UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mathematical Models of Blood Clot Deformation and Post-Translational Regulation of Enzymes in Metabolic Networks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mathematics

by

Samuel Britton

June 2020

Dissertation Committee:

Dr. Mark Alber, Chairperson
Dr. William Cannon
Dr. Weitao Chen

The Dissertation of Samuel Britton is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I first want to convey my appreciation and gratitude to Dr. Alber for directing my research, constantly encouraging me, advising my career and lending support through GSRs. Additionally, I wish to thank Dr. Cannon for investing in and shaping my career, widening my research horizons, and believing in me. Because of his efforts I was able to spend time at PNNL, pursue exciting projects, as well as receive career and life advice. I would not be where I am without you both. Next I want to thank my undergraduate professors Dr. Joseph DiMuro, Dr. Jason Wilson, and Dr. Janelle Aijian. Each of you were instrumental in my path.

Next, I want to thank my friends and fellow research companions: Mikahl Banwarth-Kuhn, Jolene Britton, Joshua Buli, Kevin Tsai, Daniel Collister, Christian Michael, Cecilia Duran, Alysha Toomey, and Jennifer Ambriz; thanks for making the office a great place to work, learn and enjoy life. I especially want to thank Ali Nematbakhsh, Oleg Kim and Francesco Pancaldi for contributing to my research and professional development, answering tons of my questions, and helping me understand topics that would have been unattainable on my own. Most importantly, I want to thank my family for their support: Frank and Susan Britton, Keith and Janet Houtsma, my siblings Joel, Hannah, Seth and Charles, their spouses Kate and Eli, and my wife Jolene.

I would also like to thank the National Institutes of Health which partially supported this research through grant no. UO1 HL116330. Additionally, I would like thank the U. S. Department of Energy, for supporting this work through two grants: the Office of Science Graduate Student Research grant (SCGSR) and the DOE Office of Biological and

Environmental Research project no. 74860.

The text of this dissertation, in part or in full, is a reprint of the material as it appears in: [20] - Britton, S., Kim, O. Pancaldi, F., Xu, Z. Litvinov, R., Weisel J., and Alber, M. *Contribution of nascent cohesive fiber-fiber interactions to the non-linear elasticity of fibrin networks under tensile load.* **Acta Biomaterialia**, Vol. 94, pp. 514–523, (2019); (Pancaldi, F., Britton, S.), Kim, O., Xu, Z. Litvinov, R., Weisel J., and Alber, M., *Combined modeling and experimental study of platelets-based mechanism of blood clot contraction*, **in preparation**, (2020); [19] - Britton, S., Alber, M., and Cannon, W., *Machine learning and optimal control of enzyme activities to preserve solvent capacity in the cell*, **submitted**, (2020); and [29] - Cannon, W., Britton, S., and Alber, M., *Maximal Entropy Modeling of Metabolism: Characteristics and Comparisons.*, **in preparation**, (2020). The co-authors, Mark Alber and William Cannon directed and supervised the research that forms this dissertation in addition to contributing to the research and preparation of manuscripts. Francesco Pancaldi and Zhiliang Xu provided technical advice on developing mathematical models and Francesco was one of the primary contributor to the development of the platelet model presented in Chapter 3. Oleg Kim, Rustem Litvinov from the laboratory of John Weisel at University Pennsylvania performed experiments with fibrin and platelets and supplied biological expertise. All members contributed in preparation of the listed manuscripts, respectively.

To my wife.

ABSTRACT OF THE DISSERTATION

Mathematical Models of Blood Clot Deformation and Post-Translational Regulation of Enzymes in Metabolic Networks

by

Samuel Britton

Doctor of Philosophy, Graduate Program in Mathematics
University of California, Riverside, June 2020
Dr. Mark Alber, Chairperson

Mathematical models are widely utilized to suggest and test physical and biological hypotheses for which no experimental validation is available at this time. This thesis contains three sections describing several different novel physics based mathematical models in which novel biological hypotheses are proposed and validated.

First, we develop a 3D mathematical model of a fibrin network, a material which determines the deformability and integrity of blood clots. The fiber network is simulated using Langevin dynamics, with each elastic fiber modelled using nonlinear springs. Computational implementation utilizes Nvidia GPUs. We use the model to test a novel structural mechanism of fibrin clots' response to external loads. This mechanism is based on neglected cohesive pairwise interactions between individual fibers in fibrin networks. The contribution of the fiber-fiber cohesion to the elasticity of fibrin networks is characterized in analysis of model simulations by evaluating changes in individual fiber stiffness, length, and alignment of fibers, as well as connectivity and density of the entire network.

Next, the model is extended to include a sub-model of filopodia of platelet cells

which incorporates mechanical forces exerted by filopodia on individual fibrin fibers, and adherence of platelets to one another. Model simulations show how a hypothesized mechanism based on platelets sensing fiber stiffness and thereby adapting their behavior is fundamental for the formation of the distinct contraction phases observed in experiments. Moreover, we quantify how different levels of filopodia strength in response to different stiffness alters clot stability.

Last, a statistical thermodynamic and metabolic control theory framework using hybrid optimization-reinforcement learning (RL) is used to predict the post-translational regulation of enzymes in metabolic pathways. We utilize a non-linear least squares optimization approach to obtain a steady state. The problem of regulation is posed as a Markov decision process and the optimization routine is directly incorporated into a RL environment. The model is used to investigate the hypothesis that regulation is driven by the need to maintain the solvent capacity in the cell. Predictions suggest novel general principles: (1) regulation itself causes reactions to be much further from equilibrium instead of the assumption that non-equilibrium reactions are targets for regulation; (2) regulation is used to maintain concentrations of both immediate and downstream product concentrations rather than to maintain a specific energy charge; and (3) minimal regulation needed to maintain metabolite levels at physiological concentrations also results in a maximal obtainable energy production rate.

Contents

List of Figures **xii**

List of Tables **xviii**

1 Introduction **1**
1.1 Overview . . . 1
1.2 Modelling Background . . . 5
1.2.1 Network Models . . . 5
1.2.2 Individual Cell Models . . . 8
1.2.3 Chemical Reaction Network Models . . . 10
1.2.4 Reinforcement Learning . . . 12

2 A Multiscale Model of Fibrin Fibers **15**
2.1 Introduction . . . 15
2.1.1 Biological Relevance and Importance of Fibrin Mechanics . . . 15
2.1.2 Multiscale Fibrin Mechanical Properties and their Structural Basis . . . 16
2.1.3 Synopsis . . . 18
2.2 Methods . . . 19
2.2.1 Confocal Microscopy and 3D Reconstruction of Stretched Fibrin Networks . . . 19
2.2.2 Formation and Stretching of Fibrin Clots . . . 19
2.2.3 Confocal Microscopy and 3D Reconstruction Methods . . . 20
2.2.4 Model Description . . . 20
2.2.5 Statistical Analysis . . . 23
2.3 Results . . . 24
2.3.1 Fiber Cohesion and Bending in Stretched Fibrin Network . . . 24
2.3.2 Model Validation . . . 27
2.3.3 Fiber-Fiber Cohesion as a Mechanism of Network Stiffening . . . 29
2.3.4 Impact of Fibrin Network Density on Stretched Fibrin Clot Mechanics 32
2.3.5 Fiber Cohesion Increases Fiber Alignment in Stretched Fibrin Clots . 33
2.4 Discussion . . . 35

3 A Model of Platelet Mediated Clot Contraction **41**
3.1 Introduction . . . 41
3.1.1 Biological Relevance . . . 41
3.1.2 Importance of Multiscale Modeling of Blood Clot Deformation and Contraction . . . 42
3.1.3 Synopsis . . . 43
3.2 Methods . . . 44
3.2.1 Formation of Fibrin Networks Mixed with Platelets . . . 44
3.2.2 Platelet-rich Plasma . . . 44
3.2.3 Formation of PRP-clots . . . 44
3.2.4 3D Image Reconstruction of Confocal Microscopy Images . . . 45
3.2.5 Image Analysis of Platelet Filopodia . . . 45
3.3 Model Description . . . 46
3.3.1 Filopodia Contractile Forces . . . 49
3.3.2 Platelet Surface Adhesion and Volume-exclusion . . . 51
3.4 Discussion . . . 52

4 Reinforcement Learning Based Prediction of Enzyme Regulation **54**
4.1 Introduction . . . 54
4.2 Methods . . . 57
4.2.1 Convex Optimization Approach for obtaining Metabolic Steady State. 57
4.2.2 Metabolic Regulation: A Metabolic Control Theory Approach . . . 62
4.2.3 Exploring Regulation: A Reinforcement Learning Approach . . . 67
4.2.4 Model Training . . . 71
4.2.5 Data . . . 72
4.3 Results . . . 73
4.3.1 High NAD/NADH require regulation of metabolite levels in glycolysis 77
4.3.2 High NAD/NADH & High NADP/NADPH require additional regulation in PPP . . . 82
4.3.3 Regulation of PFK Maximizes Flux Through PPP . . . 83
4.3.4 Regulation Increases Reaction Free Energies . . . 86
4.4 Discussion . . . 88

5 Conclusions **94**

Bibliography **98**

A Appendix - Model Descriptions from Chapters 2 & 3 **116**
A.1 Model Description . . . 116
A.2 Modeling Network Structure and Model Calibration . . . 119
A.2.1 Initial Network Generation . . . 119
A.2.2 Calibration of the Single Fiber Sub-Model . . . 121
A.2.3 Calibration of the Single Platelet Sub-Model . . . 123
A.2.4 Modeling Fiber-Fiber Cohesion . . . 125
A.2.5 Calculation of the Fibrin Network Stress-Strain Response . . . 126

A.3 Network Generation Algorithm . . . 127
A.3.1 Node Generation Procedure . . . 127
A.3.2 Fiber Generation Procedure . . . 128
A.3.3 Network generation error reduction . . . 128
A.4 Computational Implementation . . . 129
A.5 Supplementary Figure Descriptions . . . 129

B Appendix - Model Description from Chapter 4 135
B.1 Model Training . . . 135
B.2 Calculating Concentration control Coefficients . . . 136
B.3 Analysis of Gluconeogenesis Pathway . . . 137
B.4 Analysis of Glycolysis-TCA Pathway . . . 139
B.5 Analysis of Pathways with Proxy Data . . . 139
B.6 Computational Implementation . . . 146
B.7 Tables . . . 146

List of Figures

2.1 Schematics of in silico fibrin clot stretching experiment and description of the individual fibrin fiber model. (A) Representative image of a three-dimensional fibrin network used in simulations. Black arrows show the direction of applied force. Zoomed section shows the detailed fibers. (B) Spatial discretization of a single fiber using 6 interior sub-nodes connected by WLC springs. i and j are the main nodes of the fiber. Zoomed section illustrates bending springs in a fiber using angular springs between nodes (including *non-rotating* both main and sub-nodes). (C) Representation of two fibers and a cohesive interaction between them. 22

2.2 Visualization and quantitative structural analysis of non-deformed and stretched fibrin clots. (A) Reconstructed confocal microscopy-based 3D image of a fluorescently labeled non-deformed isotropic fibrin network. (B) Fibrin network stretched to a 20% strain in the y-direction. Fiber branch points are shown by yellow squares in (A) and fiber-fiber cohesion points are indicated by yellow circles in (B). Bent fibers are indicated by white arrows. (C,D) Quantification shows a node connectivity degree (C) and node density (D) in non-deformed fibrin clots and in the same clots stretched to a 20% strain (90 nodes analyzed in each fibrin clot prepared from 3 independent donors). The results in (C) and (D) are presented as M ± SEM. A two-tailed Mann-Whitney U test, $^*P < 0.05$. 26

2.3 Mechanical and structural changes of fibrin networks under uniaxial stretching. (A) Characterization of the mechanical and structural changes in the fibrin network in terms of the unidirectional tensile stress-strain response (M ± SD, n=10). (B) Average fiber alignment in the central 50% region of the network. (C) Percent of the stretched fibrin clot cross-section area covered by fibers. The solid line shows the nonlinear model simulation results, circles are experimental data [21]. 28

2.4 Impact of fiber-fiber cohesion on the tensile stress-strain response and fiber strain distribution in stretched fibrin clots. (A) Tensile stress-strain responses (M ± SD, n=5) for fibrin networks of cohesive and non-cohesive fibers of density 5 fiber/μm^3. (B) Corresponding strain distributions of individual fibers at the fibrin network strain $\Gamma = 3$. 31

2.5 Strain-dependence of the node connectivity for fibrin networks with cohesive fibers. The fiber density $\rho_f = 5$ fibers/μm^3. Dotted, dashed and solid lines in the columns correspond to the network connectivity degrees of 3, 4, and > 4, respectively. 32
2.6 Fibrin clot mechanical response and fiber strain probability distributions for different fiber densities. (A) Tensile stress-strain responses ($\mathrm{M} \pm \mathrm{SD}, n = 5$) for fibrin networks of different fiber densities, ρ_f. The inset image shows the stress on a logarithmic scale. (B) Individual fiber strain distributions for the network strain $\Gamma = 2$. 34
2.7 Impact of cohesive fiber-fiber interactions on fiber network alignment. Joint distributions for network alignment and network strain are shown for cohesive and non-cohesive networks at 300% strain for fiber density $1/\mu m^3$ (A, B) and $5/\mu m^3$ (C, D). Simulated networks are shown alongside the joint density distribution to illustrate the corresponding location of aligned fibers. The Y-axis represents the scaled position along the length of the fiber network, $z^* = z/z_0$. z_0 is the original clot length and z ranges over the current network length from bottom ($z = 0$) to top ($z = 3z_0$). Greyscale at each point corresponds to the relative number of fibers oriented along the direction of the strain. 36

3.1 Probability density function of platelet filopodia counts for simulated (n=10) and experimentally (n=4) obtained platelet-fibrin meshworks. 46
3.2 Schematics of in silico platelet contraction simulation and depiction of fibrin fiber model. (A) Left. Fibrin network (blue lines) and platelets (gray spheres). Right. Zones of influence of platelet forces: interaction zone (light gray), adhesive zone (gray); volume exclusion (dark gray). (B) Single fibrin fiber divided in subsections governed by Worm-Like-Chain springs and bending springs. (C) Formation of cohesive bond between sufficiently close fibers. . . . 47
3.3 Schematic of individual platelet interacting with fibrin fibers in the clot environment. Different forces (red arrows) are applied to fibers depending on the varying distance from the platelet center of mass. Fibers are pulled towards the platelet at a force, F_p, in the interaction zone (light gray). Adhesion force, F_0, is applied to fibers on the platelet surface (gray), and a Lennard-Jones force, F_{LJ}, is used for volume exclusion (dark gray). Individual filopodia, modelled as forces, act by pulling fibers in the environment outside the platelet volume but within the range of the filopodia, r_{fil} (light grey annulus). Once a fiber is pulled within the platelet volume of radius r_p, the fiber is held on the surface by an adhesive force, F_0 (grey zone annulus). Fibers are held out of the central platelet volume (dark grey circle) using a Lennard-Jones potential. 50

4.1 Schematic of *in silico* framework for learning regulation (grey box) with coupled simulation or optimization routine controlling environmental feedback. Initial framework input (green box) consists of target metabolite concentrations from experimental data. The output (red box) consists of a learned optimal enzyme regulation scheme necessary to reach the target concentrations. Learning is performed by repeatedly testing different regulation schemes and updating the value function, V, that returns a scalar value for a given set of enzyme activities. Enzyme activities, represented as states, are chosen for regulation by performing actions that are determined by a policy function. A given policy is determined by V. The new steady state metabolite concentrations resulting from applied regulation are determined by an optimization routine. Alterations in metabolite concentrations are a direct result of moving into a state s' from a state s after taking action a, i.e. performing regulation. These dynamic changes are used to define a reward function, R, that determines environmental feedback. Rewards are used to direct the agent as it explores and learns a policy that predicts optimal enzyme regulation. . . . 72

4.2 Initial steady state properties before any regulation is applied in the form of reduced activity coefficients for glycolysis-PPP-TCA cycle with high NAD/NADH and low NADP/NADPH conditions. The steady state is determined by maximizing the reaction path entropy such that the net thermodynamic driving force on each reaction is proportioned according to the governing equation for metabolite kinetics, Eqn. (4.10). (A) Unregulated reaction fluxes. (B) Unregulated reaction free energies. Reduction of activity coefficients to values less than 1.0 reduces both the steady state fluxes and the reaction free energies (Fig. 4.3-4.5). **Abbreviations:** HEX1, Hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, Fructose bisphosphatase; TPI, Triosephosphate isomerase; GAPD, Glyceraldehyde 3-phosphate dehydrogenase; PGK, Phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, Enolase; PYK, Pyruvate kinase; PYRt2m, pyruvate transporter; PDH, Pyruvate dehydrogenase; CSM, Citrate Synthase; ACONT, Aconitase; ICDH, Isocitrate dehydrogenase; AKDG, a-ketoglutarate dehydrogenase; SUCOAS, Succinyl-CoA synthetase; SUCD, Succinate dehydrogenase; FUM, Fumarase; MDH, Malate dehydrogenase; GOGAT, Glutamine oxoglutarate aminotransferase. 79

4.3 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and low NADP/NADPH conditions. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square).(B) Reaction free energy changes are no longer equally distributed across subpathways (Fig. 4.2, upper glycolysis, PPP, lower glycolysis, TCA cycle) but instead free energies are further from equilibrium at reactions where regulation is applied. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. **Abbreviations:** G6PDH, Glucose 6-phosphate dehydrogenase; PGL, Phosphogluconolactonase; GND, phosphogluconate dehydrogenase; RPI, Ribose 5-phosphate isomerase; RPE, Ribose 5-phosphate epimerase; TKT1, Transketolase 1; TALA, Transaldolase; TKT2, Transketolase 2. Note: previously used abbreviations are presented in Figure 4.2 81
4.4 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. 84
4.5 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions and PFK activity set to zero. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. The local MCA method results in zero flux (Appendix B Table B.1) and is therefore not shown. 86

A.1 Length and degree distributions. (A) Fiber length and (B) Network connectivity distributions, $P(L)$ and $P(D)$, for the simulated and experimentally obtained fibrin networks [93] are shown for the fiber density $= 1/\mu m^3$. 120
A.2 Model calibration using single fibrin fiber force-strain response curve. Symbols indicate experimental data for a single fiber measured using an atomic force microscope AFM [74]. The black line shows the model fit (Eqn. (A.4)). 123
A.3 Connection between node degree and network strain. Average density of 3-degree (A), 4-degree (B) and higher degree (C) nodes are plotted at different network strains: $\Gamma = 0.1 - 2.5$. Fiber density, ρ_f, varies from 1 to 15 fibers/ μm^3. 131

A.4 Comparison between fiber alignment and densification, with and without fiber-fiber cohesive interactions. The average alignment of fibers in the middle 50% of fibrin networks for cohesive (A) and non-cohesive (B) fibers at different fiber densities, ρ_f. Volume fraction occupied by fibers, α_f, inside the central 6 μm of the network for cohesive (C) and non-cohesive (D) fibers for the same densities, ρ_f. 132
A.5 Connection between cohesion and strain. Individual fiber strain and normalized network strain joint distributions are shown for cohesive and non-cohesive networks at 300% strain for fiber density $1/\mu m^3$ and $5/\mu m^3$. The Y-axis represents the scaled position along the length of the fiber network, $z^* = z/z_0$. z_0 is the original clot length and z ranges over the current network length from bottom ($z = 0$) to top ($z = 3z_0$). Grey scale at each point corresponds to the relative number of fibers at a given strain along the direction of network strain. 133
A.6 Stress-strain curves for different sub-node spacing, δ, at varying fiber density, ρ_f. Network density varies from between 1 (A), 5 (B), and 15 (C) fibers/μm^3. 134

B.1 Cumulative normalized rewards averaged over 10 RL runs for the hyperparameters (n, lr) which resulted in the maximal reward. 136
B.2 Gluconeogenesis cycle predictions with low NAD/NADH initial conditions. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent. 138
B.3 Glycolysis-TCA cycle predictions with high NAD/NADH initial conditions. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent. 140
B.4 Glycolysis-TCA cycle predictions with high NAD/NADH initial conditions without experimental metabolomics data. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent. 142
B.5 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and low NADP/NADPH conditions without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. 143

B.6 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. 144
B.7 Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions and PFK activity set to zero without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. The local MCA method results in zero flux as in the case when experimental data is utilized and is therefore not shown. 145

List of Tables

4.1 The set of enzymes found to be regulated in all analyses along with the associated pathway, the concentration control coefficient, $C_j^{\tilde{n}}$, of the reaction summed over all metabolites before any regulation is applied, the method predicting the regulation and the experimental evidence from the literature for predicted regulation. Abbreviations are the same as in Figure 2. PC is pyruvate carboxylase and is observed to be regulated in gluconeogenesis (Appendix B Table B.4). 93

A.1 Fibrin network model simulation parameters. 124
A.2 Platelet model simulation parameters. 125

B.1 Reaction fluxes at predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions. 147
B.2 Reaction free energy at predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions. 148
B.3 Predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions. Unlisted reactions are unregulated. 149
B.4 Predicted enzyme activities, fluxes, and free energy for gluconeogenesis pathway from MCA-local, MCA, and RL methods. 150
B.5 Predicted enzyme activities, fluxes, and free energy for glycolysis-TCA pathway from MCA-local, MCA, and RL methods. 151

Chapter 1

Introduction

1.1 Overview

In recent years, there has been a rapid increase in data of variable type and spatial scale. New approaches based on techniques such as multiscale modelling (MSM), optimization or physics-based models, have been effective tools in incorporating different ranges of spatial data as well as utilizing important biological aspects of the modelled system. By incorporating different types of experimental multiscale data, models can generate more precise predictions. This process of data incorporation followed by model prediction is frequently maintained through a constant iterative process between computational and experimental scientists and has become a powerful investigative tool in helping better understand complex biological systems. The focus of this thesis is on two systems: blood clotting and metabolism. Each system combines significant challenges in applied mathematics approaches, computational methods, systems biology modeling and understanding of biological processes.

In Chapters 2 and 3 mechanical spatio-temporal dynamics of biogels are studied.

Two mechanical models are presented that are utilized to quantify specific mechanisms of blood clot contraction, namely the role of fiber-fiber cohesion and the filopodia-fiber response that drives platelet aggregation. Blood clots are biological structures that form as a result of a complex series of chemical reactions known as the coagulation cascade. The coagulation cascade is usually initiated by an injury to a blood vessel inducing the release of coagulation factors in the bloodstream. These factors activate platelets present in the plasma, thereby changing their mechanical and adhesive properties. The change in platelet adhesion is characterized by a transition from discoidal shape to star shaped configuration, along with the formation of numerous protrusions called filopodia extending from the main body of the platelets. Platelets then start to merge, forming aggregates that assume the role of nucleation sites for the polymerization of fibrinogen converting into fibrin fibers [191, 208, 92]. The process eventually leads to the formation of a fully formed thrombus (blood clot) at the site of the vessel injury, the role of which is to prevent internal bleeding and promote healing of the damaged epithelium. The MSM approaches presented here combine representations of different properties and components including the properties of platelets and fibrin fibers, and the mechanical forces exerted by the platelets due to their activation. Each model presents novel insights into the biological system that would not, at this current time, be experimentally possible.

Even so, MSM is generally a computationally expensive approach and is unable to incorporate all relevant or new data without additional model features. Furthermore, with the plethora of usable data (e.g. imaging), metabolomics, proteomics and transcriptomics data it has been very difficult to integrate this data into multi-scale mechanistic models in

ways that elucidate the underlying biological processes. Models also tend to be specific or tailored towards a single organism or experimental setup, thereby increasing model precision but consequently reducing model re-usability and theoretical predictive power. For these reasons, many have turned to machine learning (ML) approaches to supplement MSM methods. ML, however, has frequently been relegated to a preprocessing step, as in the case of image analysis, PCA, or clustering algorithms, or used as a standalone method which ignores nearly all physical and biological knowledge that was present in the mechanical model. Individually, each modelling approach (physics-based MSM vs ML) can be applied to a relatively narrow range of problems. A challenging problem in the mathematical biology community is to combine the theoretical strengths of physics-based modeling methods with the flexibility and power of new ML techniques. Instead of using MSM methods and ML as separate approaches, ML methods should be used in conjunction with physics-based methods to form a hybrid approach. This hybrid approach combines the strengths for physics-based and ML in separate model components (see, for details [3, 153]).

In Chapter 4, one of the first hybrid models combining optimization and ML is presented. The model is applied to study metabolism in order to predict enzyme regulation necessary to control metabolite concentrations as well as understand novel theoretical implications related to regulation. Metabolism is the process by which cells dissipate available energy. In order to do this, cells produce and utilize a myriad of metabolites for movement, growth, and survival. The biological mechanism for the necessary conversion from one type of metabolite to another is catalyzed by a collection of proteins called enzymes. This collection of enzyme catalyzed reactions comprise what is known as a chemical reaction

network. In order to produce and reduce metabolite concentrations as well as respond to changes in the environment, cells have developed the ability to regulate specific enzymes for this directed purpose. However, experimental measurement or computational inference of the enzyme regulation needed in a metabolic pathway is difficult and costly to perform. Consequently, regulation is known only for well-studied reactions of central metabolism in various model organisms. Here, a reinforcement learning based approach coupled to an optimization routine is presented in order to form a hybrid optimization-RL model (Chapter 4). The optimization component includes the theoretical knowledge and thermodynamics from the chemical reaction network while the RL component provides the ability to learn regulation policies that match physiological levels of metabolites while minimizing the heat loss and maximizing free energy.

The methods presented here are one the first attempts in combining two rich fields of MSM and ML that have previously been fragmented. In the following sections, the necessary modelling background and previous approaches for discrete network models (Section 1.2.1), single cell models (Section 1.2.2, chemical reaction network models (Section 1.2.3) and reinforcement learning techniques (Section 1.2.4) are described. Detailed mechanical models of fibrin network and platelets are presented in Chapters 2 and 3. The hybrid optimization-RL model is described in Chapter 4. The work presented here is utilized to construct a hybrid ML-MSM environment where the ML predictions iteratively inform development and calibration of large-scale mechanistic models. The MSM model allows for an accurate representation of the biological system, while the ML model offers the ability to adapt to different data types and produce predictions based on model simulations.

1.2 Modelling Background

1.2.1 Network Models

Several theoretical and computational frameworks have been recently developed for modeling fibrin structural mechanics at different spatial scales ranging from molecular to macroscale [9, 104, 181, 109, 164, 215, 103, 43, 36, 155, 91, 213, 182, 154]. At the molecular scale, fibrin mechanics is defined by the properties of monomeric fibrin, an elementary structural unit that shares structural and mechanical similarity with fibrinogen, a blood plasma protein, converted enzymatically to monomeric fibrin [134]. To model the dynamics of human fibrin(ogen) undergoing forced elongation, molecular dynamics (MD) simulations were used to characterize the α-to-β transition in α-helical coiled-coil connectors of the fibrin molecule, revealing distinct elastic, plastic, and non-linear regimes in force-extension profiles [222]. Zhmurov et al. [221] used a model to elucidate structural mechanisms of forced elongation of fibrin molecules based on stepwise unfolding of γ nodules concomitant with partial stretching and contraction of α-helical connectors.

In order to explain the strain-stiffening behavior of stretched fibrin networks, two conceptually different types of models of cross-linked filamentous networks were previously developed and applied [86]. The first type assumed the existence of semiflexible filaments that undergo thermal fluctuations [182, 125]. One entropic approach used the Worm-Like-Chain (WLC) to model the force-strain profiles of fibrin fibers under stretching, with fitting parameters obtained in atomic force microscopy experiments [74, 76]. The second type modeled filaments as elastic rods that can bend and stretch but do not exhibit thermal fluctuations [47, 145, 32]. Both types of models were capable of capturing the strain-stiffening

behavior of fibrin networks, but gave different predictions for the degree and onset of stiffening, suggesting that further studies are required to develop these entropic and athermal models to the mechanical properties of fibrin hydrogels [86].

At the macroscale fibrin network level, discrete and continuous modeling approaches have been used to account for elastic responses of fibrin networks to external tensile and shear deformations inducing stiffening [86, 193, 88] as well as to suggest mechanisms of softening-stiffening behavior of compressed fibrin networks [91, 115, 194, 192]. Several continuous models of fibrin networks, including a three-chain model [77], an eight-chain model [80], and an isotropic network model [182], were used to predict the force-strain response of stretched fibrin clots [21]. All these models were shown to correctly reproduce fibrin network behavior under tension in the linear regime. However, at large strains, the results significantly deviated from experimental data [21]. All three models simulated isotropic networks and assumed affine network deformations. Meanwhile, biological networks such as those formed by fibrin are frequently anisotropic and their deformation is non-affine [211].

More recently, a phase transition method has been used to predict the shear viscoelastic response of compressed networks, which revealed a remarkable softening-stiffening behavior due to bent fibers and network densification [91]. Several discrete models have been developed to consider the formation of connections between individual fibers. One of such models, based on a bead-spring representation of individual fibers, was used to determine fibrin network elastic modulus for different network structures [88]. A similar model was developed to study how network connectivity affects the mechanical properties and structural integrity of tissue [175]. This model was simplified to construct a minimal 2D lattice

model that was used to show that fiber-fiber interactions could influence clot stiffness in compressed fibrin networks [195]. However, the extent to which such interactions contribute to overall clot stiffness has not been quantified.

Another model, based on the beam theory approach and network geometries obtained using a random walk algorithm, was recently developed to study the behavior of layered biomaterials such as electrospun polymeric scaffolds with a particular focus on studying orientation and bending properties of the fibers as well as their initial intersections within one layer and the immediately adjacent layers [32, 42, 31],. Additionally, continuous models have also been developed to efficiently simulate deformation of 2D layered materials at a mesoscale [220, 212]. Both types of models were used to investigate how the structure of electrospun single 2D scaffolds affect their macroscopic mechanical properties. Authors successfully simulated multiple layers and found limited difference between monolayered and multilayered biomaterials under in-plane load [32]. While both approaches are novel and important for studying layered materials, these types of models are not designed to capture the impact of fiber cohesion on the dynamical changes of the 3D structure of fibrin networks under stretching [91, 88].

To summarize, most existing models do not consider bending of individual fibers or physical contacts between them, which can significantly alter the mechanical response of the entire fibrin network. These components are included in the novel Cohesive Fibrin-Fibrin Crisscrossing Model (CFFCM) described in Chapter 2 and used to study and quantify the contribution of fiber cohesion to local and global strain stiffening.

1.2.2 Individual Cell Models

The stability and mechanisms of clot contraction in hemostasis and thrombosis are underappreciated and under studied despite their importance. This is partly due to the lack of methodology to quantitatively assess the mechanical and structural basis of this complex multiscale process. This research is meant to motivate insights into fundamental aspects of cell biology, including cell motility and interaction of cells with extracellular matrix (ECM), as well as into translational and clinical medicine in relation to blood clotting and its disorders leading to heart attack, stroke, and venous thromboembolism. After an injury occurs, a blood clot is formed which is followed by wound healing. The fundamental aspects of the clot formation are the fibrin structure, a viscoelastic proteinaceous polymer that determines the deformability and integrity of clots, and the interaction of activated platelets with the forming fibrin network. Essentially, the fibrin network facilitates platelet's ability to control changes in its environment. Platelets collectively change the shape and structure of the clot by coordinating how they pull on individual fibers. Once attached, platelets undergo contraction that causes shrinkage and stability of the blood clot. This work focuses on two primary areas: the features of fibrin that are mechanically responsible for clot stability, and how the behavior of platelets in a dynamic environment affects overall clot contraction.

Most detailed modelling approaches do not incorporate impact from individual cells within the clot structure. In order to include the mechanical properties of single cells, several models have been developed to investigate various aspects of thrombus contraction, including platelet aggregation in viscous flow [61], forces exerted by platelets spreading on a substrate [66], and remodelling of ECM due to filipodial forces [127]. Individual platelet

cells have been modeled in high resolution under activated conditions [89], under fluid flow [187], and with adhesion to the blood vessel wall structure [210].

On a larger scale, dissipative particle dynamics (DPD) models have been developed to account for recruitment of inactive platelets by activated platelets in viscous shear flow [61]. Similarly an understanding of blood flow when mediated by platelet transport through small blood vessels was gained by modelling multiple platelets [214]. Recently, modelling efforts have been made towards coupling high resolution platelets with elastic networks to quantify how contraction induces large scale stress and how the resulting stiffness of the surrounding network or stimulates cell invasion [127, 65]. A similar modelling approach utilized large networks to include myosin II motors that traversed along actin filaments [126]. Continuous modelling approaches have also been utilized to understand the dynamics of cells interacting with ECM, revealing unexpected regimes in myosin aggregation [143]. A different approach utilized perturbation theory to study actin treadmilling, an essential mechanism for contraction [143].

Previous models, while investigating important components of blood clot contraction, do not provide a detailed fibrin structure with multiple cells. For this reason, the restructuring of the clot due to the presence of multiple platelets within the fibrin network is not well understood. In the model presented in Chapter 3, the forces exerted by individual filopodia in addition to the spatial distribution of platelets, the number of filopodia per platelet, and the reaction of individual platelets to different substrate stiffness within the clot are considered. This provides a detailed model where macroscopic properties emerge due to the local alterations in individual platelets. Specifically, the formation of platelet aggregates

are measured along with, the restructuring of the fibrin network, and the colocalization of fibrin on the surface of the individual platelets.

1.2.3 Chemical Reaction Network Models

Chemical reaction networks are commonly modelled using mass action kinetics or flux based methods. Any reaction network can be described in general by Z reactions and M metabolites, where a single enzymatic reaction from some molecular species A to another species B at forward and reverse rates k_1 and k_{-1} can be written as:

$$\nu_{\mathrm{A},1}\mathrm{n_A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \nu_{\mathrm{B},1}\mathrm{n_B}. \tag{1.1}$$

Here v_A and v_B represent the unsigned stoichiometric coefficients that maintain mass balance. According to the law of mass action, the flux through the reaction is defined as $J_{net} = k_1 n_A^{v_A} - k_{-1} n_B^{v_B}$. Systems of reactions are frequently solved through time using ODE methods until a steady state is reached, i.e. $\frac{dn_A}{dt} = -k_1 n_A + k_{-1} n_B$ is small. This approach has several issues in formulating the ODEs since the values of the rate constants are not known [41]. Even when rate constants are known, they can vary several orders of magnitude across species [174]. For this reason, flux-based methods have been popularized. In this formulation, the time derivative for the vector of metabolic species, n, is written as follows: $\frac{dn}{dt} = SJ$, where SJ represents the multiplication between the stoichiometric matrix S and vector J [101, 70, 73]. The method known as flux balance analysis (FBA) [147], has been widely used to calculate the flow of metabolites by representing individual reactions in matrix format via stoichiometry, constraining fluxes, and subsequently calculating the largest possible flux via linear programming given $SJ = 0.0$. In essence,the net flux is determined

by Kirchhoff's flux law in conjunction with an optimization criteria. This is related to mass and energy consevevation as shown in Qian et al. [158]. Since the metabolic networks that are used in FBA are represented by genome scale models [49], and these models have been extensively developed over the last 20 years [48, 52, 54, 53, 163, 141, 189], FBA has grown in popularity [114]. Even more, its usability in the form of available packages in MATLAB and Python programming languages has made it the primary choice for many modellers [110, 166, 170, 95, 84].

Even so, flux-based methods are not able to produce statistically satisfactory information regarding metabolite concentrations, the direction of reactions, or the energy requirements of the pathway [101, 70, 73]. While Beard et al. [12] developed a method to relate forward and reverse flux to the free energy for a reaction at steady state, there is still a pressing need for quantitative metabolimics data and experimental measurement of thermochemical properties of the pathways. These are necessary for FBA to accurately identify reactions which limit flux and produce realistic values of chemical potentials without assuming non-equilibrium reactions are reversible [45, 139].

Another methodology for modeling chemical reaction networks makes use of discrete state based, continuous time Markov processes. While not as popular as FBA, this approach is still widely utilized [225, 172, 131, 161, 216]. Here, states of a Markov process are simulated, where states consist of counts of molecular species along with some partition function or master equation. This provides a rigorous mathematical framework in which the relations between closed systems, detailed balance, and microscopic reversal can be studied [196, 159]. However, open systems and nonequilibrium steady states (NESS) are of primary

modeling interest since life occurs away from equilibrium [157]. In this setting as well, the use of Markov models enables detailed calculations of system variables, such as the joint density function of metabolite concentrations [71]. Some have utilized this method to study nonequilibrium steady states, and have succeeded in simulating larger pathways [190].

More recently, Cannon et al. [30], developed a method similar to FBA wherein the law of mass action is reformulated as a maximum entropy problem, and consequently the time derivative can be written as follows: $\frac{dn}{dt} = SJ = S(KQ^{-} - K^{-}Q)$, where K is the ratio of the rate constants, Q is the reaction quotient. This implies that KQ^{-} and $K^{-}Q$ are the thermodynamic odds of forward and reverse reactions, respectively. Given enough fixed metabolites, a similar minimization problem can be posed as in FBA, but instead of altering fluxes, metabolite concentrations can be directly perturbed, which in turn alters the reaction quotient. This method is further developed in several recent works [29, 19]. While allowing for direct computation of metabolite concentrations, these methods also are able to make use of a wealth of metabolic control analysis [39, 13] (MCA) techniques to understand reaction control. MCA is a technique which measures the control imposed by a metabolic network on the flux through each reaction [169, 162, 85, 55, 59, 87, 69]. Additionally, rate constants for the maximum entropy solution can then be determined algebraically, and non-maximum entropy solutions can then be found by sampling rate constants randomly around the maximum entropy solution [30].

1.2.4 Reinforcement Learning

Reinforcement learning is frequently posed as a midpoint between supervised and unsupervised learning. While supervised learning provides inference from labelled data to

via classification or regression, unsupervised learning provides infer patterns from unlabeled data. At the heart of each learning task are function approximators such as linear representations, support vector machines [40], decision trees, and neural networks [171, 108]. These approximations provide the necessary flexibility to represent high-dimensional data sets or fit complex underlying patterns. Reinforcement learning (RL), however, is unlike both other types of learning in that it trains an agent to learn good or optimal behavior with respect to some environmental queue. For this reason, RL has gained enormous attention, especially when applied to systems where the number of decisions needed is practically countless for machine purposes. While more brute force tree search methods can be used by combining massive parallelism with alpha-beta pruning, as in the chess program Deep Blue [25], RL methods perform well with much less computational resources.

