins fold and function within a complex cellular environment, where they interact both with the cytosol (the aqueous medium) and with other proteins or molecules in the cell. These interactions are classified into two broad classes: (a) hydrodynamic interactions (HI) of the protein with the solvent; and (b) macromolecular crowding interactions.

Overall, the water that is found inside a cell has a dual impact on a folding protein. On one hand, the aqueous solvent causes hydrophobic collapse (Section 1.4) of the protein's side chains, which in turn initiates the folding process. On the other hand, the motions of protein particles are not independent, due to solute-solvent coupling between the protein and the water. When the solute (protein) particles move, they create a flow in the solvent, which in turn, affects the motion of the other solute particles. This mutual interaction between the solute and solvent particles yields long-range effects between solute particles that propagate throughout the solvent, known as hydrodynamic interactions (HI), and have been shown to play a critical role in protein dynamics [89, 149]. In the simple case of homopolymers in *in vitro* settings, HI accelerate the process of polymer collapse [115]. However, in the case of proteins, which are heteropolymers formed from up to 20 different types of amino acid monomers, one has to take into account other long-range interactions that are possible within the polymer residues, including both Van der Waals and electrostatic interactions, as well as interactions that occur from hydrodynamic interactions. Before the work reported herein, different groups published seemingly contradictory findings using computer-based simulations to address whether HI facilitate [46, 79, 116] or deter [135] protein folding. Surprisingly, the analysis of the temperature dependence of these findings was inadequate, although temperature effects are expected

to be significant due to the known temperature-dependence of the phase transition from folded to unfolded protein states [167, 169]. This aforementioned literature motivated this work to reconcile the differences in the reported results by studying HI effects over a broad range of temperatures, with the goal of learning how HI affects the dynamics and the dependency of the temperature on the folding of the chymotrypsin inhibitor 2 (CI2) protein, which has 64 amino acid residues [233]. Overall, this work suggests that despite the inconsistent results obtained from the work of previous research groups [46, 79, 116, 135], each of the results may be correct at the particular temperature investigated. The protein targeted for investigation, CI2, is a two-state (unfolded-folded) protein, whose folding mechanisms were extensively studied in various contexts [58, 40, 222]. The simulations in this work that were conducted by Fabio Zegarra, had two types: (a) simulations of the protein polymer with Brownian dynamics; and (b) simulations of the protein polymer with incorporated HI by implementing the Ermak-McCammon scheme [69, 194], which emulates the effect of HI on solute particles by introducing a configuration-dependent diffusion tensor.

This work revealed that the effect of HI on folding rates can both facilitate protein folding at a temperature lower than the folding temperature and impede protein folding at a temperature higher than the folding temperature. Therefore, the results reported herein were able to arbitrate the dispute in previously reported results on the temperature-dependent effects of HI on protein folding.

6.2 Methods

6.2.1 Coarse-grained protein model

The coarse-grained [213] model used to investigate the dynamics and structure of the CI2 protein (Fig. 6.1(b)) represented the protein as a chain of beads (Fig. 6.1(a)) placed at the residue's α-carbon position of each amino acid. The native (folded) structure of the protein in the coarse-grained representation of the chain of beads is defined as Γ_0. Then, the Gō-like Hamiltonian, which is an energy-based loss function to compare a configuration Γ of the chain with Γ_0 (i.e., the native structure of the protein). Because the contact map that represents Γ_0 is a binary contact map

(adjacency matrix), and each native interaction weighs the same in the energy calculation, the Hamiltonian reads as follows [49]:

$$
\begin{aligned}
\mathcal{H}\left(\Gamma,\Gamma^{0}\right) = & \\
& \sum_{i<j} \varepsilon\, k_r \left(r_{ij}-r_{ij}^{0}\right)^2 \delta_{j,i+1} + \sum_{l\,\in\,\text{angles}} \varepsilon\, k_\theta \left(\theta_l-\theta_l^{0}\right)^2 \\
& + \sum_{i\,\in\,\text{dihedral}} \varepsilon\, k_\phi \left\{ \left[1-\cos\left(\phi_i-\phi_i^{0}\right)\right] + \frac{1}{2}\left[1-\cos\left(3\left(\phi_m-\phi_m^{0}\right)\right)\right] \right\} \\
& + \underbrace{\sum_{\substack{|i-j|>3 \\ i,j\in\text{native}}} \varepsilon \left[5\left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^{0}}{r_{ij}}\right)^{10} \right]}_{\text{nonbonded term}} + \underbrace{\sum_{\substack{|i-j|>3 \\ i,j\notin\text{native}}} \varepsilon \left(\frac{\sigma}{r_{ij}}\right)^{12}}_{\text{nonbonded term}},
\end{aligned}
\tag{6.1}
$$

where a configuration of a chain with N beads is defined as $\Gamma := \left(\{\theta_i\}_{i=1}^{N}, \{\phi_i\}_{i=1}^{N}, \{r_{ij}\}_{i,j=1}^{N}\right)$, and $r_{ij} \equiv |r_i - r_j|$. The r_{ij} term is the pairwise Euclidean distance between the ith and the jth beads; θ_l is the lth angle defined between three consecutive beads, and ϕ_m is the dihedral angle between four consecutive beads and δ is the Kronecker delta function.

In the simulations, the constants were set to: $\varepsilon = 0.6\,\text{kcal}\,\text{mol}^{-1}$, $k_r = 100\,\text{nm}^{-2}$, $k_\theta = 20\,\text{rad}^{-2}$, and $k_\phi = 1$ [49].

The native state configuration value of the CI2 protein was obtained from its crystal structure [96]: $\Gamma^0 = \left(\{\theta_i^0\}_{i=1}^{N}, \{\phi_i^0\}_{i=1}^{N}, \{r_{ij}^0\}_{i,j=1}^{N}\right)$. In order to choose the native contacts, the contacts of structural units (CSU) software was used [205]. The nonbonded terms comprise the Lennard–Jones (LJ) potential between native bead pairs, and a pairwise excluded volume interaction between non-native beads. For the native pairs, the 10–12 LJ potential was used to allow shorter-range interactions and to enrich the population of the unfolded ensemble, compared with the traditional 6–12 LJ potential. For the non-native contact repulsive potential in Eq. (6.1), the LJ constant was set as $\sigma = 0.4\,\text{nm}$ [49].

The CI2 protein has one α-helix packed against three β-strands, and a 3_{10} helix, with the overall structure of CI2 shown schematically in Fig. 6.1(c). The two key cores in CI2 are the hydrophobic

core and the minicore, and their key residues [112, 106] are emphasized in Fig. 6.1(b) .

6.2.2 Brownian dynamics with or without HI

Fabio Zegarra conducted two sets of simulations: (a) systems with both Brownian dynamics and hydrodynamic interactions (BDHI); and (b) systems with only Brownian dynamics in the absence of HI (BD). To simulate BDHI, the Ermak-McCammon method [69] and its software implementation that was developed by Ando et al. [11] were used, as a far-field HI approximation without periodic boundary conditions. The equation of motion of the system is defined below:

$$\mathbf{x}_i(t+dt) = \mathbf{x}_i(t) + \frac{\sum_j \mathbf{D}_{ij}\mathbf{F}_j}{k_{\mathrm{B}}T}dt + \mathbf{G}_i(dt), \tag{6.2}$$

where $\mathbf{x}_i(t)$ is the position vector of the ith bead in coarse-grained chain model of the protein, and dt is the integration time step. $\mathbf{D}_{ij}$ is the 3×3 submatrix in the 4-rank diffusion tensor $\mathbf{D}$, which is an $N \times N$ super matrix whose ith row and jth column is $\mathbf{D}_{ij}$. In the case of BDHI, the entire hydrodynamic coupling between all possible pairs of beads was computed in the diffusion tensor $\mathbf{D}$, whereas in the case of BD, all the off-diagonal 3×3 submatrices in $\mathbf{D}$ vanish. $\mathbf{F}_j$ is the total force acting on the jth particle. Finally, the $\mathbf{G}_i$ term in Eq. (6.2) is the random displacement that mimics the stochastic effects on a bead due to the implicit solvent considered in the system.

In order to construct the diffusion tensor explicitly, the diffusion tensors for both the BD and BDHI scenarios were derived, according to the equations shown in Eq. (6.3) and Eq. (6.4), respectively. Let k_B be the Boltzmann constant, η be the viscosity of the aqueous solvent at temperature T, and $a = 0.53\,\mathrm{nm}$ [79] be the hydrodynamic radius of the beads [11]. For both BD and BDHI, the diagonal terms $\mathbf{D}_{ii}$ were calculated from the Stokes-Einstein relation between temperature and mobility. In the presence of HI, the off diagonal elements of $\mathbf{D}_{ij}$ were obtained from the Rotne-Prager-Yamakawa approximation, which is one of the most common ways to account for HI [183, 230]. As such, the diffusion tensors in the BD and BDHI cases can be defined according

to the equation listed below:

$$\mathbf{D}_{ij} = \begin{cases} \frac{k_B T}{6\pi\eta a}\mathbf{I}_3 & i = j \\ \\ \mathbf{0}_3 & i \neq j \end{cases}, \tag{6.3}$$

and as

$$\mathbf{D}_{ij} = \begin{cases} \frac{k_B T}{6\pi\eta a}\mathbf{I}_3 & i = j \\ \frac{k_B T}{8\pi\eta r_{ij}}\left[\left(1 + \frac{2}{3}\frac{a^2}{r_{ij}^2}\right)\mathbf{I}_3 + \left(1 - \frac{2a^2}{r_{ij}^2}\right)\left(\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{r_{ij}^2}\right)\right] & i \neq j \wedge r_{ij} \geq 2a \\ \frac{k_B T}{6\pi\eta a}\left[\left(1 - \frac{9}{32}\frac{r_{ij}}{a}\right)\mathbf{I}_3 + \frac{3}{32}\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{a r_{ij}}\right] & i \neq j \wedge r_{ij} < 2a \end{cases}, \tag{6.4}$$

where $\mathbf{I}_3$ represents the 3×3 identity matrix and $\mathbf{0}_3$ represents the 3×3 null matrix. The $\otimes$ operator represents the tensor product between two vectors. The random displacement $\mathbf{G}_i(dt)$ and the diffusion tensors are linked by the condition in Eq. (6.5), which guarantees that the fluctuation-dissipation theorem holds:

$$\langle \mathbf{G}_i(dt) \otimes \mathbf{G}_j(dt) \rangle = 6\mathbf{D}_{ij}dt \;\; \wedge \;\; \langle \mathbf{G}_i(dt) \rangle = 0, \tag{6.5}$$

where $\langle \cdot \rangle$ represents the ensemble average.

6.2.3 Equilibrium thermodynamics simulations

The free energy profiles of the CI2 protein were obtained by Fabio Zegarra for both the BDHI and BD simulation conditions. In order to enhance the sampling of conformations, the replica exchange method (REM) [210] was used, and was equipped with the Metropolis–Hastings acceptance criteria for exchange between two consecutive replicas i and j at the temperatures T_i and T_j:

$$\min\left(1, \exp\left\{\left[\beta_i - \beta_j\right]\left[\mathcal{H}\left(\Gamma_i\right) - \mathcal{H}\left(\Gamma_j\right)\right]\right\}\right), \tag{6.6}$$

where $\beta := \frac{1}{k_B T}$ and $\mathcal{H}(\Gamma)$ is the Hamiltonian from Eq. (6.1). Ultimately, the free energy profiles were estimated with the weighted histogram analysis method (WHAM) [44].

6.2.4 Non-equilibrium kinetic simulations

In protein folding, the activation barrier that exists between the folded and unfolded states is temperature dependent and can be represented as a non-Arrhenius chemical reaction [206]. For these simulations, a configuration Γ was considered in a folded state if the corresponding sum of the nonbonded terms in Eq. (6.1) was no greater than $0.9\mathcal{H}(\Gamma^0, \Gamma^0)$. Once the non-Arrhenius plot was created, two working temperatures were selected for below the "folding temperature" ($0.95T_f^{CI2}$) and above the "folding temperature" ($1.06T_f^{CI2}$), where T_f^{CI2} denotes the "folding temperature", defined as the temperature at which the unfolded and folded are equally populated in the energy landscape profile (see Fig. D.1).