Learning from sequential decision making with some type of feedback from the environment in the form of rewards was first formalized by Bellman [16], but has inspired many others. The RL method utilizes the formalism of a Markov decision process [15], which is commonly represented as a tuple $\{S, A, P, R\}$, where S represents the set of possible states, A represents the set of possible actions, P represents the transitional probabilities between states, and R represents the reward function or environmental feedback. Many people, including Sutton [186] and Barto [11] continued developing these ideas in his work that provides the backbone for many algorithms behind recent advancements and fundamentals [185].

The most basic form of learning is performed by iteratively updating a policy function, $\pi : S \rightarrow A$, using environmental feedback [185, 200]. A policy returns the action,

or probability distribution of actions that should be taken given that the agent is in some state. This is done by associating some value with respect to each state. For example, a state, s_t can be given value by the infinite horizon formulation: $\sum_{k=0}^{\infty} E_\pi[\gamma^k r_{t+k}|s_t]$, which is the expected diminishing return to be received under the current policy, π, with $\gamma \in [0, 1]$ being the discount factor. Each reward, $r_t = R(s_{t-1}, a, s_t)$, represents the feedback from moving into state s_t from s_{t-1} after taking some action a. Policies can be learned directly [185] or indirectly as in temporal difference algorithms [183, 186] or Q-learning [199]. Each method makes use of the fundamental recursive property, first understood by Bellman [14].

Frequently, the dimension of the state or action space requires flexible function approximators [108]. If the transitional probabilities, P, are known, then many multiple versions of bootstrapping algorithms such as n-step SARSA [184, 167] can be used to learn optimal policies via value function iteration. Q learning [199] was recently popularized because it does not require knowledge of transitional probabilities, nor does it need to learn them through a model. It side-steps this important feature and thereby led to many techniques such as actor-critic [97], deep Q-learning [133], double Q-learning [67], and dueling network architecture [171]. These techniques, and many more, have been applied to perform tasks such as optimize sepsis treatment [96], play the game of Go [177, 178], perform complex robotics tasks [146], optimize control of power grids [5], and many others games [133]. This is in part due to the success of the algorithms as well as their usability via packages such as Tensorflow [1] and PyTorch [150].

Chapter 2

A Multiscale Model of Fibrin Fibers

2.1 Introduction

2.1.1 Biological Relevance and Importance of Fibrin Mechanics

Fibrin is an end product of blood clotting and a proteinaceous polymeric component of intra- or extravascular blood clots that form at the sites of injury. Fibrin provides jelly-like blood clots with elasticity that is important for their biological functions. As a major component of extracellular matrix (ECM), it also participates in various cellular processes, including adhesion, migration, proliferation and differentiation, wound healing, angiogenesis, inflammation, and others. Formation of the fibrin gel in blood vessels in vivo is one of the key events halting bleeding (hemostasis) and impairing blood flow by obstructive pathological blood clots (thrombosis) [205]. The mechanical response of fibrin to stresses generated by blood flow, deformations of the pulsating vessel wall, during platelet-driven clot contraction, diaphragmatic excursions, and gut motility determine the course and outcomes

of thrombotic and hemostatic disorders, such as heart attack and ischemic stroke. In addition, fibrin is also employed in biomedicine as a versatile biomaterial of unique mechanical properties utilized in a variety of clinical and laboratory applications, including hemostatic fibrin sealants and scaffolds for tissue engineering [119]. Despite the vital clinical implications of the biomechanics of fibrin, both as the skeleton of blood clots and thrombi and as a widely used biomaterial, the structural mechanisms underlying the mechanical properties of fibrin remain largely unknown.

2.1.2 Multiscale Fibrin Mechanical Properties and their Structural Basis

Fibrin is a hydrogel built of a biopolymer that self-assembles to form porous 3D filamentous networks with mechanical properties substantially different from synthetic polymers [78, 135]. The mechanical responses of fibrin gels to shear, tensile, and compressive loads are known to exhibit a highly nonlinear response known as strain-stiffening [21, 207, 122, 93]. This bulk nonlinear behavior has a number of structural mechanisms at various spatial levels spanning six orders of magnitude, including molecular unfolding, interactions within and between individual fibers, spatial rearrangement of the filamentous network, and other mechanisms that are not fully understood [120].

3D fibrin networks consist of branched fibrin fibers resulting from self-assembly of fibrin monomers and oligomers further stabilized via intermolecular covalent isopeptide bonds. Formation of these bonds (fibrin crosslinking) is catalyzed by an active transglutaminase that circulates in the blood as an inactive proenzyme named clotting factor XIII; the active form of factor XIII is designated factor XIIIa. Following elongation of fibrin oligomers, they undergo lateral aggregation mediated, in part, by long unstructured polypeptide "arms"

named αC-regions. The αC-regions comprise 400-residue-long C-terminal portions of the α-chains extending from fibrin molecules and connecting fibrin oligomers (protofibrils) within a fiber. Multiple highly flexible αC-regions stretched between protofibrils can interact with each other to form αC-polymers followed by covalent crosslinking by factor XIIIa, making the αC-polymers stable and contributing to the elastic properties of fibrin. The entire cross-linked fibrin clots or individual fibers are known to show extremely large extensibility, quadrupling their length before breaking [21, 93]. Mechanical properties of individual fibrin fibers were probed experimentally by atomic force microscopy (AFM) [122, 121] and optical trapping [38], revealing the elastic moduli to be 1.7 and 14.5 MPa for uncross-linked and factor XIIIa cross-linked fibers, respectively. Studies of tensile properties of whole fibrin clots and individual fibers at large deformation revealed the molecular unfolding of fibrin molecules, suggesting the molecular mechanism of fibrin extreme extensibility [21, 122, 74, 156].

The forced molecular unfolding in fibrin(ogen) has been demonstrated and analyzed both experimentally and computationally [21, 22, 117, 62, 222, 118]. Forced elongations of several regions of the fibrin molecule have been shown to be a potential source of fibrin's extensibility, namely unfolding of the γ chain nodules [62], extension of the α-helical coiled-coil connectors [22, 117, 62, 222, 118], and unraveling of the unstructured αC-polymers within a fiber [37, 9, 50, 75]. Micromechanical properties of individual fibrin fibers were shown to determine the macroscopic characteristics of the fibrin network. In particular, the high strain of single fibers was shown to affect the overall strength of the network [121, 74, 9]. Even though aspects of the tension behavior of fibrin networks has been studied in detail, it is still

not clear how microscopic structural characteristics, such as fiber connectivity, alignment, and fiber-fiber cohesion impact the mechanical response of the entire network, and how stress propagation within the network is affected by changes in its structural organization.

2.1.3 Synopsis

Newly formed oblique cohesive pairwise interactions (crisscrossing) between fibers were found to play a role in the viscoelastic response of compressed fibrin networks [93, 88]. Moreover, increased fiber-fiber cohesion is hypothesized to contribute to the mechanical response of fibrin gels under tensile load. In this work, this hypothesis is tested and proven by using a combination of experiments and model simulations obtained with a newly developed and calibrated three-dimensional (3D) computational nonlinear model of a fibrin network.

Using 3D confocal microscopy of stretched fibrin gels, the fiber-fiber crisscrossing at relatively low strains, which provided a basis for including these structural features in a model for the mechanical response of fibrin to stretching deformations are visualized directly. The main novel feature of the model introduced here is that it considers cohesive pairwise interactions between fibrin fibers, physically calibrated mechanical bending and stretching properties of individual fibers, as well as describes structural properties of the entire fibrin network, including fiber connectivity, fiber and node density, and fiber alignment. The model is used to quantify the mechanical effects of fiber-fiber cohesion on the stress-strain response of the fibrin network, as well as spatial redistribution of the internal stress and network structural changes after the external tensile load is imposed on the network. Model-based simulations of fiber alignment and densification, as well as computed stress-strain relations for stretched fibrin clots, are shown to be in good agreement with

experimental data. Altogether, the results obtained demonstrate that increasing fiber-fiber cohesion provides an important structural mechanism of fibrin clot stiffening in response to tensile load as it increases the fibrin network connectivity and enhances the distribution of stresses through the network.

2.2 Methods

2.2.1 Confocal Microscopy and 3D Reconstruction of Stretched Fibrin Networks

2.2.2 Formation and Stretching of Fibrin Clots

Fibrin clots were prepared from pooled human citrated platelet-poor-plasma (PPP) by mixing with $CaCl_2$ (26 mM final concentration) and thrombin (0.3 U/mL final concentration, Sigma-Aldrich, USA). To visualize fibrin in a fluorescence microscope, Alexa-Fluor 488-labeled human fibrinogen (Molecular Probes, Grand Island, NY) was added to plasma samples (0.08 mg/mL final concentration). Plasma clots were formed for 2 hours at 37°C in 4×60mm cylindrical plastic tube with its internal surface pre-coated with 4% (v/v) Triton X-100 in phosphate-buffered saline (PBS) to prevent adhesion of fibrin fibers to the walls. The clots were slipped out of the tubes into 1× PBS and cut by a surgical knife into 30 mm-long pieces that were held in the grips of a home-built stretching device and extended in uniaxial tension to different extents. The clots in the stretched state were fixed by immersing them for 20 minutes into 2% glutaraldehyde dissolved in PBS, rinsing them three times with PBS, and cutting them into 4mm × 4mm pieces used for imaging.

2.2.3 Confocal Microscopy and 3D Reconstruction Methods

Samples of fixed fluorescently labeled fibrin clots (non-deformed or stretched) were applied to a coverslip and imaged in the middle, away from the clamped ends of the clot using a Zeiss LSM880 laser confocal microscope with Plan Apo 40× (NA1.2) water immersion objective lens to acquire a series of high-resolution 212.5 μm × 212.5 μm × 20 μm z-stack images using the Airyscan mode. The distance between z-stack planes was 0.2 μm. An Argonne laser with a 488-nm wavelength was used for fluorescent imaging. 3D structures of fibrin clots were reconstructed from confocal microscopy z-stack images and analyzed using Imaris software.

2.2.4 Model Description

In this work general bead and spring modeling approach is utilized to simulate single fibrin fibers, a method commonly used to model a single polymer or a system of polymers [109, 43, 90]. Specifically, a previously developed model [88] is extended by introducing interactions between individual fibers to represent fiber-fiber cohesion and potential fiber bending to study their biomechanical impacts on the fibrin network under stretching (Figure 2.1A). Though fibrin is a viscoelastic polymer, the elastic component is generally about an order of magnitude higher than the viscous component, although the viscous component increases at higher rates of deformation. In simulations and with parametrization of fibrin mechanics at quasi-static rates, the viscous properties or plasticity of fibrin networks were not addressed. Each fiber in a network is represented as an elastic segment between two nodes (branch points) containing a series of sub-nodes connected by springs. The network

branch points are referred to as main nodes while using the term sub-nodes to indicate the nodes dividing the interior of a fiber (Figure 2.1B). Furthermore the mass of the individual fibrin fiber is assumed to be lumped together at the main and sub-nodes in the model.

The sub-nodes along a single fiber are placed equidistant to each other in order to represent a uniform distribution of mass and physical properties of the fiber. Moreover, a fixed spacing of sub-nodes serves to equally distribute possible fiber-fiber cohesion sites and points of fiber bending (see 2.1B). Deformation is slowly applied to the network in order to keep the network at a quasi-equilibrium state. This state is achieved when the maximal normalized node speed is less than the width of a single fibrin fiber. A more detailed description of the quasi-equilibrium state is described in Appendix A. The dynamics of the fibrin network is formulated in terms of Langevin equations for each i^{th} node of the network as follows,

$$m_i \ddot{x}_i = F_i - \eta \dot{x}_i + F_i^B, \tag{2.1}$$

where F_i is the deterministic force, $\eta \dot{x}_i$ is the viscous dampening force, and F_i^B is the Brownian force satisfying the Fluctuation-Dissipation Theorem [100]. The inertial term is neglected and the system (2.1) is discretized using a Forward Euler scheme,

$$x_i^{n+1} = x_i^n + \frac{dt}{\eta} F_i^n + F_i^{n,B}, \tag{2.2}$$

where the superscripts n and $n+1$ refer to the vector quantitates for the i^{th} node at subsequent time points n and $n+1$ and dt is the time step. η represents the drag coefficient. A detailed description of the model forces, coefficients, and calibration is continued in Section A.2.2 of Appendix A.

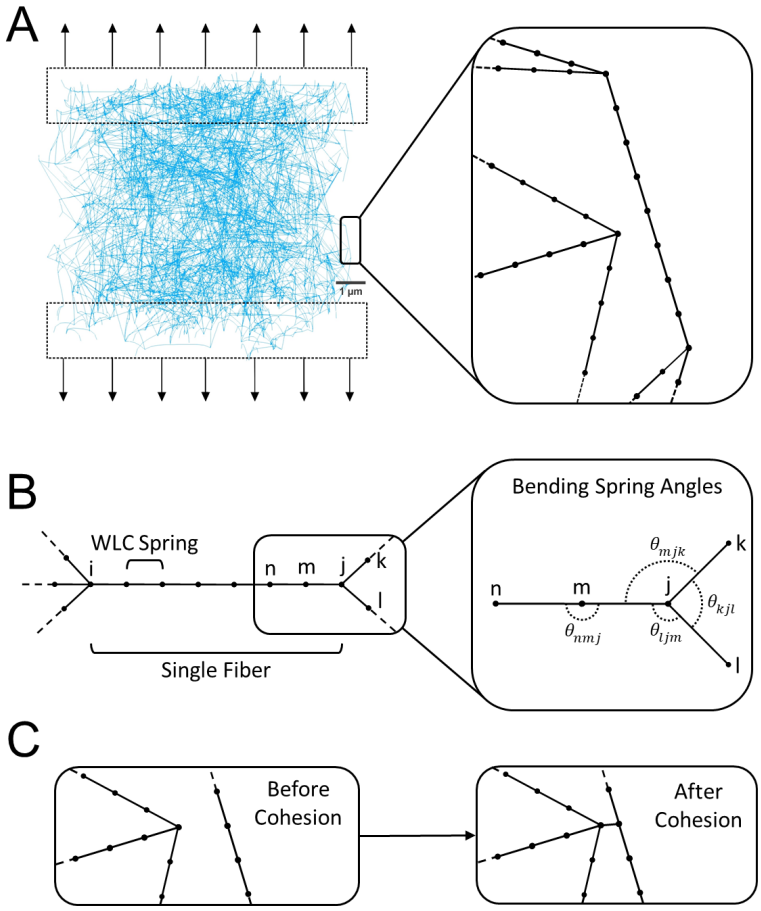

Figure 2.1: Schematics of in silico fibrin clot stretching experiment and description of the individual fibrin fiber model. (A) Representative image of a three-dimensional fibrin network used in simulations. Black arrows show the direction of applied force. Zoomed section shows the detailed fibers. (B) Spatial discretization of a single fiber using 6 interior sub-nodes connected by WLC springs. i and j are the main nodes of the fiber. Zoomed section illustrates bending springs in a fiber using angular springs between nodes (including *non-rotating* both main and sub-nodes). (C) Representation of two fibers and a cohesive interaction between them.

2.2.5 Statistical Analysis

Statistical analysis of experimental data was performed using GraphPad 7 software. Statistical significance between groups of samples was determined using Mann-Whitney U-test with a 95% confidence level. Results are presented as mean±standard deviation, unless otherwise indicated. 90 nodes of three independent donors were analyzed for network connectivity quantification.

In model simulations, statistical analysis was done by performing 5-10 simulations on independently generated fibrin networks. For model validation, Section 2.3.2, 10 simulations were performed while 5 simulations were used for model prediction statistics in the remainder of the chapter. 3-degree polynomial curves are used to fit the stress-strain data in Figures 2.3A, 2.4A, and 2.6A by performing a non-linear least squares optimization from the Python library SciPy [144]. The shaded error bars are calculated using the standard deviations in each polynomial coefficient given by $\sigma_i = \sqrt{cov_{ii}}$ where cov_{ii} is the diagonal element of the estimate covariance matrix of the optimized values of the estimated coefficients. In Figures 2.5 and A.3, for each independent simulation, averages of node types for different levels of network strain are calculated. At each point of network strain, the mean and standard deviation are calculated from the resulting 5 average values. The error bars represent a single standard deviation from the mean.

2.3 Results

2.3.1 Fiber Cohesion and Bending in Stretched Fibrin Network

A 3D reconstruction of a hydrated fibrin network obtained using confocal microscopy revealed that non-deformed fibrin is an isotropic network of branched fibers, typically with three fibers joined at each branch point (Figure 2.2A), as observed in previous studies [88, 10]. To observe structural alterations of fibrin networks upon application of a unidirectional tensile load, a cylindrical fibrin clot prepared as described in Methods is stretched to 20% strain and imaged the altered fibrin network structure using high-resolution fluorescence confocal microscopy. Subsequent 3D reconstruction of the stretched fibrin network structure revealed several specific structural changes (Figure 2.2B). First, most fibers were oriented in the direction of network stretching, which is consistent with earlier observations [21, 207, 86, 58]. Second, some individual fibrin fibers were crisscrossed and formed cohesive contacts that were revealed by areas of increased fluorescence intensity (Figure 2.2B, yellow circles). The spots with increased fluorescence intensity confirmed that these fibers formed physical contacts between each other and not projected in separate planes.

There were usually four fibers radiating from each of the crisscrossing points, indicating that formation of the cohesive contacts results in an increase of the number of 4-degree nodes in the network. The angles of crisscrossing were variable but most of them deviated from 90°; therefore, these inter-fiber contacts were mostly oblique. Further quantification of the reconstructed non-deformed and stretched three-dimensional fibrin networks revealed that stretching by 20% increased significantly the fraction of crisscrossing points (4-degree nodes) and nodes with connectivity degrees higher than 4 by a factor of 2.3 and 10

respectively, while the fraction of regular branching points (3-degree nodes) decreased by a factor of 1.3 (Figure 2.2C). Decrease of the fraction of regular branching points with strain can be both due to increase of crisscrossing points as well as due to interaction of 3-degree nodes with individual fibers or network nodes leading to increase of the fraction of nodes with connectivity degree higher than 4. It was also found that the density of fibers increased 5-fold upon stretching at 20% strain (Figure 2.2D), implying that the pores between fibers were smaller and fibers became closer to each other. Finally, in the stretched clots (unlike what is observed in the non-stretched networks) a fraction of bent fibers (indicated by white arrows in Figure 2.2B) existed, suggesting the possibility of non-affine deformations of the fibrin network upon stretching [211, 206]. The latter points out the fact that application of theoretical modeling approaches which assume an affine origin of network deformations might not be valid in certain conclusions.

To the best of our knowledge, this is the first time that fiber-fiber cohesion and bending were directly visualized in fibrin networks under unidirectional external tensile load. These observations provide an experimental basis for a hypothesis that fiber-fiber cohesive crisscrossing may contribute to the mechanical response of the network. To test this hypothesis, a novel multi-scale fibrin network model is used to computationally test the impact of fibrin fiber's cohesive crisscrossing on the mechanical behavior of stretched fibrin networks.

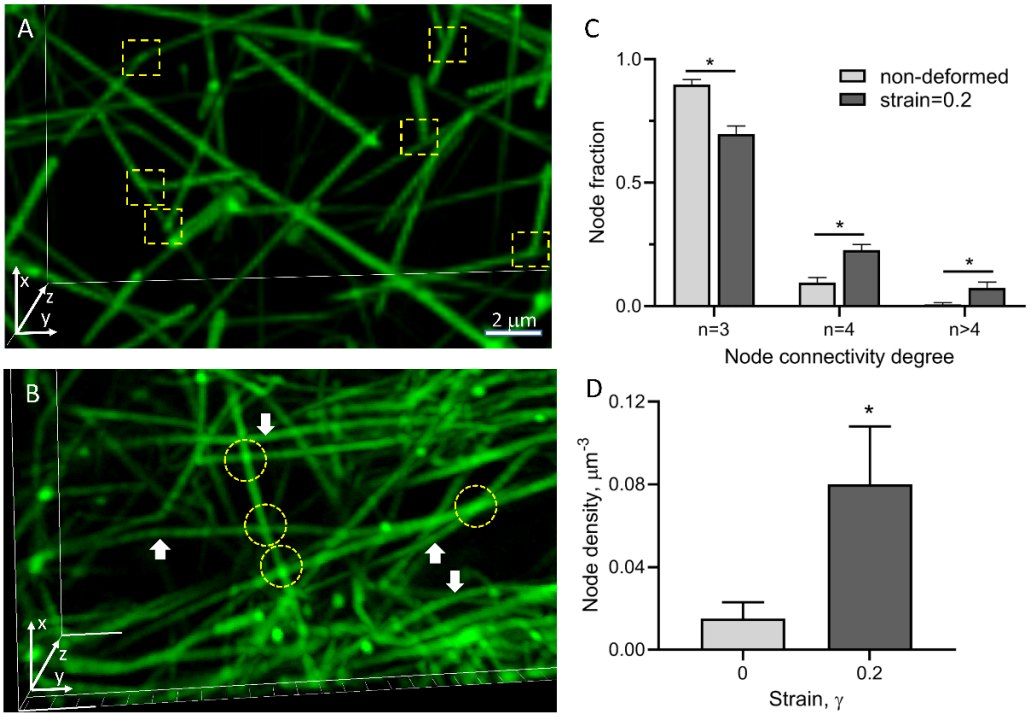

Figure 2.2: Visualization and quantitative structural analysis of non-deformed and stretched fibrin clots. (A) Reconstructed confocal microscopy-based 3D image of a fluorescently labeled non-deformed isotropic fibrin network. (B) Fibrin network stretched to a 20% strain in the y-direction. Fiber branch points are shown by yellow squares in (A) and fiber-fiber cohesion points are indicated by yellow circles in (B). Bent fibers are indicated by white arrows. (C,D) Quantification shows a node connectivity degree (C) and node density (D) in non-deformed fibrin clots and in the same clots stretched to a 20% strain (90 nodes analyzed in each fibrin clot prepared from 3 independent donors). The results in (C) and (D) are presented as $\mathrm{M} \pm \mathrm{SEM}$. A two-tailed Mann-Whitney U test, $^*P < 0.05$.

2.3.2 Model Validation

To validate the computational model, mechanical and structural responses of a high density fibrin network (30 fibers/μm^3) are simulated by slowly applying a unidirectional stretching load and compared the response with experimental results [21]. The model was calibrated by fitting spring parameters and network structural features to microscopic experimental data as shown in Section A.2 of Appendix A. In Figure 2.3, a comparison between model simulation results and macroscopic experimental data is presented. Comparison between experiment and simulation shows that simulation results fall within or near the range of experimentally observed network behavior in several different data fields. Simulations were performed by applying a linearly increasing external force to the opposite ends of a three-dimensional cubic fibrin network with volume 1000 μm^3. The mechanical model correctly reproduces the linear and nonlinear parts of the stress-strain response curve measured in the experiments and here shown as blue dots (Figure 2.3A). As the strain increases from zero to 100%, the stress increases linearly, which is followed by a non-linear 545% increase in stress up to 0.236±0.063 MPa at 220% strain (Figure 2.3A).

To quantify changes in fibrin fiber orientation, average fibrin fiber alignment was calculated as $cos(2\theta_i)$, where θ_i is the smallest angle between the i^{th} fiber and the axis to which the force is applied. An angle of $\theta = 0.0$ occurs when the fiber is parallel to the direction of applied force with a resulting value of $cos(2\theta) = 1.0$. Similarly, an angle of $\theta = \pi/2$ occurs when the fiber is perpendicular to the direction of applied force. The perpendicular alignment results in a value of $cos(2\theta) = -1.0$. For fibers with no preferred alignment, the value of $cos(2\theta)$ will take a value between 1.0 and -1.0 respectively. The

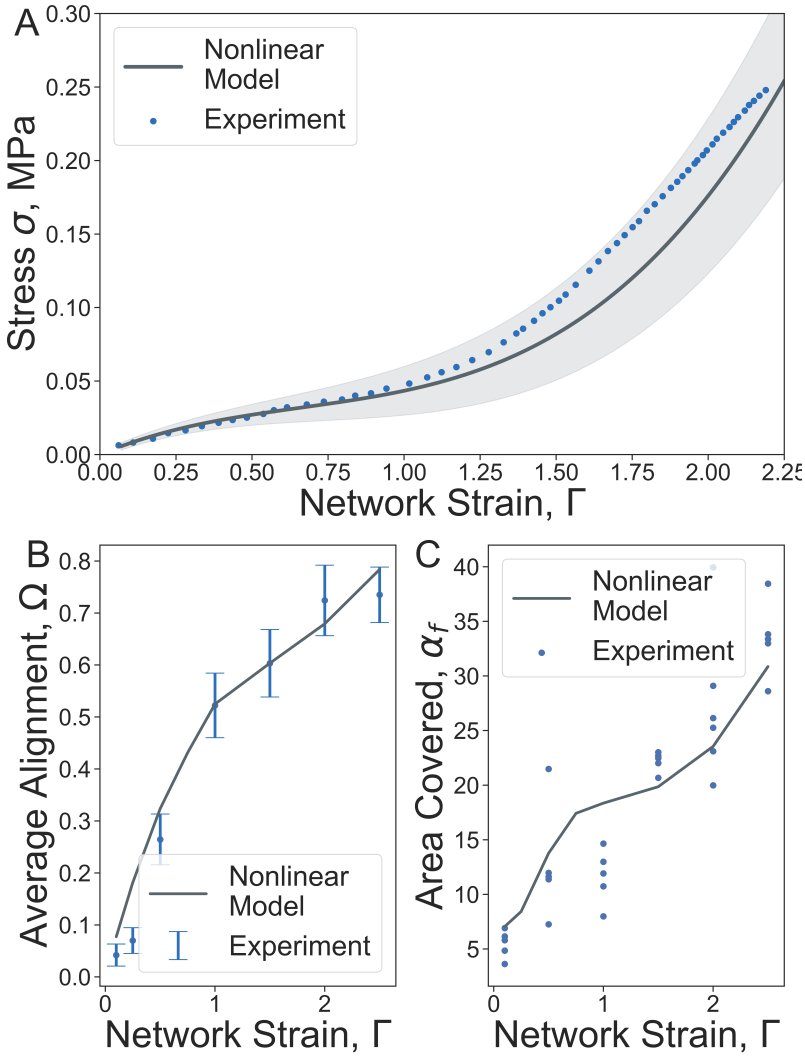

Figure 2.3: Mechanical and structural changes of fibrin networks under uniaxial stretching. (A) Characterization of the mechanical and structural changes in the fibrin network in terms of the unidirectional tensile stress-strain response ($M \pm SD$, n=10). (B) Average fiber alignment in the central 50% region of the network. (C) Percent of the stretched fibrin clot cross-section area covered by fibers. The solid line shows the nonlinear model simulation results, circles are experimental data [21].

average orientation over all fibers in a randomly generated network, $cos(2\theta)$, will be near 0.0. As the network is stretched, the fibers re-orient and the average orientation increase, but remains less than 1.0. In figures, the average alignment is referred to as Ω. Model simulations reveal a monotonic increase of the average fiber alignment with the increase of strain providing very good agreement with the experimentally measured orientation of fibrin fibers in stretched clots [93] (Figure 2.3B). The average alignment calculated from simulations is shown by dark lines while the experimental data mean and standard deviations are shown by blue dots and blue intervals respectively.

To find the percentile of the area covered by individual fibers, the middle cross-section area, S_c, of the clot is considered, and the relative area of fibers' cross section is calculated as $S_f = n_f s_f$, where n_f is the total number of fibers in the cross-section, and $s_f = \pi d_f^2$ is the fiber cross-section area. The calculated experimental percentage of the area covered by fibers from five different experiments are shown by blue dots, while the simulation average is shown by black solid line (Figure 2.3C). The simulations correctly predict the increase of the relative area covered by fibers, $\alpha_f = S_f/S_c$, revealing a 4-fold increment in α_f at 220% strain (Figure 2.3C).

2.3.3 Fiber-Fiber Cohesion as a Mechanism of Network Stiffening

To assess the effect of fiber-fiber interaction on fibrin network mechanics, the stress-strain response curve was calculated for 5 different networks stretched up to 300% strain in the presence and absence of fiber-fiber cohesion. In Section 2.3.2, a high density (30 fibers/μm^3) simulated fibrin network was used. However, experimental work has shown wide ranges for fibrin density [21, 93]. Therefore, a different fibrin network density of 5

fiber/μm^3 is utilized to evaluate the impact of fiber cohesion on the mechanical response of the network to stretching loads. The stress-strain relation, $\sigma(\Gamma)$ consisted of a linear portion for $\Gamma < 1$ and a non-linear stiffening regime for $\Gamma > 1$ (Figure 2.4A). In contrast, in the absence of fiber-fiber cohesion, the stress-strain curve did not reveal stiffening and the tension in the network was 5 times lower than in the network with cohesive fibrin fibers at a maximum strain of 300%.

To evaluate the effect of cohesion on the strain of individual fibers within the networks, the fiber strain distribution for the two types of networks were calculated (Figure 2.4B). The fiber strain distribution of non-cohesive fibers revealed a distinct peak at 0.1 strain, with the fiber strain ranging from -0.5 to 2, where negative strain value corresponds to fiber compression. In cohesive networks, the peak of fiber strain shifted to 0.17 strain, and the number of fibers with strain greater than 100% increased more than 4-fold. Thus, the cohesion of fibers resulted in dramatic changes in the fiber strain distribution, leading to an increase in the strain of individual fibers in the network.

The change in stress distribution is due to network remodeling at the microscale. To quantify the change in network structure, differences in node degree were calculated as the networks were stretched. 3, 4, and greater than 4-degree nodes were counted for different network strains. Figure 2.5 shows a decrease by 1/3 in 3-degree nodes and a corresponding 2-fold and a 7-fold increase in 4 and >4-degree nodes, respectively, which is shown to be in a good agreement with the experimental data presented in Figure 2.22C.

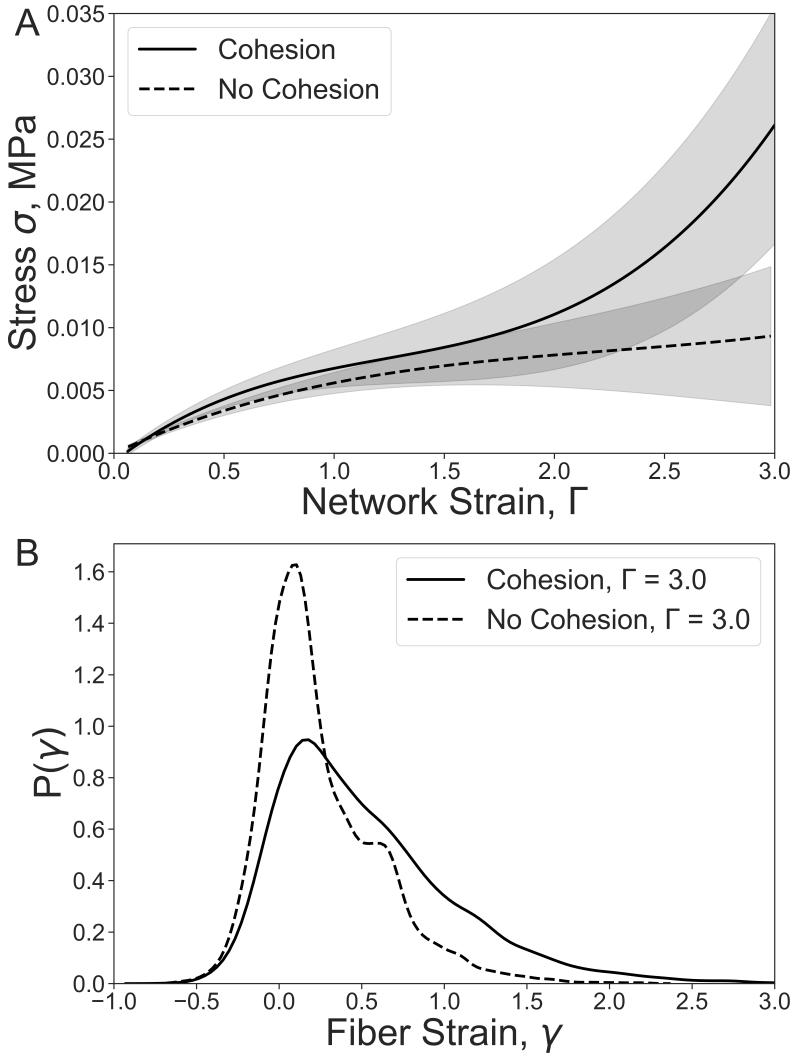

Figure 2.4: Impact of fiber-fiber cohesion on the tensile stress-strain response and fiber strain distribution in stretched fibrin clots. (A) Tensile stress-strain responses (M $\pm$ SD, n=5) for fibrin networks of cohesive and non-cohesive fibers of density 5 fiber/μm^3. (B) Corresponding strain distributions of individual fibers at the fibrin network strain $\Gamma = 3$.

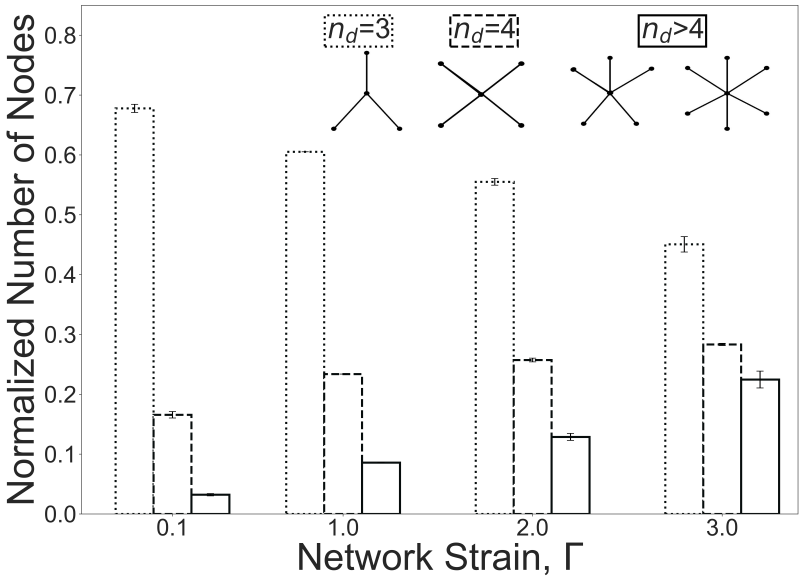

Figure 2.5: Strain-dependence of the node connectivity for fibrin networks with cohesive fibers. The fiber density $\rho_f = 5$ fibers/μm^3. Dotted, dashed and solid lines in the columns correspond to the network connectivity degrees of 3, 4, and > 4, respectively.

2.3.4 Impact of Fibrin Network Density on Stretched Fibrin Clot Mechanics

Tensile deformation of a fibrin clot has been shown to result in a strain-dependent increase of network compaction manifesting as rising fiber and node density (Figure 2.22D). Fiber-fiber cohesion is correlated with network density such that high density increases the chance of fiber-fiber cohesion while low density decreases the chance of fiber-fiber cohesion.

To quantify exactly how changes of the fibrin clot density impact clot mechanical response to unidirectional stretching, the stress-strain curves were calculated for simulations for different clot densities (1-15 fiber/μm^3) for strains from 0 to 250% (Figure 2.6). Increase in fibrin network density from 1 to 15 fiber/μm^3 results in a significant 25-fold increase in the network stress at the maximal network strain of 250%. Denser networks of 15 fiber/μm^3

demonstrated a distinct transition from a linear to a nonlinear regime at a network strain of 1.5. In contrast, sparse networks of low densities (< 5 fiber/μm^3) did not reveal stiffening behavior.

Analysis of fiber strain distributions revealed that there were more fibers at high strain ($\gamma > 1.0$) in the networks of higher density, i.e. the mechanism of fiber cohesion is coupled to fiber density. As the network density increased from 1 to 15 fiber/μm^3, the percent of high strained fibers ($\gamma > 1.0$) increased by 200%, while the percent of low-strained fibers ($0 < \gamma < 0.8$) dropped by 15%. Thus, an increase in fibrin network density results in stiffening of the fibrin clot under tensile load accompanied by an increase in local strains within the clot.

2.3.5 Fiber Cohesion Increases Fiber Alignment in Stretched Fibrin Clots

As fibrin networks are stretched, individual fibers begin to align along the axis of applied stress. When force is applied to fibers at the edge of the clot, they immediately align with other neighboring fibers over the clot volume. Because fiber-fiber cohesion increases the connectivity of the stretched network (Figure 2.5) it triggers quicker and more efficient propagation of fiber alignment than that observed in non-cohesive networks. To evaluate the effect of fiber cohesion on fiber alignment upon stretching of fibrin clots, the spatial distribution of fiber orientation is analyzed in cohesive and non-cohesive networks undergoing stretching deformations for low (1 fiber/μm^3) and intermediate (5 fibers/μm^3) fiber densities. First, the average fiber alignment in the central 50% region of clots stretched to 300% strain in the presence of fiber cohesion was found to increased by 22% and 43% when

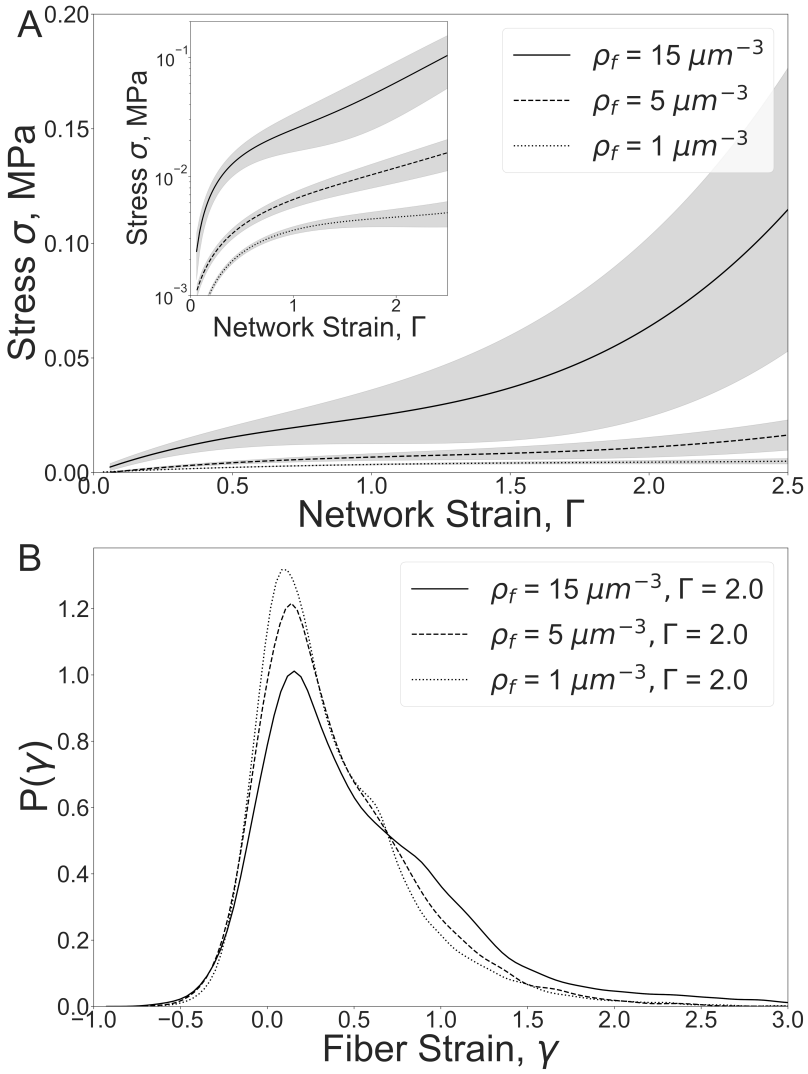

Figure 2.6: Fibrin clot mechanical response and fiber strain probability distributions for different fiber densities. (A) Tensile stress-strain responses ($\mathrm{M} \pm \mathrm{SD}, n = 5$) for fibrin networks of different fiber densities, ρ_f. The inset image shows the stress on a logarithmic scale. (B) Individual fiber strain distributions for the network strain $\Gamma = 2$.

compared to non-cohesive networks with fiber densities of 1, and 5 fibers/μm^3, respectively (Figure A.4 A, B). Furthermore, in higher density networks of 15 fibers/μm^3 the average alignment increased by 57% at 250% strain. Next, cohesion was observed to induce a more uniform spatial alignment of fibers. In the presence of cohesion, fibers in the central portion of the clot (20% of the clot length) were found to be 45% more aligned in the direction of stretching than fibers in clots without cohesion at a concentration of 1 fiber/μm^3 clot at 300% strain. Moreover, in the cohesive networks with higher fiber densities of 5 fibers/μm^3, the same estimates yielded larger spatial increase in average fiber alignment of over 200% (Figure 2.7).

Taken together, these simulation results suggest that cohesion of fibrin fibers enhances their alignment in clots undergoing stretching, which positively correlates with the increase of the initial clot density.

2.4 Discussion

Fibrin is a hydrogel with unique mechanical properties that determine behavior of blood clots and thrombi in the highly dynamic intra- and extravascular environment. In addition to the pathophysiological implications of fibrin clots, the rapidly developing field of bioengineering uses fibrin gels as a versatile biomaterial with tunable mechanical properties. Despite the great increase in our knowledge regarding the mechanics of fibrin, much about the structural mechanisms of fibrin's viscoelasticity remains unknown. In this chapter, a combined experimental and computational approach was utilized to further study the structural mechanics of fibrin networks. This approach analyses experimental structural

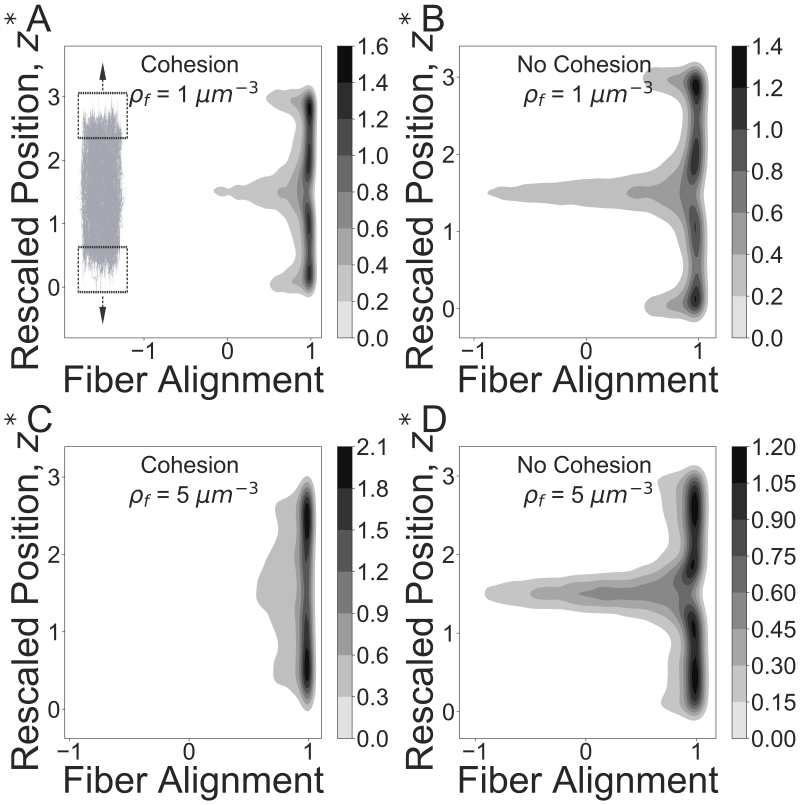

Figure 2.7: Impact of cohesive fiber-fiber interactions on fiber network alignment. Joint distributions for network alignment and network strain are shown for cohesive and non-cohesive networks at 300% strain for fiber density $1/\mu m^3$ (A, B) and $5/\mu m^3$ (C, D). Simulated networks are shown alongside the joint density distribution to illustrate the corresponding location of aligned fibers. The Y-axis represents the scaled position along the length of the fiber network, $z^* = z/z_0$. z_0 is the original clot length and z ranges over the current network length from bottom ($z = 0$) to top ($z = 3z_0$). Greyscale at each point corresponds to the relative number of fibers oriented along the direction of the strain.

data extracted from fiber networks and uses it for calibration of a novel computational model based on basic physical principles that incorporates fiber-fiber cohesion during tensile deformation of fibrin networks.

Mechanical properties of fibrin have been studied across multiple spatial scales and under various types of deformation, including stretching, compression, and shear pressure [182, 194, 192]. One of the most general mechanical properties of fibrin is a non-linear mechanical response known as strain-stiffening behavior which has been explored extensively but still does not have a comprehensive structural explanation.

Although the structural basis for fibrin mechanics has been analyzed at various spatial scales, from the sub-molecular up to macroscopic levels, the mechanical behavior of the whole fibrin gel is governed largely by the properties of single fibers and their ensembles [120]. Stiffening of individual fibers plays a crucial role in the large scale elastic response of the entire network by equitably distributing the strain through the network [121, 76]. Another important mechanism of strain-stiffening is the reorganization of the network architecture [76, 204]. This architectural reformation includes fiber densification and bundling [21], but other structural alterations have been proposed, such as the natural inclination of fibers to adhere to one another [195, 102].

In this chapter, a novel stiffening mechanism based on fiber-fiber cohesion in stretched fibrin networks is successfully tested. First, high-resolution optical microscopy are performed that followed structural changes of stretched fibrin networks at the level of individual fibers. In addition to the well-known fiber alignments along the direction of strain, the observed structural alterations included the formation of fiber-fiber cohesive con-

tacts named fiber crisscrossing and fiber densification (Figure 2.2). One consequence of the crisscrossing was shortening of fibrin fibers and their partial bending. Similar structural changes were revealed earlier in compressed fibrin networks. They were much more pronounced and their contribution to the non-linear mechanical response was straightforward [93]. However, such fiber-fiber cohesion has never been associated with tensile deformations of fibrin networks and the role of this structural mechanism has never been evaluated.

To establish mechanistic and quantitative relationships between the fiber-fiber cohesion and elasticity of individual fibers and bulk fibrin gel, a novel computational model was developed and calibrated based on comparing simulation results with the experimentally obtained stress-strain profile, alignment, and density data. (The structure of the novel biologically calibrated model and its distinction from previous modeling approaches for studying fibrin mechanics are summarized in Section 2.2.4.)

New experimental data presented in this work (Figure 2.2) suggests that fiber cohesion also correlates with clot stretching in addition to fiber alignment and densification. The computational model simulations were used to test the hypothesis that cohesion provides a novel mechanism for fiber alignment and densification. To ensure that this behavior was due to fiber cohesion, the simulated network node connectivity was quantified and shown to vary as a function of network density and strain (Figure 2.5 and A.3). Specifically, as network density increased from 1 to 15 fibers/μm^3, a more rapid increase in nodes with connectivity degrees > 4 (Figure A.3C) was shown to be associated with a more rapid increase in average alignment (A.4A-B) and densification (A.4C-D). Using the model, the following structural and mechanical relationships are established that could not be explored experimentally at

this time. (i) Fiber cohesion is shown to induce strain-stiffening in fibrin networks (3.3). This is supported by simulations in the absence of cohesive fiber-fiber interactions, which did not reveal stiffening of the networks and failed to predict the experimental data (Figure 2.4). (ii) Fiber cohesion is shown to alter the network structure (3.3). Since this has not yet been quantified in experiments, network alterations are predicted and quantified in terms of node degree (Figure 2.5). (iii) Networks experience more alterations in the presence of high fiber density (3.4). This adds to more rapid network stiffening (Figure 2.6), which is associated with increased fiber alignment (Figure 2.7). (iv) The change in fiber alignment due to a change in fiber density is predicted (3.5). This mechanism is further validated in Appendix A by showing that fiber strain is distributed more equitably throughout the network (Figure A.5).

Notice that these experiments and model have some objective limitations. It is not feasible at this time to experimentally study fibrin clots in which fiber-fiber crisscrossing is selectively prevented. Therefore, the impact of fiber cohesion has been confirmed by using a computational model. Although these simulations were done in the overdamped regime using one-way coupling with fluid, the model permits a two-way coupling extension for network-fluid interactions. Such two-way coupling may be important for modeling large compressive deformations of blood clots which are not considered here and are beyond the scope of the current study. The model can be also extended in the future to account for other types of network deformations including twisting behavior of fibers. These limitations do not affect the main conclusions of the chapter and these results clearly show that newly formed cohesive fiber-fiber interactions make an important contribution to the mechanical response of fibrin

networks under stretching deformation. Remarkably, model simulations incorporating fiber-fiber cohesion correctly describe the general behavior of real fibrin clots under unidirectional tension [93], including the stress-strain response as well as fiber orientation and fiber density changes. This suggests that fiber-fiber cohesive bond formation is an important mechanism contributing to fibrin clot stiffening and alignment.

Although the molecular mechanism of fiber-fiber cohesive bond formation is not known, various covalent and non-covalent bonds might form between fibers in contact. Covalent binding can be potentially mediated by isopeptide bonds formed by factor XIII as the fibers are brought in contact. However, a recent study [102] showed that blocking factor XIII by an inhibitor did not change the interaction force between fibrin fibers suggesting a non-covalent interaction between the fibers in contact, probably mediated by αC-regions, which allow for interactions between fibrin protofibrils and fibers. Vos et al. [195] estimated the force between two interacting fibers to be 760 pN, strong enough to maintain junction integrity, which supports the modeling assumption that fibers brought in contact form irreversible bonds perhaps originating from multiple non-covalent interactions.

Chapter 3

A Model of Platelet Mediated Clot Contraction

3.1 Introduction

3.1.1 Biological Relevance

Blood clots are complex biological structures that are initiated by injuries to blood vessels. They form as a result of the coagulation cascade which involves fibrinogen, thrombin, platelets, and coagulant factors (e.g. factor XIIIa), and. The structure of a clot is composed of fibrin network, platelets, red blood cells, and plasma. Clot contraction, mediated by activated platelets [191], is essential for proper wound healing and restoration of blood flow in an injured blood vessel [208]. Complications in the contraction process directly affect the healing of the injured vessel and the stability of the clot which result in complications such as heart attacks, strokes, and deep vein thrombosis (DVT). Even so, the interdepen-

dence between clot contraction and individual platelet activation is not well understood. In particular, individual platelet filopodia dynamics are fundamental to understanding clot contraction in addition to a multitude of other biological processes, including tissue healing and development [209, 180, 219, 218, 98], phagocytosis [176], and cancer development [160]. Despite these implications, a quantitative study of the emergent properties of aggregates of platelets has not been performed.

3.1.2 Importance of Multiscale Modeling of Blood Clot Deformation and Contraction

Many characteristic features of blood clots cannot be measured experimentally at this time. For example, platelet activation involves the formation of filopodia for each platelet, the attachment to other nearby platelets and fibrin, the retraction and resulting forces of such filopodia, and the reaction to the substrate stiffness [34, 24] of individual platelets [105]. In order to investigate these features independently, a multiscale computational approach is needed. Modelling approaches offer the ability to study each component contributing to the thrombus structure separately and quantify the resulting impact on the clot contraction.

This work aims to provide insight in how platelets control the clot contraction environment and influence the overall dynamics of the process through alterations in the mechanical forces exerted by their filopodia. In particular, a computational approach is employed in which platelet activation can be directly controlled by altering the platelets: number of filopodia, magnitude of force exerted, and reaction to fiber stiffness. This allows for quantification of the roles of different platelet activation scenarios and comparison of the

emergent properties of the clot such as platelet aggregation and the rate of clot contraction.

3.1.3 Synopsis

The hypothesis that platelet mechanosensing of substrate stiffness in addition to filopodia heterogeneity contributes to increased fibrin colocalization, clot stability and distinct contraction phases is investigated. This hypothesis is tested using a computational model that includes multiple platelets and their dynamic interaction with a detailed fibrin network. The model is validated using data from previous in vitro experimental results [92].

The effects of different levels of activation in platelets have been difficult or impossible to study in experiments due to the many interdependent factors controlled by the activation phenomenon. The computational model presented here implements each component as a separate submodel, allowing for the analysis of the interdependence between platelet activation and quantitative contraction dynamics. Specifically, how changes in platelet activation, expressed through the number of filopodia, level of filopodia force, and the manner in which platelets respond to fibrin strain are investigated. The ability of platelets to individually, and collectively, form well-defined contraction phases observed experimentally is quantified.

3.2 Methods

3.2.1 Formation of Fibrin Networks Mixed with Platelets

3.2.2 Platelet-rich Plasma

Blood was drawn by venipuncture from healthy volunteers not taking aspirin or other medications affecting platelet function for at least 10 days. Informed consent was obtained in accordance with a protocol approved by the University of Pennsylvania Institutional Review Board. Platelet-rich plasma (PRP) was prepared from whole blood drawn into 3.8% trisodium citrate ($9:1v/v$). To obtain PRP, the blood treated with citrate was centrifuged at 210×g at 25°C for 15 minutes. The supernatant plasma containing platelets was transferred to another sterile tube.

3.2.3 Formation of PRP-clots

To label fibrin and platelets, PRP samples were preincubated with Alexa-594 labeled human fibrinogen and calcein for 10 min at 37°C. To induce clotting, PRP samples were recalcified with $CaCl_2$ (29 mM final concentration) and mixed with thrombin (1U/mL, final concentration). A sample was immediately applied on a microscopic glass surface of a PELCO cell culture dish inside the environmental chamber of a confocal microscope. The glass surface was precoated with 4%(v/v) TritonX-100 to prevent attachment of fibrin to glass.

3.2.4 3D Image Reconstruction of Confocal Microscopy Images

Platelet-rich-plasma clots were imaged using Zeiss LSM710 scanning confocal microscope with Plan Apo x40 (NA1.2) water immersion lens to acquire serial 35 mm-thick z-stack images of the clots during the course of contraction (40 min). The distance between slices of the z-stack images was 0.8 mm; each image was taken at 1024 x 1024 pixels resolution. Fluorescently labeled fibrin and platelets were excited using 594-nm wavelength helium-neon and 488-nm wavelength argon laser beams.

3.2.5 Image Analysis of Platelet Filopodia

The number of filopodia generated by individual platelets has been manually measured in the acquired confocal microscopy z-stacks of contracting PRP clots. To calculate the number of filopodia, each picture was divided into 50 square domains of 10 micron-thick z-stacks [173], and the number of filopodia for individual platelets was counted.

Experimentally derived filopodia counts averages over 4 experiments containing an average of 100 platelets are presented below. Resulting distributions (Figure 3.1) are shown in light gray and are fitted with a lognormal distribution (μ=1.85, σ=0.27). Simulated platelet filopodia counts are calculated by sampling the lognormal distribution. Averaged simulated platelet filopodia counts are shown in dark grey (Figure 3.1). The filopodia distribution obtained from this experimental data is used in the model to determine the number of filopodia for each single platelet in a simulated clot.

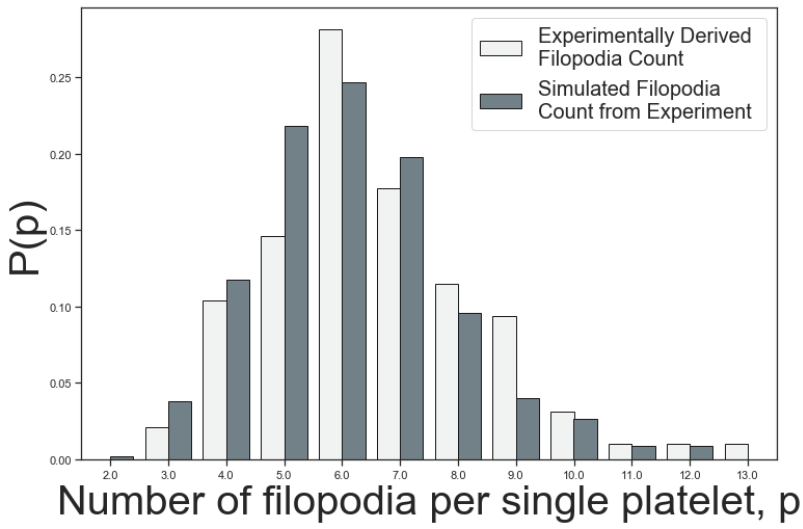

Figure 3.1: Probability density function of platelet filopodia counts for simulated (n=10) and experimentally (n=4) obtained platelet-fibrin meshworks.

3.3 Model Description

Here a coupling of the 3D model of fibrin network described in Chapter 2 with a sub-model of individual platelets pulling on individual fibers through individual filopodia is described. The fibrin network is represented by a mass and spring modeling approach which represents fiber stretching, bending, and fiber-fiber cohesive components (Figure 3.2B-C).

The platelet sub-model describes pulling, adhesion, and exclusion forces applied to the center of mass of each platelet (Figure 3.2A). Each platelet interacts with other platelets and fibrin fibers in several different ways. First, platelets exert pulling forces on other platelets and fiber nodes in the environment through filopodia. When an object (platelet or fibrin nodes) in the environment is near to the platelet surface, adhesion force is applied. Finally, a volume exclusion force ensures the platelets and environment do not overlap in the space they occupy. Platelets and fibrin dynamics are described in terms of

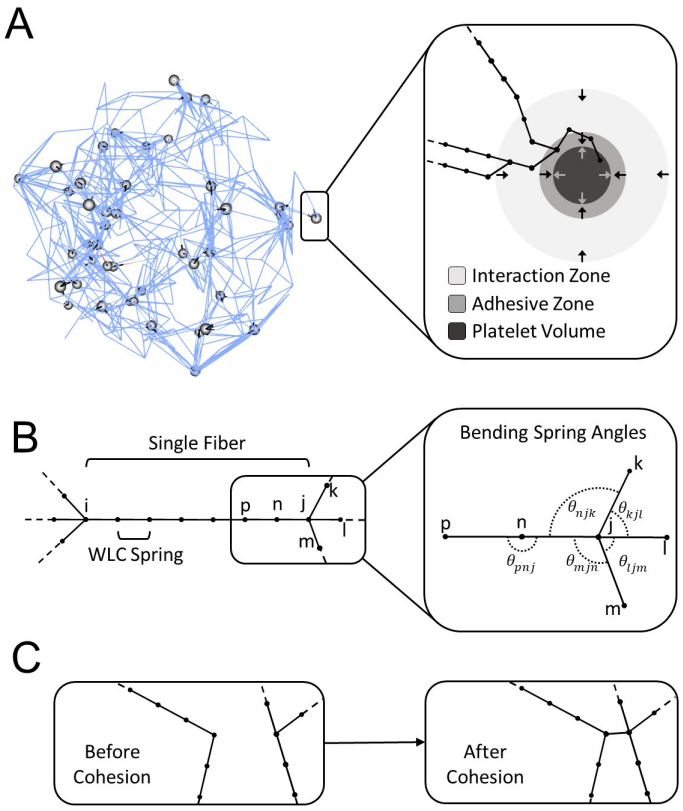

Figure 3.2: Schematics of in silico platelet contraction simulation and depiction of fibrin fiber model. (A) Left. Fibrin network (blue lines) and platelets (gray spheres). Right. Zones of influence of platelet forces: interaction zone (light gray), adhesive zone (gray); volume exclusion (dark gray). (B) Single fibrin fiber divided in subsections governed by Worm-Like-Chain springs and bending springs. (C) Formation of cohesive bond between sufficiently close fibers.

the following Langevin equations. Namely, motion of the center of mass of the i^{th} platelet is described by the following stochastic ordinary differential equation:

$$m_{p,i}\ddot{x}_{p,i} = F_{p,i} - \eta_p \dot{x}_{p,i} + F^B_{p,i}, \tag{3.1}$$

where $F_{p,i}$ is the deterministic force applied to the i^{th} platelet, $\eta_p \dot{x}_{p,i}$ is the viscous damping force, and $F^B_{p,i}$ is the Brownian force due to thermal fluctuation [100]. As in Chapter 2 an overdamped regime is assumed due to the low Reynolds number. The inertial term is therefore neglected and the system (3.1) is discretized in time using a Forward Euler scheme:

$$x^{n+1}_{p,i} = x^n_{p,i} + \frac{dt}{\eta_p} F^n_{p,i} + F^{n,B}_{p,i}, \tag{3.2}$$

where the superscripts n and $n+1$ refer to the vector quantities at time steps n and $n+1$ for the i^{th} platelet and dt is the time step. η_p represents the drag coefficient for a platelet with fixed volume in plasma. Application of the Stokes estimation, $\eta_p = 6\pi\mu r_p$ results in the value $\eta_p = 85.2\ nN.s.m^{-1}$. Similarly, the fiber nodes follow the same discretization as in Chapter 2:

$$m_{f,i}\ddot{x}_{f,i} = F_{f,i} - \eta_f \dot{x}_{f,i} + F^B_{f,i}. \tag{3.3}$$

Assuming fibers maintain a diameter of 100nm, the value of η_f is $3.77\ nN.s.m^{-1}$. The Brownian force in three dimensions is calculated using the Einstein relation between the diffusion coefficient in a medium and the temperature [132]. Because the mean squared displacement of a particle following Brownian dynamics is known to be $\langle x \rangle = \frac{6k_BT}{\eta} dt$ over a single time step, the random force on an individual platelet is calculated as:

$$F^{n,B}_{p,i} = \xi_i \sqrt{\frac{6k_BT}{\eta_p} dt}, \tag{3.4}$$

where ξ_i is sampled from a standard normal distribution. Random force are calculated similarly for fiber nodes but utilize the drag term, η_f.

For each platelet, the deterministic force, $F_{p,i}$, is generated by (1) pulling forces from filopodia, (2) adhesive forces and (3) volume exclusion forces. The total force can be calculated using the negative gradient of the energy, $F_{p,i} = -\nabla E_i^{PLT}$, where E_i^{PLT} is the energy associated to the i^{th} platelet and is separated into three parts,

$$E_i^{PLT} = E_i^{FIL} + E_i^{ADH} + E_i^{EXC}, \tag{3.5}$$

with E_i^{FIL}, E_i^{ADH}, and E_i^{EXC} representing the filopodia, adhesion, and volume exclusion energies. The explicit formulation of each energy is written as follows:

$$E_i^{FIL} = F_p \frac{1}{2} x^2, \tag{3.6}$$

$$E_i^{ADH} = F_0 \frac{1}{2} x^2, \tag{3.7}$$

$$E_i^{EXC} = \epsilon \left(\frac{\sigma^{12}}{x^{12}} - 2\frac{\sigma^6}{x^6}\right), \tag{3.8}$$

where $x = x_i - x_j$ is the distance between the i^{th} platelet and the j^{th} platelet/fiber, $\sigma = \frac{2}{2^{1/6}} f_r r_p$, and $\epsilon = F_0^2$. The calibration and choices of parameters are described in Section A.2.3.

3.3.1 Filopodia Contractile Forces

Each individual platelet is modeled as a spherical volume with multiple filopodia of length r_{fil} by which it exerts pulling forces on the fibrin network (Figure 3.3). Individual filopodia are modelled as forces exerted from the platelet center of mass onto fibers in the environment. Unless, otherwise stated, the number of filopodia for each simulated platelet

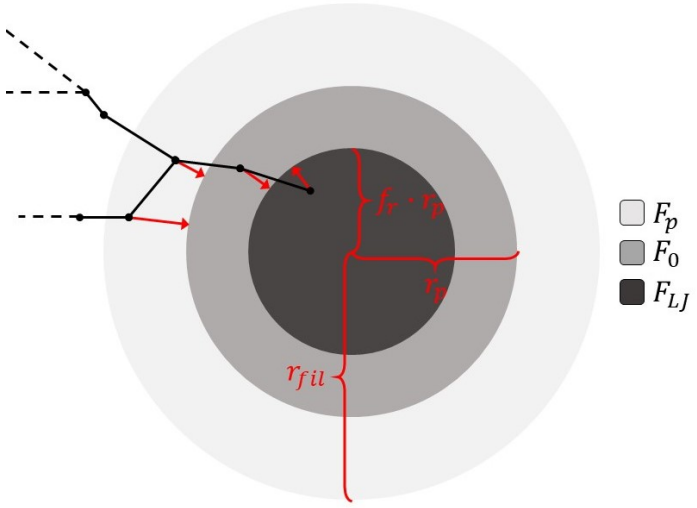

Figure 3.3: Schematic of individual platelet interacting with fibrin fibers in the clot environment. Different forces (red arrows) are applied to fibers depending on the varying distance from the platelet center of mass. Fibers are pulled towards the platelet at a force, F_p, in the interaction zone (light gray). Adhesion force, F_0, is applied to fibers on the platelet surface (gray), and a Lennard-Jones force, F_{LJ}, is used for volume exclusion (dark gray). Individual filopodia, modelled as forces, act by pulling fibers in the environment outside the platelet volume but within the range of the filopodia, r_{fil} (light grey annulus). Once a fiber is pulled within the platelet volume of radius r_p, the fiber is held on the surface by an adhesive force, F_0 (grey zone annulus). Fibers are held out of the central platelet volume (dark grey circle) using a Lennard-Jones potential.

is extracted from the simulated distribution described in Section 3.2.5 (see also Figure 3.2). Force is applied from the i^{th} platelet at a given position, $x_{p,i}$, to other j^{th} platelets and/or fibers in the environment at position x_j in the same transversal direction observed in Kim et al [92]. Typical applied force directions are shown as red arrows (Figure 3.3). Nearby platelets and fibers with respective radii $r_p = 1.13\ \mu m$ and $r_f = 0.5\ \mu m$ are pulled with a magnitude, F_p, if the i^{th} platelet has an unused filopodia and the j^{th} platelet/fiber is within range, i.e. the distance $d = ||x_{p,i} - x_j||_2^2$, is such that $2r_p < d < r_{fil}$ in the case that the j^{th} point represents the center of mass of a platelet, or $r_p + r_f < d < r_{fil}$ when the j^{th} point represents a fiber node. Model simulations were performed for different magnitudes of

pulling force, F_p, and under assumptions of its dependence on the fiber strain being constant or provided by different functions. Specifically, the average strain of the fiber being pulled, $\bar{\gamma}$, was incorporated into the magnitude of force via linear (Section A.2.3), constant, and discontinuous step function. The step function is based on evidence in Qui et al. [160] where the magnitude of forces exerted by platelets on the environment were observed to triple after the stiffness of the substrate is increased. Given an initial magnitude of force, F_0, the strain dependent magnitude pulling force can be expressed as follows:

$$F_p = \begin{cases} F_0, \bar{\gamma} < \gamma_{th} \\ 3F_0, \bar{\gamma} \geq \gamma_{th}, \end{cases} \tag{3.9}$$

where γ_{th} represents the fibrin stiffness threshold. As the i^{th} platelet approaches the j^{th} platelet/fiber, the distance can decrease such that $d < r_p + r_f$ or $d < 2r_p$ respectively. In this case, a surface adhesion force (Section 3.3.2) is applied to the j^{th} platelet/fiber. The filopodia belonging to the i^{th} platelet is then free to pull on a platelet or fiber that is already connected to the j^{th} platelet/fiber, simulating the hand-to-hand pulling previously observed [94]. If no such platelet or fiber exists, the filopodia is left free to connect with other platelets and fibers whose distance is less than r_{fil} from the center of mass of the i^{th} platelet.

3.3.2 Platelet Surface Adhesion and Volume-exclusion

During clot contraction, each platelet pulls fibers towards its center of mass and subsequently attaches the fibers to the adhesive surface. To model adhesion, a fraction of the platelet radius, $f_r = 0.9$, is denoted as an adhesive zone wherein a force, F_0, directed towards the platelet center of mass at location $x_{p,i}$ is continually applied to all other platelet centers of mass and fibers at points x_j such that the distance, d, is within the adhesive zone,

i.e. $2r_p f_r < d < 2r_p$ or $f_r r_p + r_f/2 < d < r_p + r_f/2$ for platelets and fibers, respectively. The adhesion annulus is shown in dark gray (Figure 3.3). The magnitude of force applied to the fiber or platelet at location x_j is F_0. Once $d < 2r_p f_r$ for platelets or $d < f_r r_p + r_f/2$ for fibers, a volume exclusion force is applied in the opposite direction to the adhesive force, namely $x_{p,i} - x_j$. The volume is illustrated by the dark circle (Figure 3.3A). The exclusion is implemented using a standard Lennard-Jones potential based force $F_{LJ} = \epsilon(\frac{\sigma^{12}}{x^{13}} - \frac{\sigma^6}{x^7})$, where $x = x_{p,i} - x_j$.

3.4 Discussion

Individual platelet activation first results in change in the number of filopodia occurring on the cell surface [107] which later determines the size of platelet aggregates, and the force exerted on the surrounding fibrin network by individual platelets. Platelets have been observed to dynamically respond to variations in the clot microenvironment, such as alterations in substrate stiffness and locally applied forces [160, 105, 136]. While individual cellular reactions may vary, general consensus reveals a correlation between increased substrate stiffness and increased expression of integrin αIIbβ3, platelet spreading and contractile force [160]. Here, several different responses of platelets due to changes in the stiffness of the surrounding environment are tested. The model will be validated using several different metrics and making comparison to experimental data from Kim et al [92]. The metrics utilized are platelet-colocalized fibrin, f_p, fibrin densification, f_d, and change in fibrin densification, Δf_d. When applicable, the superscript t is used to denote time in each metric.

To quantify the amount of platelet colocalized fibrin at time t, the amount of fibrin

volume within distance r_p for each individual platelet position, $x_{p,i}$. Since the diameter of a fiber is assumed to be 0.1 μm, the volume of an individual fiber segment of length 0.3 μm is $V_f = 0.3 \cdot \pi r_f^2$. Therefore, the amount of platelet compacted fibrin is determined by calculating $f_p^t = \{\sum_j V_f | \text{dist}(x_{p,i}, x_{f,j}) < r_p, \text{ for some } i\}$, where i and j range over all platelet and fiber indices, respectively. Similarly, the fibrin density is calculated by denoting a central cubic region within the clot of side length 10 μm, denoted σ. The fibrin density at a given time is then given by $f_d^t = \{\sum_j V_f | x_{f,j} \in \sigma\}$.

This model will be utilized to perform large scale simulations of clot contraction and perform predictive measurements not possible with current experiments. Specifically, this model allows the ability to simulate the affect of anti-platelet drugs thereby informing future drug design and wound healing techniques. Individual clots can be simulated according to parameters drawn from diseased patients so that quantification of stability can inform immediate patient risk based on clot composition. Finally, this model allows the translation of the current knowledge regarding individual platelets to larger communities. It is a continuous challenge to begin with the current knowledge of processes occurring within individual cells and translate that knowledge to complex communities.

Chapter 4

Reinforcement Learning Based Prediction of Enzyme Regulation

4.1 Introduction

While our understanding of regulation of transcription and post-transcriptional processes has blossomed in the past 25 years due to advances in high-throughput experimental technologies such as RNA expression, ChIP-Seq, and mass spectrometry-based proteomics, our understanding of post-translational regulation has advanced [39, 63, 165, 123], but not as rapidly or as far.

Recent breakthroughs include work in which mass spectrometry and NMR measured metabolite and protein levels, along with fluxes modeled from 13C isotope labeling were used with Michaelis-Menten kinetics to determine whether the predicted reaction fluxes matched fluxes modeled from isotope labeling data [63]. The correlation between predicted

fluxes were evaluated with and without regulation. If the match was better with regulation, then regulation was assumed. The work was a *tour de force* in that chemostat studies were used to carefully measure both absolute and relative metabolomics data while at the same time cover as much of the proteome as possible. In addition, Michaelis-Menten kinetic models addressed multiple levels of regulation. The payoff was not only predictions of which enzymes might be regulated, but also inferences about the regulating molecule.

In addressing possible scalability (or at least cost of experimentation) in the previously mentioned study, a similarly sophisticated informatics approach was used to develop a model of small molecule regulatory networks from curated databases of enzymes, integrate the regulatory network with a metabolic model of *E. coli*, and distill information on how substrates and inhibitors contribute to metabolic flux regulation [165]. Interestingly, this work did not find support for the common notion that reactions which are furthest from equilibrium are those that are most likely regulated.

Fifty years ago it was postulated that the purpose of post-translational regulation in metabolism is to either maintain a balance of the energy charge of the adenylate pool [7], or to control solvent properties [8]. Solvent properties have long been recognized as important determinants of cellular activity and function. Atkinson recognized that the maintenance of physiological concentrations of metabolites may well be the most pressing problem of metabolic control [8]. Metabolite concentrations are exponential functions of the standard chemical potentials but only a linear function of the rate constants. Consequently, metabolite concentrations are less a function of the reaction kinetics and primarily a function of a molecule's standard chemical potential, which varies over a small range across species

because solution conditions inside a cell also vary over a small range. Interestingly, the set of enzymes which are post-translationally regulated is relatively well-conserved across species as well [165], despite the fact that the rate constants for the same enzymes can vary dramatically [81].

In addition to metabolite concentrations *per se*, solvent capacity in the cell has recently focused on molecular crowding [224, 223] and the impairment of diffusion [149]. As a cell approaches equilibrium, the cell's cytoplasm can become glassy such that diffusion is limited. At the same time, control of metabolites through regulation of enzyme activities is no longer effective near equilibrium [85]. The equilibrium constant K for a reaction is the ratio of the exponent of the standard chemical potentials. Consequently, metabolite concentrations may potentially approach values determined by their standard chemical potentials in solution, which can be quite large for highly charged metabolites like fructose 1,6-bisphosphate and acetyl-coenzyme A. Not only will metabolite levels rise, but also less water will be produced by metabolism inside the cell. In *E. coli*, up to 50% of the bulk water is produced by metabolism [99]. Even away from equilibrium, cells clearly must regulate metabolite levels to prevent high concentrations that would be detrimental to diffusional processes necessary for life.

Here, the hypothesis that the post-translational regulation of enzymes is at least in part driven by the need to maintain the solvent capacity in the cell is investigated. Evaluation of this hypothesis is performed by comparing experimental metabolomics data with steady state concentrations predicted computationally from equations for reformulated mass action kinetics. Using quantitative metabolomics data as well as physical and biolog-

ical principles, metabolic control analysis and alternatively reinforcement learning are used to predict the control of activity required to bring metabolite levels down to observed values. Consequently, the machine learning results confirm that an optimal control policy can be formulated which directly achieves minimal regulation by efficiently reducing excessive metabolite concentrations.

The predictions agree with known regulation of central metabolism in model organisms. Moreover, these results show that regulated enzymes have free energies of reaction further away from equilibrium precisely because of the regulation, turning common wisdom about enzyme regulation upside-down. Instead of highly non-equilibrium reactions being the targets for regulation in metabolic pathways [112, 137], regulation results in reactions being much further from equilibrium than non-regulated reactions. Being further away from equilibrium than other reactions is an effect, not a cause, of regulation.

4.2 Methods

4.2.1 Convex Optimization Approach for obtaining Metabolic Steady State.

For a reversible chemical reaction, the reaction is described by the chemical equation,

$$\nu_{\mathrm{A},1}\mathrm{n_A} + \nu_{\mathrm{B},1}\mathrm{n_B} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \nu_{\mathrm{C},1}\mathrm{n_C} + \nu_{\mathrm{D},1}\mathrm{n_D}, \tag{4.1}$$

where A, B, C, D represents the molecular species, the concentrations are given by n_i, $i = \{A, B, C, D\}$, and $\nu_{i,j}$ represent the unsigned stoichiometric coefficients for each molecular species i in the forward and reverse reactions $j = \{1, -1\}$.