6.2.5 The reaction coordinate Q measures the fraction of native contacts

$Q \in [0, 1]$ measures the similarity of the contact map of any protein chain configuration, Γ to the contact map of the "native state", Γ^0. According to the energy landscape theory [33, 168], protein folding can be reduced to a stochastic motion of a few collective degrees of freedoms (reaction coordinates), which are used to effectively describe protein kinetics and conformations. In the reduced space of the reaction coordinate, the probability $P(Q, t)$ of the protein to have a similarity Q to the native state at time t may be described by a Fokker-Planck equation [33, 168, 206]:

$$\frac{\partial P(Q,t)}{\partial t} = \frac{\partial}{\partial Q}\left\{ D(Q,T)\left[\frac{\partial P(Q,t)}{\partial Q} + P(Q,t)\frac{\partial \beta F(Q,T)}{\partial Q}\right]\right\}, \tag{6.7}$$

where Q is the reaction coordinate, $F(Q, T)$ is the free energy at temperature T, and $D(Q, T)$ is the diffusion, which is proportional to the transition rates from state Q' to Q, as shown in the equivalent master equation representation below [33] :

$$\frac{\partial P(Q,t)}{\partial t} = \sum_{Q'}\left[k_{Q'Q}P\left(Q',t\right) - k_{QQ'}P(Q,t)\right], \tag{6.8}$$

where $k_{Q'Q}$ is the transition rate from Q' to Q, and in which many states may contribute collectively to the rate $k_{Q'Q}$.

Using the aforementioned equations, we determined that the effect of HI on the protein chain is purely kinetic, i.e., the Hamiltonian defined in Eq. (6.1) remains the same at both the BD and BDHI conditions, while the diffusion tensors of the beads change (Eq. (6.4-6.3)). Since HI do not change the free energy profile, one would expect to observe changes in the profiles of the mean-squared displacement (MSD) of Q over a lag time t' for both BD and BDHI at $T = 0.95T_f^{CI2}$ and at $T = 1.06T_f^{CI2}$. The fraction of native contacts $Q \in [0, 1]$ is the chosen reaction coordinate, because Q captures the global configuration of the chain. From the energy landscape theory (ELT) of protein folding, Q is likely to have a diffusion coefficient $D(Q)$, which reflects the changes in the folding rates that occur due to the kinetic effect of HI on the beads [168]. The notation $\langle f(Q(t))\rangle_\Omega$ indicates an average over all simulated frames of the function $f(Q(t))$ over all the trajectories, at the same normalized time $t/t_{\max}$.

6.2.6 Data analysis

To analyze the data from trajectories and to study the effect of HI at both temperatures above T_f and below T_f, I helped Fabio Zegarra define two numerical quantities: (i) the pairwise displacement correlation matrix $C_{ij}(t)$ between all the N beads in chain, and (ii) the chance of correlation vector $\Pi(|i-j|, t)$ between beads i and j whose linear separation on the chain is $|i-j|$ at time t.

Displacement correlation matrix

Let $\mathbf{s}_k(t) := \mathbf{x}_k(t) - \mathbf{x}_k(t - dt)$ be the 3D displacement vector of the kth bead throughout the simulations. Then the displacement correlation matrix $\mathbf{C}(t)$ at time t is given by:

$$\mathbf{C}_{ij}(t) = \langle \hat{\mathbf{s}}_i(t) \cdot \hat{\mathbf{s}}_j(t) \rangle_\Omega \,, \tag{6.9}$$

where $\mathbf{s}_i(t) \cdot \hat{\mathbf{s}}_j(t)$ is the dot (inner) product between the two displacement vectors and the $\langle\cdot\rangle_\Omega$ is the ensemble average at time t.

Chance of correlation occurrence

The chance of occurrence $\Pi(|i-j|,t)$ measures is defined as the ratio of number of residues with a sequence separation $i-j>0$ whose magnitude of displacement correlation is above a threshold $\mu(t)$, to the total number residue pairs at that sequence separation:

$$\Pi(|i-j|,t)=\frac{\sum_{i'>j'}\Theta\left(\mathbf{C}_{i'j'}(t)-\mu(t)\right)\delta_{|i'-j'|,|i-j|}}{\sum_{i'>j'}\delta_{|i'-j'|,|i-j|}}, \tag{6.10}$$

where Θ is the Heaviside step function and δ is Kronecker delta function. The chosen threshold $\mu(t)$ is the average positive displacement correlation from the ensemble:

$$\mu(t)=\frac{\sum_{i>j}C_{ij}(t)\Theta\left(\mathbf{C}_{ij}(t)\right)}{\sum_{i>j}\Theta\left(\mathbf{C}_{ij}(t)\right)}. \tag{6.11}$$

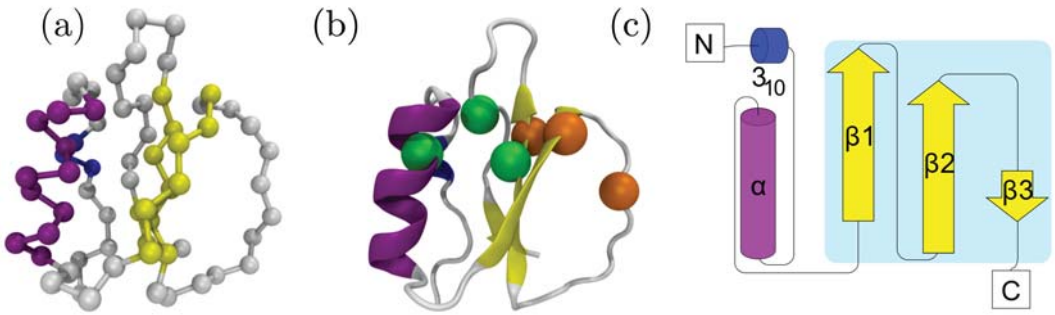

Figure 6.1 Representations of protein models of chymotrypsin inhibitor 2 (CI2). (a) Coarse-grained representation of the secondary structure of CI2 using VMD [102] in (b), representation of the same protein in which the three key residues of the hydrophobic core and three key residues of the minicore are represented with green and orange beads, respectively. (c) Protein topology cartoon created with the Pro-origami software [209].

6.3 Results

6.3.1 The impact of BDHI on the folding time depends on temperature

The folding temperature T_f is defined as the temperature at which the folded and the unfolded states have an equal population, as depicted in Fig. 6.2. Of note, at T_f this condition is met for both the BD and BDHI conditions, since the Hamiltonian is independent of the effect of the hydrodynamics interactions. Fig. 6.2 shows that the energy landscape is identical for BD and BDHI systems, at $T > T_f$ and at $T < T_f$, as expected.

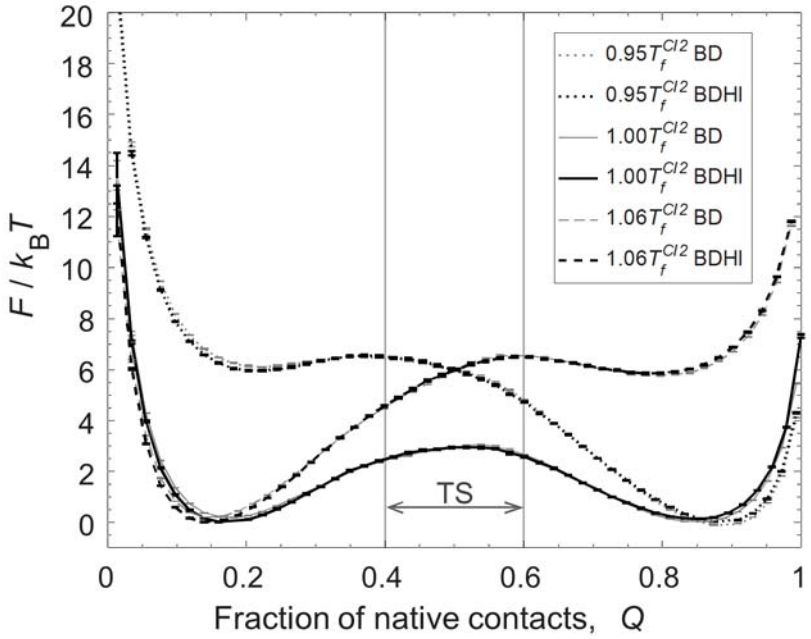

Figure 6.2 Free energy (F) in units of k_BT with respect to the fraction of native contact formation Q without or with HI (BD and BDHI, respectively) for CI2 at a temperature below the folding temperature T_f, at T_f, and above T_f . The transition state region is indicated by TS. Error bars are included.

Although the equilibrium thermodynamics characteristics profiles of BD and BDHI are identical, their dynamics profiles differ drastically. At a temperature below T_f, the folding that occurs in the

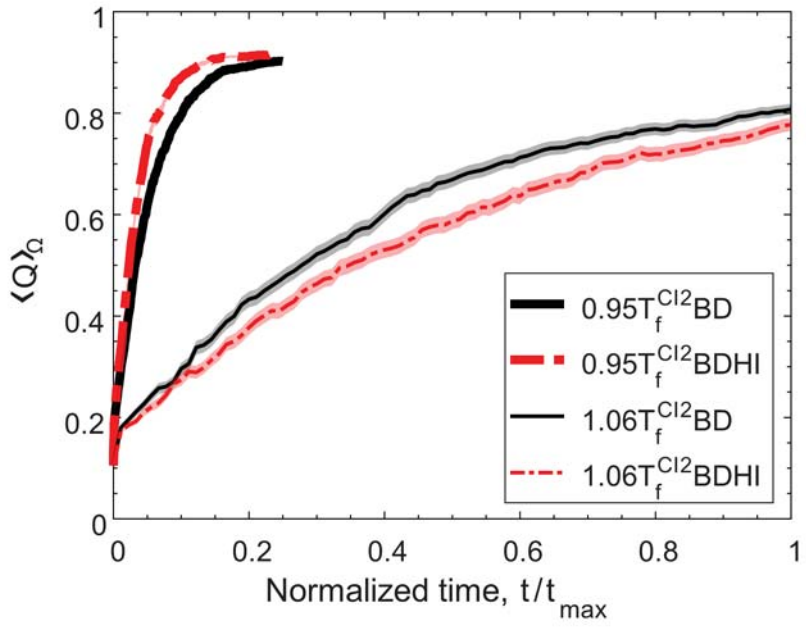

Figure 6.3 Temporal evolution of the fraction of native contact formation Q averaged over all trajectories as a function of normalized time $t/t_{\max}$ at $T < T_f$ and $T > T_f$. Shaded width of lines represents the error bar, which was calculated using a jackknife method.

presence of HI is markedly accelerated, and at temperatures $T > T_f$, the existence of HI decreases the rate of protein folding (Fig. 6.3). The results shown in Fig. 6.3 indicate the average number of native contacts formed $\langle Q \rangle_\Omega$ as a function of normalized time, which depicts the effects of HI on protein folding dynamics across various temperatures.

6.3.2 Hydrodynamic coupling of midrange and long-range contacts and their opposing impact on the general ordering of contact formation at $T < T_f$ and $T > T_f$

Knowing that the effect of HI on the folding depends on the system temperature, and knowing that HI diffusion tensor, defined in Eq. (6.4), considers pairwise distances between different beads of the protein, I sought to determine how the induced pairwise displacement correlation matrix $C_{ij}(t)$ (Eq. 6.9) influences the global secondary structure of the CI2 protein. First, with the help of Fabio Zegarra, I focused the analysis on the point where $Q = 0.4$, which corresponds to the transition