The law of mass action may be formulated in terms of chemical kinetics or thermodynamics. With respect to chemical kinetics, the law of mass action is expressed by the rate or net flux, $J_{net,1}$, of the reaction where the forward and reverse rates are proportional to the respective reactants,

$$J_{net,1} = k_1 n_A^{\nu_{A,1}} n_B^{\nu_{B,1}} - k_{-1} n_C^{\nu_{C,1}} n_D^{\nu_{D,1}}. \tag{4.2}$$

In this formulation, k_1 and k_{-1} represent the rate constants of the forward and the reverse reaction, respectively. On the other hand, the thermodynamic expression of the reaction utilizes the change in free energy, G, with respect to the extent of a reaction, ξ. The ratio of the respective reactants and products are combined to form the reaction affinity, $A_1 = \partial G / \partial \xi_1$, such that,

$$\begin{aligned} e^{A_1/R_g T} &= K_1 \frac{n_A^{\nu_{A,1}} n_B^{\nu_{B,1}}}{n_C^{\nu_{C,1}} n_D^{\nu_{D,1}}} \\ &= K_1 Q_1^{-1}, \end{aligned} \tag{4.3}$$

where $K_1 = k_1/k_{-1}$ is the equilibrium constant and Q_1 is the reaction quotient. Also, the analogous equation for the reverse reaction is the reciprocal,

$$\begin{aligned} e^{A_{-1}/R_g T} &= K_{-1} \frac{n_C^{\nu_{C,1}} n_D^{\nu_{D,1}}}{n_A^{\nu_{A,1}} n_B^{\nu_{B,1}}} \\ &= e^{-A_1/R_g T}. \end{aligned} \tag{4.4}$$

Note that Eqn. (4.2) is a purely kinetic description of the law of mass action, while Eqns. (4.3) and (4.4) are purely thermodynamic expressions. This results from the fact that the latter equations do not contain any information on the time dependence of the reaction. These formulations, however, are not mutually exclusive. Time dependence and thermodynamics can both be described in a single equation by factoring the opposing rate from each

term of Eqn. (4.2),

$$J_{net,1} = k_{-1}{n_C}^{\nu_{C,1}}{n_D}^{\nu_{D,1}}\left(\frac{k_1{n_A}^{\nu_{A,1}}{n_B}^{\nu_{B,1}}}{k_{-1}{n_C}^{\nu_{C,1}}{n_D}^{\nu_{D,1}}}\right) - k_1{n_A}^{\nu_{A,1}}{n_B}^{\nu_{B,1}}\left(\frac{k_{-1}{n_C}^{\nu_{C,1}}{n_D}^{\nu_{D,1}}}{k_1{n_A}^{\nu_{A,1}}{n_B}^{\nu_{B,1}}}\right)$$
$$= k_{-1}{n_C}^{\nu_{C,1}}{n_D}^{\nu_{D,1}}\left(K_1Q_1^{-1}\right) - k_1{n_A}^{\nu_{A,1}}{n_B}^{\nu_{B,1}}\left(K_{-1}Q_{-1}^{-1}\right), \tag{4.5}$$

where K_1 and K_{-1} are the equilibrium constants and Q_1 and Q_{-1} are the reaction quotients for reaction 1 and -1, respectively. Eqn. (4.5) is the Marcelin-de Donder equation [44, 30], which describes the forward and reverse reactions as being functions of the time independent odds of the reaction and the rate of change of the odds.

Given a metabolic model with Z reactions, M metabolic species, and N total particles, the flux through each reaction is formulated using Eqn. (4.5). In this work, the largest values of Z and M in a pathway are 29 and 47 respectively. If the rate of change of the odds are assumed to be equal and independent of concentrations, then the coupled reactions occur on the same time scale. Under these assumptions, the resulting equation for the j^{th} reaction is the Marcelin equation [128],

$$J_{net,j} = c_j\left(K_jQ_j^{-1}\right) - c_j\left(K_{-j}Q_{-j}^{-1}\right), \tag{4.6}$$

where c_j represents the time dependence of the reaction odds. Because the exponential family of distributions are always log-concave when counts are greater than or equal to zero, the energy surface on which the reactions occur is convex. This is achieved by expressing the reactions as functions of the reaction affinities via Eqns. (4.3) and (4.4),

$$J_{net,j} = c_j\left(e^{A_j/R_gT}\right) - c_j\left(e^{-A_j/R_gT}\right). \tag{4.7}$$

A vector of Z reaction fluxes $J = [J_1, ..., J_Z]^T$ can be determined from the M by Z stoichiometric matrix S and the M chemical potentials. The stoichiometric matrix consists of

elements $\gamma_{i,j}$, which are the signed stoichiometric coefficients for chemical species j in reaction i. The time dependence of the vector of counts $n = [n_1, ..., n_M]^T$ of chemical species is,

$$\begin{aligned} \frac{dn}{dt} &= SJ \\ &= S(KQ^- - K^-Q), \end{aligned} \tag{4.8}$$

where SJ is the matrix multiplication between S and J, $KQ^- = [K_1 Q_1^{-1}, ..., K_Z Q_Z^{-1}]^T$ is the vector of thermodynamic odds for the forward reactions, and $K^-Q = [K_1^{-1} Q_1, ..., K_Z^{-1} Q_Z]^T$ is the vector of thermodynamic odds for the reverse reactions. Without any constraints applied, Eqn. (4.8) will converge to an equilibrium solution, whether the equation is solved using ordinary differential equations or optimization methods. To obtain a non-equilibrium steady state, non-equilibrium boundary conditions must be applied. In this case, the non-equilibrium boundary conditions consist of boundary metabolite values representing the reactants and products of the overall chemical process that are held fixed. If there are M_V variable species and $M_B = M - M_V$ boundary (fixed) species such that $n = [n_1, ..., n_{M_V}, n_{M_V+1}, ..., n_M]^T$, then the stoichiometric matrix will contain a non-singular submatrix and Eqn. (4.8) will have unique solutions only if $M_V \leq Z$. The vector of counts n can be split into subvectors $n_V = [n_1, ..., n_{M_V}]^T$ and $n_B = [n_{M_V+1}, ..., n_M]^T$ such that $n = [n_V^T n_B^T]^T$. Likewise, the stoichiometric matrix can also be split along the rows representing metabolites to separate the entries for the variable metabolites from those for the boundary metabolites such that $S = [S_V^T S_B^T]^T$ where S_V is an M_V by Z matrix and S_B is

M_B by Z. The time dependence of each of the chemical species is given by,

$$\frac{dn}{dt} = \begin{bmatrix} \frac{dn_V}{dt} \\ \\ \frac{dn_B}{dt} \end{bmatrix} = \begin{bmatrix} S_V \\ \\ S_B \end{bmatrix} (KQ^- - K^-Q). \tag{4.9}$$

The optimization problem is to find n_V satisfying,

$$\left|\left|S_V(KQ^- - K^-Q)\right|\right|_2^2 = 0.0, \tag{4.10}$$

subject to the M_B boundary conditions. The optimization is carried out with a nonlinear least-squares approach using the Levenberg-Marquardt method [113, 130], and solves for the concentrations of the chemical species which makes up the reaction quotient, Q. When $S_V(KQ^- - K^-Q) = 0.0$, the optimization has found a kinetic steady state as well as a thermodynamically balanced state such that the net thermodynamic driving forces on all the reactions are equal for linear pathways, or for branched pathways, the net thermodynamic driving forces are proportional to the stoichiometry. If one is only interested in the thermodynamic properties, fluxes and concentrations at steady state, then there is no need to solve for the rate constants. Otherwise, rate constants can be back-calculated and used to solve for the non-equilibrium transient trajectories, for example, Eqn. (4.2). For example, setting $j = 1$, Eqn. (4.5) can be solved for $k_{\pm 1}$ as follows:

$$\begin{aligned} k_1 &= \frac{J_{1,net}}{n_A(1 - K_{-1}Q_{-1}^{-1})} \text{ and} \\ k_{-1} &= \frac{K_1}{k_1}. \end{aligned} \tag{4.11}$$

The kinetically accessible energy surface is not necessarily convex because of the introduction of the rate constants – each reaction now has its own time dependence.

The predicted concentrations from the optimization follow the multinomial Boltzmann distribution in which the concentration of each species is proportional to its standard

chemical potential, μ_i°, adjusted for aqueous solution at pH 7.0,

$$n_i \propto e^{-\mu_i^\circ / k_B T}, \tag{4.12}$$

subject to the constraints of the reaction stoichiometries and the non-equilibrium boundary conditions. The boundary conditions consist of fixed concentrations of environmental nutrients such as glucose and waste products such as CO_2, as well as some cofactors. Because the concentrations are distributed as a function of their standard chemical potentials in aqueous solution, the concentrations of highly hydrophilic charged species may be orders of magnitude above physiological values. For instance, concentrations of ATP or acetyl CoA may be on the order of ten molar or more. Such high concentrations would make the cytoplasm highly viscous such that diffusion would be slowed down significantly, and cellular activity would come to a halt. However, as will be shown, the concentrations can be brought into alignment with physiological values using enzyme activities determined from Metabolic Control Analysis [55, 169].

4.2.2 Metabolic Regulation: A Metabolic Control Theory Approach

Regulation is applied to reactions by changing the scalar valued activity of the j^{th} enzyme, $\alpha_j \in [0.0, 1.0]$, where activity values of 0.0 and 1.0 represent complete reaction regulation and no enzyme regulation, respectively. The activity for each reaction j is represented by a multiplier to the net reaction flux J_j such that,

$$J_j = \alpha_j (K_j Q_j^{-1} - K_{-j} Q_{-j}^{-1}), \tag{4.13}$$

and likewise,

$$\frac{dn}{dt} = \begin{bmatrix} \frac{dn_V}{dt} \\ \\ \frac{dn_B}{dt} \end{bmatrix} = \begin{bmatrix} S_V \\ \\ S_B \end{bmatrix} \left[\alpha \circ (KQ^- - K^-Q)\right], \tag{4.14}$$

where $\circ$ represents the Hadamard element-wise product. Since any reaction may be regulated, the state of the system can be described by the activity vector, α, steady state fluxes, J, and steady state metabolite concentrations n. Because the latter two state variables can be determined from a fixed set of activities via the optimization routine, system states can be defined simply by the activity vector α instead of the tuple (α, J, n).

In Metabolic Control Analysis (MCA), the sensitivity of a concentration n_i to the activity α_j of enzyme j is defined as the concentration control coefficient,

$$C^n_{i,j} = \frac{\partial \log n_i}{\partial \log \alpha_j}. \tag{4.15}$$

Concentration control coefficients can be used to determine how much to reduce the activities of an enzyme to bring the predicted concentrations into alignment with physiological values observed from experimental metabolomics assays. The detailed calculation is described in Appendix B. If concentrations n_i for a metabolite i have not been measured, then target values are assumed to be 1.0 millimolar, which ensures that concentrations stay reasonable even for metabolites whose concentrations have not been measured. When predicted values exceed the measured or target values, regulation is applied to reactions by changing the scalar valued activity of the j^{th} enzyme, α_j.

Which reaction to regulate is determined from examining the concentration control coefficients with regard to the metabolites whose concentrations are higher than is observed in experiment. The set of such metabolites is denoted as $M' = \{i | \tilde{n}_i > n_i\}$. An activ-

ity is then selected to be reduced based on the influence that the activity has on these concentrations,

$$C_j^n = \sum_{i \in M'} \max\left(C_{i,j}^n, 0.0\right). \tag{4.16}$$

Because activities are reduced from initial values of 1.0 (full activity), only $C_{i,j}^n > 0.0$ are considered in the sum so that reduction in activity correlates with reduction in concentration. A component cost function, L_i, is defined as the division of the predicted concentrations or counts to the measured concentrations or counts, $L_i = \log(\tilde{n}_i/n_i)$. In order to determine the point where steady state metabolite levels are 'in caliber', a stopping criteria function is utilized which returns a positive scalar if any $L_i > 0.0$ and returns zero once $L_i \leq 0.0$ for all i. The cost function is defined as follows:

$$L = \sum_{i=1}^{M} \max(L_i, 0.0). \tag{4.17}$$

The maximum of L_i or zero is used because the model only predicts metabolite populations that are free in solution, but the experimentally measured concentrations are in principle those that are both enzyme-bound and free in solution. Thus, concentrations from predictions are assumed to be 'in caliber' with experimental data if the predicted concentrations are less than or equal to experimentally measured concentrations ($L_i \leq 0.0$).

In practice, the activity that reduces the cost function, L, the greatest amount is chosen for regulation and is again determined using MCA. In MCA, the concentration control coefficient for metabolite i due to control by reaction j is defined by Eqn. (4.15). Consequently, the change in concentration or counts due to a change in activity of reaction j is,

$$d\tilde{n}_i = C_{i,j}^{\tilde{n}} \frac{d\alpha_j}{\alpha_j} \tilde{n}_i. \tag{4.18}$$

For metabolite i with predicted concentration $\tilde{n}_i$ and a target concentration of n_i, the estimated change in the costs, $\Delta L_{i,j}$, of metabolite due to a change in activity α_j of reaction j is:

$$\begin{aligned}
\Delta L_{i,j} &= \log \frac{\tilde{n}_i(\alpha_j)}{n_i} - \log \frac{\tilde{n}_i(\alpha_j - d\alpha_j)}{n_i} \\
&= \log \frac{\tilde{n}_i(\alpha_j)}{n_i} - \log \frac{\tilde{n}_i(\alpha_j) - d\tilde{n}_i(\alpha_j, d\alpha_j))}{n_i} \\
&= \log \frac{\tilde{n}_i(\alpha_j)}{\tilde{n}_i(\alpha_j) - d\tilde{n}_i(\alpha_j, d\alpha_j))} \\
&= -\log \frac{\tilde{n}_i(\alpha_j) - d\tilde{n}_i(\alpha_j, d\alpha_j))}{\tilde{n}_i(\alpha_j)} \\
&= -\log \left(1 - \frac{d\tilde{n}_i(\alpha_j, d\alpha_j))}{\tilde{n}_i(\alpha_j)}\right) \\
&= -\log \left(1 - C_{i,j}^{\tilde{n}} \frac{d\alpha_j}{\alpha_j}\right).
\end{aligned} \tag{4.19}$$

The change in total costs over all metabolites due to a change in activity of reaction j is calculated by summing over metabolites that are out of 'caliber' with respect to the experimentally observed concentrations. The total cost is calculated as follows:

$$\begin{aligned}
\Delta L_j &= \sum_{i \in \mathbb{M}} \Delta L_{i,j} \\
&= -\sum_{i \in \mathbb{M}} \log \left(1 - C_{i,j}^{\tilde{n}} \frac{d\alpha_j}{\alpha_j}\right),
\end{aligned} \tag{4.20}$$

where $\mathbb{M}$ represents the set of reactions able to be regulated or controlled. Finally, the question of which enzymes should be allowed to be control points must be addressed. Two approaches were taken with MCA: an unrestricted control approach in which any enzyme could be a regulator for any metabolite, and a restricted approach in which only enzymes whose immediate products exceeded the target values could be considered as a regulator. The latter is referred to as a local-control approach (MCA Local) since an enzyme's immediate products (and possibly other metabolites) are being controlled. Regulation is then

applied at the reaction maximizing,

$$\arg \max_{j \in \{1,\ldots,Z\}} (\Delta L_j)\,. \tag{4.21}$$

Once a reaction j is chosen, the activity α_j is changed by an appropriate amount (Appendix B). When all metabolite values are brought into agreement with experimental observations, rate constants can be determined, if desired, using Eqn. (4.11). Alternately, the influence of the activities can directly be incorporated into the rate constants. For example, given $j = 1$, the resulting rate constant is,

$$k_1 = c\frac{J_{1,net}}{\alpha_1 n_A(1 - K_{-1}Q_{-1}^{-1})}. \tag{4.22}$$

However, there is an important conceptual difference between solving mass action rate laws with parameters based on the approach provided by Eqn. (4.11) compared to Eqn. (4.22). While the former assumes regulation is needed to bring concentrations under control, the latter assumes no regulation is needed and control is hardwired into rate constants. The advantages of the former are two-fold: (1) under different nutrient conditions, enzyme activities can be altered to control metabolite concentrations; and (2) enzyme activities are adjusted away from the maximal entropy distribution only enough to bring concentrations into alignment with observed values, resulting in a more favorable total free energy of the system. A lower total free energy also would reduce the cost of replicating of metabolism. The actual balance between these two approaches will likely be a middle ground between the reliance on activity coefficients as opposed to rate constants. It is unlikely that enzymes can evolve such that the ideal rate constants, i.e. those implied by Eqn. 4.11, are possible for every reaction. Instead, rate constant values will be limited by constraints due to the physics of the catalytic process.

4.2.3 Exploring Regulation: A Reinforcement Learning Approach

The MCA method for bringing the predicted concentrations in alignment with observed concentrations is a deterministic approach based on an assumption that metabolite concentrations depend linearly on the enzyme activities. It is feasible that the assumption of linearity used in the MCA analysis (Appendix B) results in sub-optimal regulation. Optimal regulation has been hypothesized, based on empirical data, as regulation that maintains a high energy charge, defined in terms of ATP, ADP and AMP [7]. A less *ad hoc* definition of optimal regulation would be the maximization of the entropy production rate, which has also long been hypothesized as an objective of biological systems [124, 142]. Neither of these concepts are addressed in the MCA approaches discussed above. For steady state systems, the entropy production rate (EPR) is the negative of the free energy dissipation rate [29, 57],

$$\text{EPR} = -\frac{dG}{dt} = R_g T \sum_{j}^{Z} \alpha_j \left[r_j \log K_j Q_j^{-1} - r_{-j} \log K_j Q_j^{-1} \right]. \tag{4.23}$$

Given a goal of maximizing the EPR, it is not clear which MCA protocol above, if either, would maximize the entropy production rate. On one hand, the unrestricted MCA method uses less regulation and therefore often results in higher reaction fluxes, which would increase the EPR (Eqn. (4.23)). On the other hand, entropy is maximized when the value of the argument of the logarithms are distributed as uniformly as possible, which is the opposite of what occurs when a minimal set of enzymes are chosen to be regulated. In order to explore the regulation space more completely to investigate these issues, a machine learning method that avoids the linearity assumption by directly testing multiple future states is utilized and is directly rewarded for maximizing the EPR.

A Reinforcement Learning (RL) framework can be used to address decision prob-

lems that are otherwise combinatorially intractable. Even a small metabolic network may have on the order of 20-50 reactions. To explore the state space fully using the deterministic MCA approach, on the order of 100-500 decisions need to be made as to which reaction to regulate depending on the state of the system. The search space is then approximately between 20^{100} and 50^{500}, a number much too large to tackle by an exhaustive search or Monte Carlo approach.

The RL method (Figure 4.1) formulates the problem of regulation in terms of a Markov decision process [15], which is commonly represented as a tuple $\{S, A, P, R\}$, where S represents the set of possible states (enzyme activities for each reaction), A represents the set of possible actions (reactions to regulate), P represents the transitional probabilities between states, and R represents the reward function. Reinforcement learning is utilized to obtain an optimal regulation scheme by learning from delayed environmental feedback [185, 200]. Figure 4.1 illustrates how reactions are chosen using a policy function which returns the reaction to be regulated (action) given the current enzyme activities (state). Learning is performed by iteratively updating the state value function using environmental feedback (rewards) from solving the optimization routine.

In this framework, optimal regulation of a metabolic network requires that the EPR be maximized while satisfying a stopping criteria: $L = 0.0$. A diverse set of reaction regulation schemes represented by enzyme activity values, $\{\alpha_1, ..., \alpha_Z\}$, satisfy the stopping criteria, but each scheme results in a different EPR (Fig. 4.3C-4.5C, grey dots). Thus, a hybrid optimization-RL approach to iteratively searches for the best regulation scheme. (A hybrid simulation-RL approach can also be used.) States correspond to the value of enzyme

activities while actions correspond to regulating a specified reaction. Therefore, the state space, S, is defined as the subset of $\mathbb{R}^Z$ using range of each enzyme activity, $[0.0, 1.0]^Z$, and the action space as the set of reactions, $A = \{1, 2, ..., Z\}$. A subset of S where learning terminates is denoted as, $S_T = \{s \in S | L(s) = 0.0\}$.

Because regulating the j^{th} enzyme results in a deterministic step-size, $\Delta\alpha_j$, the resulting state is given by the following set of enzyme activities: $\{\alpha_1, ..., \alpha_j - \Delta\alpha_j, ..., \alpha_Z\}$. The goal of Reinforcement Learning is to learn an optimal policy, $\pi^* : S \to A$, which results in a regulation scheme that maximize some defined notion of rewards, $R : S \times A \times S \to \mathbb{R}$. In other words, learning the optimal policy corresponds to learning the regulation scheme for the chemical reaction network that results in the largest reward.

Each reaction that is regulated results in a scalar valued reward, or feedback, from the environment based on an action/regulation (Figure 4.1) that indirectly defines optimal regulation schemes. Each regulation decision alters the steady state metabolite concentrations, which are obtained from optimization or simulation of Eqn. (4.14), and used to calculate rewards using a loss function, Λ, specified by

$$\Lambda = \log(\sum_{i}^{M} \frac{\tilde{n}_i}{n_i}). \tag{4.24}$$

The formulation of Λ emphasizes regulation of reactions that affect metabolites which are furthest from being in caliber with experimental measurements.

The environmental feedback, or reward function, R, is constructed as follows:

$$R\left(s, a, s'\right) = \begin{cases} \frac{\Lambda(s) - \Lambda(s')}{\eta}, \ L\left(s'\right) \neq 0.0 \\ \text{EPR}\left(s'\right) \ + \ \frac{\Lambda(s) - \Lambda(s')}{\eta}, \ L\left(s'\right) = 0.0. \end{cases} \tag{4.25}$$

Intermediate rewards are calculated by the reduction in Λ scaled by a positive factor η.

Once a terminal state is found, the final reward consists of the final change in the scaled loss function as well as the entropy production rate calculated at the final state, $\mathrm{EPR}(s') + \frac{\Lambda(s)-\Lambda(s')}{\eta}$. Thus, the agent aims to both increase the value of $\mathrm{EPR}(s)$ for $s_t \in S_T$ while satisfying the constraint $L = 0.0$ and regulating as many reactions as is necessary.

Learning is conducted by iteratively updating the current policy function, $\pi : S \rightarrow A$, that determines the agent behavior. The policy function determines which reaction $j \in Z$ should be regulated based on the current enzyme activities, $\{\alpha_1, ..., \alpha_Z\} \in S$. Here, an n-step SARSA algorithm [167] is utilized to perform fitted value function iteration. An optimal policy is therefore learned by iteratively updating the value function, $V : S \rightarrow \mathbb{R}$, which is defined as the expected rewards to be received by following a fixed policy from a specified state, $V^{\pi}(s_t) = E_{\pi}[r_{t:t+n}|s_t]$. In an n-step algorithm, the value function is meant to predict the discounted reward, $r_{t:t+n}$, for n future steps. The n-step reward experienced by the agent is defined as $r_{t:t+n} = r_t + \gamma r_{t+1} + ... + \gamma^{n-1} r_{t+n-1} + \gamma^n V(s_{t+n})$, where $\gamma \in [0.0, 1.0]$ is the discount factor. Each reward, $r_t = R(s_{t-1}, a, s_t)$, represents the feedback from moving into state s_t from s_{t-1} after taking some action a. The first n steps represent the rewards experienced, while the term $V(s_{t+n})$ represents the future rewards. Once a terminal state is less than n steps away, the n-step reward is truncated to the appropriate length.

Learning the value function implicitly improves the policy. The relation between the value of a state and the policy is given by an ϵ-greedy policy, which is defined as:

$$\pi(s) = \begin{cases} \arg\max_{a \in A} \left(R(s, a, s') + \gamma V(s') \right), & \xi \geq \varepsilon \\ \textit{random choice}, & \xi < \varepsilon, \end{cases} \tag{4.26}$$

where ξ is a uniform random number between 0.0 and 1.0. As the value function is better estimated, the policy determines reactions to regulate that lead to the greatest cumulative

reward. Exploration is imposed by randomly choosing reactions to regulate, allowing the policy to escape local minima. As the agent learns, ε is slowly annealed to reduce exploration and fluctuations in the value function. During each training episode, the initial state is $s = \{1.0, ..., 1.0\}$, such that all enzyme activities are unregulated. Trajectories through state space are stopped when the stopping criteria $L = 0.0$ is satisfied. This condition requires that all reactions have cost function values at or below zero before the reinforcement learning ends and the predictions are in caliber with the experimental values.

Finally, the state value function is estimated by using a neural network implemented in PyTorch [150] with a single hidden layer and hyperbolic tangent activation functions. Updates to the value function are performed by optimizing the neural network using stochastic gradient descent. This is done by backpropagating the squared loss between the predicted value and the n-step reward, $[V(s_t) - r_{t:t+n}]^2$.

4.2.4 Model Training

Prediction of network regulation was performed using a trained neural network to estimate the value function. Network weights were adjusted using stochastic gradient descent with a learning rate, $lr \in \{10^{-4}, 10^{-5}, 10^{-6}\}$. Each algorithm learned and generated data using an ϵ-greedy policy with initial $\epsilon = 0.5$ or 0.2 depending on the size of the pathway. ϵ was annealed by dividing by a factor of two every 25 learning episodes.

For each pathway, 10 agents are trained for each different value of $n \in \{2, 4, ..., 12\}$ and each learning rate. The resulting average of 10 RL runs for the glycolysis-PPP-TCA pathway (Figure S1) show the mean reward for the 350 training episodes. Optimal regulation

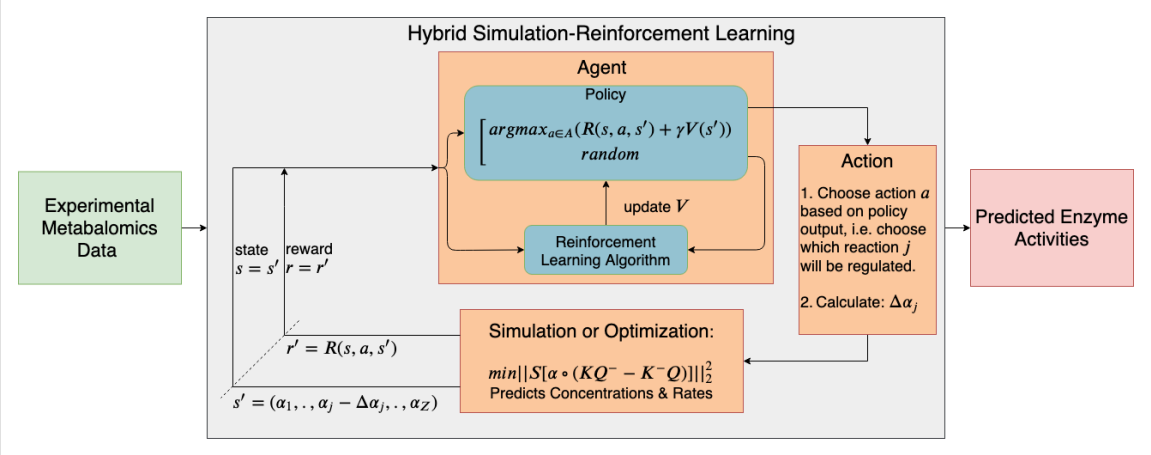

Figure 4.1: Schematic of *in silico* framework for learning regulation (grey box) with coupled simulation or optimization routine controlling environmental feedback. Initial framework input (green box) consists of target metabolite concentrations from experimental data. The output (red box) consists of a learned optimal enzyme regulation scheme necessary to reach the target concentrations. Learning is performed by repeatedly testing different regulation schemes and updating the value function, V, that returns a scalar value for a given set of enzyme activities. Enzyme activities, represented as states, are chosen for regulation by performing actions that are determined by a policy function. A given policy is determined by V. The new steady state metabolite concentrations resulting from applied regulation are determined by an optimization routine. Alterations in metabolite concentrations are a direct result of moving into a state s' from a state s after taking action a, i.e. performing regulation. These dynamic changes are used to define a reward function, R, that determines environmental feedback. Rewards are used to direct the agent as it explores and learns a policy that predicts optimal enzyme regulation.

is prescribed by analyzing the agent with the largest cumulative reward averaged over the last 50 terminal states.

4.2.5 Data

The metabolomics data used in this study was from *E. coli* studies by Bennett, et al. [17], and Park, et al [148]. Briefly, *E. coli* cells were grown in isotope-labeled media and then extracted in organic solvent containing unlabeled internal standards in known concentrations. Metabolites were extracted in cold solvent and analyzed using chromatography-MS,

and concentrations relative to the known standard concentrations were obtained using peak ratios of the labeled samples to unlabeled standards.

If no experimental data is available, the analysis is carried out using estimates of metabolite concentrations (Appendix B). For this purpose, an estimate of 0.1 mM for each metabolite that is variable is used. For fixed metabolites that form the boundary conditions, specific values are required that induce appropriate non-equilibrium boundary conditions.

Standard free energies of reaction were calculated using eQuilibrator and the eQuilibrator API [56]. eQuilibrator uses well-curated gold standard data on the thermodynamics of reactions from the National Institute of Standards and Technology [60], which is the basis for subsequently adjusting reference free energies for pH and ionic strength. For reactions for which experimental data are not available, free energies are estimated using reliable reaction comparison methods [140] or electronic structure calculations [82].

4.3 Results

We solve the prediction problem of which enzyme to regulate by a novel combination of methods from statistical thermodynamics, control theory and reinforcement learning (RL). The initial step is to determine steady state concentrations without applying regulation by using numerical optimization of the respective ordinary differential equations on a convex energy surface. The convex energy surface for metabolic dynamics is obtained by assuming that the time dependence is the same for all reactions in the Marcelin-de Donder dynamical force equation for mass action kinetics [30]. Due to the assumption that the reactions all occur on the same time scale, the thermodynamic odds of each reaction

(Methods, Eqn. (4.3)) are similar in value in upper glycolysis, lower glycolysis and the TCA cycle, though varying by a factor of two due to stoichiometry. Such a configuration results in at least a local maximum of the reaction path entropy [46, 27]. Figure 4.2 shows the resulting steady state reaction fluxes and reaction free energies for the glycolysis-PPP-TCA cycle under high NAD/NADH and low NADP/NADPH conditions.

If there are no constraints, the configuration also results in a higher entropy distribution of metabolites. However, the metabolites will be constrained to be away from the equilibrium distribution if there are non-equilibrium boundary conditions. Since the initially predicted concentrations will then be proportional to their Boltzmann probabilities, the initially predicted concentrations may be exceedingly high [8] compared to experimentally observed values from isotope-labeled, mass spectrometry measurements [17, 148]. However, these high concentrations allow for highly effective inference of regulation to control the concentrations. The predicted concentrations, $\tilde{n_i}$, are brought into alignment with experimental observations, n_i, by applying regulation. Regulation is determined using either a Metabolic Control Analysis (MCA) approach, or a hybrid optimization-reinforcement learning (RL) approach (Methods). In both cases, regulation is applied in the form of an activity coefficient, α_j, that scales the reaction flux for reaction j, where $\alpha_j = 1.0$ indicates no regulation while $\alpha_j = 0.0$ indicates complete regulation.

In the two MCA based methods that were developed, reactions are regulated based on the sensitivity of the predicted concentrations to the activity coefficient that modulates each reaction, which is carried out by a specific enzyme. The sensitivity of the i^{th} metabolite with concentration n_i (observed or predicted) to the activity, α_j, of enzyme j, is described

by the concentration control coefficient, $C^n_{i,j}$. When using predicted concentrations, $\tilde{n}_i$, the notation $C^{\tilde{n}}_{i,j}$ is utilized to specify the concentration control coefficient for predicted metabolite concentrations. A loss function defined as the logarithm of the division of the predicted concentrations or counts is used to the measured concentrations or counts, $L_i = \log(\tilde{n}_i/n_i)$. The change in the loss function due to a change in the activity of reaction j is

$$\Delta L_{i,j} = \log \tilde{n}_i - \log\left(\tilde{n}_i - \Delta\tilde{n}_i(\Delta\alpha_j)\right). \tag{4.27}$$

The reaction j selected for regulation is the one whose change in activity results in the largest change in the loss functions of all metabolites whose predicted concentrations exceed the experimentally observed concentrations, as determined by $\Delta L_j = \sum_i \Delta L_{i,j}$. Regulation is considered complete when predicted metabolite concentrations are brought into agreement with experimental measurements.

The three different regulation approaches are compared by statistically characterizing the rate of energy flow across the reactions. The rate that energy is produced in metabolism has long been known to be one of the most significant factors in metabolic regulation [7]. The sum of the rate of free energy generated across all reactions is the free energy dissipation rate, or equivalently the negative of the entropy production rate. The free energy of the j^{th} reaction at steady state is $\Delta G_j = -R_g T \log(K_j Q_j^{-1})$, where R_g is the gas constant, and T is the temperature, K_j is the equilibrium constant and Q_j is the reaction quotient. The free energy dissipation rate is defined as the rate at which free energy is dissipated [29, 57],

$$\frac{dG}{dt} = -R_g T \sum_j^Z \alpha_j [r_j \log K_j Q_j^{-1} - r_{-j} \log K_j Q_j^{-1}]. \tag{4.28}$$

In the maximum path entropy formulation (Section 4.2, Eqn. (4.6)), the rate r_j is proportional to the thermodynamic driving force on the reaction, $K_j Q_j^{-1}$. The free energy change for a reaction j can be broken down into two components, an energy change, $\Delta E_j = -R_g T \log K_j$, and a configurational entropy change, $T\Delta S_j = R_g T \log Q_j$ [27]. As the reactions occur, the system moves towards equilibrium, decreasing the reactants and increasing the products, which results in a change in the configurational entropy due to changes in the reaction quotients. In a steady state or pseudo-steady state system, the steady state is replenished by additional nutrients such that the reaction quotients, Q_j, return to their steady state values. Replenishing the steady state, however, requires work. Since the net entropy change in a pseudo-steady state system must be zero, the measure of work available for processes other than maintaining the steady state, such as replication, is,

$$\frac{dE}{dt} = -R_g T \sum_{j}^{Z} \alpha_j [r_j \log K_j - r_{-j} \log K_j]. \tag{4.29}$$

Both dG/dt and dE/dt (the energy dissipation rate) are important metrics of the rate of work produced by metabolism. When regulating reactions, a biological system must find a balance between a free energy dissipation rate that extracts energy from the environment as quickly as possible and a low rate of entropy change to maintain the pseudo-steady state. In principle, any individual or species in a pseudo-steady state that maximizes the rate of usable work, dE/dt, will outcompete those with lower rates of net work and will be the organism selected by nature.

Three different versions of *E. coli* central metabolism are evaluated under four different nutrient conditions. The three different versions of metabolism were (1) gluconeogenesis, (2) glycolysis and the TCA cycle, and (3) glycolysis, the pentose phosphate pathway

(PPP) and the TCA cycle (glycolysis-PPP-TCA). Metabolite concentration data used in the analysis were from *E. coli* in exponential growth with glucose as the carbon source [17, 148]. As an alternative to experimentally measured metabolite concentrations, rough estimates of concentrations can be used as well that give qualitatively similar results (see Appendix B). In all cases, the predicted regulation matched known regulation points in central metabolism or were adjacent to known regulation points.

Below, changes in the largest network, glycolysis-PPP-TCA, are discussed under two identical nutrient conditions except for the NADP/NADPH ratio, which is held fixed but at different values throughout each analysis. In condition 1, the NAD/NADH ratio is high (31.3) and the NADP/NADPH ratio is low (0.02), which favors flux through upper glycolysis rather than PPP. In condition 2, the NADP/NADPH ratio is also high such that NADP/NADPH = NAD/NADH = 31.3 [17]. The latter condition favors increased flux through PPP. Analyses of gluconeogenesis and glycolysis and the TCA cycle are included in the Appendix A (Figures B.2 and B.3). In all conditions, regulation that is found by the reinforcement learning method is compared with that found by deterministic methods using only MCA.

4.3.1 High NAD/NADH require regulation of metabolite levels in glycolysis

Prediction of enzyme activities using MCA methods are deterministic. Given the conditions for fixed metabolites in which the NAD/NADP ratio is high and the NADP/NADPH ratio is low, flux is favored through upper glycolysis over PPP, and the local MCA method predicts (Figure 4.3A, red 'plus') that five reactions in glycolysis are regulated due to the en-

zymes hexose kinase (HEX1), phosphofructokinase (PFK), glyceraldehyde-3-phosphate dehydrogenase (GAPD), phosphoglycerate kinase (PGK), and pyruvate dehydrogenase (PDH), while one enzyme in PPP is regulated, phosphogluconolactonase (PGL), near the beginning of the pathway. It is known that regulation of PPP occurs one enzyme up from PGL at glucose 6-phosphate dehydrogenase (G6PDH) instead. But the metabolite that is over produced and is predicted to have high concentration without regulation is phosphogluconate, the product of the PGL reaction. In practice, PGL may be a hard reaction to allosterically regulate since it is a unimolecular ring opening reaction that may be catalyzed significantly by binding alone [28].

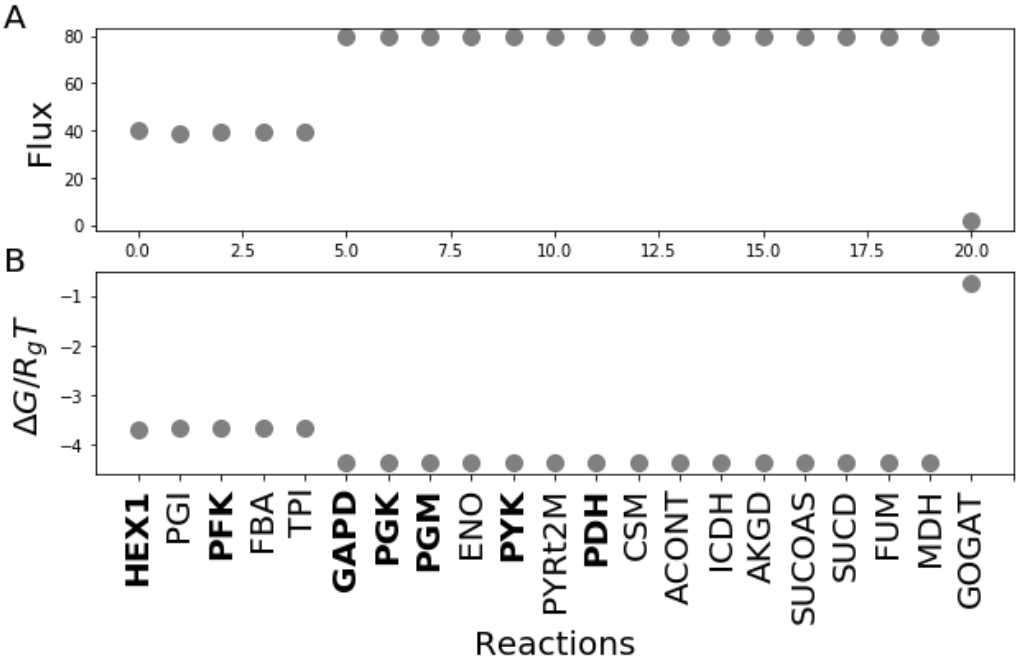

Figure 4.2: Initial steady state properties before any regulation is applied in the form of reduced activity coefficients for glycolysis-PPP-TCA cycle with high NAD/NADH and low NADP/NADPH conditions. The steady state is determined by maximizing the reaction path entropy such that the net thermodynamic driving force on each reaction is proportioned according to the governing equation for metabolite kinetics, Eqn. (4.10). (A) Unregulated reaction fluxes. (B) Unregulated reaction free energies. Reduction of activity coefficients to values less than 1.0 reduces both the steady state fluxes and the reaction free energies (Fig. 4.3-4.5). **Abbreviations:** HEX1, Hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, Fructose bisphosphatase; TPI, Triosephosphate isomerase; GAPD, Glyceraldehyde 3-phosphate dehydrogenase; PGK, Phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, Enolase; PYK, Pyruvate kinase; PYRt2m, pyruvate transporter; PDH, Pyruvate dehydrogenase; CSM, Citrate Synthase; ACONT, Aconitase; ICDH, Isocitrate dehydrogenase; AKDG, a-ketoglutarate dehydrogenase; SUCOAS, Succinyl-CoA synthetase; SUCD, Succinate dehydrogenase; FUM, Fumarase; MDH, Malate dehydrogenase; GOGAT, Glutamine oxoglutarate aminotransferase.

The RL and unrestricted MCA methods both predict the same minimal regulation at HEX1 and GAPD to achieve the same goal of maintaining the predicted concentrations at or below the experimentally observed values. The RL method, however, additionally regulates PGK, pyruvate kinase (PYK), the pyruvate mitochondrial transporter (PYRt2m) and PDH to obtain a similar energy dissipation rate. As shown in Figure 4.3A, four of these enzymes were also regulated in the local MCA method. The difference is that HEX1 and GAPD are more extensively regulated in the RL and the unrestricted MCA methods. Despite these differences in regulation, each regulated enzyme with the exception of the pyruvate transporter are known sites of regulation (known sites of regulation are highlighted in bold). Regulation of the pyruvate transporter was only predicted in the stochastic RL approach. It is likely that this regulation should be assigned to PYK or PDH as it was in the deterministic MCA approach.

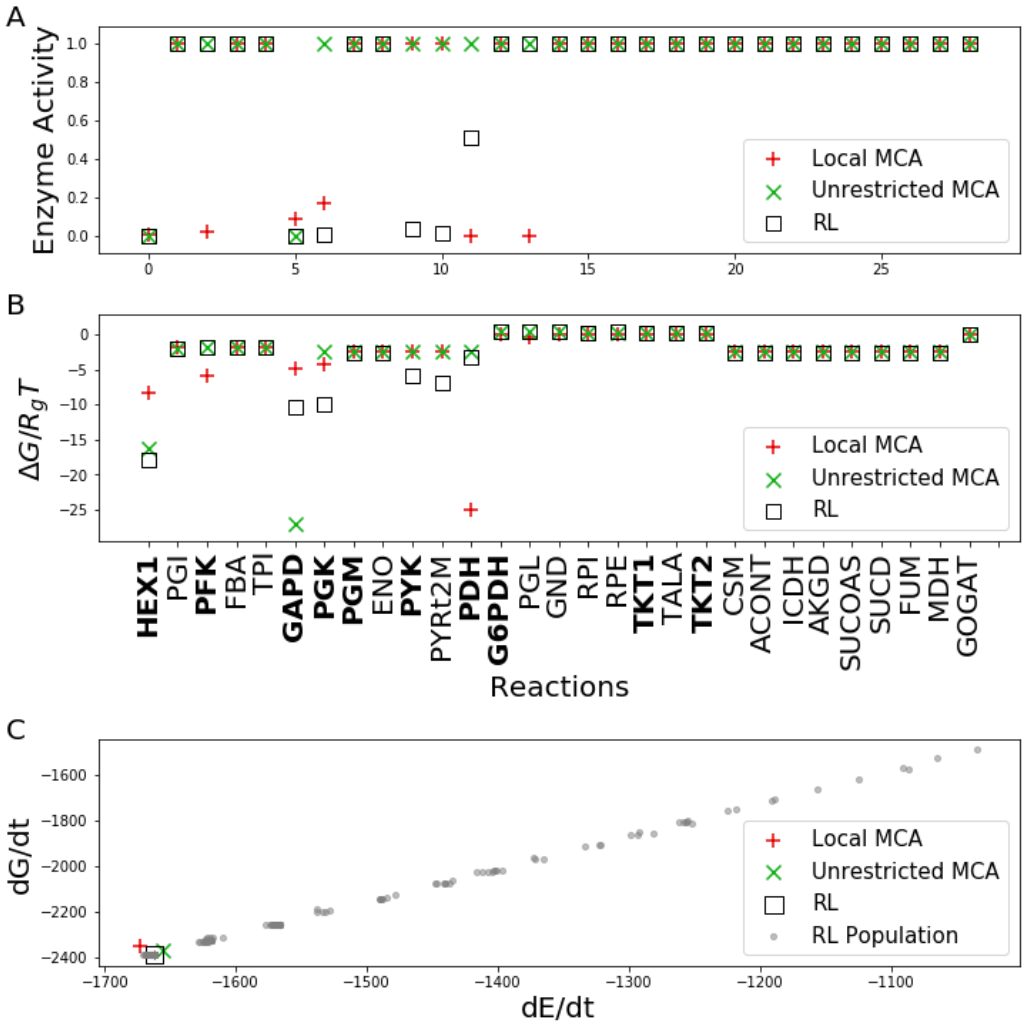

Figure 4.3: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and low NADP/NADPH conditions. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square).(B) Reaction free energy changes are no longer equally distributed across subpathways (Fig. 4.2, upper glycolysis, PPP, lower glycolysis, TCA cycle) but instead free energies are further from equilibrium at reactions where regulation is applied. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. **Abbreviations:** G6PDH, Glucose 6-phosphate dehydrogenase; PGL, Phosphogluconolactonase; GND, phosphogluconate dehydrogenase; RPI, Ribose 5-phosphate isomerase; RPE, Ribose 5-phosphate epimerase; TKT1, Transketolase 1; TALA, Transaldolase; TKT2, Transketolase 2. Note: previously used abbreviations are presented in Figure 4.2

As shown in Figure 4.3B, whenever regulation is applied in the form of reducing the activity coefficient, the free energy of the reaction becomes more favorable compared to reactions in the same pathway (e.g., compare to the consistency of free energy changes in upper glycolysis, PPP, lower glycolysis and TCA cycle in Fig 4.2). Reducing the activity of an enzyme in a non-equilibrium setting will cause the reactants to increase in concentration and the products to decrease in concentration, resulting in reaction free energies being further away from equilibrium. Despite the different sites of regulation and the difference in reaction free energies for the three methods, the free energy and energy dissipation rates are similar and are the most favorable rates found (Figure 4.3C).

4.3.2 High NAD/NADH & High NADP/NADPH require additional regulation in PPP

In the second set of conditions, the NADP/NADPH ratio is also high, which in principle favors more flux through PPP. The resulting regulation is similar to the first conditions in which NADP/NADPH is low with a few exceptions (Figure 4.4A and 4.4B). The local MCA method additionally regulated G6PDH, the entry point into the PPP as well as transketolase (TKT), while the RL method no longer regulated PYK and regulated the pyruvate mitochondrial transporter (PYRt2m) rather than PDH. The latter is likely incorrect, but the fact that the method was trying to regulate pyruvate concentrations suggests that PYK might be the true target of regulation. Like the local MCA method, the RL method also regulated HEX1, GAPD and PGK.

In contrast, the unrestricted MCA method regulated the same reactions as in the low NADP/NADPH conditions, HEX1 and GAPD. The regulation under a high NADP/NADPH

ratio is similar to the conditions in which NADP/NADPH is low primarily because increasing the NADP/NADPH ratio alone is insufficient to drive much flux through PPP. Because of less total regulation, the unrestricted MCA and RL methods result in significantly higher energy dissipation rates than the local MCA method and are thus likely to be more optimal regulation schemes.

4.3.3 Regulation of PFK Maximizes Flux Through PPP

Increased flux can be channeled through the PPP if PFK activity is regulated to a greater extent or is turned off completely. Then significant flux flows through PPP instead of upper glycolysis and does so in a cyclical manner. There is experimental support for this as well. In *Neurospora crassa*, glycolysis and the PPP are circadianly regulated, with the PPP being regulated 180 degrees out of phase with upper glycolysis. In the extreme case when PFK activity is turned off in the model, then the cyclical operation of the PPP is such that three carbons are lost from each glucose molecule as CO_2 before all the carbon reaches lower glycolysis as glyceraldehyde 3-phosphate.

In the case when PFK activity is set to zero, all methods apply regulation to HEX1. This is enough for the unrestricted MCA and RL methods to bring concentrations to within the observed experimental range, and both methods result in maximal energy dissipation rates (Figure 4.5). In contrast, the local MCA method additionally requires regulation in PPP at G6PDH, PGL and TKT. But even in this case, the local MCA method fails to completely bring sedoheptulose 7-phosphate into the range of the experimental observations. In attempting to control sedoheptulose 7-phosphate, the applied regulation is extensive

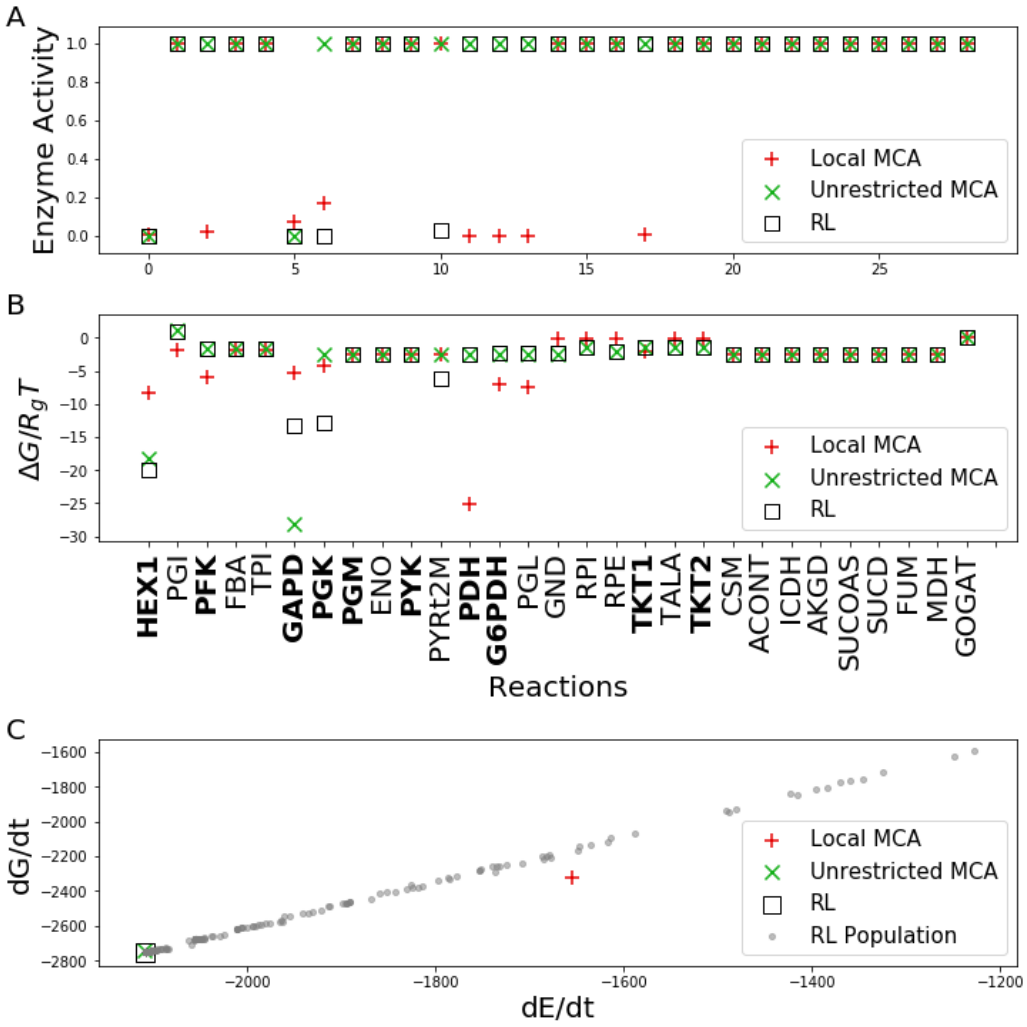

Figure 4.4: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent.

enough such that the net flux through glycolysis, the pentose phosphate pathway and the TCA cycle approaches zero. Thus, the local MCA method fails to obtain control. In several cases involving the local MCA method, the concentration of sedoheptulose 7-phosphate and sometimes 6-phospho D-gluconate become uncontrollable resulting in concentrations

higher than what is observed experimentally. The reason for this is that the respective reactions producing these compounds approach equilibrium; it is known that when a reaction approaches equilibrium, the concentrations of the products are no longer controllable [85].

In these cases, the reactions and their metabolites are effectively uncoupled from the non-equilibrium reactions. Lack of control may result in the respective metabolites reaching high concentrations in the cytoplasm, and the cytoplasm consequently becoming glassy and diffusion limited. Experiments support this principle. Recent reports provide evidence that active metabolism promotes cytoplasmic fluidization while inactive metabolism results in a glass-like cytoplasm with limited diffusion in both bacteria [149, 99] and eukaryotes [68].

However, it is not clear that the failure to maintain control when using the local MCA method reflects poorly on the concept of modularity whereby enzymes use local control. The failure to obtain control of sedoheptulose 7-phosphate can also be due to the incomplete nature of the model of metabolism used here. It may be that in a more extensive model of metabolism, such as the inclusion of purine and pyrimidine biosynthesis pathways branching off of D-ribose 5-phosphate, control of sedoheptulose 7-phosphate by the local MCA method may be possible. This possibility is presented because Transketolase (TKT), the enzyme producing sedoheptulose 7-phosphate is a key post-translational regulation point into purine synthesis [168].

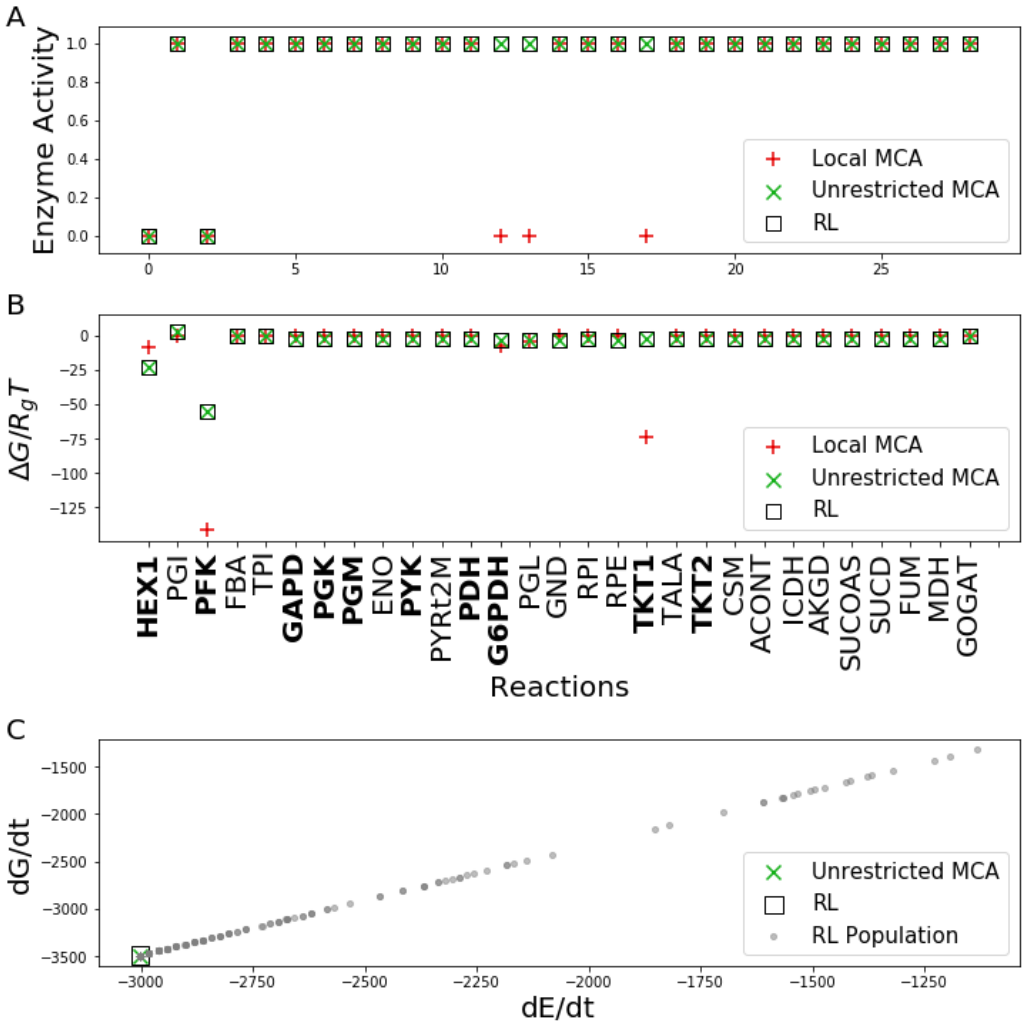

Figure 4.5: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions and PFK activity set to zero. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. The local MCA method results in zero flux (Appendix B Table B.1) and is therefore not shown.

4.3.4 Regulation Increases Reaction Free Energies

In all cases of regulation, whenever a reaction is regulated significantly, the reaction free energy is significantly different from the neighboring reactions which are not regulated,

as shown in Figures 4.3B, 4.4B, and 4.5B. It is the act of regulating each reaction that causes the respective reactants to build up and products to become relatively depleted, which causes the free energy change of the reaction to increase in magnitude. This observation has wide support in the literature [112, 137], but the cause has been misinterpreted as being such that reactions are selected for regulation because they are far from equilibrium, rather than reactions being far from equilibrium because they are regulated.

The reasoning for assuming that highly non-equilibrium reactions are selected for regulation has to do with the established principle that biological systems activate metabolites for reactivity by covalently attaching high potential groups such as coenzyme A and phosphates. These reactions will then have much higher standard free energies of reaction than they would otherwise.

However, the use of such activators as phosphoryl groups and coenzyme A to drive a reaction will not just result in the respective reaction being further from equilibrium, but all reactions in the pathway will be further from equilibrium because increased product formation of the activated reaction will result in increased reactant concentration for the next reaction, and so forth, as the effect propagates down the pathway until a steady state is reached. As a result of the highly non-equilibrium nature of reactions in the pathway, many reaction products may be produced in biologically unreasonable concentrations. This problem is solved by reducing the activity of either the enzyme catalyzing the reaction or upstream enzymes that have control of the flow of material into the pathway. The reactions that have the most control can be determined using concentration control coefficients and thermodynamics.

4.4 Discussion

All predicted schemes discussed above enforce regulation on enzymes that are known to be regulation sites. Nine of these 11 enzymes are known to be sites of post-translational regulation in glycolysis and the pentose phosphate pathway, either allosterically or through chemical modification (Table 4.1): hexokinase, phospho-fructokinase, glyceraldehyde phosphate dehydrogenase, phosphoglycerate kinase, pyruvate kinase, pyruvate dehydrogenase, glucose 6-phosphate dehydrogenase, transketolase, and pyruvate carboxylase (Appendix B). Only the pyruvate mitochondrial transporter (PYRt2m) and phosphogluconolactonase (PGL) are not known to be regulated. The regulation assigned to the pyruvate transporter was done stochastically by the reinforcement learning and likely should be assigned to PYK or PDH, as it was the deterministic MCA approaches. PGL presumably would be hard to control since it catalyzes a highly favorable ring opening which may only require desolvation in the enzyme active site. It is worth noting the enzymes that are known to be regulated but were not indicated as being regulated in this study. Foremost among these is fructose bisphosphatase (FBA), an enzyme that is well-known to be regulated in gluconeogenesis. Under the limited number of conditions used in the study of gluconeogenesis herein, levels of fructose 6-phosphate or other downstream products never rose high enough to require regulation. Likewise, the products of enolase, phosphoglucose isomerase, PEP carboxykinase, glucose 6-phosphatase never rose to the level that these needed to be regulated, but it would be reasonable to expect that the respective enzymes may need to be controlled under conditions that were not tested here.

Of the 11 enzymes predicted to be regulated, outsized roles were played by the

branch points of each of the pathways, as quantified by the influence of the enzyme activity coefficients, C_j^n, on the respective reactants or products (Table 4.1). The summary concentration control coefficient reports the total influence of the activity of the enzyme on all metabolites exceeding the experimentally observed values. The C_j^n values reported in Table 4.1 are consistent with recent experimental measurements on the effect of changes in expression levels of glycolytic enzymes on the concentrations of the same metabolites [188].

Hexokinase, the entry point into the model and entry point into upper glycolysis and the pentose phosphate pathway, had the largest role with $C_j^n = 12.4$, meaning that hexokinase effectively had 100% control over 12.4 reactions. It is worth noting that both the RL and unrestricted MCA methods achieved successful control by consistently regulating hexokinase, which is again consistent with recent experimental observations of glycolysis [188]. In the experimental studies, increased expression of hexokinase lead to increases in down stream phosphorylated compounds, including fructose 1,6-bisphosphate, sedoheptulose 7-phosphate, sedoheptulose 1,7-bisphosphate, and 6-phosphogluconate, just as predicted here. Not surprisingly, increased levels of these metabolites due to increased hexokinase expression were correlated with a decrease of glycolytic rate, as one would expect if cytoplasmic solubility were adversely affected.

Likewise, glucose 6-phosphate dehydrogenase, the entry point into the PPP, had effectively 100% influence over 16.8 reactions, although this value is only seen this high when the phosphofructokinase activity is set to 0.0 such that the PPP acts cyclically and three circuits around the cycle are made for each glucose metabolized. Likewise, for lower glycolysis the main control point, glyceraldehyde 3-phosphate dehydrogenase, is the entry

into the pathway which is also where upper glycolysis and PPP converge. No regulation was needed for the TCA cycle under the conditions studied.

While the predictions align well with known sites of post-translational regulation, the predictions offer no information on whether the regulation would be due to allosteric interactions or chemical modification as might be inferred from more complex and expensive approaches that utilize (and require) absolute metabolite concentrations, fluxes inferred from isotope labeling studies, MS proteomics analyses and detailed kinetic models that include explicit enzyme binding, catalysis and product release [63]. The regulation predictions provided here, however, were done purely *in silico* with the optional use of absolute metabolite concentrations, if available. Although the regulatory effector can't yet be inferred from this approach, it would seem reasonable to assume that control of metabolite concentrations would be due to allosteric regulation since allosteric interactions work on a faster time scale than post-translational modifications. It is likely that post-translational modifications act to redirect flux when either degradation of enzyme would be too slow, or when degradation and later resynthesis of the enzyme would be too costly [179], which is not the scenario addressed here.

Both MCA approaches were based only on adjusting the activities of enzymes that would have the most influence on reducing concentrations to physiological values. Only the RL approach rewards regulation schemes for maximizing the entropy production rate (Eqn. (4.23)). Even though the RL and MCA methods have different aims, both maximized the energy dissipation rate, dE/dt, a principle alluded to by Lotka [124]. Furthermore, while the unrestricted MCA approach and the RL performed similarly, the local MCA approach did

not always find a solution, which could reflect the incompleteness of the metabolic network that is modeled, or may simply indicate that modular regulation to this degree is insufficient. In addition, in at least one case the local MCA approach did not produce solutions with the highest energy dissipation rates. However, the set of enzymes predicted by the local MCA approach covers many more of the enzymes known to be classically regulated, as shown in Table 4.1.

Consequently, we have shown how post-translational regulation results in the emergence of the general principle of maximal, entropy production rate for metabolism, and we can now also include the principle of maximization of the energy production rate, dE/dt, for pseudo-steady state phenotypes. When these principles are used to make predictions, each prediction must also take into account the physicochemical constraints on the system, such as the inherent constraints on the maximal rates of enzymes and thermodynamic costs and benefits, not simply metabolite solubilities [179]. These additional physicochemical constraints can explain the observed upper limit to free energy dissipation in microbial systems [138].

The observation of an upper limit to free energy dissipation is related to the concept of maintaining the adenylate energy charge ratio. The adenylate energy charge rule widely found in textbooks was defined in terms of concentrations as [(ATP) + 0.5 (ADP)]/ [(ATP) + (ADP) + (AMP)]. It was proposed that biological systems maintain values of the energy charge between 0.75 and 0.90. There are now many known exceptions to this proposed rule that it can no longer be regarded as a rule but as an emergent property, just as the maximization of energy production rates is an emergent property due to natural selection.

The simulation-based predictions of enzyme activities presented in this work advance both the practice and theory of biology. The ability to predict from simulation or infer the free energy changes and control coefficients (in addition to fluxes) for each reaction allows the use of control theory and learning to analyze and explore the operations of the cell. In synthetic biology the development of cell lines often requires additional circuits and can result in unforeseen consequences or lower cell growth rates. Simulation of cells with engineered or deleted circuits will allow prediction of the effects in place of difficult trial and error in experiments.

Finally, it is important to understand the principles behind post-translational regulation because regulation of metabolism is precisely what controls a cell's energetic behavior. From bacterial growth and reproduction, to developing cells or even halting the growth of cancer cells, regulation plays the central role. Learning how cells regulate and control themselves is essential for designing new organisms that have an intended purpose (synthetic biology), developing new strategies to target and control microbial and metabolic diseases (medicine), and understanding design principles of biology (fundamental science). Currently no other experimental or computational approach has been shown to identify points of regulation in metabolism in a rapid manner.

Enzyme	Pathway	$C_j^{\tilde{n}}$	Prediction Method	Evidence
HEX1	Upper glycolysis	12.4	RL, L-MCA, MCA	[112]
PFK	Upper glycolysis	4.6	L-MCA	[112, 137]
GAPD	Lower glycolysis	4.6	RL, L-MCA, MCA	[137, 2]
PGK	Lower glycolysis	3.8	RL, L-MCA	[137, 106]
PYK	Lower glycolysis	1.7	RL	[203, 202, 201]
PYRt2m	Lower glycolysis	1.1	RL	–
PDH	Lower glycolysis	0.6	RL, L-MCA	[72]
G6PDH	Pentose phosphate	16.8	RL, L-MCA	[197]
PGL	Pentose phosphate	16.0	L-MCA	–
TKT	Pentose phosphate	5.0	L-MCA	[168]
PC	Gluconeogenesis	3.7	RL, L-MCA, MCA	[83]
Modeled and known to be regulated but not observed				
PGM	Lower glycolysis	3.0	–	[64]
FBP	Gluconeogenesis	0.1	–	[112, 137]

Table 4.1: The set of enzymes found to be regulated in all analyses along with the associated pathway, the concentration control coefficient, $C_j^{\tilde{n}}$, of the reaction summed over all metabolites before any regulation is applied, the method predicting the regulation and the experimental evidence from the literature for predicted regulation. Abbreviations are the same as in Figure 2. PC is pyruvate carboxylase and is observed to be regulated in gluconeogenesis (Appendix B Table B.4).

Chapter 5

Conclusions

Modern biology raises unique challenges and questions that cannot be addressed by experiments alone. Multiscale modelling and machine learning have proven to be worthy supplements to experimental approaches. This is especially true when models are combined with experiments in order to computationally generate, test, explain, and validate novel biological hypotheses. The novel mathematical modelling approaches presented here have many important biological applications such as the study of deformation and contraction of blood clots, cancer invasion, HIV virus budding, bacterial behavior, and early development of animal tissues. In addition, this work has brought together different fields of applied and computational mathematics and data science in addition to answering exciting questions in biology from a mathematical modeling perspective.

The nascent cohesive crisscrossing of fibers in stretched fibrin networks described in Chapter 2, comprise a novel structural mechanism that, in combination with other structural rearrangements, underlies stiffening of fibrin gels upon tensile deformation. Notice that the

newly described mechanism involves two spatial levels of fibrin mechanics, namely the non-linear elasticity of individual fibrin fibers as well as bulk strain-stiffening of the entire fibrin gel.

One of the most likely physiological conditions where fiber-fiber cohesion may play a role in determining fibrin elasticity is clot deformation under hydrodynamic blood flow. The analysis of the network orientation in clots formed in vitro under different flow conditions revealed that fibrin fibers orientations were not random. Instead, fibers were found to be aligned in the direction of the shear stress [58, 26] and their alignment is associated with an increased number of fiber-fiber contacts and cohesive interactions. Such networks must be more resistant to stretching deformations in the direction of alignment as the initial degree of alignment increases, suggesting that the mechanical response and structural stability of a blood clot are greatly affected by the flow shear. In other words, clot breakage and formation of thrombotic emboli in the regions of the circulation system with high shear, can be mechanically regulated due to increased stiffness of the aligned fibrin clots. The newly developed model and the extension including platelet cells detailed in Chapter 3 can be further applied to other types of network deformations such as shear and compression, emphasizing its universality and applicability to mechanics of natural biopolymers and for designing biomaterials in tissue engineering. Moreover, this modelling framework allows for the ability to computationally test the necessary inhibitors and/or activators in platelets to prevent hyperactivity and clotting complications that cause life-threatening conditions, such as ischemic stroke, or pulmonary embolism. These predictions could directly impact future drug design, wound healing techniques, and immediate patient risk based on the analysis of

clot composition.

The findings described in Chapters 2 and 3 can be extended to other hydrogels with filamentous scaffolds such as collagen, fibronectin, actin, and others that may undergo interfilamentous interaction upon deformations as well as contribute to our understanding of how single cells influences cell-ECM behavior.

While multiscale models, as presented in Chapters 2 and 3, can also make use of machine learning techniques at different spatial scales, hybrid multiscale models informed by machine learning are still rare. At this point, machine learning is frequently applied to biological systems without utilizing approaches from physics and mathematics. Chapter 4 introduces a hybrid RL framework that incorporates a thermodynamic description of the system into the environment and reward function. This is a new methodology where the machine learning predictions iteratively inform large-scale mechanistic or multiscale models. This type of approach is necessary because multiscale problems require both experimental data and theoretical description of biological mechanisms at different spatial scales. While understanding regulation itself is essential for organism function, disease identification and basic biological processes, the model flexibility allows for many extensions. Specifically, the direct incorporation of physical principles as well as mathematical methods from theory of optimization and control theory into the RL framework allows for an accurate representation of the biological system and yields novel biological insights.

In the future, regulation determined by machine learning analysis of metabolism could be applied to mechanical models in several contexts such as bacteria chemotactic motion, platelet contraction, and many others. This will contribute to predictive understanding

for the metabolism in single cells interacting with their environment. More importantly, this an important first step in extending predictive understanding of specific biological processes occurring within individual cells to communities of cells.

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv [cs.DC]*, 2016.

[2] H. N. Aithal, M. M. Walsh-Reitz, and F. G. Toback. Regulation of glyceraldehyde-3-phosphate dehydrogenase by a cytosolic protein. *Am J Physiol*, 249(1 Pt 1):C111–6, 1985.

[3] Mark Alber, Adrian Buganza Tepole, William R. Cannon, Suvranu De, Salvador Dura-Bernal, Krishna Garikipati, George Karniadakis, William W. Lytton, Paris Perdikaris, Linda Petzold, and Ellen Kuhl. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine*, 2(1):115, 2019.

[4] N. O. Alieva, A. K. Efremov, S. Hu, D. Oh, Z. Chen, M. Natarajan, H. T. Ong, A. Jégou, G. Romet-Lemonne, J. T. Groves, M. P. Sheetz, J. Yan, and A. D. Bershadsky. Myosin iia and formin dependent mechanosensitivity of filopodia adhesion. *Nat. Commun.*, 10(1):3593, 2019.

[5] R. N. Anderson, A. Boulanger, W. B. Powell, and W. Scott. Adaptive stochastic control for the smart grid. *Proc. IEEE*, 99(6):1098–1115, 2011.

[6] Ashley L. Arthur, Livia D. Songster, Helena Sirkia, Akash Bhattacharya, Carlos Kikuti, Fernanda Pires Borrega, Anne Houdusse, and Margaret A. Titus. Optimized filopodia formation requires myosin tail domain cooperation. *Proc. Natl. Acad. Sci. U. S. A.*, 116(44):22196–22204, 2019.

[7] D. E. Atkinson. The energy charge of the adenylate pool as a regulatory parameter. interaction with feedback modifiers. *Biochemistry*, 7(11):4030–4034, 1968.

[8] Daniel E. Atkinson. *Limitation of Metabolite Concentrations and the Conservation of Solvent Capacity in the Living Cell*, volume 1, pages 29–43. Academic Press, 1969.

[9] Rodney D. Averett, Bryant Menn, Eric H. Lee, Christine C. Helms, Thomas Barker, and Martin Guthold. A modular fibrinogen model that captures the stress-strain behavior of fibrin fibers. *Biophys. J.*, 103(7):1537–1544, 2012.

[10] T. C. Baradet, J. C. Haselgrove, and J. W. Weisel. Three-dimensional reconstruction of fibrin clot networks from stereoscopic intermediate voltage electron microscope images and analysis of branching. *Biophys. J.*, 68(4):1551–1560, 1995.

[11] A. G. Barto, R. S. Sutton, and C. W. Anderson. Neuron like adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5):834–846, 1983.

[12] D. A. Beard and H. Qian. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS One*, 2(1):e144, 2007.

[13] Daniel A. Beard and Hong Qian. *Chemical biophysics : quantitative analysis of cellular systems*. Cambridge texts in biomedical engineering. Cambridge University Press, Cambridge ; New York, 2008.

[14] Richard Bellman. The theory of dynamic programming. *Bull. Am. Math. Soc.*, 60(6):503–515, 1954.

[15] Richard Bellman. A markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

[16] Richard Bellman and Rand Corporation. *Dynamic programming*. A Rand Corporation research study. Princeton University Press, Princeton, 1957.

[17] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nature chemical biology*, 5(8):593–9, 2009.

[18] Ginevra Biino, Iolanda Santimone, Cosetta Minelli, Rossella Sorice, Bruno Frongia, Michela Traglia, Sheila Ulivi, Augusto Di Castelnuovo, Martin Gögele, Teresa Nutile, Marcella Francavilla, Cinzia Sala, Nicola Pirastu, Chiara Cerletti, Licia Iacoviello, Paolo Gasparini, Daniela Toniolo, Marina Ciullo, Peter Pramstaller, Mario Pirastu, Giovanni de Gaetano, and Carlo L. Balduini. Age- and sex-related variations in platelet count in italy: a proposal of reference ranges based on 40987 subjects' data. *PLoS One*, 8(1):e54289, 2013.

[19] Samuel Britton, Alber, and W. R. M. Cannon. Machine learning and optimal control of enzyme activities to preserve solvent capacity in the cell. *Submitted*, 2020.

[20] Samuel Britton, Oleg Kim, Francesco Pancaldi, Zhiliang Xu, Rustem I. Litvinov, John W. Weisel, and Mark Alber. Contribution of nascent cohesive fiber-fiber interactions to the non-linear elasticity of fibrin networks under tensile load. *Acta Biomaterialia*, 94:514 – 523, 2019.

[21] André E. X. Brown, Rustem I. Litvinov, Dennis E. Discher, Prashant K. Purohit, and John W. Weisel. Multiscale mechanics of fibrin polymer: gel stretching with protein unfolding and loss of water. *Science*, 325(5941):741–744, 2009.

[22] André E. X. Brown, Rustem I. Litvinov, Dennis E. Discher, and John W. Weisel. Forced unfolding of coiled-coils in fibrinogen by single-molecule afm. *Biophys. J.*, 92(5):L39–41, 2007.

[23] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage dna. *Science*, 265(5178):1599–1600, 1994.

[24] Joseph P. Califano and Cynthia A. Reinhart-King. Substrate stiffness and cell area predict cellular traction stresses in single cells and cells in contact. *Cell. Mol. Bioeng.*, 3(1):68–75, 2010.

[25] Murray Campbell, A. Joseph Hoane, and Feng-hsiung Hsu. Deep blue. *Artificial Intelligence*, 134(1):57–83, 2002.

[26] Robert A. Campbell, Maria Aleman, Laura D. Gray, Michael R. Falvo, and Alisa S. Wolberg. Flow profoundly influences fibrin network structure: implications for fibrin formation and clot stability in haemostasis. *Thromb. Haemost.*, 104(6):1281–1284, 2010.

[27] W. R. Cannon. Simulating metabolism with statistical thermodynamics. *PLoS One*, 9(8):e103582, 2014.

[28] W. R. Cannon and S. J. Benkovic. Solvation, reorganization energy, and biological catalysis. *J Biol Chem*, 273(41):26257–60, 1998.

[29] W. R. Cannon, Samuel Britton, and M. Alber. Maximal entropy modeling of metabolism: Characteristics and comparisons. *In preparation*, 2020.

[30] William Cannon, Jeremy Zucker, Douglas Baxter, Neeraj Kumar, Scott Baker, Jennifer Hurley, and Jay Dunlap. Prediction of metabolite concentrations, rate constants and post-translational regulation using maximum entropy-based simulations with application to central metabolism of neurospora crassa. *Processes*, 6(6):63, 2018.

[31] J. B. Carleton, A. D'Amore, K. R. Feaver, G. J. Rodin, and M. S. Sacks. Geometric characterization and simulation of planar layered elastomeric fibrous biomaterials. *Acta Biomater*, 12:93–101, 2015.

[32] James B. Carleton, Gregory J. Rodin, and Michael S. Sacks. Layered elastomeric fibrous scaffolds: An in-silico study of the achievable range of mechanical behaviors. *ACS Biomater. Sci. Eng.*, 3(11):2907–2921, 2017.

[33] Jr. Carr, M. E. and S. L. Zekert. Measurement of platelet-mediated force development during plasma clot formation. *Am. J. Med. Sci.*, 302(1):13–18, 1991.

[34] Jr. Carr, Marcus E. Development of platelet contractile force as a research and clinical measure of platelet function. *Cell Biochem. Biophys.*, 38(1):55–78, 2003.