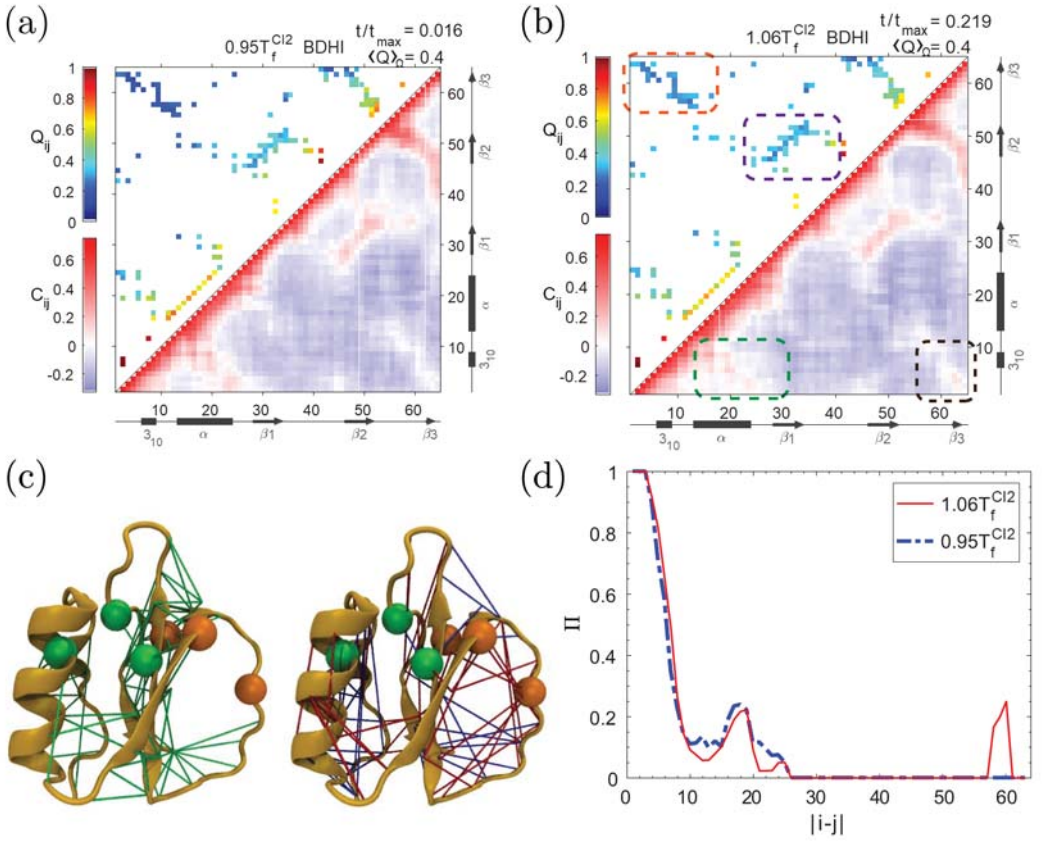

Figure 6.4 (a) The probability of contact formation of each native pair $Q_{ij}(t)$ (represented in the upper triangle) and the displacement correlation $C_{ij}(t)$ (represented in the lower triangle) at $T < T_f$ at a normalized time where $\langle Q \rangle_\Omega = 0.4$. (b) The same representation as (a) for $Q_{ij}(t)$ and $C_{ij}(t)$ at $T > T_f$. (e) The displacement correlation at $T < T_f$(blue) and $T > T_f$ projected on the 3D structure of CI2, where all the pairs are classified into three sets based on the magnitude of positive correlations. If the magnitude of the correlation is similar at both temperatures, the pair is colored with a green edge on the left structure. If the magnitude of the correlation is greater at $T < T_f$ than at $T > T_f$ a pair is colored with a blue edge on the right structure. If the magnitude of the correlation is greater at $T > T_f$ than at $T < T_f$ a pair is colored with a red edge on the right structure. Only the pairs with a sequence separation greater than 8 residues and a magnitude of displacement correlation above the threshold $\mu = 0.061$ as defined in Eq. (6.11) are considered in this representation. The key residues for the hydrophobic core and the minicore are illustrated with green and orange beads, respectively. (d) $\Pi(|ij|)$ from Eq. (6.10) for all pairs whose magnitude of the displacement correlation is above μ are organized according to the sequence separation $|ij$.

state region ($Q \in (0.4, 0.6)$) in Fig. 6.2. The analysis suggests that at both $T < T_f$ and $T > T_f$ the hydrodynamic interactions create a correlation between the native contacts in both types of

secondary structures (i.e., β-sheets and α-helices), as shown in Fig. 6.4(a-b). However, surprisingly, at high temperatures, HI retards the folding process in three ways. First, the paired residues move cooperatively in the same direction, thus adversely affecting the formation of the midrange contacts around the minicore. Second, in addition to the native pairs, the surrounding non-native pairs are more correlated and such pairs need to unfold in order for the folding process to commence. Lastly, the HI at high temperature accelerate the formation of long range native residue contacts earlier than at lower temperatures, a fact which reduces the overall ability of the protein to maneuver towards folding since its two termini are too close (Fig. 6.4(b,d)). Overall ,while the long and mid-range interaction are amplified in $T > T_f$, the short range interactions are stronger in $T < T_f$ as shown in Fig. 6.4(c), using the chance of occurrence defined in equation Eq. (6.10). To illustrate how the correlation in the case of BD where the diffusion tensor does not depend on the pairwise distances of different beads, the data represented in Fig. 6.5 is shown for both low ($T < T_f$) and high ($T > T_f$) temperatures.

6.4 Discussion

This work has settled the controversy of the extent of the impact of HI on folding kinetics by comparing the effect of HI over a broad range of temperatures instead of focusing exclusively on a single temperature [46, 79, 116, 135]. Neither of the groups who previously investigated HI in protein folding discussed temperature effects, nor did they discuss how HI change the dynamics of the formation secondary structure. In contrast, in this work, after a careful analysis of the CI2 protein, the conclusion is that HI affect protein folding depending on the system temperature. Finally, I want to conclude that the correlated motion between beads in viscous solvents, due to HI, accelerate the folding process when $T < T_f$, because it facilitates the formation of local structures, in high temperature conditions ($T > T_f$), by contrast, the correlation is counterproductive and retards the rate of protein folding by reducing the protein flexibility and its concomitant availability to find a productive folding pathway on the energy landscape. Moreover, HI, at high temperatures, add correlation between non-native pairs, which leads to misfolding of local secondary structures

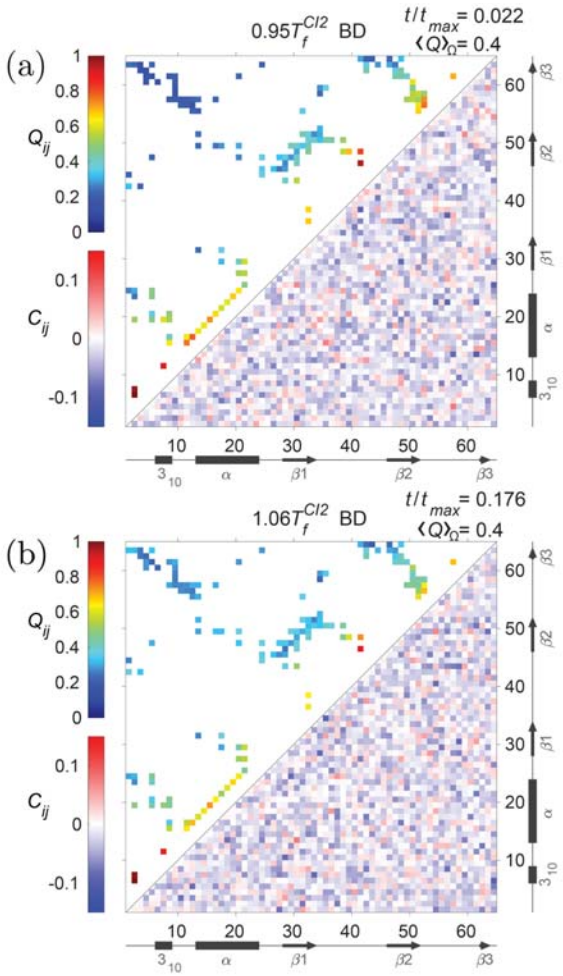

Figure 6.5 The Brownian motion of residues without HI (BD) shows small and random displacement correlation for native and non-native pairs of the CI2 protein. Upper and lower triangles represent the probability of contact formation for each native pair $Q_{ij}(t)$ and displacement correlation $C_{ij}(t)$, respectively. There are two scenarios shown: (a) $T < T_f$ and (b) $T > T_f$. Both at a normalized time where $\langle Q \rangle_\Omega = 0.4$.

that have to be undone before the completion of folding process. Ultimately, the results show that HI [203] has a minor yet measurable effect on on protein folding, which cannot be overlooked in cases where high accuracy is needed, and that temperature is a key parameter to be considered in future HI investigations.

7 Conclusion

In this dissertation, I have summarized our work on three unique examples of complex biological matter. Although complexity has many definitions, I use intuitive and well-precedented definitions in order to study the complex systems investigated herein. Results reported herein demonstrate how network theory and graph theory can be used intuitively to describe the configuration of discrete systems. Moreover, using the formalism of a graph helps in the study of systems using generally applicable techniques, as well as in the ability to visualize the system on as small of a scale as a one-grained representation. This trend of analyzing physical systems using network definitions, parameters, and simulations has been gaining attention in recent years, with a particular focus on the use of network theory for understanding real-world network physics. Moreover, the field of computer science uses a graph-based model in many contexts to describe complex problems and even to describe the nature of complexity itself (i.e. via a computational model such as a Turning machine). As deep learning becomes smarter and smarter and the computational force required for such deep learning continues to increase, there is ample opportunity for network theory to function as a language that can bridge between simple and complex systems. Comparisons also exist between such network understanding and biological systems. In one example, the human brain can be thought of as a billion-node network, including both neural networks as well graphical networks with both edges and directionality. The idea to take a big system and construct a graph from its instant configuration based on pairwise interactions makes it a computationally simple problem to solve, since graphs are fundamental data structures that are highly amenable to analysis. Moreover, network scientists can analyze these systems using network theory and graph theory; as such theories are generally applicable to a broad variety of networks. Additionally, humans will study such systems due to their inherent social nature and interest in understanding connectivity. Finally, neural network researchers are expected to be interested in these results because the graphical representations can be thought of as tensors and/or matrices, and convolutional networks may outperform humans in visual recognition tasks in the near future. The idea of using networks and

complexity has spread also to the field of many-body quantum problems using tensor networks, as well as the connection of and the consolidation between the theory of complexity, networks, and an entanglement towards a unified quantum gravity theory. The novelty of using such networks and network analysis for such complex systems is that it is intuitive for humans to formalize connections between system components, precisely in the way that network analysis requires. I humbly show only the tip of the iceberg of how simple mapping from physical systems to their contact maps can open up a new world. In fact, the striking and surprising fact is that although graph theory existed for decades in analyzing both protein folding and Hi-C experiments, we are the first who brought this theory to the analysis of actomyosin networks. In all cases, the hope remains that graph theory will be able to dramatically advance the pace of predictions, the classification of phases and morphologies, and the development of new theories. I hope that this dissertation will motivate people from wide range of scientific disciplines to consider adopting graph theory, to teach it in undergraduate science classes, and to use it as an intermediate representation to reduce the degrees of freedom of the original system and simplify such analysis. With that, I hope we will learn to let the language and tools of network theory assist us in simplifying complex matter, similar to how Feynman diagrams, which are also one type of graph, describe complex quantum subatomic interactions.

A Appendix 1

A.1 Input file of the phase separation simulations

```
units lj
timestep 0.001

dimension 2
boundary p p p
atom_style atomic
neighbor 0.3 bin
neigh_modify every 20 delay 0 check no

#       variable temp equal 0.5

lattice hex 0.5
region simbox block 0 100 0 100 -0.1 0.1
create_box 2 simbox

lattice hex 0.5
create_atoms 1 region simbox

lattice hex 0.1 origin 0.5 0.5 0
create_atoms 2 region simbox

mass 1 1
mass 2 1.0

comm_modify vel yes
velocity all create ${temp} 343432 dist gaussian
pair_style dpd ${temp} 25.0 12345132

pair_coeff 1 1 25.0 1.0
pair_coeff 2 2 25.0 1.0
pair_coeff 1 2 50.0 1.0
fix 1 all nvt temp ${temp} ${temp} 0.1


dump 1 all atom 10
outputs/temp_${temp}.lammpstrj
write_data dpd_data.all
thermo 500
run 5000
```

A.2 Input file of the Cytosim simulations

```
% Cytosim property file
% Sat Mar  7 13:46:36 2020
% pid 10707

set simul system
{
 time_step            = 0.004;
 viscosity            = 0.5;
 random_seed          = 3158880834;
 steric               = 1, 500, 100;
}

set space cell
{
 property_number      = 1;
 shape                = square;
 dimensions           = 1 1 1;
}

set fiber filament
{
 property_number      = 1;
 rigidity             = 0.075;
 segmentation         = 0.1;
 viscosity            = 0.5;
 confine              = 1, 100, first;
 display              = (line=0.5, 1; color=orange;);
}

set hand motor
{
 property_number      = 1;
 binding              = 10, 0.05;
 unbinding            = 0.1, inf;
 display              = (size=2; color=green;);
 activity             = move;
 stall_force          = 6;
 unloaded_speed       = 0.2;
}

set hand binder
{
 property_number      = 2;
```

```
 binding              = 5, 0.0175;
 unbinding            = 0.1, inf;
 display              = (size=2; color=blue;);
}

set couple crosslinker
{
 property_number      = 1;
 hand1                = binder;
 hand2                = binder;
 stiffness            = 250;
 diffusion            = 10;
 fast_diffusion       = 1;
}

set couple complex
{
 property_number      = 2;
 hand1                = motor;
 hand2                = motor;
 stiffness            = 250;
 diffusion            = 10;
 fast_diffusion       = 1;
}

set hand binder_multivalence
{
 property_number      = 3;
 binding              = 5, 0.006;
 unbinding            = 0.1, inf;
 display              = (size=2; color=blue;);
}

set single grafted
{
 property_number      = 1;
 hand                 = binder_multivalence;
 stiffness            = 200;
}

set solid blob
{
 property_number      = 1;
 viscosity            = 0.5;
 steric               = 10, 0;
```

```
 confine               = 1, 100, cell;
 display               = (style=7; coloring=0; color=0x88888888;);
}
```