[35] Allen T. Chwang and T. Yao-Tsu Wu. Hydromechanics of low-reynolds-number flow. part 2. singularity method for stokes flows. *J. Fluid Mech.*, 67(4):787–815, 1975.

[36] Adrian R. Cioroianu and Cornelis Storm. Normal stresses in elastic networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 88(5):052601, 2013.

[37] Jean-Philippe Collet, Jennifer L. Moen, Yuri I. Veklich, Oleg V. Gorkun, Susan T. Lord, Gilles Montalescot, and John W. Weisel. The alphac domains of fibrinogen affect the structure of the fibrin clot, its physical properties, and its susceptibility to fibrinolysis. *Blood*, 106(12):3824–3830, 2005.

[38] Jean-Philippe Collet, Henry Shuman, Robert E. Ledger, Seungtaek Lee, and John W. Weisel. The elasticity of an individual fibrin fiber in a clot. *Proc. Natl. Acad. Sci. U. S. A.*, 102(26):9133–9137, 2005.

[39] Athel Cornish-Bowden, María Luz Cárdenas, and North Atlantic Treaty Organization. Scientific Affairs Division. *Control of metabolic processes*. NATO ASI series Series A, Life sciences. Plenum Press, New York, 1990.

[40] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[41] M. W. Covert, C. H. Schilling, I. Famili, J. S. Edwards, II Goryanin, E. Selkov, and B. O. Palsson. Metabolic modeling of microbial strains in silico. *Trends Biochem Sci*, 26(3):179–86, 2001.

[42] A. D'Amore, N. Amoroso, R. Gottardi, C. Hobson, C. Carruthers, S. Watkins, W. R. Wagner, and M. S. Sacks. From single fiber to macro-level mechanics: A structural finite-element model for elastomeric fibrous biomaterials. *J Mech Behav Biomed Mater*, 39:146–61, 2014.

[43] Moumita Das, D. A. Quint, and J. M. Schwarz. Redundancy and cooperativity in the mechanics of compositely crosslinked filamentous networks. *PLoS One*, 7(5):e35939, 2012.

[44] T. de Donder and P. van Rysselberghe. *Thermodynamic Theory of Affinity: a Book of Principles*. Stanford university press, 1936.

[45] D. De Martino, M. Figliuzzi, A. De Martino, and E. Marinari. A scalable algorithm to explore the gibbs energy landscape of genome-scale metabolic networks. *PLoS Comput Biol*, 8(6):e1002562, 2012.

[46] R. Dewar. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal of Physics a-Mathematical and General*, 36(3):631–641, 2003.

[47] B. A. Didonna and T. C. Lubensky. Nonaffine correlations in random elastic media. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 72(6 Pt 2):066619, 2005.

[48] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6):1777–82, 2007.

[49] J. S. Edwards and B. O. Palsson. How will bioinformatics influence metabolic engineering? *Biotechnol Bioeng*, 58(2-3):162–9, 1998.

[50] Michael R. Falvo, Oleg V. Gorkun, and Susan T. Lord. The molecular origins of the mechanical properties of fibrin. *Biophys. Chem.*, 152(1-3):15–20, 2010.

[51] Dmitry A. Fedosov, Bruce Caswell, and George Em Karniadakis. A multiscale red blood cell model with accurate mechanics, rheology, and dynamics. *Biophys. J.*, 98(10):2215–2225, 2010.

[52] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

[53] A. M. Feist, M. J. Herrgard, I. Thiele, J. L. Reed, and B. O. Palsson. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–43, 2009.

[54] A. M. Feist and B. O. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using escherichia coli. *Nat Biotechnol*, 26(6):659–67, 2008.

[55] D. A. Fell and H. M. Sauro. Metabolic control and its analysis. additional relationships between elasticities and control coefficients. *Eur J Biochem*, 148(3):555–61, 1985.

[56] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo. equilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res*, 40(Database issue):D770–5, 2012.

[57] Hao Ge and Hong Qian. Physical origins of entropy production, free energy dissipation, and their mathematical representations. *Physical Review E*, 81(5):051133, 2010.

[58] K. C. Gersh, K. E. Edmondson, and J. W. Weisel. Flow rate and fibrin fiber alignment. *J. Thromb. Haemost.*, 8(12):2826–2828, 2010.

[59] C. Giersch. Control analysis of metabolic networks. 1. homogeneous functions and the summation theorems for control coefficients. *Eur J Biochem*, 174(3):509–13, 1988.

[60] R. N. Goldberg, Y. B. Tewari, and T. N. Bhat. Thermodynamics of enzyme-catalyzed reactions–a database for quantitative biochemistry. *Bioinformatics*, 20(16):2874–7, 2004.

[61] P. Gupta, P. Zhang, J. Sheriff, D. Bluestein, and Y. Deng. A multiscale model for recruitment aggregation of platelets by correlating with in vitro results. *Cell Mol Bioeng*, 12(4):327–343, 2019.

[62] Martin Guthold and Samuel S. Cho. Fibrinogen unfolding mechanisms are not too much of a stretch. *Structure*, 19(11):1536–1538, 2011.

[63] Sean R. Hackett, Vito R. T. Zanotelli, Wenxin Xu, Jonathan Goya, Junyoung O. Park, David H. Perlman, Patrick A. Gibney, David Botstein, John D. Storey, and Joshua D. Rabinowitz. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science*, 354(6311), 2016.

[64] W. C. Hallows, W. Yu, and J. M. Denu. Regulation of glycolytic enzyme phosphoglycerate mutase-1 by sirt1 protein-mediated deacetylation. *J Biol Chem*, 287(6):3850–8, 2012.

[65] Yu Long Han, Pierre Ronceray, Guoqiang Xu, Andrea Malandrino, Roger D. Kamm, Martin Lenz, Chase P. Broedersz, and Ming Guo. Cell contraction induces long-ranged stress stiffening in the extracellular matrix. *Proceedings of the National Academy of Sciences*, 115(16):4075–4080, 2018.

[66] J. Hanke, D. Probst, A. Zemel, U. S. Schwarz, and S. Koster. Dynamics of force generation by spreading platelets. *Soft Matter*, 14(31):6571–6581, 2018.

[67] Hado V. Hasselt. *Double Q-learning*, pages 2613–2621. Curran Associates, Inc., 2010.

[68] M. B. Heimlicher, M. Bachler, M. Liu, C. Ibeneche-Nnewihe, E. L. Florin, A. Hoenger, and D. Brunner. Reversible solidification of fission yeast cytoplasm after prolonged nutrient starvation. *J Cell Sci*, 132(21), 2019.

[69] R. Heinrich and T. A. Rapoport. A linear steady-state treatment of enzymatic chains. general properties, control and effector strength. *Eur J Biochem*, 42(1):89–95, 1974.

[70] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophys J*, 92(5):1792–805, 2007.

[71] W. J. Heuett and H. Qian. Grand canonical markov model: A stochastic theory for open nonequilibrium biochemical networks. *Journal of Chemical Physics*, 124(4), 2006.

[72] M. J. Holness and M. C. Sugden. Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem Soc Trans*, 31(Pt 6):1143–51, 2003.

[73] A. Hoppe, S. Hoffmann, and H. G. Holzhutter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst Biol*, 1:23, 2007.

[74] John R. Houser, Nathan E. Hudson, Lifang Ping, 3rd O'Brien, E. Timothy, Richard Superfine, Susan T. Lord, and Michael R. Falvo. Evidence that αc region is origin of low modulus, high extensibility, and strain stiffening in fibrin fibers. *Biophys. J.*, 99(9):3038–3047, 2010.

[75] Nathan E. Hudson, Feng Ding, Igal Bucay, 3rd O'Brien, E. Timothy, Oleg V. Gorkun, Richard Superfine, Susan T. Lord, Nikolay V. Dokholyan, and Michael R. Falvo. Submillisecond elastic recoil reveals molecular origins of fibrin fiber mechanics. *Biophys. J.*, 104(12):2671–2680, 2013.

[76] Nathan E. Hudson, John R. Houser, 3rd O'Brien, E. Timothy, 2nd Taylor, Russell M., Richard Superfine, Susan T. Lord, and Michael R. Falvo. Stiffening of individual fibrin fibers equitably distributes strain and strengthens networks. *Biophys. J.*, 98(8):1632–1640, 2010.

[77] Hubert M. James and Eugene Guth. Theory of the elastic properties of rubber. *J. Chem. Phys.*, 11(10):455–481, 1943.

[78] Paul A. Janmey, Jessamine P. Winer, and John W. Weisel. Fibrin gels and their clinical and bioengineering applications. *J. R. Soc. Interface*, 6(30):1–10, 2009.

[79] C. J. Jen and L. V. McIntire. The structural properties and contractile force of a clot. *Cell Motil.*, 2(5):445–455, 1982.

[80] H. Jerry Qi, Christine Ortiz, and Mary C. Boyce. Mechanics of biomacromolecular networks containing folded domains. *J. Eng. Mater. Technol.*, 128(4):509–518, 2006.

[81] Lisa Jeske, Sandra Placzek, Ida Schomburg, Antje Chang, and Dietmar Schomburg. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research*, 47(D1):D542–D549, 11 2018.

[82] A. Jinich, D. Rappoport, I. Dunn, B. Sanchez-Lengeling, R. Olivares-Amaya, E. Noor, A. Bar Even, and A. Aspuru-Guzik. Quantum chemical approach to estimating the thermodynamics of metabolic reactions. *Scientific Reports*, 4, 2014.

[83] S. Jitrapakdee and J. C. Wallace. Structure, function and regulation of pyruvate carboxylase. *Biochem J*, 340 (Pt 1):1–16, 1999.

[84] T. S. Jung, H. C. Yeo, S. G. Reddy, W. S. Cho, and D. Y. Lee. Webcoli: an interactive and asynchronous web application for in silico design and analysis of genome-scale e.coli model. *Bioinformatics*, 25(21):2850–2, 2009.

[85] H. Kacser and J. A. Burns. The control of flux. *Biochem. Soc. Trans.*, 23(2):341–366, 1995.

[86] Hyeran Kang, Qi Wen, Paul A. Janmey, Jay X. Tang, Enrico Conti, and Fred C. MacKintosh. Nonlinear elasticity of stiff filament networks: strain stiffening, negative normal stress, and filament alignment in fibrin gels. *J. Phys. Chem. B*, 113(12):3799–3805, 2009.

[87] B. N. Kholodenko and H. V. Westerhoff. The macroworld versus the microworld of biochemical regulation and control. *Trends Biochem Sci*, 20(2):52–4, 1995.

[88] Eunjung Kim, Oleg V. Kim, Kellie R. Machlus, Xiaomin Liu, Timur Kupaev, Joshua Lioi, Alisa S. Wolberg, Danny Z. Chen, Elliot D. Rosen, Zhiliang Xu, and Mark Alber. Correlation between fibrin network structure and mechanical properties: an experimental and computational analysis. *Soft Matter*, 7(10):4983–4992, 2011.

[89] M. C. Kim, J. Whisler, Y. R. Silberberg, R. D. Kamm, and H. H. Asada. Cell invasion dynamics into a three dimensional extracellular matrix fibre network. *PLoS Comput Biol*, 11(10):e1004535, 2015.

[90] Min-Cheol Kim, Jordan Whisler, Yaron R. Silberberg, Roger D. Kamm, and H. Harry Asada. Cell invasion dynamics into a three dimensional extracellular matrix fibre network. *PLoS Comput. Biol.*, 11(10):e1004535, 2015.

[91] Oleg V. Kim, Xiaojun Liang, Rustem I. Litvinov, John W. Weisel, Mark S. Alber, and Prashant K. Purohit. Foam-like compression behavior of fibrin networks. *Biomech. Model. Mechanobiol.*, 15(1):213–228, 2016.

[92] Oleg V. Kim, Rustem I. Litvinov, Mark S. Alber, and John W. Weisel. Quantitative structural mechanobiology of platelet-driven blood clot contraction. *Nat. Commun.*, 8(1):1274, 2017.

[93] Oleg V. Kim, Rustem I. Litvinov, John W. Weisel, and Mark S. Alber. Structural basis for the nonlinear mechanics of fibrin networks under compression. *Biomaterials*, 35(25):6739–6749, 2014.

[94] Taeyoon Kim, Wonmuk Hwang, Hyungsuk Lee, and Roger D. Kamm. Computational analysis of viscoelastic properties of crosslinked actin networks. *PLoS Comput. Biol.*, 5(7):e1000439, 2009.

[95] S. Klamt, J. Saez-Rodriguez, and E. D. Gilles. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Syst Biol*, 1:2, 2007.

[96] Matthieu Komorowski, Leo A. Celi, Omar Badawi, Anthony C. Gordon, and A. Aldo Faisal. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.*, 24(11):1716–1720, 2018.

[97] Vijay R. Konda and John N. Tsitsiklis. *Actor-Critic Algorithms*, pages 1008–1014. MIT Press, 2000.

[98] Holger Kress, Ernst H. K. Stelzer, Daniela Holzer, Folma Buss, Gareth Griffiths, and Alexander Rohrbach. Filopodia act as phagocytic tentacles and pull with discrete steps and a load-dependent velocity. *Proc. Natl. Acad. Sci. U. S. A.*, 104(28):11633–11638, 2007.

[99] H. W. Kreuzer-Martin, M. J. Lott, J. R. Ehleringer, and E. L. Hegg. Metabolic processes account for the majority of the intracellular water in log-phase escherichia coli cells as revealed by hydrogen isotopes. *Biochemistry*, 45(45):13622–30, 2006.

[100] Rep Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(1):255, 1966.

[101] A. Kummel, S. Panke, and M. Heinemann. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol*, 2:2006 0034, 2006.

[102] Nicholas A. Kurniawan, Bart E. Vos, Andreas Biebricher, Gijs J. L. Wuite, Erwin J. G. Peterman, and Gijsje H. Koenderink. Fibrin networks support recurring mechanical loads by adapting their structure across multiple scales. *Biophys. J.*, 111(5):1026–1034, 2016.

[103] Victor K. Lai, Spencer P. Lake, Christina R. Frey, Robert T. Tranquillo, and Victor H. Barocas. Mechanical behavior of collagen-fibrin co-gels reflects transition from series to parallel interactions with increasing collagen content. *J. Biomech. Eng.*, 134(1):011004, 2012.

[104] Spencer P. Lake, Mohammad F. Hadi, Victor K. Lai, and Victor H. Barocas. Mechanics of a fiber network within a non-fibrillar matrix: model and comparison with collagen-agarose co-gels. *Ann. Biomed. Eng.*, 40(10):2111–2121, 2012.

[105] Wilbur A. Lam, Ovijit Chaudhuri, Ailey Crow, Kevin D. Webster, Tai-De Li, Ashley Kita, James Huang, and Daniel A. Fletcher. Mechanics and contraction dynamics of single platelets and implications for clot stiffening. *Nat. Mater.*, 10(1):61–66, 2011.

[106] M. Larsson-Raznikiewicz. Kinetic studies on the reaction catalyzed by phosphoglycerate kinase. ii. the kinetic relationships between 3-phosphoglycerate, mgatp2-and activating metal ion. *Biochim Biophys Acta*, 132(1):33–40, 1967.

[107] Giang Le Minh, Alina D. Peshkova, Izabella A. Andrianova, John W. Weisel, and Rustem I. Litvinov. Differential sensitivity of various markers of platelet activation with adenosine diphosphate. *Bionanoscience*, 9(1):53–58, 2019.

[108] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–44, 2015.

[109] Byoungkoo Lee, Xin Zhou, Kristin Riching, Kevin W. Eliceiri, Patricia J. Keely, Scott A. Guelcher, Alissa M. Weaver, and Yi Jiang. A three-dimensional computational model of collagen network mechanics. *PLoS One*, 9(11):e111896, 2014.

[110] D. Y. Lee, H. Yun, S. Park, and S. Y. Lee. Metafluxnet: the management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics*, 19(16):2144–6, 2003.

[111] Dooyoung Lee, Karen P. Fong, Michael R. King, Lawrence F. Brass, and Daniel A. Hammer. Differential dynamics of platelet contact and spreading. *Biophys. J.*, 102(3):472–482, 2012.

[112] Albert L. Lehninger, David L. Nelson, Michael M. Cox, Louis B. Flexner Medical Book Fund., Alumni, Friends Memorial Book Fund., and Rosengarten Family Fund. *Lehninger principles of biochemistry.* W.H. Freeman, New York, 5th edition, 2008.

[113] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2(2):164–168, 1944.

[114] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012.

[115] Xiaojun Liang, Irina Chernysh, Prashant K. Purohit, and John W. Weisel. Phase transitions during compression and decompression of clots from platelet-poor plasma, platelet-rich plasma and whole blood. *Acta Biomater.*, 60:275–290, 2017.

[116] Xin M. Liang, Sangyoon J. Han, Jo-Anna Reems, Dayong Gao, and Nathan J. Sniadecki. Platelet retraction force measurements using flexible post force sensors. *Lab Chip*, 10(8):991–998, 2010.

[117] Bernard B. C. Lim, Eric H. Lee, Marcos Sotomayor, and Klaus Schulten. Molecular basis of fibrin clot elasticity. *Structure*, 16(3):449–459, 2008.

[118] Rustem I. Litvinov, Dzhigangir A. Faizullin, Yuriy F. Zuev, and John W. Weisel. The α-helix to β-sheet transition in stretched and compressed hydrated fibrin clots. *Biophys. J.*, 103(5):1020–1027, 2012.

[119] Rustem I. Litvinov and John W. Weisel. What is the biological and clinical relevance of fibrin? *Semin. Thromb. Hemost.*, 42(4):333–343, 2016.

[120] Rustem I. Litvinov and John W. Weisel. Fibrin mechanical properties and their structural origins. *Matrix Biol.*, 60-61:110–123, 2017.

[121] W. Liu, C. R. Carlisle, E. A. Sparks, and M. Guthold. The mechanical properties of single fibrin fibers. *J. Thromb. Haemost.*, 8(5):1030–1036, 2010.

[122] W. Liu, L. M. Jawerth, E. A. Sparks, M. R. Falvo, R. R. Hantgan, R. Superfine, S. T. Lord, and M. Guthold. Fibrin fibers have extraordinary extensibility and elasticity. *Science*, 313(5787):634, 2006.

[123] Jason W. Locasale. New concepts in feedback regulation of glucose metabolism. *Current Opinion in Systems Biology*, 8:32 – 38, 2018. • Regulatory and metabolic networks • Special Section: Single cell and noise.

[124] A. J. Lotka. Contribution to the energetics of evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 8(6):147–51, 1922.

[125] F. C. MacKintosh, J. Käs, and P. A. Janmey. Elasticity of semiflexible biopolymer networks. *Phys. Rev. Lett.*, 75(24):4425–4428, 1995.

[126] M. Mak, M. H. Zaman, R. D. Kamm, and T. Kim. Interplay of active processes modulates tension and drives phase transition in self-renewing, motor-driven cytoskeletal networks. *Nat Commun*, 7:10323, 2016.

[127] A. Malandrino, X. Trepat, R. D. Kamm, and M. Mak. Dynamic filopodial forces induce accumulation, damage, and plastic remodeling of 3d extracellular matrices. *PLoS Comput Biol*, 15(4):e1006684, 2019.

[128] R. Marcelin. The mechanics of irreversible phenomenon. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, 151:1052–1055, 1910.

[129] Olena O. Marchenko, Sulagna Das, Ji Yu, Igor L. Novak, Vladimir I. Rodionov, Nadia Efimova, Tatyana Svitkina, Charles W. Wolgemuth, and Leslie M. Loew. A minimal actomyosin-based model predicts the dynamics of filopodia on neuronal dendrites. *Mol. Biol. Cell*, 28(8):1021–1033, 2017.

[130] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[131] Donald A. McQuarrie. *Stochastic approach to chemical kinetics*. Methuen's monographs on applied probability and statistics Supplementary review series in applied probability. Methuen, London, 1967.

[132] Christina Cruickshank Miller and James Walker. The stokes-einstein law for diffusion in solution. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(740):724–749, 1924.

[133] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[134] M. W. Mosesson. Fibrinogen and fibrin structure and functions. *J. Thromb. Haemost.*, 3(8):1894–1904, 2005.

[135] Stéphanie Motte and Laura J. Kaufman. Strain stiffening in collagen i networks. *Biopolymers*, 99(1):35–46, 2013.

[136] David R. Myers, Yongzhi Qiu, Meredith E. Fay, Michael Tennenbaum, Daniel Chester, Jonas Cuadrado, Yumiko Sakurai, Jong Baek, Reginald Tran, Jordan C. Ciciliano, Byungwook Ahn, Robert G. Mannino, Silvia T. Bunting, Carolyn Bennett, Michael Briones, Alberto Fernandez-Nieves, Michael L. Smith, Ashley C. Brown, Todd Sulchek, and Wilbur A. Lam. Single-platelet nanomechanics measured by high-throughput cytometry. *Nat. Mater.*, 16(2):230–235, 2017.

[137] E. A. Newsholme and C. Start. *Regulation in Metabolism*. John Wiley & Sons, London, 1973.

[138] Bastian Niebel, Simeon Leupold, and Matthias Heinemann. An upper limit on gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*, 1(1):125–132, 2019.

[139] E. Noor, A. Bar-Even, A. Flamholz, E. Reznik, W. Liebermeister, and R. Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol*, 10(2):e1003483, 2014.

[140] E. Noor, H. S. Haraldsdottir, R. Milo, and R. M. Fleming. Consistent estimation of gibbs energy using component contributions. *PLoS Comput Biol*, 9(7):e1003098, 2013.

[141] M. A. Oberhardt, B. O. Palsson, and J. A. Papin. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*, 5:320, 2009.

[142] H. T. Odum and R. T. Pinkerton. Time's speed regulator: The optimum efficiency for maximum power output in physical and biological systems. *American Scientist*, 43(2), 1955.

[143] D. Oelz and A. Mogilner. Actomyosin contraction, aggregation and traveling waves in a treadmilling actin array. *Physica D*, 318-319:70–83, 2016.

[144] T. E. Oliphant. Python for scientific computing. *Computing in Science Engineering*, 9(3):10–20, 2007.

[145] P. R. Onck, T. Koeman, T. van Dillen, and E. van der Giessen. Alternative explanation of stiffening in cross-linked semiflexible networks. *Phys. Rev. Lett.*, 95(17):178102, 2005.

[146] OpenAi, Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, Jonas Schneider, Nikolas Tezak, Jerry Tworek, Peter Welinder, Lilian Weng, Qiming Yuan, Wojciech Zaremba, and Lei Zhang. Solving rubik's cube with a robot hand. *arXiv [cs.LG]*, 2019.

[147] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010.

[148] J. O. Park, S. A. Rubin, Y. F. Xu, D. Amador-Noguez, J. Fan, T. Shlomi, and J. D. Rabinowitz. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol*, 12(7):482–9, 2016.

[149] B. R. Parry, I. V. Surovtsev, M. T. Cabeen, C. S. O'Hern, E. R. Dufresne, and C. Jacobs-Wagner. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell*, 156(1-2):183–94, 2014.

[150] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[151] J. M. Paulus. Platelet size in man. *Blood*, 46(3):321–336, 1975.

[152] Andrew C. Pearce, Owen J. T. McCarty, Simon D. J. Calaminus, Elena Vigorito, Martin Turner, and Steve P. Watson. Vav family proteins are required for optimal regulation of plcgamma2 by integrin alphaiibbeta3. *Biochem. J*, 401(3):753–761, 2007.

[153] Grace C. Y. Peng, Mark Alber, Adrian Buganza Tepole, William R. Cannon, Suvranu De, Savador Dura-Bernal, Krishna Garikipati, George Karniadakis, William W.

Lytton, Paris Perdikaris, Linda Petzold, and Ellen Kuhl. Multiscale modeling meets machine learning: What can we learn? *Archives of Computational Methods in Engineering*, 2020.

[154] Izabela K. Piechocka, Rommel G. Bacabac, Max Potters, Fred C. Mackintosh, and Gijsje H. Koenderink. Structural hierarchy governs fibrin gel mechanics. *Biophys. J.*, 98(10):2281–2289, 2010.

[155] Izabela K. Piechocka, Karin A. Jansen, Chase P. Broedersz, Nicholas A. Kurniawan, Fred C. MacKintosh, and Gijsje H. Koenderink. Multi-scale strain-stiffening of semiflexible bundle networks. *Soft Matter*, 12(7):2145–2156, 2016.

[156] Prashant K. Purohit, Rustem I. Litvinov, Andre E. X. Brown, Dennis E. Discher, and John W. Weisel. Protein unfolding accounts for the unusual mechanical behavior of fibrin networks. *Acta Biomater.*, 7(6):2374–2383, 2011.

[157] H. Qian. Open-system nonequilibrium steady state: statistical thermodynamics, fluctuations, and chemical oscillations. *J Phys Chem B*, 110(31):15063–74, 2006.

[158] H. Qian, D. A. Beard, and S. D. Liang. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur J Biochem*, 270(3):415–21, 2003.

[159] H. Qian and H. Wang. Continuous time random walks in closed and open single-molecule systems with microscopic reversibility. *EPL*, 76(1):15, 2006.

[160] Yongzhi Qiu, Ashley C. Brown, David R. Myers, Yumiko Sakurai, Robert G. Mannino, Reginald Tran, Byungwook Ahn, Elaissa T. Hardy, Matthew F. Kee, Sanjay Kumar, Gang Bao, Thomas H. Barker, and Wilbur A. Lam. Platelet mechanosensing of substrate stiffness during clot formation mediates adhesion, spreading, and activation. *Proc. Natl. Acad. Sci. U. S. A.*, 111(40):14430–14435, 2014.

[161] L. Rabiner and B. Juang. An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16, 1986.

[162] C. Reder. Metabolic control theory: a structural approach. *J Theor Biol*, 135(2):175–201, 1988.

[163] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson. An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr). *Genome Biol*, 4(9):R54, 2003.

[164] Robbie Rens, Carlos Villarroel, Gustavo Düring, and Edan Lerner. Micromechanical theory of strain-stiffening of biopolymer networks. *arXiv [cond-mat.soft]*, 2018.

[165] Ed Reznik, Dimitris Christodoulou, Joshua E. Goldford, Emma Briars, Uwe Sauer, Daniel Segrè, and Elad Noor. Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. *Cell Reports*, 20(11), 9 2017.

[166] E. Rowe, B. O. Palsson, and Z. A. King. Escher-fba: a web application for interactive flux balance analysis. *BMC Syst Biol*, 12(1):84, 2018.

[167] Gavin A. Rummery and Mahesan Niranjan. *On-line Q-learning using connectionist systems*, volume 37. University of Cambridge, Department of Engineering Cambridge, England, 1994.

[168] A. Saha, S. Connelly, J. Jiang, S. Zhuang, D. T. Amador, T. Phan, R. B. Pilz, and G. R. Boss. Akt phosphorylation and regulation of transketolase is a nodal point for amino acid control of purine synthesis. *Mol Cell*, 55(2):264–76, 2014.

[169] H. M. Sauro, J. R. Small, and D. A. Fell. Metabolic control and its analysis. extensions to the theory and matrix method. *Eur J Biochem*, 165(1):215–21, 1987.

[170] J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. O. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2.0. *Nat Protoc*, 6(9):1290–307, 2011.

[171] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural Netw*, 61:85–117, 2015.

[172] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics*, 48(4):571–585, 1976.

[173] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nat. Methods*, 9(7):671–675, 2012.

[174] I. Schomburg, A. Chang, S. Placzek, C. Sohngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer, and D. Schomburg. Brenda in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in brenda. *Nucleic Acids Res*, 41(Database issue):D764–72, 2013.

[175] A. Sharma, A. J. Licup, K. A. Jansen, R. Rens, M. Sheinman, G. H. Koenderink, and F. C. MacKintosh. Strain-controlled criticality governs the nonlinear mechanics of fibre networks. *Nat. Phys.*, 12:584, 2016.

[176] Tsukasa Shibue, Mary W. Brooks, M. Fatih Inan, Ferenc Reinhardt, and Robert A. Weinberg. The outgrowth of micrometastases is enabled by the formation of filopodium-like protrusions. *Cancer Discov.*, 2(8):706–721, 2012.

[177] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[178] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian

Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[179] D. A. Sivak and M. Thomson. Environmental statistics and optimal regulation. *PLoS Comput Biol*, 10(9):e1003826, 2014.

[180] Jerome Solon, Aynur Kaya-Copur, Julien Colombelli, and Damian Brunner. Pulsed forces timed by a ratchet-like mechanism drive directed tissue movement during dorsal closure. *Cell*, 137(7):1331–1342, 2009.

[181] Andrew M. Stein, David A. Vader, David A. Weitz, and Leonard M. Sander. The micromechanics of three-dimensional collagen-i gels. *Complexity*, 16(4):22–28, 2011.

[182] Cornelis Storm, Jennifer J. Pastore, F. C. MacKintosh, T. C. Lubensky, and Paul A. Janmey. Nonlinear elasticity in biological gels. *Nature*, 435(7039):191–194, 2005.

[183] Richard S. Sutton. Learning to predict by the methods of temporal differences. *Mach. Learn.*, 3(1):9–44, 1988.

[184] Richard S. Sutton. *Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding*, pages 1038–1044. MIT Press, 1996.

[185] Richard S. Sutton and Andrew G. Barto. Reinforcement learning: an introduction mit press. *Cambridge, MA*, 1998.

[186] Richard Stuart Sutton. *Temporal credit assignment in reinforcement learning.* Thesis, 1984.

[187] C. R. Sweet, S. Chatterjee, Z. Xu, K. Bisordi, E. D. Rosen, and M. Alber. Modelling platelet-blood flow interaction using the subcellular element langevin method. *J R Soc Interface*, 8(65):1760–71, 2011.

[188] L. B. Tanner, A. G. Goglia, M. H. Wei, T. Sehgal, L. R. Parsons, J. O. Park, E. White, J. E. Toettcher, and J. D. Rabinowitz. Four key steps control glycolytic flux in mammalian cells. *Cell Syst*, 7(1):49–62 e8, 2018.

[189] I. Thiele and B. O. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93–121, 2010.

[190] D. G. Thomas, S. Jaramillo-Riveri, D. J. Baxter, and W. R. Cannon. Comparison of optimal thermodynamic models of the tricarboxylic acid cycle from heterotrophs, cyanobacteria, and green sulfur bacteria. *J Phys Chem B*, 2014.

[191] Valerie Tutwiler, Hailong Wang, Rustem I. Litvinov, John W. Weisel, and Vivek B. Shenoy. Interplay of platelet contractility and elasticity of fibrin/erythrocytes in blood clot retraction. *Biophys. J.*, 112(4):714–723, 2017.

[192] Mahsa Vahabi, Abhinav Sharma, Albert James Licup, Anne S. G. van Oosten, Peter A. Galie, Paul A. Janmey, and Fred C. MacKintosh. Elasticity of fibrous networks under uniaxial prestress. *Soft Matter*, 12(22):5050–5060, 2016.

[193] T. van Dillen, P. R. Onck, and E. Van der Giessen. Models for stiffening in cross-linked biopolymer networks: A comparative study. *J. Mech. Phys. Solids*, 56(6):2240–2264, 2008.

[194] Anne S. G. van Oosten, Mahsa Vahabi, Albert J. Licup, Abhinav Sharma, Peter A. Galie, Fred C. MacKintosh, and Paul A. Janmey. Uncoupling shear and uniaxial elastic moduli of semiflexible biopolymer networks: compression-softening and stretch-stiffening. *Sci. Rep.*, 6:19270, 2016.

[195] Bart E. Vos, Luka C. Liebrand, Mahsa Vahabi, Andreas Biebricher, Gijs J. L. Wuite, Erwin J. G. Peterman, Nicholas A. Kurniawan, Fred C. MacKintosh, and Gijsje H. Koenderink. Programming the mechanics of cohesive fiber networks by compression. *Soft Matter*, 13(47):8886–8893, 2017.

[196] H. Y. Wang and H. Qian. On detailed balance and reversibility of semi-markov processes and single-molecule enzyme kinetics. *Journal of Mathematical Physics*, 48(1), 2007.

[197] Y. P. Wang, L. S. Zhou, Y. Z. Zhao, S. W. Wang, L. L. Chen, L. X. Liu, Z. Q. Ling, F. J. Hu, Y. P. Sun, J. Y. Zhang, C. Yang, Y. Yang, Y. Xiong, K. L. Guan, and D. Ye. Regulation of g6pd acetylation by sirt2 and kat9 modulates nadph homeostasis and cell survival during oxidative stress. *EMBO J*, 33(12):1304–20, 2014.

[198] Tomonobu M. Watanabe, Hiroshi Tokuo, Kohsuke Gonda, Hideo Higuchi, and Mitsuo Ikebe. Myosin-x induces filopodia by multiple elongation mechanism. *J. Biol. Chem.*, 285(25):19605–19614, 2010.

[199] Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Mach. Learn.*, 8(3):279–292, 1992.

[200] Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. 1989.

[201] E. B. Waygood, J. S. Mort, and B. D. Sanwal. The control of pyruvate kinase of escherichia coli. binding of substrate and allosteric effectors to the enzyme activated by fructose 1,6-bisphosphate. *Biochemistry*, 15(2):277–82, 1976.

[202] E. B. Waygood, M. K. Rayman, and B. D. Sanwal. The control of pyruvate kinases of escherichia coli. ii. effectors and regulatory properties of the enzyme activated by ribose 5-phosphate. *Can J Biochem*, 53(4):444–54, 1975.

[203] E. B. Waygood and B. D. Sanwal. The control of pyruvate kinases of escherichia coli. i. physicochemical and regulatory properties of the enzyme activated by fructose 1,6-diphosphate. *J Biol Chem*, 249(1):265–74, 1974.

[204] J. W. Weisel. Structure of fibrin: impact on clot stability. *J. Thromb. Haemost.*, 5 Suppl 1:116–124, 2007.

[205] John W. Weisel and Rustem I. Litvinov. Mechanisms of fibrin polymerization and clinical implications. *Blood*, 121(10):1712–1719, 2013.

[206] Qi Wen, Anindita Basu, Paul A. Janmey, and A. G. Yodh. Non-affine deformations in polymer hydrogels. *Soft Matter*, 8(31):8039–8049, 2012.

[207] Qi Wen, Anindita Basu, Jessamine P. Winer, Arjun Yodh, and Paul A. Janmey. Local and global deformations in a strain-stiffening fibrin gel. *New J. Phys.*, 9(11):428, 2007.

[208] William Wood, Antonio Jacinto, Richard Grose, Sarah Woolner, Jonathan Gale, Clive Wilson, and Paul Martin. Wound healing recapitulates morphogenesis in drosophila embryos. *Nat. Cell Biol.*, 4(11):907–912, 2002.

[209] Sarah Woolner, Antonio Jacinto, and Paul Martin. The small gtpase rac plays multiple roles in epithelial sheet fusion–dynamic studies of drosophila dorsal closure. *Dev. Biol.*, 282(1):163–173, 2005.

[210] Ziheng Wu, Zhiliang Xu, Oleg Kim, and Mark Alber. Three-dimensional multi-scale model of deformable platelets adhesion to vessel wall in blood flow. *Philos. Trans. A Math. Phys. Eng. Sci.*, 372(2021), 2014.

[211] Adam R. Wufsus, Kuldeepsinh Rana, Andrea Brown, John R. Dorgan, Matthew W. Liberatore, and Keith B. Neeves. Elastic behavior and platelet retraction in low- and high-density fibrin gels. *Biophys. J.*, 108(1):173–183, 2015.

[212] Bin Xu, Ming-Jay Chow, and Yanhang Zhang. Experimental and modeling study of collagen scaffolds with the effects of crosslinking and fiber alignment. *Int. J. Biomater.*, 2011:172389, 2011.

[213] Shixin Xu, Zhiliang Xu, Oleg V. Kim, Rustem I. Litvinov, John W. Weisel, and Mark Alber. Model predictions of deformation, embolization and permeability of partially obstructive blood clots under variable shear flow. *J. R. Soc. Interface*, 14(136), 2017.

[214] Alireza Yazdani and George Em Karniadakis. Sub-cellular modeling of platelet transport in blood flow through microchannels with constriction. *Soft Matter*, 12(19):4339–4351, 2016.

[215] S. Yesudasan and R. D. Averett. Multiscale network model for fibrin fibers and fibrin clot with protofibril binding mechanics. *arXiv preprint arXiv:1808.04036*, 2018.

[216] Byung-Jun Yoon. Hidden markov models and their applications in biological sequence analysis. *Curr. Genomics*, 10(6):402–415, 2009.

[217] Peter Young. *Recursive Least Squares Regression Analysis*, pages 42–54. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[218] Jennifer Zanet, Asier Jayo, Serge Plaza, Tom Millard, Maddy Parsons, and Brian Stramer. Fascin promotes filopodia formation independent of its role in actin bundling. *J. Cell Biol.*, 197(4):477–486, 2012.

[219] Jennifer Zanet, Brian Stramer, Thomas Millard, Paul Martin, François Payre, and Serge Plaza. Fascin is required for blood cell migration during drosophila embryogenesis. *Development*, 136(15):2557–2565, 2009.

[220] Will Zhang, Rana Zakerzadeh, Wenbo Zhang, and Michael S. Sacks. A material modeling approach for the effective response of planar soft tissues for efficient computational simulations. *J. Mech. Behav. Biomed. Mater.*, 89:168–198, 2019.

[221] Artem Zhmurov, Andre E. X. Brown, Rustem I. Litvinov, Ruxandra I. Dima, John W. Weisel, and Valeri Barsegov. Mechanism of fibrin(ogen) forced unfolding. *Structure*, 19(11):1615–1624, 2011.

[222] Artem Zhmurov, Olga Kononova, Rustem I. Litvinov, Ruxandra I. Dima, Valeri Barsegov, and John W. Weisel. Mechanical transition from α-helical coiled coils to β-sheets in fibrin(ogen). *J. Am. Chem. Soc.*, 134(50):20396–20402, 2012.

[223] H. X. Zhou, G. N. Rivas, and A. P. Minton. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics*, 37:375–397, 2008.

[224] S. B. Zimmerman and S. O. Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli. *J Mol Biol*, 222(3):599–620, 1991.