A.3 Chemical reaction input file for MEDYAN

```
############################ SPECIES ############################
####_Rep

SPECIESDIFFUSING: A 14500 80.0 0.0 0.0 REG

#Branching
SPECIESDIFFUSING: BD 300 80.0 1.0 0.0 REG

SPECIESDIFFUSING: MD _M_ 0.8 10.0 0.0 REG
SPECIESDIFFUSING: LD _L_ 8.0 0.0 0.0 REG

SPECIESFILAMENT: AF 0

SPECIESPLUSEND: PA 0

SPECIESMINUSEND: MA 0

SPECIESBOUND: LMBEA 0

SPECIESMOTOR: MOA 0
SPECIESLINKER: LA 0
SPECIESBRANCHER: BA 0

LINKERBINDINGSITE: LMBEA 0
MOTORBINDINGSITE: LMBEA 0
BRANCHERBINDINGSITE: LMBEA 0

########################## REACTIONS ##########################

#Normal polymerization
POLYMERIZATIONREACTION: 0 A:DIFFUSING + PA:PLUSEND -> AF:FILAMENT + PA:PLUSEND
↪  0.151
POLYMERIZATIONREACTION: 0 A:DIFFUSING + MA:MINUSEND -> AF:FILAMENT + MA:MINUSEND
↪  0.017

#Normal depolymerization
DEPOLYMERIZATIONREACTION: 0 AF:FILAMENT + PA:PLUSEND -> A:DIFFUSING + PA:PLUSEND
↪  1.4
```

```
DEPOLYMERIZATIONREACTION: 0 AF:FILAMENT + MA:MINUSEND -> A:DIFFUSING +
↪  MA:MINUSEND 0.8

#Motor and linker binding and unbinding
MOTORREACTION: 0 LMBEA:BOUND:1 + LMBEA:BOUND:2 + MD:DIFFUSING <-> MOA:MOTOR:1 +
↪  MOA:MOTOR:2 0.2 1.7 175.0 225.0
LINKERREACTION: 0 LMBEA:BOUND:1 + LMBEA:BOUND:2 + LD:DIFFUSING <-> LA:LINKER:1 +
↪  LA:LINKER:2 0.009 0.3 30.0 40.0

#Motor walking
MOTORWALKINGREACTION: 0 MOA:MOTOR:N + LMBEA:BOUND:N+1 -> MOA:MOTOR:N+1 +
↪  LMBEA:BOUND:N 0.2

#Branch binding and unbinding
BRANCHINGREACTION: 0 BD:DIFFUSING + A:DIFFUSING + LMBEA:BOUND <-> BA:BRANCHER +
↪  PA:PLUSEND 0.0001 0.00000000001 ALL 0.0

#Destruction
DESTRUCTIONREACTION: 0 PA:PLUSEND + MA:MINUSEND -> A:DIFFUSING + A:DIFFUSING 1.0
```

A.4 System input file for MEDYAN

```
####################################################
##################### GEOMETRY #####################

NDIM:                       3

NX:                         2
NY:                         2
NZ:                         2

COMPARTMENTSIZEX:           500.0
COMPARTMENTSIZEY:           500.0
COMPARTMENTSIZEZ:           500.0

MONOMERSIZE:                2.7

CYLINDERSIZE:               27.0

BOUNDARYSHAPE:              CUBIC

####################################################
################### MECHANICS ######################

### ALGORITHM ###
```

```
CONJUGATEGRADIENT:           POLAKRIBIERE

GRADIENTTOLERANCE:           1.0

MAXDISTANCE:                 0.05
LAMBDAMAX:                   0.005

### FILAMENTS ###
FSTRETCHINGFFTYPE:           HARMONIC
FSTRETCHINGK:                100.0

FBENDINGFFTYPE:              COSINE
FBENDINGK:                   2690.0
FBENDINGTHETA:               0.0

VOLUMEFFTYPE:               REPULSION
VOLUMECUTOFF:               54.0

VOLUMEK:                    1E5

### MOTORS ###
MSTRETCHINGFFTYPE:           HARMONIC
MSTRETCHINGK:                2.5

### LINKERS ###
LSTRETCHINGFFTYPE:  HARMONIC
LSTRETCHINGK:       8.0

### BOUNDARY ###
BOUNDARYFFTYPE:             REPULSIONEXP
BOUNDARYCUTOFF:             200.0

BOUNDARYINTERACTIONK:       41.0
BOUNDARYSCREENLENGTH:       2.7

### BRANCHING ###
BRSTRETCHINGFFTYPE: HARMONIC
BRSTRETCHINGK:      100.0
BRSTRETCHINGL:      6.0

BRBENDINGFFTYPE:     COSINE
BRBENDINGK:             100.0
BRBENDINGTHETA:         1.22

BRDIHEDRALTYPE:       COSINE
```

```
BRDIHEDRALK:         100.0

BRPOSITIONFFTYPE: COSINE
BRPOSITIONK:      100.0

##################################################
################### CHEMISTRY ####################

CHEMISTRYFILE:               chemistryinput.txt

CALGORITHM:                  NRM
NUMTOTALSTEPS:
RUNTIME:                     2000.0

SNAPSHOTTIME:                10.0

MINIMIZATIONTIME:   0.05
NEIGHBORLISTTIME:   0.05

NUMDIFFUSINGSPECIES:        4
NUMBULKSPECIES:             0

NUMFILAMENTTYPES:           1

NUMFILAMENTSPECIES:          1
NUMPLUSENDSPECIES:           1
NUMMINUSENDSPECIES:          1

NUMBOUNDSPECIES:             1
NUMLINKERSPECIES:            1
NUMMOTORSPECIES:             1
NUMBRANCHERSPECIES:          1

NUMBINDINGSITES:             1

NUMMOTORHEADSMIN:   15
NUMMOTORHEADSMAX:   30
MOTORSTEPSIZE:      6.0

##################################################
############### DYNAMIC RATE CHANGING ############

DFPOLYMERIZATIONTYPE:   BROWRATCHET
DFPOLYMERIZATIONLEN:    2.7
```

```
DMUNBINDINGTYPE:    LOWDUTYCATCH
DMUNBINDINGFORCE:   12.62

DMWALKINGTYPE:  LOWDUTYSTALL
DMWALKINGFORCE: 15.0

DLUNBINDINGTYPE:    SLIP
DLUNBINDINGLEN:     0.24


################################################
################## INITIALIZATION ##################

NUMFILAMENTS:   50
FILAMENTLENGTH: 1
FILAMENTTYPE:   0
```

B Twitter university ranking table

Rank	University	Handle	Followers
1	Massachusetts Institute of Technology	mit	1,065,024
2	Harvard University	harvard	1,040,326
3	Stanford University	stanford	732,453
4	Yale University	yale	494,248
5	Princeton University	princeton	382,973
6	Columbia University in the City of New York	columbia	349,614
7	The Ohio State University	OhioState	338,224
8	Louisiana State University	lsu	318,282
9	Arizona State University	asu	288,268
10	Cornell University	Cornell	287,788
11	Texas A&M University	tamu	279,081
12	Michigan State University	michiganstateu	243,938
13	University of Michigan	umich	231,723
14	The University of Texas at Austin	utaustin	225,164
15	Indiana University Bloomington	IUBloomington	219,030
16	Penn State University	penn_state	212,145
17	University of Houston	UHouston	206,284
18	University of Florida	UF	198,140
19	University of South Carolina	UofSC	194,890
20	University of California, Los Angeles	ucla	191,775
21	University of Wisconsin-Madison	UWMadison	187,046
22	West Virginia University	westvirginiau	185,783
23	Brown University	BrownUniversity	184,195
24	University of California, Berkeley	UCBerkeley	183,199
25	University of Central Florida	UCF	179,031
26	Georgia Institute of Technology	georgiatech	175,895
27	Texas Tech University	texastech	170,602
28	The University of Oklahoma	uofoklahoma	169,538
29	The University of Alabama	UofAlabama	168,226
30	University of Notre Dame	NotreDame	163,396
31	University of Washington	UW	163,004
32	The University of Arizona	uarizona	162,965
33	Temple University	templeuniv	158,654
34	University of Southern California	usc	152,953
35	University of Pennsylvania	Penn	152,592
36	Texas Christian University	tcu	147,834
37	Mississippi State University	msstate	147,502
38	University of Oregon	uoregon	144,507
39	Rutgers, The State University of New Jersey	RutgersU	134,638
40	The University of Tennessee, Knoxville	UTKnoxville	132,461
41	New York University	nyuniversity	132,384

Table B.1 continued from previous page

Rank	University	Handle	Followers
42	University of Missouri	Mizzou	130,858
43	Auburn University	auburnu	126,804
44	University of North Carolina at Chapel Hill	UNC	126,491
45	Grand Canyon University	gcu	124,898
46	North Carolina State University	ncstate	122,371
47	Clemson University	clemsonuniv	118,308
48	Florida State University	floridastate	117,666
49	University of South Florida	USouthFlorida	117,341
50	University of Iowa	uiowa	116,093
51	The University of Utah	uutah	115,456
52	University of Georgia	universityofga	113,467
53	University of Kentucky	universityofky	108,280
54	United States Military Academy	westpoint_usma	106,704
55	University of Minnesota	UMNews	106,654
56	The University of Texas MD Anderson Cancer Center	mdandersonnews	106,480
57	University of Connecticut	uconn	103,011
58	Purdue University	lifeatpurdue	102,457
59	San Diego State University	sdsu	101,086
60	Johns Hopkins University	johnshopkins	100,199
61	Boston University	BU_Tweets	98,829
62	Baylor University	baylor	98,019
63	University of Maryland	UofMaryland	97,891
64	University of Mississippi	olemissrebels	95,939
65	Texas State University	txst	94,341
66	University of Cincinnati	UofCincy	94,103
67	Liberty University	libertyu	92,823
68	University of Miami	univmiami	91,902
69	University of Virginia	uva	89,660
70	The New School	thenewschool	85,764
71	Duke University	DukeU	84,985
72	University of North Carolina at Charlotte	UNCCharlotte	84,926
73	California State University, Fullerton	csuf	83,838
74	University of Colorado Boulder	cuboulder	83,551
75	University of Akron	uakron	83,424
76	United States Air Force Academy	AF_Academy	82,284
77	University of Illinois at Urbana-Champaign	Illinois_Alma	81,778
78	Howard University	HowardU	81,759
79	California Institute of Technology	caltech	80,732
80	Syracuse University	syracuseu	80,239
81	University of New Mexico	unm	79,393
82	San Josa State University	SJSU	77,750
83	The University of Texas at Arlington	utarlington	77,729
84	Dartmouth College	dartmouth	76,408

Table B.1 continued from previous page

Rank	University	Handle	Followers
85	Oregon State University	oregonstate	75,835
86	Virginia Polytechnic Institute and State University	virginia_tech	75,579
87	Oklahoma State University	okstate	74,558
88	California State University, Fresno	fresno_state	74,077
89	United States Naval Academy	NavalAcademy	73,593
90	Ohio University	ohiou	72,921
91	The University of Texas at San Antonio	utsa	72,859
92	Kansas State University	KState	71,411
93	Boston College	BostonCollege	70,072
94	Virginia Commonwealth University	VCU	69,632
95	Brigham Young University	byu	68,026
96	Georgetown University	georgetown	67,833
97	University of Toledo	utoledo	66,559
98	Wichita State University	wichitastate	65,753
99	University of California, Irvine	UCIrvine	65,571
100	University of Arkansas	uarkansas	64,335
101	Berklee College of Music	BerkleeCollege	64,103
102	Colorado State University	ColoradoStateU	63,738
103	Kennesaw State University	kennesawstate	63,434
104	Northwestern University	northwesternu	62,372
105	Missouri State University	missouristate	59,810
106	University of California, San Francisco	ucsf	59,461
107	The University of Memphis	uofmemphis	59,092
108	University of Chicago	UChicago	58,344
109	Grand Valley State University	gvsu	58,200
110	The Culinary Institute of America	CIACulinary	57,955
111	East Carolina University	EastCarolina	57,425
112	Georgia Southern University	GeorgiaSouthern	57,348
113	James Madison University	JMU	57,072
114	Carnegie Mellon University	carnegiemellon	57,010
115	George Washington University	Gwtweets	55,530
116	University of Delaware	Udelaware	54,938
117	North Carolina Agricultural and Technical State University	ncatsuaggies	54,595
118	Miami University	miamiuniversity	52,925
119	University of California, Davis	ucdavis	52,316
120	Vanderbilt University	vanderbiltu	52,306
121	University of Massachusetts Amherst	umassamherst	52,094
122	California State University, Northridge	csunorthridge	51,650
123	Appalachian State University	appstate	51,623
124	Florida International University	fiu	50,828
125	Marquette University	MarquetteU	50,126
126	Emory University	emoryuniversity	49,429
127	Illinois State University	IllinoisStateU	48,522

Table B.1 continued from previous page

Rank	University	Handle	Followers
128	University of Pittsburgh	PittTweet	48,267
129	Hillsdale College	hillsdale	47,752
130	Drake University	drakeuniversity	47,673
131	University of Rochester	UofR	47,585
132	Wayne State University	waynestate	47,422
133	Ball State University	ballstate	47,240
134	University at Albany, State University of New York	ualbany	46,967
135	Arkansas State University	ArkansasState	46,397
136	American University	AmericanU	45,904
137	University of Dayton	univofdayton	44,791
138	Tufts University	TuftsUniversity	44,787
139	Florida Atlantic University	FloridaAtlantic	44,429
140	Florida Agricultural and Mechanical University	FAMU_1887	44,211
141	University of California, Santa Barbara	ucsantabarbara	44,112
142	College of Charleston	CofC	43,691
143	Washington State University	WSUPullman	43,685
144	Northeastern University	northeastern	43,541
145	Georgia State University	georgiastateu	43,492
146	Kent State University	kentstate	43,258
147	Western Michigan University	WesternMichU	43,062
148	Morehouse College	morehouse	42,621
149	University of North Texas	untnews	42,383
150	The University of Texas at El Paso	utep	42,045
151	University of Nebraska-Lincoln	UNLincoln	42,012
152	Middle Tennessee State University	mtsunews	41,808
153	Northern Illinois University	NIUlive	41,528
154	Savannah College of Art and Design	SCADdotedu	41,228
155	Iowa State University of Science and Technology	IowaStateU	40,961
156	University of North Carolina Wilmington	UNCWilmington	40,341
157	University of California, Riverside	UCRiverside	40,189
158	University of Nebraska at Omaha	unomaha	39,565
159	Northern Arizona University	nau	39,277
160	Southern Methodist University	smu	39,274
161	Spelman College	spelmancollege	39,257
162	Bowling Green State University	bgsu	38,678
163	California State University, Sacramento	sacstate	38,537
164	North Dakota State University	NDSU	38,389
165	Jackson State University	jacksonstateU	38,379
166	University of Northern Iowa	northerniowa	38,136
167	University of California, San Diego	ucsandiego	37,847
168	Loyola University Chicago	LoyolaChicago	37,649
169	Central Michigan University	CMUniversity	37,432
170	Lehigh University	lehighu	37,254

Table B.1 continued from previous page			
Rank	University	Handle	Followers
171	University of Wisconsin-Milwaukee	uwm	37,213
172	Butler University	butleru	37,194
173	Indiana University - Purdue University Indianapolis	iupui	37,156
174	Valdosta State University	valdostastate	36,932
175	California Polytechnic State University, San Luis Obispo	calpoly	36,257
176	Quinnipiac University	QUINNIPIACU	36,152
177	Rice University	RiceUniversity	35,903
178	Oakland University	oaklandu	35,705
179	Washington University in St. Louis	wustl	35,225
180	University of California, Santa Cruz	ucsc	34,981
181	Villanova University	VillanovaU	34,094
182	Stony Brook University	stonybrooku	33,847
183	University of North Carolina at Greensboro	uncg	33,724
184	Binghamton University, State University of New York	binghamtonu	33,437
185	University of Louisiana at Lafayette	ULLafayette	33,141
186	Towson University	TowsonU	33,064
187	College of William & Mary	williamandmary	32,980
188	Hampton University	_hamptonu	32,742
189	University of Rhode Island	universityofri	32,400
190	University of Nevada, Las Vegas	UNLV	32,061
191	Wake Forest University	WakeForest	31,914
192	The University of Tampa	UofTampa	30,862
193	Pratt Institute	prattinstitute	30,680
194	University of Vermont	uvmvermont	30,628
195	The Juilliard School	juilliardschool	30,435
196	University of Louisville	uofl	30,383
197	Purdue University Global	purdueglobal	30,264
198	University of Denver	uofdenver	30,231
199	Rowan University	RowanUniversity	30,192
200	University of Nevada, Reno	unevadareno	29,702

Table B.1 a list of all twitter account that was created based on some this list was created based on `http://www.4icu.org/`, as obtained on May 30, 2020.

C Pool-testing algorithm to enhance COVID-19 detection in mass populations

C.1 Pooled Testing Strategy

Testing of multiple samples, also known as pooled testing, [60] allows one to significantly reduce the number of tests performed when screening a susceptible population, resulting in a quicker screening and using less resources compared to screening each individual in that population. In order to devise an optimal pooling strategy for such testing, researchers first need to estimate the viral prevalence in the population in question [225].

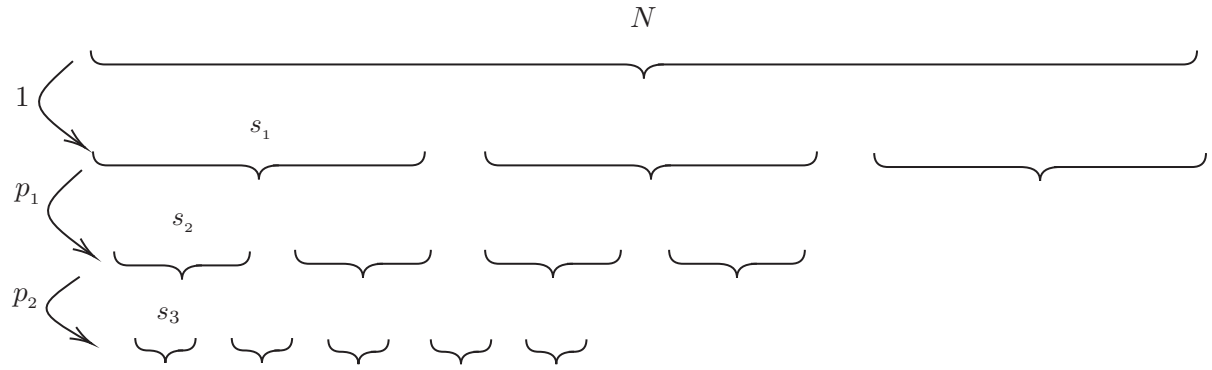

Figure C.1 Nested pooling schematics. A population N is divided into pools containing s_1, s_2 and s_3 samples per pool in each stage of the pooling. Each pooling stage reduces the population by a factor p_i to a presumptive positive population used in the next stage of pooling. The final pooling consists of individual testing, $s_3 = 1$.

C.2 Optimal Pooled Testing Strategy

In a population of size N with a prevalence r_0 of the virus, the number of expected tests following a nested pooling strategy with n pools containing $s_1, s_2, .., s_n$ pools is given by

$$E_n(r_0, \{s_i\}_{i=1}^n) = N\left[\sum_{i=1}^{n}\left(\frac{1}{s_i}\prod_{k=0}^{i-1} p_k\right) + \prod_{k=1}^{n} p_k\right]. \tag{C.1}$$

where $p_0 = 1$, $p_{k>0} = 1 - (1 - r_{k-1})^{s_k}$ is the positive fraction in pool k, and $r_{k>0} = \frac{r_{k-1}}{p_k}$ is the

effective prevalence.

The two most common examples in practice are the 1- and 2-pool testing. Using Eq. (C.1) with $n = 1$ and $n = 2$ we get,

$$E_1(r_0, \{s_1\}) = N\left(\frac{1}{s_1} + p_1\right), \tag{C.2}$$

and,

$$E_2(r_0, \{s_1, s_2\}) = N\left(\frac{1}{s_1} + \frac{p_1}{s_2} + p_1 p_2\right), \tag{C.3}$$

where $p_1 = 1 - (1 - r_0)^{s_1}$, $p_2 = 1 - \left(1 - \frac{r_0}{p_1}\right)^{s_2}$.

We note that for a test with a non-zero false positive rate ϕ, the above expressions of p_k are readily adjusted with the replacement below [10],

$$p_k \to 1 - (1 - \phi)(1 - r_{k-1})^{s_k}, \tag{C.4}$$

where (1-ϕ) is also known as the specificity of the test.

In Fig. C.2, we show the optimal pool sizes for 1-step and 2-step pool-testing and the expected number of total tests given as a percentage of the population $T = E_n/N$. As Eq. (C.1) indicates, the pooling strategy is independent of the size of the population, with both small and large populations using the same general strategy; only the prevalence of the virus in the population is used for strategy development.

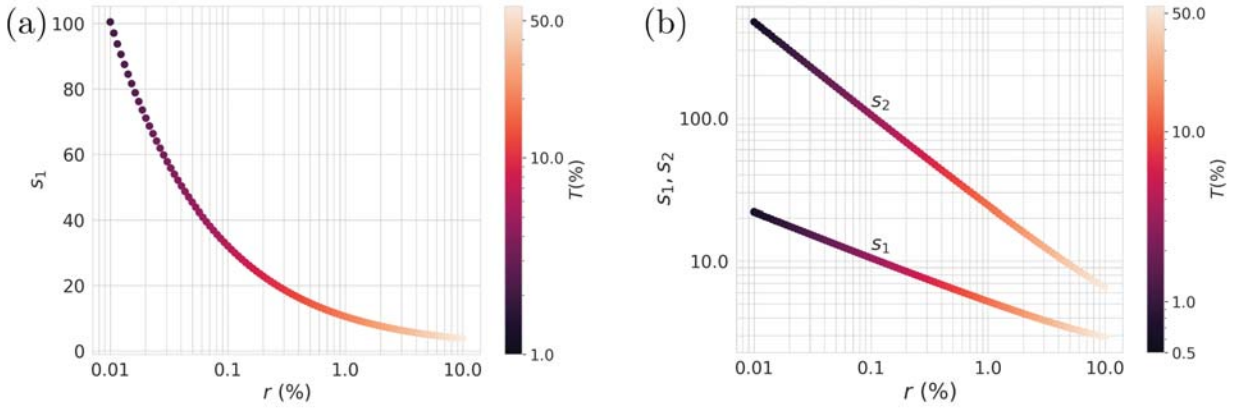

Figure C.2 Simulation of the optimal sizes of the pools in a (a) single-step pooling and (b) double-step pooling framework. The color bar of T represents the total number of tests performed (shown as a percentage of the whole population) to screen the population. r is prevalence of the virus, and s_1, s_2 are the sizes of each pool in the first and the second steps, respectively.

D Appendix 4

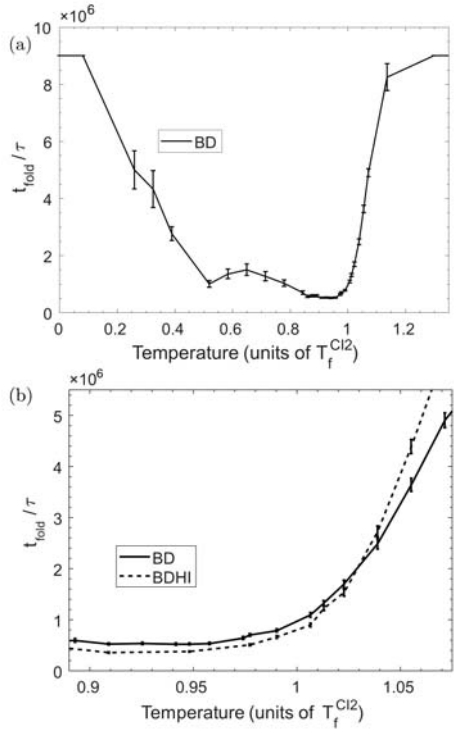

Figure D.1 The average folding time t_{fold} in units of reduced time t/τ with respect to temperature for CI2 in the presence or in the absence of hydrodynamic interactions (HI). Panel (a) shows t_fold over a broad range of temperatures using the Brownian dynamics (BD) without hydrodynamics interactions (HI). The temperature is expressed in units of their corresponding folding temperature T_f . Note the U-shaped dependence of the folding time (non-Arrhenius behavior). t_{fold} using BD with HI (BDHI) is compared to t_{fold} using only BD in panel (b). The crossover occurs when the two curves intersect. Error bars are calculated using the jackknife method.

E Appendix 5

E.1 Actomyosin and memory formation

Dendritic spines in neurons

Neurons are brain cells that communicate using electrical and chemical signals through action potential signaling and neurotransmitter release. Such communication, of note, occurs without any degree of physical contact. At the junction between the two neurons, called a synapse, the presynaptic neuron releases a chemical signal that interacts with and affects the postsynaptic neuron, allowing for signal propagation. Dendritic spines are protrusions that arise from postsynaptic neurons, defined as neurons which are on the "receiver" end of the synapse. The spine volume ranges between $0.001\,\mu m^3$ and $1\,\mu m^3$ [161], and each neuron may grow thousands of spines. Consistent with the Hebbian theory of learning and memory, spines expand during long-term potentiation (LTP) and shrink during long-term depression (LTD) [171]. Experiments of high-resolution imaging have shown that LTP induction causes morphological changes in dendritic spines [231, 235]. Moreover, besides the shrinking or expanding of spines that occur during LTD and LTP, the generation and destruction of spines have also been observed [161]. Of note, the spine is rich with ABPs such as CaMKII, which are crucial in maintaining the information and structural integrity of a spine [187, 111].