[225] Péter Érdi and J. Tóth. *Mathematical models of chemical reactions : theory and applications of deterministic and stochastic models*. Princeton University Press, Princeton, N.J., 1989.

Appendix A

Appendix - Model Descriptions from Chapters 2 & 3

A.1 Model Description

Points in the simulation follow the Langevin equations for each node i,

$$m_i \ddot{x}_i = F_i - \eta \dot{x}_i + F_i^B, \tag{A.1}$$

where F_i is the deterministic force, $\eta \dot{x}_i$ is the dampening term. Because the network is in a fluid environment with a low flow Reynolds number ($<< 1$), we incorporate the term x_i to account for interaction between the network and the fluid using the Stokes force estimation. Because the network is at a quasi-equilibrium state, we neglect the inertial term, $m_i x_i$, and discretize Eqn. (A.1) in time using a forward Euler scheme:

$$x_i^{n+1} = x_i^n + \frac{dt}{\eta} F_i^n + F_i^{n,B}. \tag{A.2}$$

Here, the superscripts n and $n+1$ refer to the vector quantities for the i^{th} node at subsequent time points n and $n+1$ and dt is the time step. The drag coefficient, η, is calculated as follows. Using the Stokes force on a particle of radius r, η can be estimated by the relation $\eta = 6\pi\mu r$ [35]. Therefore, assuming the following blood viscosity $\mu = 0.004\ Pa.s$ and $r = 0.05\ \mu m$, the related value of is $3.77\ nN.s.m^{-1}$. The choice of r is based on the assumed fiber diameter of 100nm.

The deterministic forces, F_i, acting on each main node or sub-node of the network can be segregated into (1) forces applied externally to the network which generate (2) individual fiber elastic forces and (3) bending forces as a response. The deterministic force is calculated in the form of the negative gradient of the energy, $F_i = -\nabla E_i$, where E_i is the energy associated with each node or sub-node (Eqn. (A.3)). Each fiber in the system is modeled as a nonlinear spring using a generalized Worm-Like-Chain (WLC) model.

We justify this assumption of modeling fibrin fibers using stiff springs by following the previously developed theory [74, 23]. The theory claims that the structural origin of overall fibrin fiber mechanical behavior is a result of the microscopic constituents, namely stretching of unstructured αC-polymers. The generalized WLC estimates the macroscopic force response of fibers by combining the effects of thousands of unstructured polypeptide chains in series and parallel that are present in the fiber [74, 23].

To account for bending deformations of individual fibers, bending springs are then introduced at each triplet of nodes. We, therefore, define the energy of the i^{th} node as follows:

$$E_i = E_i^{EXT} + E_i^{WLC} + E_i^{BEND}, \tag{A.3}$$

with E_i^{EXT}, E_i^{WLC}, and E_i^{BEND} representing the external, WLC, and bending energies, respectively. The energies can be explicitly written as follows [74, 51]:

$$E_i^{WLC} = \sum_j \frac{Nk_BTC}{4P} \frac{3x^2 - 2x^3}{1-x}, x = \frac{x_i - x_j}{C}, \tag{A.4}$$

$$E_i^{BEND} = \sum_{n \in S(i)} \frac{k_b}{2}(\theta_{jik} - \theta_0)^2, \tag{A.5}$$

where k_B is the Boltzmann constant, T is the absolute temperature, C and P are the scaled persistence and contour lengths, x is the strain between the i^{th} and j^{th} node [74]. Scaling the model is done by setting N as the number of fibrin monomers in parallel, and C as the normalized contour length representing the number of fibrin monomers in series following Houser et al [74]. The sum is taken over the neighboring j nodes connected to the i^{th} node. For bending energy, the sum is taken over all triplets, $S(i)$, including node i. jik represents the current angle between node j, k, and i, where $(j, i, k) \in S(i)$, and 0 represents the preferred angle. k_b represents the spring constant.

Stretching of a fiber segment corresponds to change in E_i^{WLC} due to change in x of a WLC model (Eqn. (A.4)). Bending is modeled by using bending springs added to any node with two or more neighbors and characterized by a preferred angle 0. If nodes move, causing a deviation from their original angle, then a change in energy occurs in E_i^{BEND} due to a change in j (Eqn. (A.5)). Figure 2.1 shows an example of a fibrin network (A) about to be pulled by both ends in addition to individual fiber segmentation (Figure 2.1B). The black arrows in (A) show the direction of force applied to the network.

A single fiber between nodes i and j is subdivided using three internal segments between newly introduced sub-nodes (Figure 2.1B). The bending springs connecting nodes in the dotted box indicated in Figure 2.1B are shown in Figure 2.1C. The energy of each

bending spring is due to the three main or sub-nodes involved. θ_{mjk} represents the bending spring between nodes m, j, and k.

A.2 Modeling Network Structure and Model Calibration

A.2.1 Initial Network Generation

Computer generation of the initial fibrin network structure is done by producing a random graph with structural properties extracted from confocal microscopy images of fibrin clots [93]. The matched structural properties, taken from experiments, include network connectivity, fiber length, and fiber density [93, 21].

The details of the network generation algorithm are described in detail in Section A.3.2. Main nodes (branch points) of the fibrin network are uniformly distributed in a cubic domain according to a prescribed fiber density and fibers are stochastically formed between main nodes according to the rules of the algorithm. This stochastic method accurately generates networks of any size over 10 μm^3, shape, or density with the error less than 5%. The error calculation is explained in detail in Section A.3.3.

Fiber lengths and network connectivity in a typical simulated network are compared in Figure A.1 with those in experimentally derived fibrin clots with the density of 1 fiber/μm^3. Two data sets demonstrate close agreement between simulated networks and experimental data.

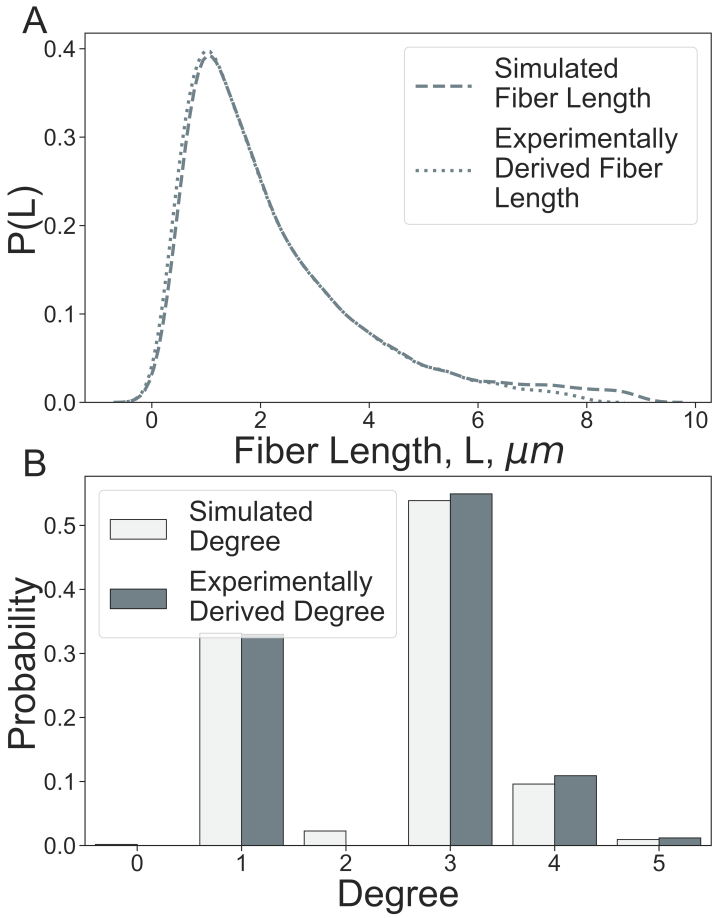

Figure A.1: Length and degree distributions. (A) Fiber length and (B) Network connectivity distributions, $P(L)$ and $P(D)$, for the simulated and experimentally obtained fibrin networks [93] are shown for the fiber density $= 1/\mu m^3$.

A.2.2 Calibration of the Single Fiber Sub-Model

To incorporate fiber-fiber interaction and maintain volume exclusion between fibers, sub-nodes are placed along the length of each fiber at short intervals. The sub-nodes, as well as main-nodes, interact with each another if the distance between them is less than a fiber diameter. Specifically, a new spring is placed between the two interacting nodes. This interaction simulates two properties. First, fibers are not able to occupy the same location since the newly formed spring prevents them from moving into each other's volume. Second, fibers interact with one another in a cohesive manner since the new WLC spring between two nodes prevents separation of the interacting fibers. Distance between each pair of adjacent sub-nodes along a fiber is assumed to be 0.3 μm. If the distance between sub-nodes is larger than 0.3 μm, fibers can pass through one another because the fiber thickness is smaller than the sub-node space. Conversely, decreasing the distance between adjacent sub-nodes increases the overall node count in the simulated network and the computational cost of each simulation. Therefore, the distance of 0.3 μm was chosen as a midpoint to maintain computationally tractable simulations while satisfying the physical constraint of fiber volume exclusion. Conversely, decreasing the distance between adjacent sub-nodes increases the overall node count in the simulated network and the computational cost of each simulation. Therefore, the distance of 0.3 μm was chosen as a midpoint to maintain computationally tractable simulations while satisfying the physical constraint of fiber volume exclusion. Additionally, we performed simulations of fibers under stretching with different numbers of the sub-nodes (with distances between sub-nodes being both smaller and larger than 0.3 μm). We performed simulations using sub-node distances of 0.1-0.5 μm for fiber

densities of 1, 5, and 15 fibers/μm^3. The 5-fold variation in sub-node spacing resulting in a near 2.16, 2.35, and 2.56 fold change in stiffness for the respective fiber densities (Figure A.6).

Elastic properties of the fibers were calibrated as follows. The general WLC modeling approach is used in this paper to simulate the dynamic extension of individual fibrin fibers. Such an approach has been previously shown to provide an adequate analytical expression for modeling the single fibrin fiber mechanics [76]. Using the experimental stress-strain curves for a single fibrin fiber obtained using atomic force microscopy (AFM) [74], the response of an individual fiber to applied stretching force was simulated by fitting the AFM data to the extended WLC-based model.

Values of two parameters, persistence length, and contour length multiplier, in the WLC model were determined by using a recursive least-squares algorithm: persistence length and contour length multiplier [217]. Both the experimental data and resulting fitting curve for a single fiber are shown in Figure A.2.

Table A.1 lists ranges of parameter values that were calculated under the assumption that the temperature is fixed at 300 Kelvin and 200-1100 monomers can fit in the cross-sectional area of a fiber with a diameter 0.1 μm based on an estimated diameter of 3-7 nm for a single monomer [74].

To model fiber bending and cohesion, we add bending springs between sub-nodes to all fibers (see Figure 2.1B). A bending spring is placed at every node triplet (see Figure 2.1C). The force applied to each sub-node in the triplet is calculated using bending spring potential (Eqn. (A.5)). The spring constant is calculated using the relation $B_s = EI$, in

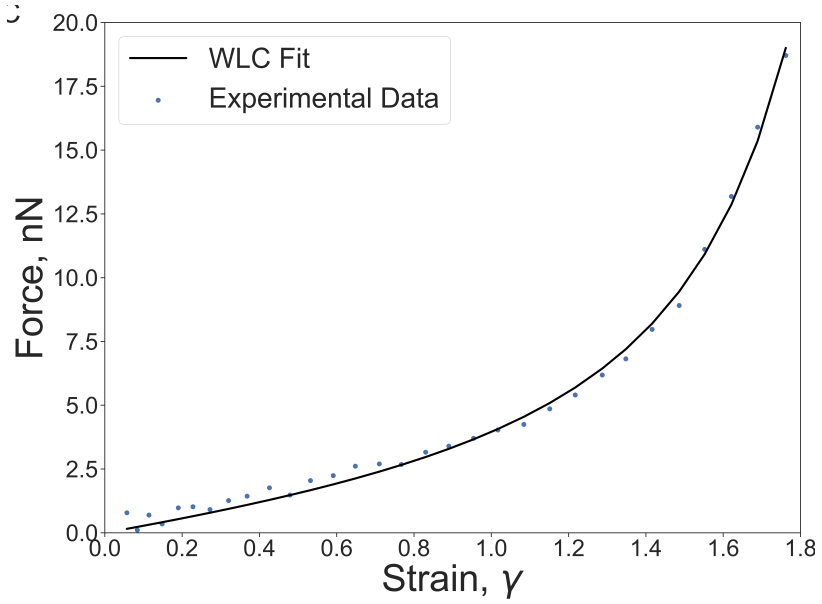

Figure A.2: Model calibration using single fibrin fiber force-strain response curve. Symbols indicate experimental data for a single fiber measured using an atomic force microscope AFM [74]. The black line shows the model fit (Eqn. (A.4)).

which E is Young's modulus, extracted from the WLC fit and I, the moment of inertia for a circle, is calculated using the assumption that all fiber diameters are 100 nm. This calculation yields $B_s = 2.252\,10^{-3} nN\mu m^2$. The same bending constant is utilized at both fiber branch points and along individual fibers due to the similar strength of the two [102].

A.2.3 Calibration of the Single Platelet Sub-Model

Various properties of individual platelets are calibrated using experimental results (Table A.2). The minimal and maximal force range (2.1-29 nN) was chosen based on the lower and upper values of the range of forces measured experimentally [116, 217, 79, 33]. Platelet responses to substrate (fiber) stiffness were incorporated since platelets are known to exert higher forces on stiffer substrates [211, 105]. Several different responses were tested,

Parameter	**Value**
Monomers Per Fiber, Nm	200-1100
Monomer Persistence Length, L_p	0.005-1.074
Absolute Temperature, T	300
Normalized Contour Length Multiplier, C	2.350

Table A.1: Fibrin network model simulation parameters.

ranging from a step function (Section 3.3.1), a constant function and linear function. The linear force response was formulated as a function of the average strain, $\bar{\gamma}$, of the individual fiber being pulled by a filopodia:

$$F = F_0 + \frac{\bar{\gamma}(F_{max} - F_0)}{C}. \tag{A.6}$$

In Eqn. A.6, F_0 represents the minimum force exerted by the platelet and F_{max} represents the force exerted when the fiber strain is equal to the contour length factor, C. The choice of linear response is dependent on the average fiber strain so that when $\bar{\gamma} = 0$ we have $F = F_0$, but when $\bar{\gamma} = C$ we have $F = F_{max}$. Parameters values are listed in Table A.2.

Each filopodia primarily consists of actin and myosin. Actin forms a network distributed in the interior of the platelet that extends up to the membrane to the initial point of each filopodia where myosin is then recruited and stimulates the formation and subsequent elongation of filopodia. Myosin is the motor inducing extension and retraction of the filopodia. Since myosin is independently accumulated at the location of each filopodium [198, 4, 129, 6], the force exerted by each filopodium is assumed to be independent of the

Parameter	Value
Platelet Radius, r_p	1.13 μm [151]
Platelets Number of Filopodia, n_{fil}	2-10 [111, 152]
Platelet Range, r_{fil}	2.18 [92]
Platelet Minimum Force, F_0	2.1 [116, 217, 79, 33]
Platelet Maximum Force, F_{max}	29 [116, 217, 79, 33]
Platelet Volume Fraction, V_p	8.0e-4 per μm^3 [18]

Table A.2: Platelet model simulation parameters.

that exerted by other filopodium.

A.2.4 Modeling Fiber-Fiber Cohesion

Formation of cohesive bonds between fibrin fibers has been previously observed in compressed fibrin clots [93]. As the compression strain increased, the density of the fibrin network increased, the fibers got close to each other, crisscrossed and conglutinated causing a dramatic mechanical reinforcement of the entire clot. Experimental data indicate that stretching of fibrin clots result in fiber-fiber cohesion due to progressive fibrin network densification observed in scanning and transmission electron microscopy images in both lateral and transverse cross-sections of the clots [21], suggesting that the distance between fibrin fibers decreases with the clot strain. Although the molecular mechanism of bond formation between crisscrossed fibers is unknown, the likely mechanisms are electrostatic fibrin-fibrin interactions at lower strains and hydrophobic interactions at higher strains due

to exposure of core residues during the forced unfolding of protein [21]. To model interactions between fibers brought in oblique contact, it is assumed that a cohesive linking is established between the nodes of two neighboring fibers with achievement of a minimal distance. This formed link is assumed to be permanent, which is in agreement with recent experiments [102], and it is characterized by the same set of WLC parameters used for the main and sub-nodes of individual fibers.

The model representation of the fiber-fiber link formation consists of several stages. At each timestep, the distance between disconnected fiber nodes is compared. If the distance between these nodes is less than the fiber diameter of 0.1 μm, a flexible link is established between the nodes to represent a fiber-fiber cohesive bond.

A.2.5 Calculation of the Fibrin Network Stress-Strain Response

To quantify individual fibrin fiber strain, we made use of the initial discretization of fibers. Once the network is generated, individual fibers are divided into small segments by placing sub-nodes along the fiber. For each fiber, the strain is calculated as the average strain over all segments that comprise the fiber. The resulting strain of a fiber, f_i, is then: $s_{f_i} = \frac{1}{N}\sum_{k=1}^{N}\frac{l_c^k - l_0^k}{l_0^k}$, where N is the number of segments comprising the fiber, l_c^k and l_0^k are the current and original length of the k^{th} segment in the i^{th} fiber.

To quantify the stress-strain response of a simulated fibrin network, the external tensile force is applied in a step-wise manner with increments of 0.25 nN at the top and bottom 20% of network nodes at the boundaries of the network (see Figure 2.1A). The subsequent force increment is applied only when the entire network relaxation state is reached. The network relaxation state is found by calculating the maximum network node velocity

defined as $v_m = \max \frac{|x_i^{n+1} - x_i^n|}{dt}$, where the maximum is taken over all network nodes. Since the mean squared displacement of a Brownian particle over a single time step is known to be $< x^2 >= \frac{k_B T}{\eta} dt$, we define the maximal displacement as $B = \phi_i \sqrt{\frac{k_B T}{\eta} dt}$, where ϕ_i is the standard normal sample drawn for node i on the current iteration. The system is in a relaxed state when the maximal velocity is low, i.e. $v_m dt < B + \delta$, where δ is an order of magnitude smaller than the diameter of a single fibrin fiber: 0.01 μm.

A.3 Network Generation Algorithm

First, the domain size is set and the fiber density ρ_f is chosen. Next, a distribution Y is chosen for the fiber connectivity and it is used to calculate the total number of nodes, N. Last, a length distribution X and maximum fiber length L are specified. Each node n_i is given an associated density, ρ_i, by counting the number of nodes n_j with distances satisfying condition: $\text{dist}(n_i, n_j) < L$. In this work, experimentally derived distributions of Y and X were utilized [93].

A.3.1 Node Generation Procedure

Nodes n_i are then ordered in accordance with the decrease in ρ_f. Than each node is given a preferred degree d_i^p from sampling Y. Nodes with the higher ρ_f are given a higher preferred degree while nodes with the lower neighbor density are given a lower preferred degree. A node is considered in need of a neighbor if its current degree d_i^c is such that $d_i^p > d_i^c$.

A.3.2 Fiber Generation Procedure

Next, a fixed fiber density parameter $0.0 < E < 1.0$ is chosen which represents the percentage of the total number of fibers from the distribution Y that will be created. Setting $E = 1.0$ results in $d_i^p = d_i^c$ for all i and without any error in node degree or fiber density. In the network generation algorithm, we set $E = 0.96$, resulting in a less than 5% error in the node degree and fiber density. Namely, fibers will be created until $E \sum_i d_i^p < \sum_i d_i^c$ according to the following algorithm. Node n_i with $d_i^p > d_i^c$ is chosen at random to start a fiber. An end node must then be chosen to complete the fiber. All points n_j with $d_i = \text{dist}(n_i, n_j) < L$ are weighted with weight $X(d_i)$. Next, a single sample s is drawn from X. The node n_j satisfying $|X(d_i) - s|$ is chosen to end the fiber if the fiber between n_i and n_j has not yet been established. If the fiber has already been established, we set $X(d_i) = 0$ and choose a new sample s^* from X, resulting in a new end point for the fiber beginning at n_i. If no node can be found, a new random starting node n_i is chosen again from the nodes such that $d_i^p > d_i^c$. This process of creating fibers continues until the desired density of fibers is reached. Choosing fibers in this manner serves to approximate the target distributions for experimental fiber length and node degree.

A.3.3 Network generation error reduction

To increase the accuracy, nodes are perturbed randomly within the domain. A random move serves to slightly change the associated fiber lengths. To determine fiber length error, a sample is drawn representing each fiber from the experimental distribution. Samples are binned with width (= 0.1) and errors are determined by comparing the frequency in each

bin. Using this method, a node perturbation is accepted if the new fiber lengths lowers the fiber length error. If the fiber alignment distribution Z is prescribed, a condition is added so that a node perturbation must also either decrease the alignment error or maintain a 5% error threshold. Fiber alignment is calculated using the same error method with bins of size 10 degrees. The process continues until both the fiber length and alignment distributions are below 5% error.

A.4 Computational Implementation

The code for this work was implemented in CUDA using the Thrust parallel algorithms library and C++. MATLAB was used to generate initial network structures. Post processing and visualization was performed with a Python statistical visualization library, Seaborn. Code for each project is available. [†] [*]

A.5 Supplementary Figure Descriptions

Figure A.3 below quantifies the change in node degree $(3, 4, > 4)$ for different fiber densities ($\rho_f = 1, 5, 15$). Each node degree type is calculated at various values of the network strain ($\Gamma = 0.1, 1.0, 2.0, 2.5$). At lower fiber densities, $\rho_f = 1, 5$, the fraction of 3-degree nodes decrease as 4 and > 4 degree nodes increase. The reason for this behavior is the new bonds forming between fibers that add 4 and > 4 degree nodes. However, at high density, $\rho_f = 15$, a sharp decrease is seen in 4-degree nodes. This occurs when > 4 degree nodes

[†] https://github.com/sambritton/Fibrin_Network

[*] https://github.com/sambritton/Fibrin_Platelet_Cmake

are created from 4-degree nodes, thereby lowering the fraction of 4-degree nodes. Because of the large number of interacting fibers necessary to convert 4 to higher degree nodes, the behavior only emerges at high density.

Figure A.4 quantifies the change in average fiber alignment and densification as networks are stretched under strain from 0.1 to 3.0. Initial range of the fiber densities range is as follows: ($\rho_f = 1, 5, 15$). Figure A.5 shows the joint distribution of fiber strain and rescaled network position in cohesive and non-cohesive networks for fiber densities $1/\mu m^3$ and $5/\mu m^3$. Figure A.6 variations in stress-strain profiles for different sub-node spacing, δ. Several fiber densities ($\rho_f = 1, 5, 15$) are tested to measure sensitivity of the model due to the spacing parameter.

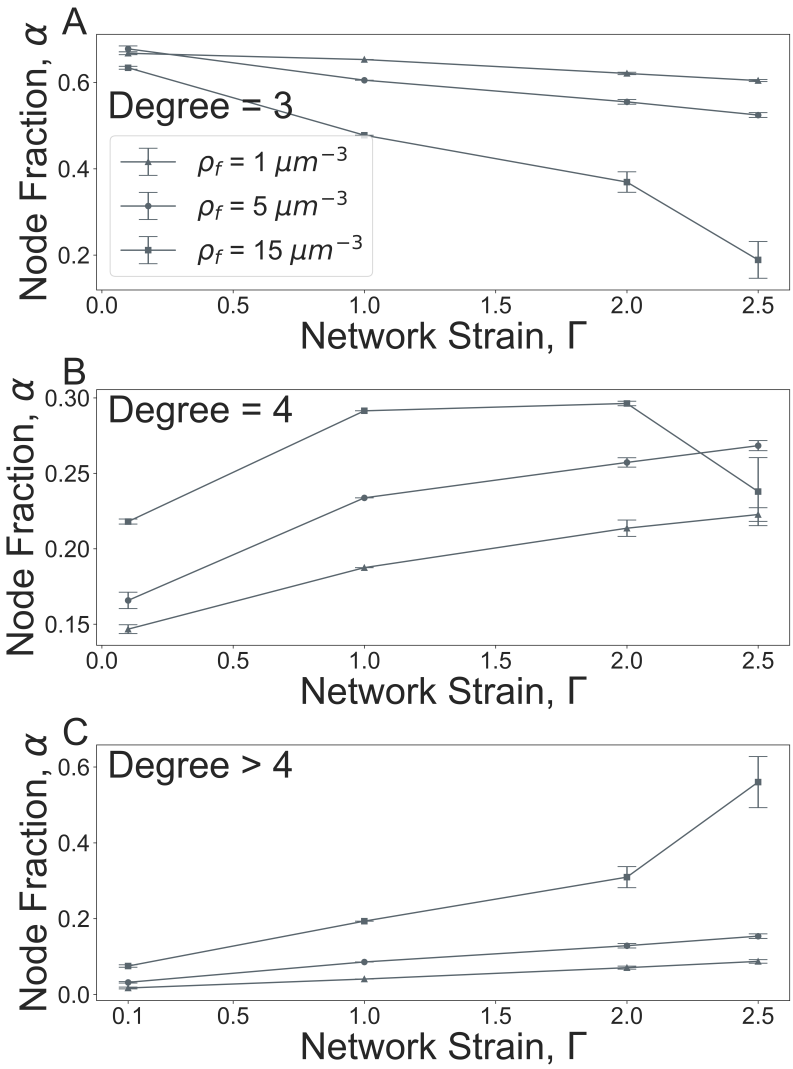

Figure A.3: Connection between node degree and network strain. Average density of 3-degree (A), 4-degree (B) and higher degree (C) nodes are plotted at different network strains: $\Gamma = 0.1 - 2.5$. Fiber density, ρ_f, varies from 1 to 15 fibers/ μm^3.

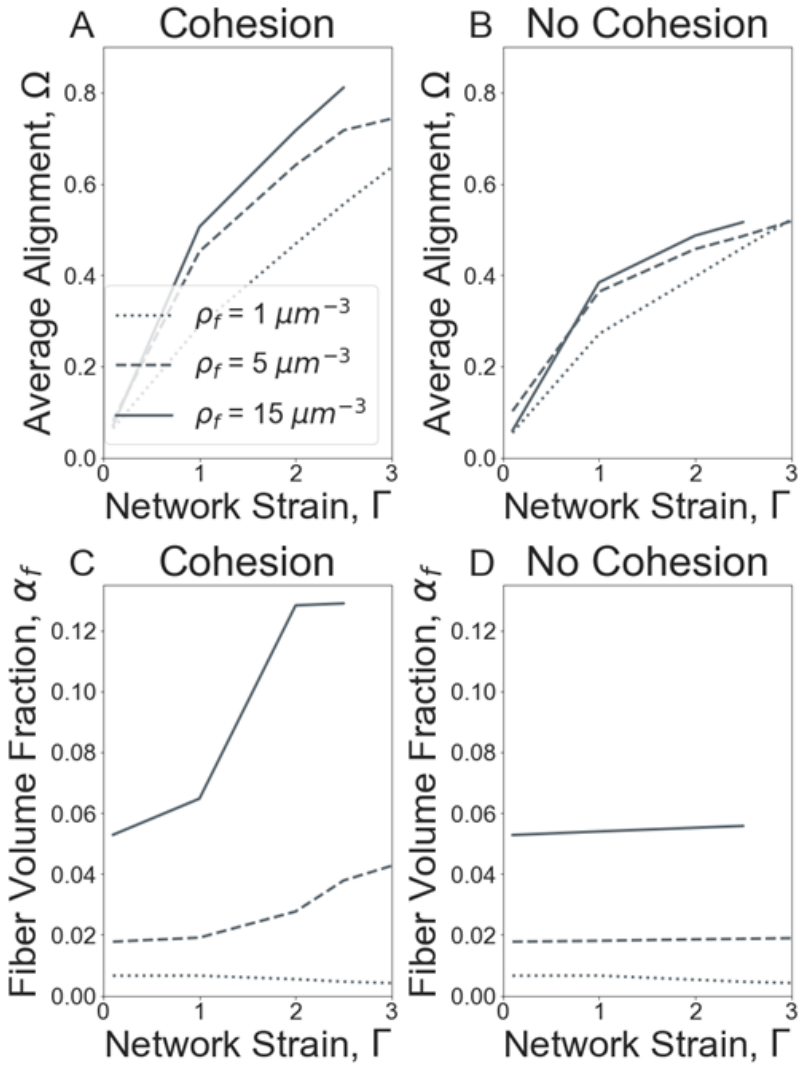

Figure A.4: Comparison between fiber alignment and densification, with and without fiber-fiber cohesive interactions. The average alignment of fibers in the middle 50% of fibrin networks for cohesive (A) and non-cohesive (B) fibers at different fiber densities, ρ_f. Volume fraction occupied by fibers, α_f, inside the central 6 μm of the network for cohesive (C) and non-cohesive (D) fibers for the same densities, ρ_f.

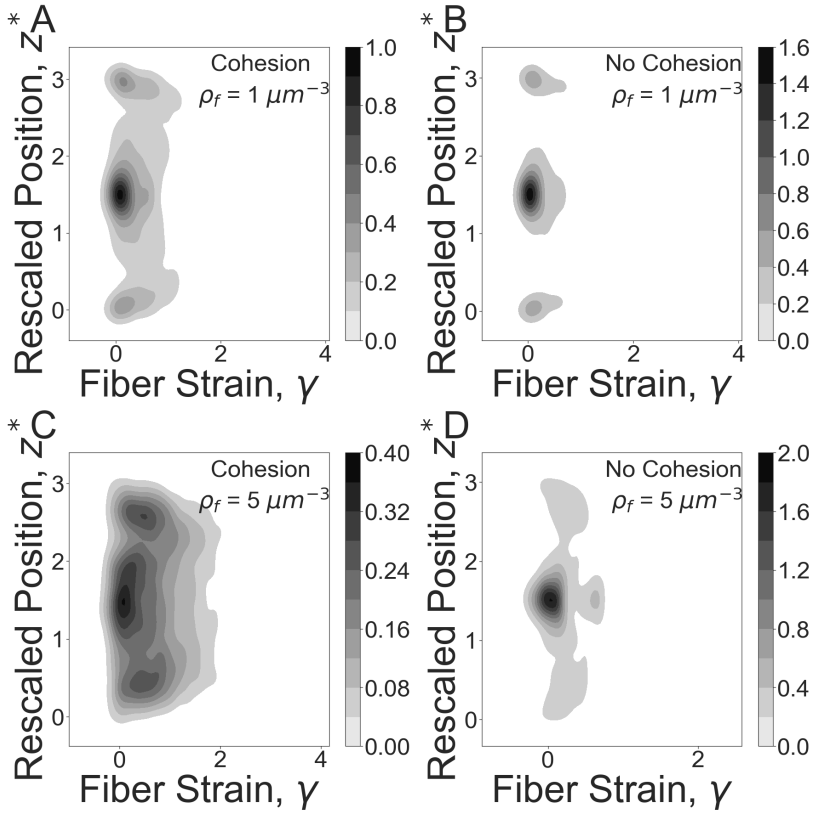

Figure A.5: Connection between cohesion and strain. Individual fiber strain and normalized network strain joint distributions are shown for cohesive and non-cohesive networks at 300% strain for fiber density $1/\mu m^3$ and $5/\mu m^3$. The Y-axis represents the scaled position along the length of the fiber network, $z^* = z/z_0$. z_0 is the original clot length and z ranges over the current network length from bottom ($z = 0$) to top ($z = 3z_0$). Grey scale at each point corresponds to the relative number of fibers at a given strain along the direction of network strain.

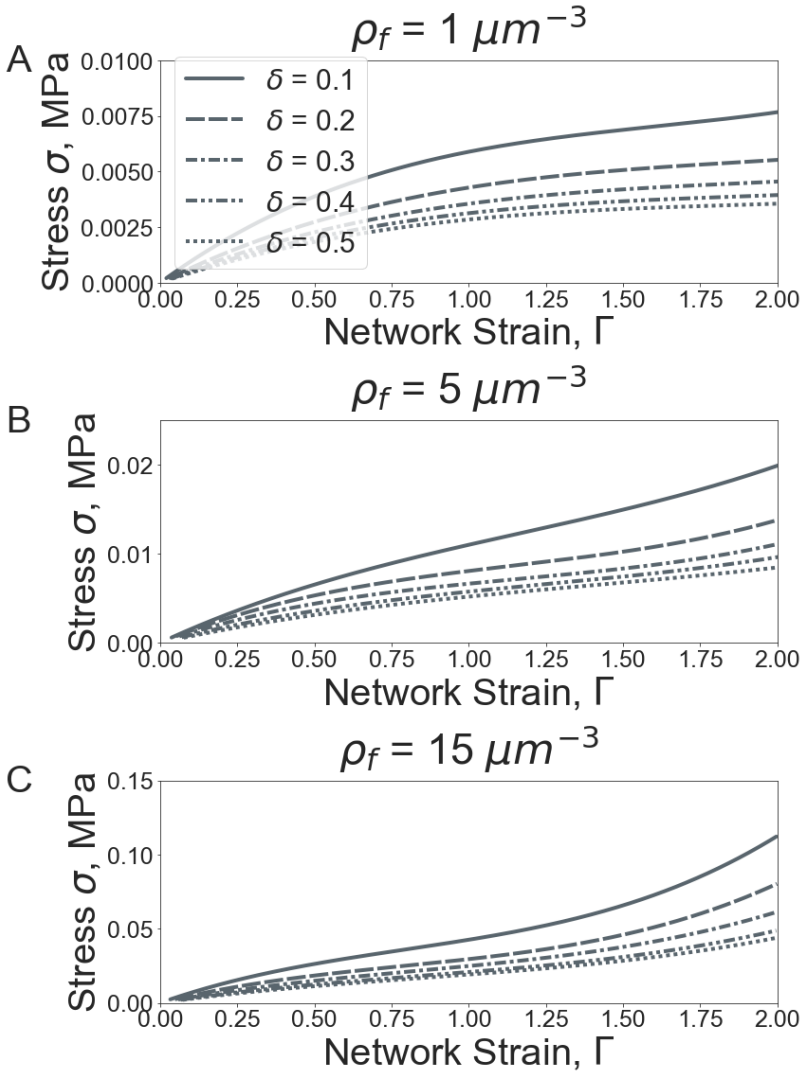

Figure A.6: Stress-strain curves for different sub-node spacing, δ, at varying fiber density, ρ_f. Network density varies from between 1 (A), 5 (B), and 15 (C) fibers/μm^3.

Appendix B

Appendix - Model Description from Chapter 4

B.1 Model Training

Reinforcement learning agents are trained by iteratively learning the value function of each state, s_t. The value at a state represents the expected reward to be achieved from following the current policy: $V(s_t) = E[r_{t:t+n}|s_t]$. At each state with $t \geq n$ the squared error between the value of the state, $V(s_t)$, and the experienced rewards, $r_{t:t+n}$, is back-propagated to calculate appropriate changes in the neural network weights. As agents explore different possible regulation schemes, rewards are accumulated and averaged over each episode of training. Average rewards per episode are shown over the 350 training episodes for the gluconeogenesis and glycolysis-TCA pathways (Figure SB.1A) as well as the glycolysis-PPP-TCA pathway for each of the environmental conditions (Figure SB.1B).

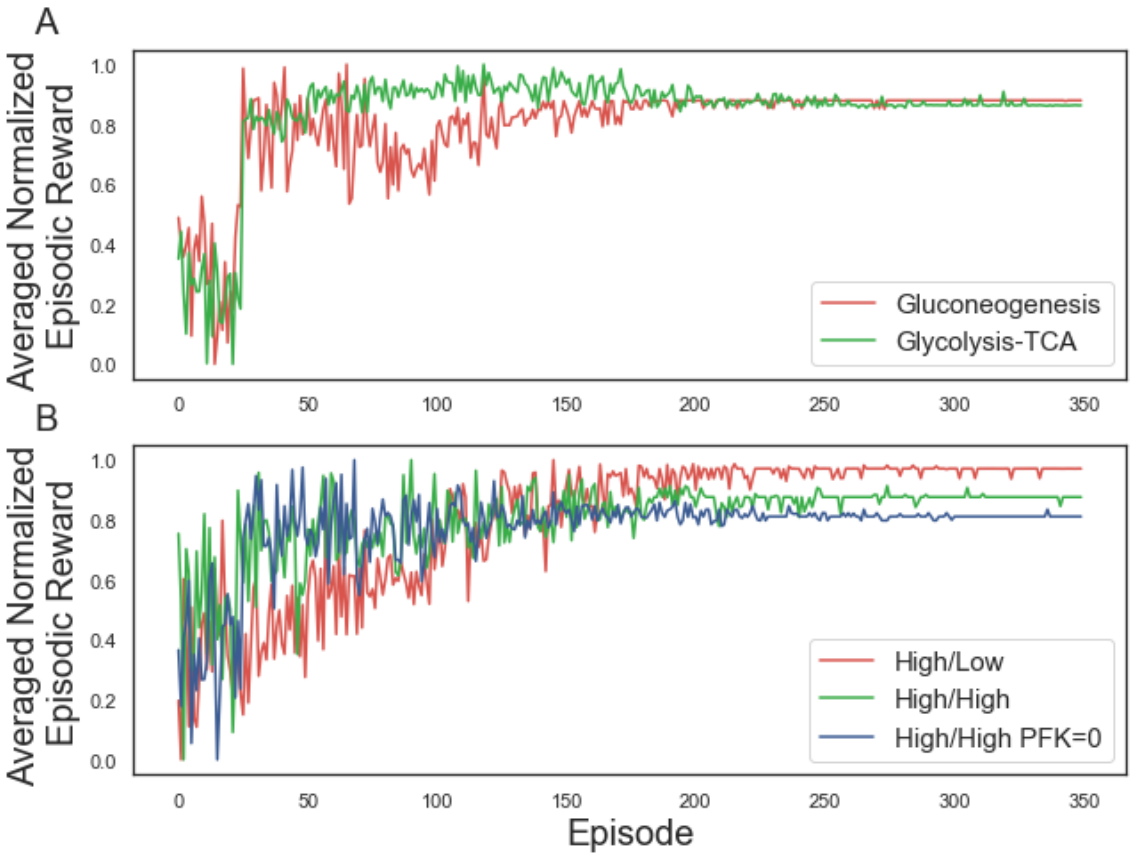

Figure B.1: Cumulative normalized rewards averaged over 10 RL runs for the hyper-parameters (n, lr) which resulted in the maximal reward.