CaMKII, morphogenesis, and synaptic plasticity

The Ca^{2+}/calmodulin-dependent kinase II [20] (CaMKII) holoenzyme protein serves a dual role as a Ca^{2+} signaling decoder and as a structural agent in directing calcium signals to change the makeup of actomyosin networks in a spine. It is an extremely common protein, which accounts for up to 2% of the mammalian brain proteome [20, 70]. Even beyond mammals, this protein is found in vertebrates in four major CaMKII isoforms observed in over 40 different splice variants. Overall, these variants are genetically expressed in diverse tissue types.

The two most prevalent isoforms in the brain are CaMKIIα and CaMKIIβ isoforms [114, 113] and the number of the protein domains in these multimeric complexes varies from 12 to 14. Each

unit is made up of several connected domains that serve distinctive functions [220]. Besides being a biochemical catalyzer and having autophosphorylation capabilities [19] in one domain of the CaMKII protein, each monomeric unit is capable of binding actin filaments. To date, the highest actin-binding valency observed for CaMKII is six for the dodecamer (12-mer) CaMKIIβ isoform [113].

Bibliography

[1] Aaronson, S. *Quantum computing since Democritus*. Cambridge University Press, 2013.

[2] Aaronson, S., Kuperberg, G., and Granade, C. The complexity zoo. `https://complexityzoo.uwaterloo.ca`, 2005.

[3] Ajtai, M. Determinism versus nondeterminism for linear time RAMs with memory restrictions. *The Journal of Computer and System Sciences 65*, 1 (2002), 2.

[4] Albert, R., and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics 74*, 1 (2002), 47.

[5] Alberti, S. Phase separation in biology. *Current Biology 27*, 20 (2017), R1097.

[6] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Keith Roberts, P. W., et al. *Molecular biology of the cell*. Taylor and Francis Group, 2018.

[7] Alencar, A. M., Ferraz, M. S. A., Park, C. Y., Millet, E., Trepat, X., Fredberg, J. J., and Butler, J. P. Non-equilibrium cytoquake dynamics in cytoskeletal remodeling and stabilization. *Soft Matter 12*, 41 (2016), 8506.

[8] Alnaes, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N. The FEniCS project version 1.5. *Archive of Numerical Software 3*, 100 (2015), 9.

[9] Alon, U. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

[10] Amos, C. I., Frazier, M. L., and Wang, W. DNA pooling in mutation detection with reference to sequence analysis. *The American Journal of Human Genetics 66*, 5 (2000), 1689.

[11] Ando, T., and Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proceedings of the National Academy of Sciences 107*, 43 (2010), 18457.

[12] Andreas, B., Azuma, Y., Bartl, G., Becker, P., Bettin, H., Borys, M., Busch, I., Fuchs, P., Fujii, K., Fujimoto, H., et al. Counting the atoms in a 28Si crystal for a new kilogram definition. *Metrologia 48*, 2 (2011), S1.

[13] Arora, S., and Barak, B. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[14] Bak, P. *How nature works: the science of self-organized criticality*. Springer Science & Business Media, 2013.

[15] Barabási, A.-L. *Linked: The new science of networks*. American Association of Physics Teachers, 2003.

[16] Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences 101*, 11 (2004), 3747.

[17] Barzel, B., and Barabási, A.-L. Universality in network dynamics. *Nature Physics 9*, 10 (2013), 673.

[18] Bastian, M., Heymann, S., Jacomy, M., et al. Gephi: an open source software for exploring and manipulating networks. *Third International AAAI Conference on Weblogs and Social Media 8* (2009), 361.

[19] Baucum, A. J., Shonesy, B. C., Rose, K. L., and Colbran, R. J. Quantitative proteomics analysis of CaMKII phosphorylation and the CaMKII interactome in the mouse forebrain. *ACS Chemical Neuroscience 6*, 4 (2015), 615.

[20] Bayer, K. U., and Schulman, H. CaM kinase: still inspiring at 40. *Neuron 103*, 3 (2019), 380.

[21] Beggs, J. M., and Plenz, D. Neuronal avalanches in neocortical circuits. *The Journal of Neuroscience 23*, 35 (2003), 11167.

[22] Belmonte, J. M., Leptin, M., and Nédélec, F. A theory that predicts behaviors of disordered cytoskeletal networks. *Molecular Systems Biology 13*, 9 (2017), 941.

[23] Bendix, P. M., Koenderink, G. H., Cuvelier, D., Dogic, Z., Koeleman, B. N., Brieher, W. M., Field, C. M., Mahadevan, L., and Weitz, D. A. A quantitative analysis of contractility in active cytoskeletal protein networks. *Biophysical Journal 94*, 8 (2008), 3126.

[24] Berger, B., and Leighton, T. Protein folding in the hydrophobic-hydrophilic (HP) is NP-complete. In *Proceedings of the Second Annual International Conference on Computational Molecular Biology* (1998), p. 30.

[25] Betts, J. G., Young, K. A., Wise, J. A., Johnson, E., Poe, B., Kruse, D. H., Korol, O., Johnson, J. E., Womble, M., and DeSaix, P. *Anatomy and Physiology*. OpenStax, 2013.

[26] Bikadi, Z., Demko, L., and Hazai, E. Functional and structural characterization of a protein based on analysis of its hydrogen bonding network by hydrogen bonding plot. *Archives of Biochemistry and Biophysics 461*, 2 (2007), 225.

[27] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *The Journal of Statistical Mechanics: Theory and Experiment*, 10 (2008), P10008.

[28] Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. Protein phase separation: a new phase in cell biology. *Trends in Cell Biology 28*, 6 (2018), 420.

[29] Bonacich, P. Power and centrality: A family of measures. *American Journal of Sociology 92*, 5 (1987), 1170.

[30] Bonev, B., and Cavalli, G. Organization and function of the 3D genome. *Nature Reviews Genetics 17*, 11 (2016), 661.

[31] Brylinski, M., Konieczny, L., and Roterman, I. Hydrophobic collapse in (in silico) protein folding. *Computational Biology and Chemistry 30*, 4 (2006), 255.

[32] Bryngelson, J. D., and Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences 84*, 21 (1987), 7524.

[33] Bryngelson, J. D., and Wolynes, P. G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *The Journal of Physical Chemistry 93*, 19 (1989), 6902.

[34] Bunde, A., and Havlin, S. *Fractals and disordered systems*. Springer Science & Business Media, 2012.

[35] Cahn, J. W. On spinodal decomposition. *Acta Metallurgica 9*, 9 (1961), 795.

[36] Cahn, J. W., and Hilliard, J. E. Free energy of a nonuniform system. III. nucleation in a two-component incompressible fluid. *The Journal of Chemical Physics 31*, 3 (1959), 688.

[37] Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E., and Strogatz, S. H. Are randomly grown graphs really random? *Physical Review E 64*, 4 (2001), 041902.

[38] Callaway, E. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature News 525*, 7568 (2015), 172.

[39] Cameron, L. A., Giardini, P. A., Soo, F. S., and Theriot, J. A. Secrets of actin-based motility revealed by a bacterial pathogen. *Nature Reviews Molecular Cell Biology 1*, 2 (2000), 110.

[40] Camilloni, C., Sutto, L., Provasi, D., Tiana, G., and Broglia, R. A. Early events in protein folding: Is there something more than hydrophobic burst? *Protein Science 17*, 8 (2008), 1424.

[41] Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (1986), 679.

[42] Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., and Skelton, N. J. *Protein NMR spectroscopy: principles and practice*. Elsevier, 1995.

[43] Chandler, D. *Introduction to modern statistical mechanics*. Oxford University Press, 1987.

[44] Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C., and Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *The Journal of Chemical Theory and Computation 3*, 1 (2007), 26.

[45] Chong, S.-H., and Ham, S. folding free energy landscape of ordered and intrinsically disordered proteins. *Scientific Reports 9*, 1 (2019), 1.

[46] Cieplak, M., and Niewieczerzał, S. Hydrodynamic interactions in protein folding. *The Journal of Chemical Physics 130*, 12 (2009), 124906.

[47] Cingolani, L. A., and Goda, Y. Actin in action: the interplay between the actin cytoskeleton and synaptic efficacy. *Nature Reviews Neuroscience 9*, 5 (2008), 344.

[48] Clauset, A., Newman, M. E., and Moore, C. Finding community structure in very large networks. *Physical Review E 70*, 6 (2004), 066111.

[49] Clementi, C., Nymeyer, H., and Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *The Journal of Molecular Biology 298*, 5 (2000), 937.

[50] Cohen, R., and Havlin, S. *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.

[51] Cook, S. A., and Reckhow, R. A. Time bounded random access machines. *The Journal of Computer and System Sciences 7*, 4 (1973), 354.

[52] Cooper, J. A., Wear, M. A., and Weaver, A. M. Arp2/3 complex: advances on the inner workings of a molecular machine. *Cell 107*, 6 (2001), 703.

[53] Coveney, P. V., and Novik, K. E. Computer simulations of domain growth and phase separation in two-dimensional binary immiscible fluids using dissipative particle dynamics. *Physical Review E 54*, 5 (1996), 5134.

[54] Crucitti, P., Latora, V., and Porta, S. Centrality measures in spatial networks of urban streets. *Physical Review E 73*, 3 (2006), 036125.

[55] Dill, K. A., and MacCallum, J. L. The protein-folding problem, 50 years on. *Science 338*, 6110 (2012), 1042.

[56] Dill, K. A., and Shortle, D. Denatured states of proteins. *Annual Review of Biochemistry 60*, 1 (1991), 795.

[57] Dima, R. I., and Thirumalai, D. Asymmetry in the shapes of folded and denatured states of proteins. *The Journal of Physical Chemistry B 108*, 21 (2004), 6564.

[58] Dokholyan, N. V., Li, L., Ding, F., and Shakhnovich, E. I. Topological determinants of protein folding. *Proceedings of the National Academy of Sciences 99*, 13 (2002), 8637.

[59] Donald, A. M., Windle, A. H., and Hanna, S. *Liquid crystalline polymers*. Cambridge University Press, 2006.

[60] Dorfman, R. The detection of defective members of large populations. *The Annals of Mathematical Statistics 14*, 4 (1943), 436.

[61] Drenth, J. *Principles of protein X-ray crystallography*. Springer Science & Business Media, 2007.

[62] DuBoff, B., Götz, J., and Feany, M. B. Tau promotes neurodegeneration via DRP1 mislocalization in vivo. *Neuron 75*, 4 (2012), 618.

[63] Eliaz, Y., and Cheung, M. Graph-based ginzburg-landau energy functional theory for soft matter. *Bulletin of the American Physical Society 63* (2018).

[64] Eliaz, Y., Nedelec, F., Morrison, G., Levine, H., and Cheung, M. S. Multivalent actin-binding proteins augment the variety of morphologies in actomyosin networks. *arXiv preprint arXiv:2006.06503* (2020).

[65] ENCODE Project Consortium, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature 489*, 7414 (2012), 57.

[66] Engler, A. J., Sen, S., Sweeney, H. L., and Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell 126*, 4 (2006), 677.

[67] Erard, R. E. If it walks like a duck: A case of confirmatory bias. *Psychological Injury and Law 9*, 3 (2016), 275.

[68] Erdős, P., and Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci 5*, 1 (1960), 17.

[69] Ermak, D. L., and McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics 69*, 4 (1978), 1352.

[70] Erondu, N. E., and Kennedy, M. B. Regional distribution of type II Ca2+/calmodulin-dependent protein kinase in rat brain. *The Journal of Neuroscience 5*, 12 (1985), 3270.

[71] Espanol, P., and Warren, P. Statistical mechanics of dissipative particle dynamics. *EPL (Europhysics Letters) 30*, 4 (1995), 191.

[72] Espanol, P., and Warren, P. B. Perspective: Dissipative particle dynamics. *The Journal of Chemical Physics 146*, 15 (2017), 150901.

[73] Fan, Y., Iwashita, T., and Egami, T. Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. *Nature Communications 8*, 1 (2017), 1.

[74] Feng, J., Levine, H., Mao, X., and Sander, L. M. Nonlinear elasticity of disordered fiber networks. *Soft Matter 12*, 5 (2016), 1419.

[75] Feynman, R. P., Leighton, R. B., and Sands, M. *The Feynman lectures on physics*, vol. 1. Addison-Wesley, 1963.

[76] Fontela, Y. C., Kadavath, H., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. Multivalent cross-linking of actin filaments and microtubules through the microtubule-associated protein Tau. *Nature Communications 8*, 1 (2017), 1.

[77] Fortunato, S. Community detection in graphs. *Physics Reports 486*, 3 (2010), 75.

[78] Fortune, S., and Wyllie, J. Parallelism in random access machines. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing* (1978), p. 114.

[79] Frembgen-Kesner, T., and Elcock, A. H. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *The Journal of Chemical Theory and Computation 5*, 2 (2009), 242.

[80] Frette, V., Christensen, K., Malthe-Sørenssen, A., Feder, J., Jøssang, T., and Meakin, P. Avalanche dynamics in a pile of rice. *Nature 379*, 6560 (1996), 49.

[81] Fruchterman, T. M., and Reingold, E. M. Graph drawing by force-directed placement. *Software: Practice and experience 21*, 11 (1991), 1129.

[82] Garboczi, E., Snyder, K., Douglas, J., and Thorpe, M. Geometrical percolation threshold of overlapping ellipsoids. *Physical Review E 52*, 1 (1995), 819.

[83] Gardner, M. Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American 223*, 4 (1970), 120.

[84] Gebhardt, J. C. M., Clemen, A. E.-M., Jaud, J., and Rief, M. Myosin-V is a mechanical ratchet. *Proceedings of the National Academy of Sciences 103*, 23 (2006), 8680.

[85] Gibson, M. A., and Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A 104*, 9 (2000), 1876.

[86] Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *The Journal of Computational Physics 22*, 4 (1976), 403.

[87] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry 81*, 25 (1977), 2340.

[88] Gittes, F., Mickey, B., Nettleton, J., and Howard, J. Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape. *The Journal of Cell Biology 120*, 4 (1993), 923.

[89] Goldtzvik, Y., Zhang, Z., and Thirumalai, D. Importance of hydrodynamic interactions in the stepping kinetics of kinesin. *The Journal of Physical Chemistry B 120*, 8 (2016), 2071.

[90] Griffith, L. C. Regulation of calcium/calmodulin-dependent protein kinase II activation by intramolecular and intermolecular interactions. *The Journal of Neuroscience 24*, 39 (2004), 8394.

[91] Groot, R. D., and Warren, P. B. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics 107*, 11 (1997), 4423.

[92] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al. Recent advances in convolutional neural networks. *Pattern Recognition 77* (2018), 354.

[93] Gupta, M., Sarangi, B. R., Deschamps, J., Nematbakhsh, Y., Callan-Jones, A., Margadant, F., Mège, R.-M., Lim, C. T., Voituriez, R., and Ladoux, B. Adaptive rheology and ordering of cell cytoskeleton govern matrix rigidity sensing. *Nature Communications 6*, 1 (2015), 1.

[94] Hadziselimovic, N., Vukojevic, V., Peter, F., Milnik, A., Fastenrath, M., Fenyves, B. G., Hieber, P., Demougin, P., Vogler, C., Dominique, J.-F., et al. Forgetting is regulated via Musashi-mediated translational control of the Arp2/3 complex. *Cell 156*, 6 (2014), 1153.

[95] Hagberg, A., Swart, P., and S Chult, D. Exploring network structure, dynamics, and function using NetworkX. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[96] Harpaz, Y., Elmasry, N., Fersht, A., and Henrick, K. Direct observation of better hydration at the N terminus of an alpha-helix with glycine rather than alanine as the N-cap residue. *Proceedings of the National Academy of Sciences 91*, 1 (1994), 311.

[97] Hingorani, K. S., and Gierasch, L. M. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Current Opinion in Structural Biology 24* (2014), 81.

[98] Hoffman, L., Farley, M. M., and Waxham, M. N. Calcium-calmodulin-dependent protein kinase II isoforms differentially impact the dynamics and structure of the actin cytoskeleton. *Biochemistry 52*, 7 (2013), 1198.

[99] Honkura, N., Matsuzaki, M., Noguchi, J., Ellis-Davies, G. C., and Kasai, H. The subspine organization of actin fibers regulates the structure and plasticity of dendritic spines. *Neuron 57*, 5 (2008), 719.

[100] Hoogerbrugge, P., and Koelman, J. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *EPL (Europhysics Letters) 19*, 3 (1992), 155.

[101] Hubel, D. H., and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology 195*, 1 (1968), 215.

[102] Humphrey, W., Dalke, A., Schulten, K., et al. VMD: visual molecular dynamics. *The Journal of Molecular Graphics 14*, 1 (1996), 33.

[103] Hwa, T., and Kardar, M. Avalanches, hydrodynamics, and discharge events in models of sandpiles. *Physical Review A 45*, 10 (1992), 7002.

[104] Hyeon, C., Dima, R. I., and Thirumalai, D. Size, shape, and flexibility of RNA structures. *The Journal of Chemical Physics 125*, 19 (2006), 194905.

[105] Hyman, A. A., Weber, C. A., and Jülicher, F. Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology 30* (2014), 39.

[106] Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *The Journal of Molecular Biology 254*, 2 (1995), 260.

[107] Izhikevich, E. M., Conway, J. H., and Seth, A. Game of life. *Scholarpedia 10*, 6 (2015), 1816.

[108] Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters 78*, 14 (1997), 2690.

[109] Jaynes, E. T. Information theory and statistical mechanics. *Physical Review 106*, 4 (1957), 620.

[110] Jülicher, F., Ajdari, A., and Prost, J. Modeling molecular motors. *Reviews of Modern Physics 69*, 4 (1997), 1269.

[111] Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A., and Noguchi, J. Structural dynamics of dendritic spines in memory and cognition. *Trends in Neurosciences 33*, 3 (2010), 121.

[112] Kazmirski, S. L., Wong, K.-B., Freund, S. M., Tan, Y.-J., Fersht, A. R., and Daggett, V. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proceedings of the National Academy of Sciences 98*, 8 (2001), 4349.

[113] Khan, S., Conte, I., Carter, T., Bayer, K. U., and Molloy, J. E. Multiple CaMKII binding modes to the actin cytoskeleton revealed by single-molecule imaging. *Biophysical Journal 111*, 2 (2016), 395.

[114] Khan, S., Downing, K. H., and Molloy, J. E. Architectural dynamics of CaMKII-actin networks. *Biophysical Journal 116*, 1 (2019), 104.

[115] Kikuchi, N., Gent, A., and Yeomans, J. Polymer collapse in the presence of hydrodynamic interactions. *The European Physical Journal E 9*, 1 (2002), 63.

[116] Kikuchi, N., Ryder, J., Pooley, C., and Yeomans, J. Kinetics of the polymer collapse transition: the role of hydrodynamics. *Physical Review E 71*, 6 (2005), 061804.

[117] Kim, I. H., Racz, B., Wang, H., Burianek, L., Weinberg, R., Yasuda, R., Wetsel, W. C., and Soderling, S. H. Disruption of Arp2/3 results in asymmetric structural plasticity of dendritic spines and progressive synaptic and behavioral abnormalities. *The Journal of Neuroscience 33*, 14 (2013), 6081.

[118] Kim, I. H., Rossi, M. A., Aryal, D. K., Racz, B., Kim, N., Uezu, A., Wang, F., Wetsel, W. C., Weinberg, R. J., Yin, H., et al. Spine pruning drives antipsychotic-sensitive locomotion via circuit control of striatal dopamine. *Nature Neuroscience 18*, 6 (2015), 883.

[119] Kim, J., Feng, J., Jones, C. A., Mao, X., Sander, L. M., Levine, H., and Sun, B. Stress-induced plasticity of dynamic collagen networks. *Nature Communications 8*, 1 (2017), 1.

[120] Koller, D., and Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

[121] Kovács, M., Wang, F., Hu, A., Zhang, Y., and Sellers, J. R. Functional divergence of human cytoplasmic myosin II kinetic characterization of the non-muscle IIa isoform. *The Journal of Biological Chemistry 278*, 40 (2003), 38132.

[122] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (2012), p. 1097.

[123] Lamprecht, R. The roles of the actin cytoskeleton in fear memory formation. *Frontiers in Behavioral Neuroscience 5* (2011), 39.

[124] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278.

[125] Lehn, J.-M. Toward self-organization and complex matter. *Science 295*, 5564 (2002), 2400.

[126] Lehn, J.-M. Towards complex matter: supramolecular chemistry and self-organization. *European Review 17*, 2 (2009), 263.

[127] Lemons, D. S., and Langevin, P. *An introduction to stochastic processes in physics*. JHU Press, 2002.

[128] Levitt, M., and Warshel, A. Computer simulation of protein folding. *Nature 253*, 5494 (1975), 694.

[129] Levy, Y., and Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct. 35* (2006), 389.

[130] Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y.-H., and Reinberg, D. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. *Molecular Cell 38*, 1 (2010), 41.

[131] Li, M., Vitányi, P., et al. *An introduction to Kolmogorov complexity and its applications*, vol. 3. Springer, 2008.

[132] Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science 326*, 5950 (2009), 289.

[133] Liman, J., Bueno, C., Eliaz, Y., Schafer, N. P., Waxham, M. N., Wolynes, P. G., Levine, H., and Cheung, M. S. The role of the Arp2/3 complex in shaping the dynamics and structures of branched actomyosin networks. *Proceedings of the National Academy of Sciences 117*, 20 (2020), 10825.

[134] Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. How fast-folding proteins fold. *Science 334*, 6055 (2011), 517.

[135] Lipska, A. G., Seidman, S. R., Sieradzan, A. K., Giełdoń, A., Liwo, A., and Scheraga, H. A. Molecular dynamics of protein a and a ww domain with a united-residue model including hydrodynamic interaction. *The Journal of Chemical Physics 144*, 18 (2016), 184110.

[136] Lisman, J., Yasuda, R., and Raghavachari, S. Mechanisms of CaMKII action in long-term potentiation. *Nature Reviews Neuroscience 13*, 3 (2012), 169.

[137] Lloyd, S. Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine 21*, 4 (2001), 7.

[138] Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Scott, M. P., Bretscher, A., Ploegh, H., Matsudaira, P., et al. *Molecular cell biology*. Macmillan, 2008.

[139] Logg, A., and Wells, G. N. DOLFIN: Automated finite element computing. *ACM Transactions on Mathematical Software (TOMS) 37*, 2 (2010), 1.

[140] López-Alamilla, N., Jack, M. W., and Challis, K. Reconstructing free-energy landscapes for nonequilibrium periodic potentials. *Physical Review E 97*, 3 (2018), 032419.

[141] Loughlin, R., Heald, R., and Nédélec, F. A computational model predicts Xenopus meiotic spindle organization. *The Journal of Cell Biology 191*, 7 (2010), 1239–1249.

[142] Lu, L., Liu, X., Huang, W.-K., Giusti-Rodríguez, P., Cui, J., Zhang, S., Xu, W., Wen, Z., Ma, S., Rosen, J. D., et al. Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural development and diseases. *Molecular Cell* (2020).

[143] Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Aa resolution. *Nature 389*, 6648 (1997), 251.

[144] Magnasco, M. O. Forced thermal ratchets. *Physical Review Letters 71*, 10 (1993), 1477.

[145] Malik-Garbi, M., Ierushalmi, N., Jansen, S., Abu-Shah, E., Goode, B. L., Mogilner, A., and Keren, K. Scaling behaviour in steady-state contracting actomyosin networks. *Nature Physics 15*, 5 (2019), 509.

[146] Mallinson, J., Shirai, S., Acharya, S., Bose, S., Galli, E., and Brown, S. Avalanches and criticality in self-organized nanoscale networks. *Science Advances 5*, 11 (2019), eaaw8438.

[147] Mandelbrot, B. *Fractals and chaos: the Mandelbrot set and beyond*. Springer Science & Business Media, 2013.

[148] McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[149] Mikhailov, A. S., and Kapral, R. Hydrodynamic collective effects of active protein machines in solution and lipid bilayers. *Proceedings of the National Academy of Sciences 112*, 28 (2015), E3639.

[150] Milo, R. What is the total number of protein molecules per cell volume? a call to rethink some published values. *Bioessays 35*, 12 (2013), 1050.

[151] Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., and Weigt, M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences 108*, 49 (2011), E1293.

[152] Mueller, J., Szep, G., Nemethova, M., De Vries, I., Lieber, A. D., Winkler, C., Kruse, K., Small, J. V., Schmeiser, C., Keren, K., et al. Load adaptation of lamellipodial actin networks. *Cell 171*, 1 (2017), 188.

[153] Mullins, R. D., Heuser, J. A., and Pollard, T. D. The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proceedings of the National Academy of Sciences 95*, 11 (1998), 6181.

[154] Murrell, M., Oakes, P. W., Lenz, M., and Gardel, M. L. Forcing cells into shape: the mechanics of actomyosin contractility. *Nature Reviews Molecular Cell Biology 16*, 8 (2015), 486.

[155] Nakagawa, T., Engler, J. A., and Sheng, M. The dynamic turnover and functional roles of α-actinin in dendritic spines. *Neuropharmacology 47*, 5 (2004), 734.

[156] Nedelec, F., and Foethke, D. Collective langevin dynamics of flexible cytoskeletal fibers. *New Journal of Physics 9*, 11 (2007), 427.

[157] Newman, M. *Networks*. Oxford University Press, 2018.

[158] Newman, M. E. Assortative mixing in networks. *Physical Review Letters 89*, 20 (2002), 208701.

[159] Newman, M. E. Mixing patterns in networks. *Physical Review E 67*, 2 (2003), 026126.

[160] Newman, M. E. The structure and function of complex networks. *SIAM Review 45*, 2 (2003), 167.

[161] Nimchinsky, E. A., Sabatini, B. L., and Svoboda, K. Structure and function of dendritic spines. *Annual Review of Physiology 64*, 1 (2002), 313.

[162] Novick-Cohen, A. The cahn–hilliard equation. *Handbook of Differential Equations: Evolutionary Equations 4* (2008), 201.

[163] Novik, K. E., and Coveney, P. V. Spinodal decomposition of off-critical quenches with a viscous phase using dissipative particle dynamics in two and three spatial dimensions. *Physical Review E 61*, 1 (2000), 435.

[164] Ohno, M., Priest, D. G., and Taniguchi, Y. Nucleosome-level 3d organization of the genome. *Biochemical Society Transactions 46*, 3 (2018), 491.

[165] Okamoto, K.-I., Narayanan, R., Lee, S. H., Murata, K., and Hayashi, Y. The role of CaMKII as an F-actin-bundling protein crucial for maintenance of dendritic spine structure. *Proceedings of the National Academy of Sciences 104*, 15 (2007), 6418.

[166] Olins, A. L., and Olins, D. E. Spheroid chromatin units (ν bodies). *Science 183*, 4122 (1974), 330.

[167] Oliveberg, M., and Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Quarterly Reviews of Biophysics 38*, 3 (2005), 245.

[168] Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry 48*, 1 (1997), 545.

[169] Onuchic, J. N., Nymeyer, H., García, A. E., Chahine, J., and Socci, N. D. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advances in Protein Chemistry 53* (2000), 87.

[170] Otmakhov, N., and Lisman, J. Measuring CaMKII concentration in dendritic spines. *The Journal of Neuroscience Methods 203*, 1 (2012), 106–114.

[171] Park, M., Salgado, J. M., Ostroff, L., Helton, T. D., Robinson, C. G., Harris, K. M., and Ehlers, M. D. Plasticity-induced growth of dendritic spines by exocytic trafficking from recycling endosomes. *Neuron 52*, 5 (2006), 817.

[172] Parsons, J. T., Horwitz, A. R., and Schwartz, M. A. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology 11*, 9 (2010), 633.

[173] Patterson, D. A., and Hennessy, J. L. *Computer Organization and Design ARM Edition: The Hardware Software Interface*. Morgan Kaufmann, 2016.

[174] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research 12* (2011), 2825.

[175] Petsko, G. A., and Ringe, D. *Protein structure and function*. New Science Press, 2004.

[176] Phillips, R., Kondev, J., Theriot, J., and Garcia, H. *Physical biology of the cell*. Garland Science, 2012.

[177] Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. Tech. rep., Sandia National Labs., Albuquerque, NM (United States), 1993.

[178] Pollard, T. D. Rate constants for the reactions of ATP-and ADP-actin with the ends of actin filaments. *The Journal of Cell Biology 103*, 6 (1986), 2747.

[179] Pollard, T. D., and Borisy, G. G. Cellular motility driven by assembly and disassembly of actin filaments. *Cell 112*, 4 (2003), 453.

[180] Popov, K., Komianos, J., and Papoian, G. A. MEDYAN: Mechanochemical simulations of contraction and polarity alignment in actomyosin networks. *PLoS Computational Biology 12*, 4 (2016), e1004877.

[181] Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell 159*, 7 (2014), 1665.

[182] Robinson, R. C., Turbedsky, K., Kaiser, D. A., Marchand, J.-B., Higgs, H. N., Choe, S., and Pollard, T. D. Crystal structure of Arp2/3 complex. *Science 294*, 5547 (2001), 1679.

[183] Rotne, J., and Prager, S. Variational treatment of hydrodynamic interaction in polymers. *The Journal of Chemical Physics 50*, 11 (1969), 4831.

[184] Rotty, J. D., Wu, C., and Bear, J. E. New insights into the regulation and cellular functions of the Arp2/3 complex. *Nature Reviews Molecular Cell Biology 14*, 1 (2013), 7.

[185] Rowley, M. J., and Corces, V. G. Organizational principles of 3D genome architecture. *Nature Reviews Genetics 19*, 12 (2018), 789.

[186] Rubio, M. D., Johnson, R., Miller, C. A., Huganir, R. L., and Rumbaugh, G. Regulation of synapse structure and function by distinct myosin II motors. *The Journal of Neuroscience 31*, 4 (2011), 1448.

[187] Ryu, J., Liu, L., Wong, T. P., Wu, D. C., Burette, A., Weinberg, R., Wang, Y. T., and Sheng, M. A critical role for myosin IIb in dendritic spine morphology and synaptic function. *Neuron 49*, 2 (2006), 175.

[188] Sammis, C. G., and Sornette, D. Positive feedback, memory, and the predictability of earthquakes. *Proceedings of the National Academy of Sciences 99*, suppl 1 (2002), 2501.

[189] Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., and Kertesz, J. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E 75*, 2 (2007), 027105.

[190] Savage, J. E. *Models of computation*, vol. 136. Addison-Wesley Reading, 1998.

[191] Sayood, K. *Introduction to data compression*. Morgan Kaufmann, 2017.

[192] Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature 436*, 7047 (2005), 138.

[193] Scheffer, M. P., Eltsov, M., and Frangakis, A. S. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences 108*, 41 (2011), 16992.

[194] Schlick, T. *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*, vol. 21. Springer Science & Business Media, 2010.

[195] Schoenfelder, S., and Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics* (2019), 1.

[196] Schrodinger, R., Schrödinger, E., and Dinger, E. S. *What is life?: With mind and matter and autobiographical sketches*. Cambridge University Press, 1992.

[197] Seifert, U. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical Review Letters 95*, 4 (2005), 040602.

[198] Seifert, U. Stochastic thermodynamics: principles and perspectives. *The European Physical Journal B 64*, 3-4 (2008), 423.

[199] Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics 75*, 12 (2012), 126001.

[200] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W., Bridgland, A., et al. Improved protein structure prediction using potentials from deep learning. *Nature 577*, 7792 (2020), 706.

[201] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 3 (1948), 379.

[202] Sipser, M. *Introduction to the Theory of Computation*. Cengage Learning, 2012.

[203] Skolnick, J. Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *The Journal of Chemical Physics 145*, 10 (2016), 100901.

[204] Smith, B. A., Daugherty-Clarke, K., Goode, B. L., and Gelles, J. Pathway of actin filament branch formation by Arp2/3 complex revealed by single-molecule imaging. *Proceedings of the National Academy of Sciences 110*, 4 (2013), 1285.

[205] Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics (Oxford, England) 15*, 4 (1999), 327.

[206] Socci, N., Onuchic, J. N., and Wolynes, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *The Journal of Chemical Physics 104*, 15 (1996), 5860.

[207] Stam, S., Freedman, S. L., Banerjee, S., Weirich, K. L., Dinner, A. R., and Gardel, M. L. Filament rigidity and connectivity tune the deformation modes of active biopolymer networks. *Proceedings of the National Academy of Sciences 114*, 47 (2017), E10037.

[208] Standish, R. K. Concept and definition of complexity. In *Intelligent Complex Adaptive Systems*. IGI Global, 2008, p. 105.

[209] Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C., and Stuckey, P. J. Automatic generation of protein structure cartoons with pro-origami. *Bioinformatics 27*, 23 (2011), 3315.

[210] Sugita, Y., and Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters 314*, 1 (1999), 141.

[211] Suraneni, P., Rubinstein, B., Unruh, J. R., Durnin, M., Hanein, D., and Li, R. The arp2/3 complex is required for lamellipodia extension and directional fibroblast cell migration. *The Journal of Cell Biology 197*, 2 (2012), 239.

[212] Svitkina, T. M., and Borisy, G. G. Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *The Journal of Cell Biology 145*, 5 (1999), 1009.

[213] Taketomi, H., Ueda, Y., and Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation: I. the effect of specific amino acid sequence represented by specific inter-unit interactions. *International Journal of Peptide and Protein Research 7*, 6 (1975), 445.

[214] Talkner, P., and Weber, R. O. Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Physical Review E 62*, 1 (2000), 150.

[215] Tan, T. H., Malik-Garbi, M., Abu-Shah, E., Li, J., Sharma, A., MacKintosh, F. C., Keren, K., Schmidt, C. F., and Fakhri, N. Self-organized stress patterns drive state transitions in actin cortices. *Science Advances 4*, 6 (2018), eaar2847.

[216] Tilney, L. G., Connelly, P. S., Vranich, K. A., Shaw, M. K., and Guild, G. M. Why are two different cross-linkers necessary for actin bundle formation in vivo and what does each cross-link contribute? *The Journal of Cell Biology 143*, 1 (1998), 121.

[217] Tuckerman, M. *Statistical mechanics: theory and molecular simulation*. Oxford University Press, 2010.

[218] Vicsek, T. Complexity: The bigger picture. *Nature 418*, 6894 (2002), 131.

[219] Wachsstock, D. H., Schwartz, W., and Pollard, T. D. Affinity of alpha-actinin for actin determines the structure and mechanical properties of actin filament gels. *Biophysical Journal 65*, 1 (1993), 205.

[220] Wang, Q., Chen, M., Schafer, N. P., Bueno, C., Song, S. S., Hudmon, A., Wolynes, P. G., Waxham, M. N., and Cheung, M. S. Assemblies of calcium/calmodulin-dependent kinase II with actin and their dynamic regulation by calmodulin in dendritic spines. *Proceedings of the National Academy of Sciences 116*, 38 (2019), 18937.

[221] Wasserman, S., Faust, K., et al. *Social network analysis: Methods and applications*, vol. 8. Cambridge University Press, 1994.

[222] Weikl, T. R., and Dill, K. A. Folding kinetics of two-state proteins: effect of circularization, permutation, and crosslinks. *The Journal of Molecular Biology 332*, 4 (2003), 953.

[223] Weirich, K. L., Banerjee, S., Dasbiswas, K., Witten, T. A., Vaikuntanathan, S., and Gardel, M. L. Liquid behavior of cross-linked actin bundles. *Proceedings of the National Academy of Sciences 114*, 9 (2017), 2131.

[224] Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics 15*, 2 (1967), 70.

[225] Williams, B. G. Optimal pooling strategies for laboratory testing. *arXiv preprint arXiv:1007.4903* (2010).

[226] Winkelman, J. D., Suarez, C., Hocky, G. M., Harker, A. J., Morganthaler, A. N., Christensen, J. R., Voth, G. A., Bartles, J. R., and Kovar, D. R. Fascin-and α-actinin-bundled networks contain intrinsic structural features that drive protein sorting. *Current Biology 26*, 20 (2016), 2697.

[227] Wloka, C., Vallen, E. A., Thé, L., Fang, X., Oh, Y., and Bi, E. Immobile myosin-ii plays a scaffolding role during cytokinesis in budding yeast. *The Journal of Cell Biology 200*, 3 (2013), 271–286.

[228] Wolfram, S. *A new kind of science*, vol. 5. Wolfram Media Champaign, IL, 2002.

[229] Xue, Z., and Sokac, A. M. Back-to-back mechanisms drive actomyosin ring closure during drosophila embryo cleavage. *The Journal of Cell Biology 215*, 3 (2016), 335.

[230] Yamakawa, H. Transport properties of polymer chains in dilute solution: hydrodynamic interaction. *The Journal of Chemical Physics 53*, 1 (1970), 436.

[231] Yang, Y., Wang, X.-b., Frerking, M., and Zhou, Q. Spine expansion and stabilization associated with long-term potentiation. *The Journal of Neuroscience 28*, 22 (2008), 5740.

[232] Yi, J., Wu, X. S., Crites, T., and Hammer III, J. A. Actin retrograde flow and actomyosin ii arc contraction drive receptor cluster dynamics at the immunological synapse in jurkat t cells. *Molecular Biology of the Cell 23*, 5 (2012), 834.

[233] Zegarra, F. C., Homouz, D., Eliaz, Y., Gasic, A. G., and Cheung, M. S. Impact of hydrodynamic interactions on protein folding rates depends on temperature. *Physical Review E 97*, 3 (2018), 032402.

[234] Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434* (2018).

[235] Zito, K., Knott, G., Shepherd, G. M., Shenolikar, S., and Svoboda, K. Induction of spine growth and synapse formation by regulation of the spine actin cytoskeleton. *Neuron 44*, 2 (2004), 321.