B.2 Calculating Concentration control Coefficients

Enzyme activities begin from a value of 1.0, i.e. the enzyme is unregulated. The current value of the activity is adjusted using Metabolic Control Analysis (MCA) [13]. In MCA, the concentration control coefficient is found by first computing the M by M symmetric linear stability matrix, A^{n^*}, given by,

$$A_{ij}^{n^*} = n_j^* \sum_{k=1}^{Z} S_{ik} \frac{\partial J_k}{\partial n_j}|_{n=n*}. \tag{B.1}$$

Here n_i^* is the current concentration of metabolite i, S is the stoichiometric matrix and J is the vector of reaction fluxes. The concentration control coefficient for metabolite i due to reaction j is then,

$$\begin{aligned} C_{i,j}^n &= \frac{\partial \log n_i}{\partial \log \alpha_j} \\ &= \frac{\alpha_j}{n_i^*} \frac{\Delta n_i}{\Delta \alpha_j} \\ &= -(BS)_{ij} J_j, \end{aligned} \tag{B.2}$$

where $B = A^{-1}$. Note that the calculation of $C_{i,j}^n$ assumes metabolite concentrations depend linearly on enzyme activities. This assumption can be used to isolate the change in activity, j, needed to make a change in the product concentration n_i:

$$\Delta \alpha_j = \alpha_j \frac{\Delta n_i}{n_i^*} \left(- \sum_{k=1}^{Z} B_{ik} S_{kj} J_j \right). \tag{B.3}$$

In practice, when $C_{i,j}^n >> 0.0$ the assumption of a small change n_i used in MCA is no longer valid and instead the current activity j is instead updated using $\alpha_{j,new} = \alpha_{j,current}/5$. As the cost function L (Section 4.2 Eqn. (4.17)) approaches zero, then Eqn. (B.3) can be applied.

B.3 Analysis of Gluconeogenesis Pathway

The gluconeogenesis pathway is analyzed at low NAD/NADH ratio (0.02). The pathway has two known regulation sites fructose 1,6-bisphosphatase (FBP) and pyruvate carboxylase (PC). While both can be utilized to bring steady state metabolite concentrations into agreement with experimentally observed values, regulation of pyruvate carboxylase results in a larger energy dissipation rate (dE/dt). Regulation of alternative enzymes results

in lower flux through the pathway and less energy available for use. Optimal predicted enzyme activities are in agreement for each method (Figure SB.2A). Table B.4 lists the complete reaction activity, flux and free energy for each respective prediction method.

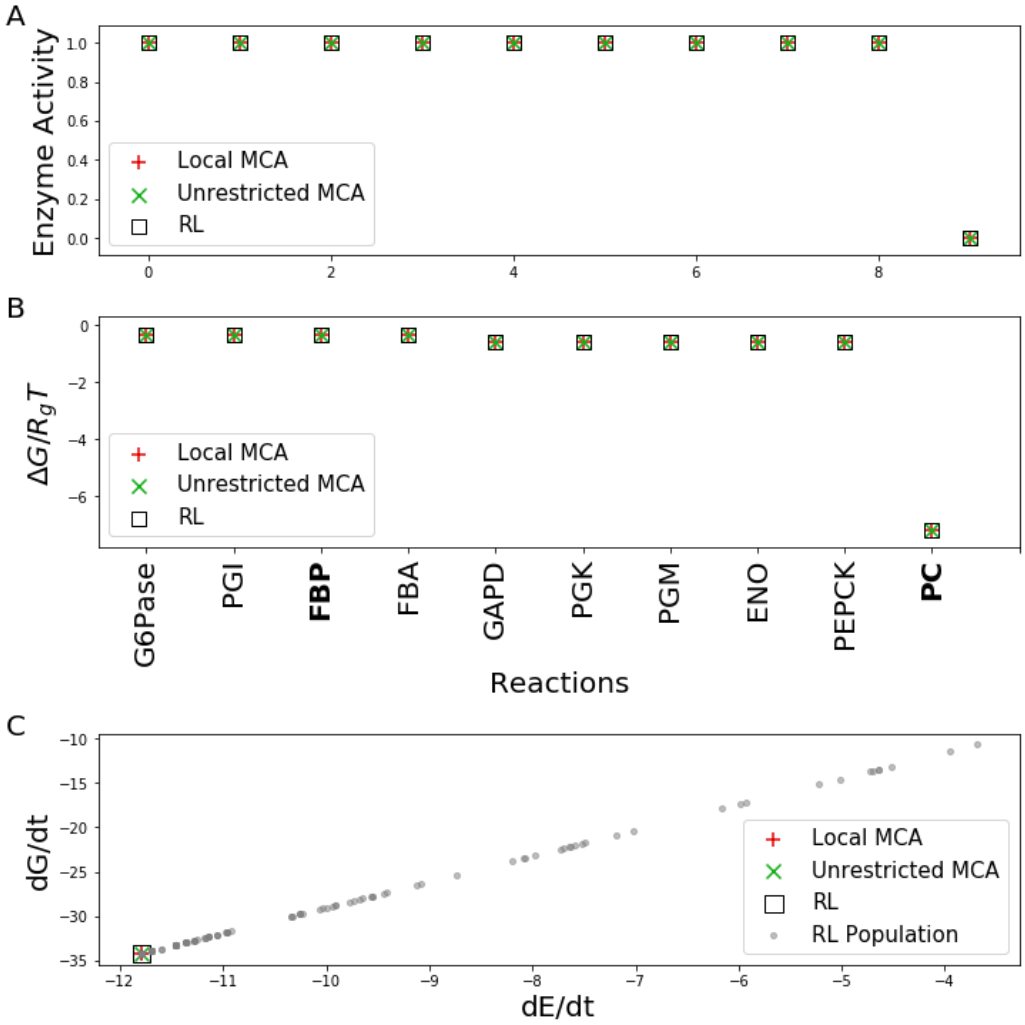

Figure B.2: Gluconeogenesis cycle predictions with low NAD/NADH initial conditions. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent.

B.4 Analysis of Glycolysis-TCA Pathway

The glycolysis-TCA pathway is a subset of the larger glycolysis-PPP-TCA pathway discussed in Chapter 4 which includes the pentose phosphate pathway (PPP). Reducing the number of reactions limits the possible regulation schemes. When utilizing the same initial metabolite concentrations as the larger pathway when appropriate, i.e. high NAD/NADH (31.3), the regulation schemes for various methods show closer agreement. Both HEX1 and GAPD are regulated by every method as in the glycolysis-PPP-TCA pathway. The local MCA method, however, regulates PFK, PGK, and PDH, while the RL method additionally regulates PGI. Both methods regulate more reactions than the unrestriced MCA method and therefore result in a lower energy dissipation rate. Table B.5 lists the complete reaction activity, flux and free energy for each respective prediction method.

B.5 Analysis of Pathways with Proxy Data

When no metabolomics data is available, the methods presented here are still able to perform accurate predictive measurements in terms of enzyme regulation, steady-state metabolite concentrations and reaction flux. Instead of utilizing known metabolomics data measurements, we instead assume the target values of previously measured variable metabolites are fixed at 0.1 mM. Predictive learning is performed for the glycolysis-TCA cycle (Figure B.4) and glycolysis-PPP-TCA pathway under the same three initial conditions (Figures B.5, B.6 and B.7). In all initial conditions, for both pathways, the unrestricted MCA method maintains the same regulation of enzymes GAPD and HEX1. Variations occur only

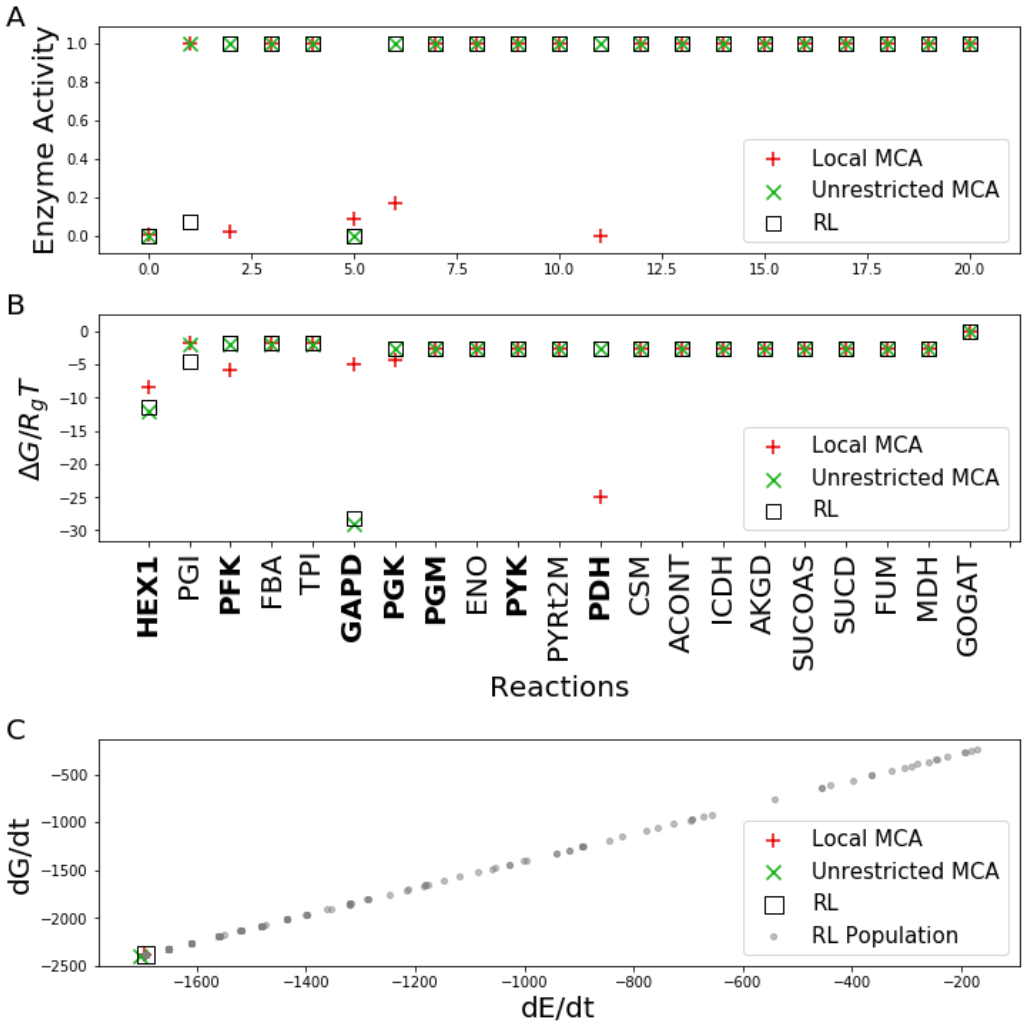

Figure B.3: Glycolysis-TCA cycle predictions with high NAD/NADH initial conditions. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent.

in the amount of regulation applied to the respective reactions. The other two methods show more variation. In the glycolysis-TCA cycle, the local MCA method regulates PYK in addition to the reactions previously regulated, while the RL method predicts additional regulation to PFK and PGM (Figure B.4).

The glycolysis-PPP-TCA pathway, on the other hand, shows more variation. Specifically, in the high NAD/NADH and low NADP/NADPH ratio initial condition, the local MCA regulates PYK in addition to the reactions previously regulated. The RL method predicts increased regulation to PGI and PGM but no longer regulates PDH (Figure B.5). Under the high NAD/NADH and high NADP/NADPH ratio initial conditions, both the local MCA and RL methods predict regulation schemes with additional reduction in activity of PYK. The RL additionally regulates G6PDH, while neither method regulates TKT1 or PYRt2m (Figure B.6). Only slight alterations are observed in enzyme activity when PFK has zero activity (Figure B.7).

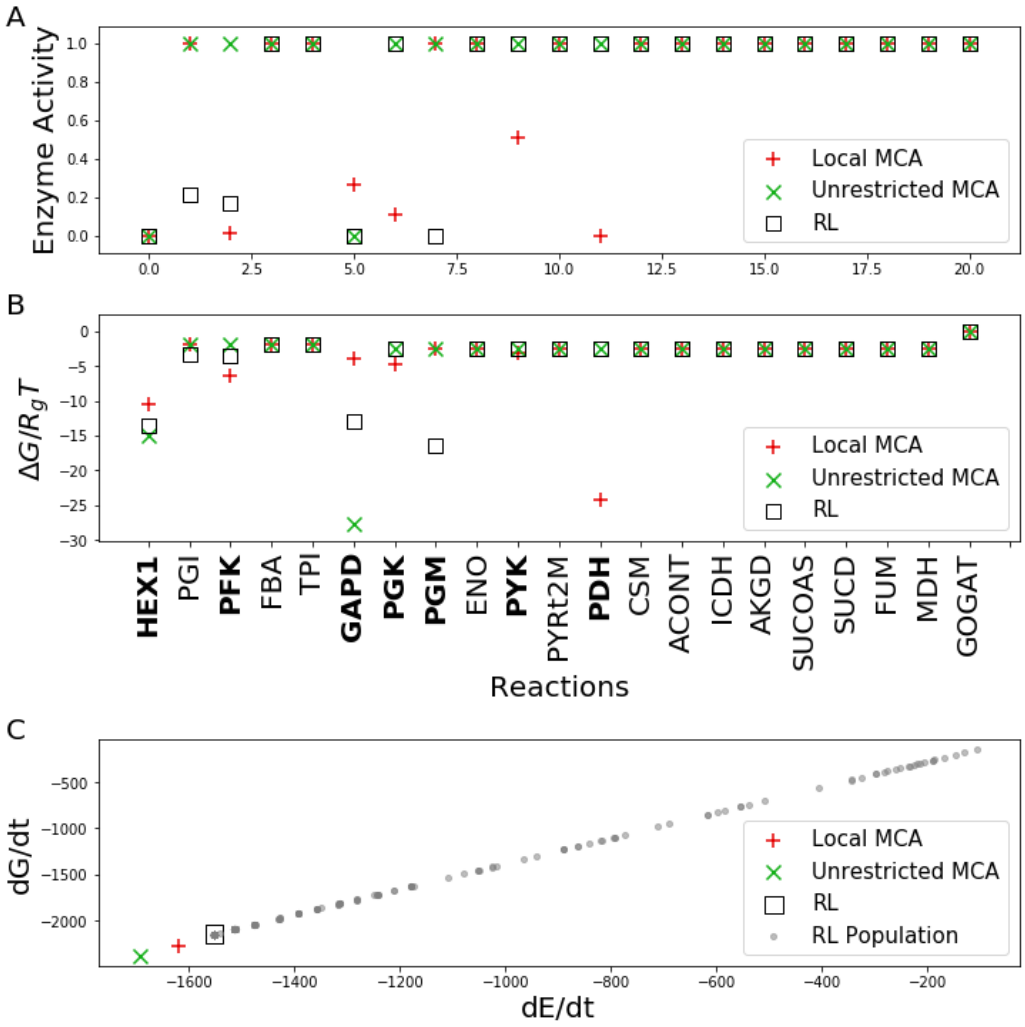

Figure B.4: Glycolysis-TCA cycle predictions with high NAD/NADH initial conditions without experimental metabolomics data. Predicted enzyme activities (A) and free energy (B) at terminal states are calculated using concentration control theory, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). Grey dots (C) represent the population of terminal states found while training the RL agent.

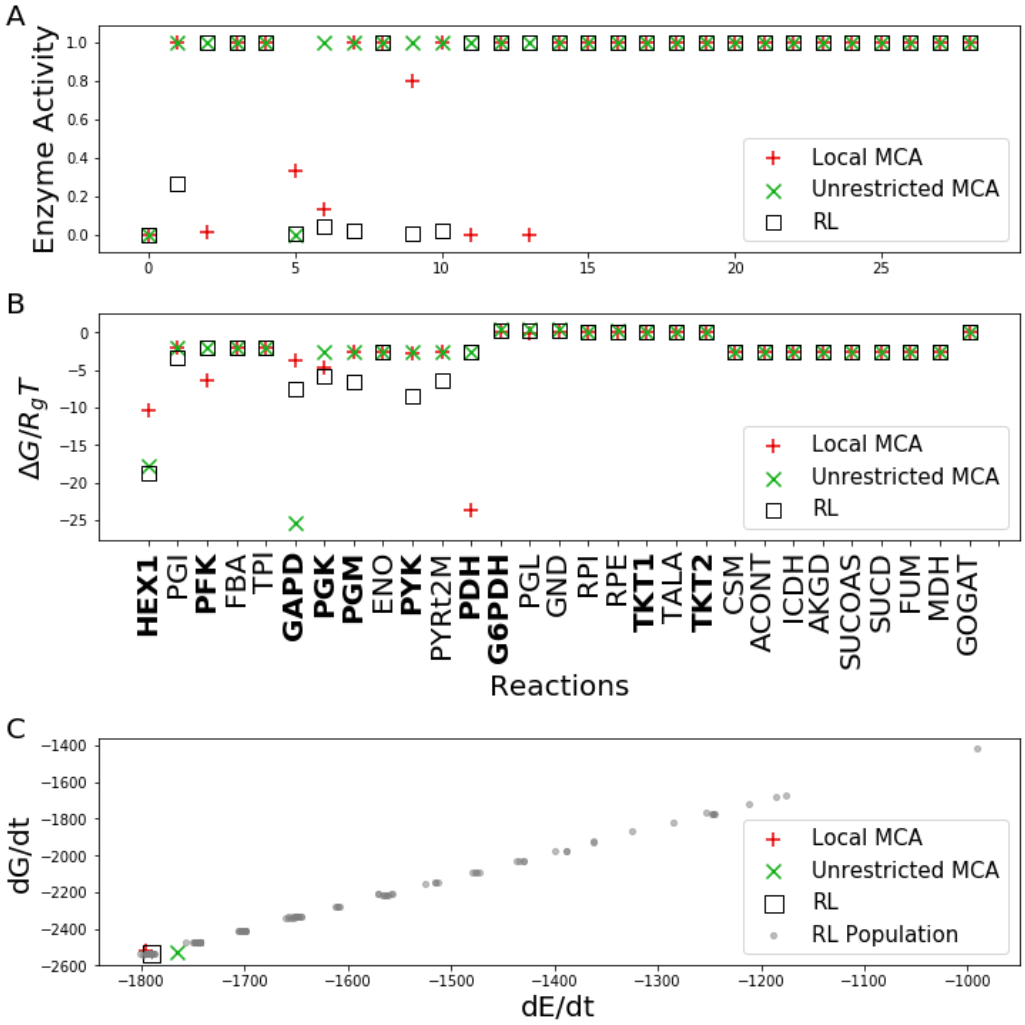

Figure B.5: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and low NADP/NADPH conditions without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent.

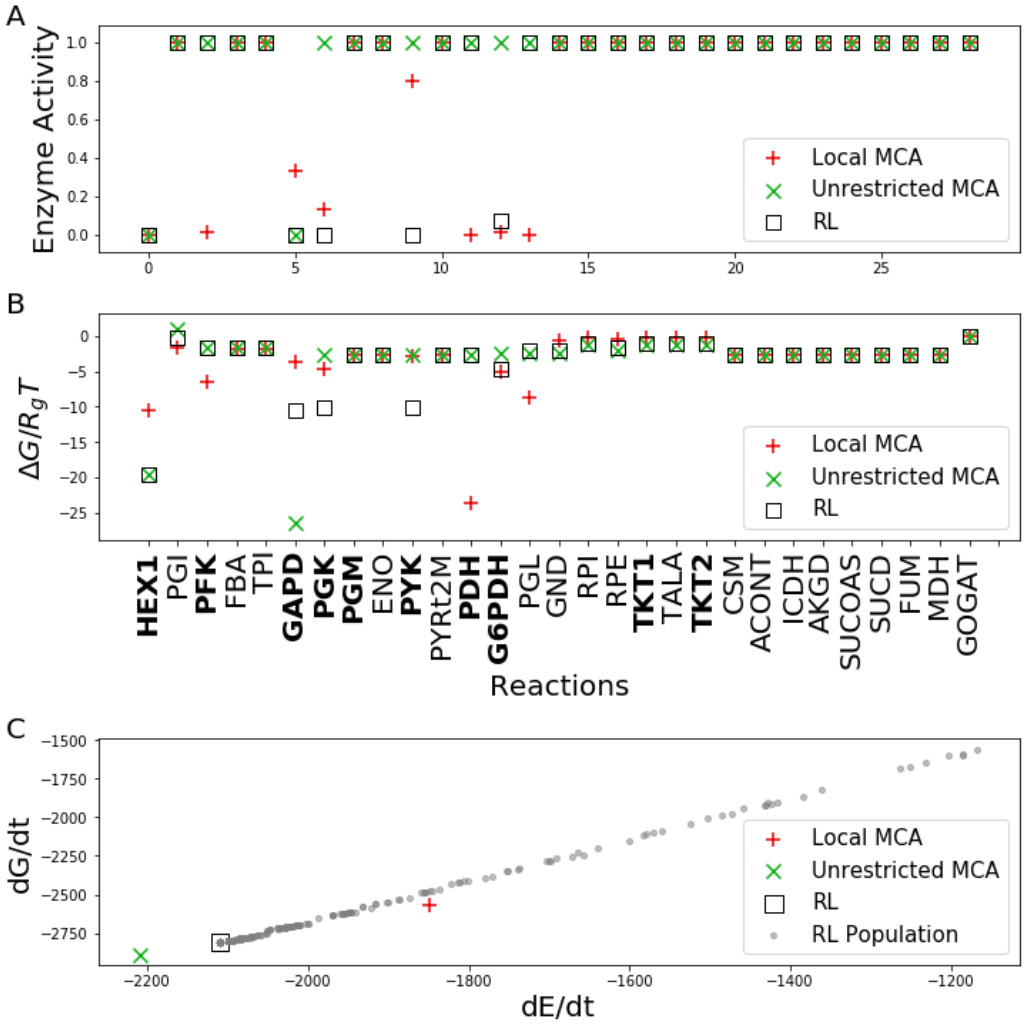

Figure B.6: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent.

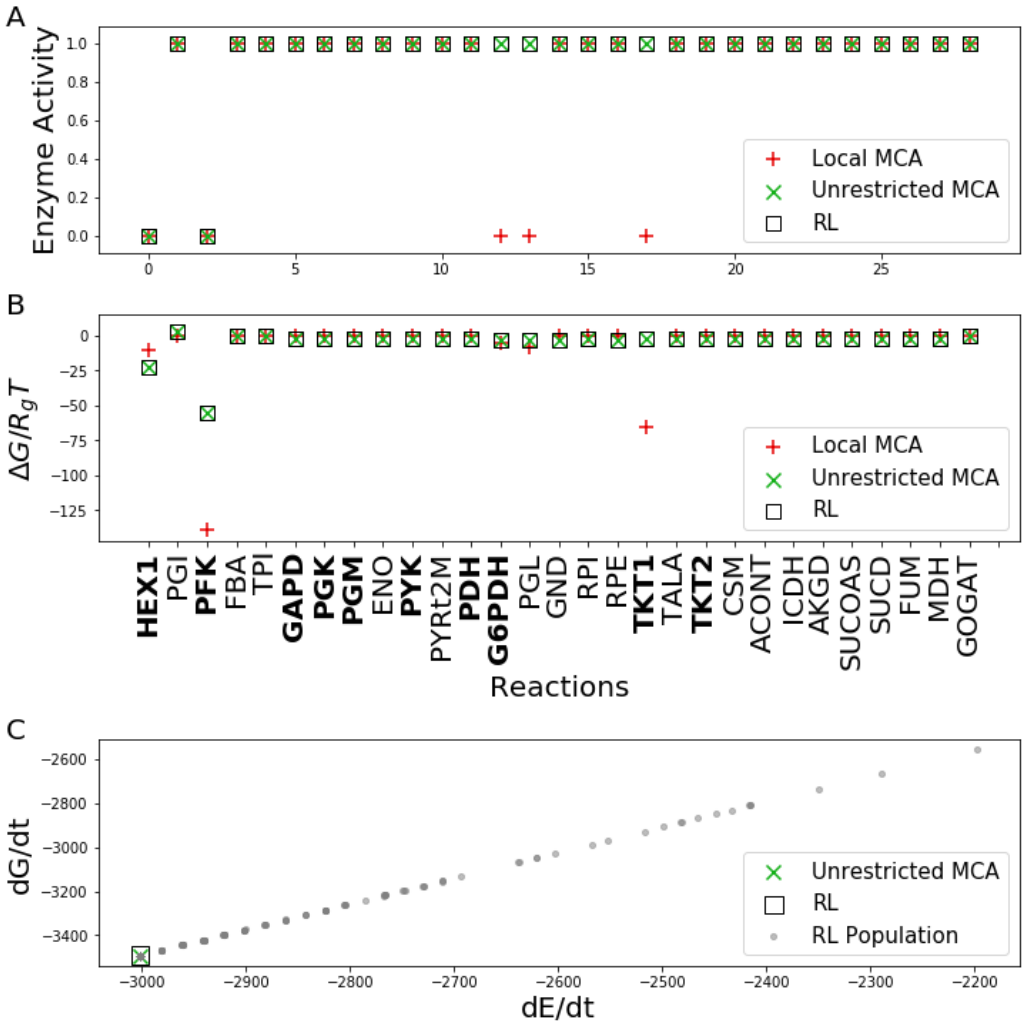

Figure B.7: Glycolysis-PPP-TCA cycle predictions with high NAD/NADH and high NADP/NADPH conditions and PFK activity set to zero without experimental metabolomics data. (A) Predicted enzyme activities at terminal states are calculated using Metabolic Control Analysis, shown as red 'plus's and green 'X's, respectively. Results are compared to those found using a RL approach (black square). (B) Reaction free energies. (C) Free energy and energy dissipation rates. Grey dots represent the population of terminal states found while training the RL agent. The local MCA method results in zero flux as in the case when experimental data is utilized and is therefore not shown.

B.6 Computational Implementation

The code for this work was implemented in Python using the Multiprocessing package for parallelism. Neural networks were implemented using PyTorch [150]. Jupyter notebooks for each respective pathway are available.*

B.7 Tables

Data for steady-state flux, reaction free energy, and enzyme activities at final predicted values is presented for the gluconeogenesis, glycolysis-TCA, and glycolysis-PPP-TCA pathways for various methods of prediction (MCA-Local, MCA, RL).

*https://github.com/sambritton/Max_Entropy_Python

NAD/NADH // NADP/NADPH Ratio		High//Low			High//High			High//High no PFK		
Method		MCA Local	MCA	RL	MCA Local	MCA	RL	MCA Local	MCA	RL
Glycolysis	HEX1	6.34E+00	6.25E+00	6.28E+00	6.27E+00	8.23E+00	8.22E+00	5.76E-03	1.22E+01	1.22E+01
	PGI	6.34E+00	7.05E+00	7.21E+00	6.19E+00	-2.80E+0	-2.40E+00	-1.15E-02	-2.44E+01	-2.44E+01
	PFK	6.34E+00	6.52E+00	6.59E+00	6.24E+00	4.56E+00	4.68E+00	0.00E+00	0.00E+00	0.00E+00
	FBA	6.34E+00	6.52E+00	6.59E+00	6.24E+00	4.56E+00	4.68E+00	1.05E-14	-3.80E-15	-3.80E-15
	TPI	6.34E+00	6.52E+00	6.59E+00	6.24E+00	4.56E+00	4.68E+00	-7.80E-16	6.22E-15	-7.80E-16
	GAPD	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	PGK	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	PGM	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	ENO	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	PYK	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	PYRt2m	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	PDH	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
PPP	G6PDH	2.56E-14	-7.98E-01	-9.30E-01	8.09E-02	1.10E+01	1.06E+01	1.73E-02	3.66E+01	3.66E+01
	PGL	5.27E-33	-7.98E-01	-9.30E-01	8.09E-02	1.10E+01	1.06E+01	1.73E-02	3.66E+01	3.66E+01
	GND	3.18E-14	-7.98E-01	-9.30E-01	8.09E-02	1.10E+01	1.06E+01	1.73E-02	3.66E+01	3.66E+01
	RPI	8.10E-15	-2.66E-01	-3.10E-01	2.70E-02	3.68E+00	3.54E+00	5.76E-03	1.22E+01	1.22E+01
	RPE	1.42E-14	-5.32E-01	-6.20E-01	5.39E-02	7.36E+00	7.08E+00	1.15E-02	2.44E+01	2.44E+01
	TKT1	4.88E-15	-2.66E-01	-3.10E-01	2.70E-02	3.68E+00	3.54E+00	5.76E-03	1.22E+01	1.22E+01
	TALA	9.77E-15	-2.66E-01	-3.10E-01	2.70E-02	3.68E+00	3.54E+00	5.76E-03	1.22E+01	1.22E+01
	TKT2	1.07E-14	-2.66E-01	-3.10E-01	2.70E-02	3.68E+00	3.54E+00	5.76E-03	1.22E+01	1.22E+01
TCA	CSM	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	ACONT	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	ICDH	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	AKGD	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	SUCOAS	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	SUCD	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	FUM	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	MDH	1.27E+01	1.28E+01	1.29E+01	1.25E+01	1.28E+01	1.29E+01	5.76E-03	1.22E+01	1.22E+01
	GOGAT	6.99E-15	6.99E-15	-1.10E-16	-1.10E-16	-1.10E-16	-1.10E-16	-1.10E-16	-1.10E-16	-1.10E-16

Table B.1: Reaction fluxes at predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions.

NAD/NADH // NADP/NADPH Ratio		High//Low			High//High			High//High no PFK		
Method		MCA Local	MCA	RL	MCA Local	MCA	RL	MCA Local	MCA	RL
Glycolysis	HEX1	-8.32E+00	-1.63E+01	-1.79E+01	-8.31E+00	-1.82E+01	-2.00E+01	-8.23E+00	-2.28E+01	-2.28E+01
	PGI	-1.87E+00	-1.97E+00	-1.99E+00	-1.85E+00	1.14E+00	1.01E+00	5.76E-03	3.19E+00	3.19E+00
	PFK	-5.86E+00	-1.90E+00	-1.91E+00	-5.85E+00	-1.56E+00	-1.59E+00	-1.41E+02	-5.49E+01	-5.49E+01
	FBA	-1.87E+00	-1.90E+00	-1.91E+00	-1.86E+00	-1.56E+00	-1.59E+00	-5.30E-15	1.78E-15	1.78E-15
	TPI	-1.87E+00	-1.90E+00	-1.91E+00	-1.86E+00	-1.56E+00	-1.59E+00	3.33E-16	-3.10E-15	3.33E-16
	GAPD	-4.99E+00	-2.71E+01	-1.04E+01	-5.20E+00	-2.82E+01	-1.33E+01	-2.88E-03	-2.51E+00	-2.51E+00
	PGK	-4.32E+00	-2.55E+00	-9.92E+00	-4.31E+00	-2.55E+00	-1.28E+01	-2.88E-03	-2.51E+00	-2.51E+00
	PGM	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	ENO	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	PYK	-2.55E+00	-2.55E+00	-5.90E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	PYRt2m	-2.55E+00	-2.55E+00	-7.02E+00	-2.53E+00	-2.55E+00	-6.13E+00	-2.88E-03	-2.51E+00	-2.51E+00
	PDH	-2.51E+01	-2.55E+00	-3.23E+00	-2.51E+01	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
PPP	G6PDH	-1.30E-14	3.89E-01	4.50E-01	-7.08E+00	-2.41E+00	-2.37E+00	-7.10E+00	-3.60E+00	-3.60E+00
	PGL	-3.84E-01	3.89E-01	4.50E-01	-7.53E+00	-2.41E+00	-2.37E+00	-4.20E+00	-3.60E+00	-3.60E+00
	GND	-1.60E-14	3.89E-01	4.50E-01	-4.04E-02	-2.41E+00	-2.37E+00	-8.64E-03	-3.60E+00	-3.60E+00
	RPI	-4.00E-15	1.33E-01	1.54E-01	-1.35E-02	-1.37E+00	-1.34E+00	-2.88E-03	-2.51E+00	-2.51E+00
	RPE	-7.10E-15	2.63E-01	3.05E-01	-2.70E-02	-2.01E+00	-1.98E+00	-5.76E-03	-3.19E+00	-3.19E+00
	TKT1	-2.40E-15	1.33E-01	1.54E-01	-1.98E+00	-1.37E+00	-1.34E+00	-7.34E+01	-2.51E+00	-2.51E+00
	TALA	-4.90E-15	1.33E-01	1.54E-01	-1.35E-02	-1.37E+00	-1.34E+00	-2.88E-03	-2.51E+00	-2.51E+00
	TKT2	-5.30E-15	1.33E-01	1.54E-01	-1.35E-02	-1.37E+00	-1.34E+00	-2.88E-03	-2.51E+00	-2.51E+00
TCA	CSM	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	ACONT	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	ICDH	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	AKGD	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	SUCOAS	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	SUCD	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	FUM	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	MDH	-2.55E+00	-2.55E+00	-2.56E+00	-2.53E+00	-2.55E+00	-2.56E+00	-2.88E-03	-2.51E+00	-2.51E+00
	GOGAT	-3.60E-15	-3.60E-15	1.11E-16	1.11E-16	1.11E-16	1.11E-16	1.11E-16	1.11E-16	1.11E-16

Table B.2: Reaction free energy at predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions.

NAD/NADH // NADP/NADPH Ratio		High//Low			High//High			High//High no PFK		
Method		MCA Local	MCA	RL	MCA Local	MCA	RL	MCA Local	MCA	RL
Glycolysis	HEX1	1.55E-03	5.02E-07	1.05E-07	1.55E-03	1.55E-03	1.77E-08	1.23E-06	1.52E-09	1.52E-09
	PFK	1.80E-02	1.00E+00	1.00E+00	1.80E-02	1.00E+00	1.00E+00	0.00E+00	0.00E+00	0.00E+00
	GAPD	8.59E-02	2.19E-11	4.06E-04	6.87E-02	7.17E-12	2.23E-05	1.00E+00	1.00E+00	1.00E+00
	PGK	1.68E-01	1.00E+00	6.34E-04	1.68E-01	1.00E+00	3.48E-05	1.00E+00	1.00E+00	1.00E+00
	PYK	1.00E+00	1.00E+00	3.52E-02	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
	PYRt2m	1.00E+00	1.00E+00	1.15E-02	1.00E+00	1.00E+00	2.81E-02	1.00E+00	1.00E+00	1.00E+00
	PDH	1.63E-10	1.00E+00	5.12E-01	1.63E-10	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
PPP	G6PDH	1.00E+00	1.00E+00	1.00E+00	6.81E-05	1.00E+00	1.00E+00	1.43E-05	1.00E+00	1.00E+00
	PGL	5.36E-33	1.00E+00	1.00E+00	4.36E-05	1.00E+00	1.00E+00	2.60E-04	1.00E+00	1.00E+00
	TKT1	1.00E+00	1.00E+00	1.00E+00	3.78E-03	1.00E+00	1.00E+00	7.72E-35	1.00E+00	1.00E+00

Table B.3: Predicted enzyme activities from MCA-local, MCA, and RL methods for the glycolysis-PPP-TCA pathway under different boundary conditions. Unlisted reactions are unregulated.

		Activity			Flux			Energy		
Method		MCA Local	MCA	RL	MCA Local	MCA	RL	MCA Local	MCA	RL
Gluconeogenesis	G6Pase	1.00E+00	1.00E+00	1.00E+00	6.33E-01	6.33E-01	6.33E-01	-3.11E-01	-3.11E-01	-3.11E-01
	PGI	1.00E+00	1.00E+00	1.00E+00	6.33E-01	6.33E-01	6.33E-01	-3.11E-01	-3.11E-01	-3.11E-01
	FBP	1.00E+00	1.00E+00	1.00E+00	6.33E-01	6.33E-01	6.33E-01	-3.11E-01	-3.11E-01	-3.11E-01
	FBA	1.00E+00	1.00E+00	1.00E+00	6.33E-01	6.33E-01	6.33E-01	-3.11E-01	-3.11E-01	-3.11E-01
	GAPD	1.00E+00	1.00E+00	1.00E+00	1.27E+00	1.27E+00	1.27E+00	-5.97E-01	-5.97E-01	-5.97E-01
	PGK	1.00E+00	1.00E+00	1.00E+00	1.27E+00	1.27E+00	1.27E+00	-5.97E-01	-5.97E-01	-5.97E-01
	PGM	1.00E+00	1.00E+00	1.00E+00	1.27E+00	1.27E+00	1.27E+00	-5.97E-01	-5.97E-01	-5.97E-01
	ENO	1.00E+00	1.00E+00	1.00E+00	1.27E+00	1.27E+00	1.27E+00	-5.97E-01	-5.97E-01	-5.97E-01
	PEPCK	1.00E+00	1.00E+00	1.00E+00	1.27E+00	1.27E+00	1.27E+00	-5.97E-01	-5.97E-01	-5.97E-01
	PC	9.90E-04	9.90E-04	9.90E-04	1.27E+00	1.27E+00	1.26E+00	-7.15E+00	-7.15E+00	-7.15E+00

Table B.4: Predicted enzyme activities, fluxes, and free energy for gluconeogenesis pathway from MCA-local, MCA, and RL methods.

		Activity			Flux			Energy		
Method		MCA Local	MCA	RL	MCA Local	MCA	RL	MCA Local	MCA	RL
Glycolysis	HEX1	1.55E-03	3.48E-05	6.81E-05	6.42E+00	6.45E+00	6.41E+00	-8.33E+00	-1.21E+01	-1.15E+01
	PGI	1.00E+00	1.00E+00	6.87E-02	6.42E+00	6.45E+00	6.41E+00	-1.88E+00	-1.89E+00	-4.54E+00
	PFK	1.80E-02	1.00E+00	1.00E+00	6.42E+00	6.45E+00	6.41E+00	-5.88E+00	-1.89E+00	-1.88E+00
	FBA	1.00E+00	1.00E+00	1.00E+00	6.42E+00	6.45E+00	6.41E+00	-1.88E+00	-1.89E+00	-1.88E+00
	TPI	1.00E+00	1.00E+00	1.00E+00	6.42E+00	6.45E+00	6.41E+00	-1.88E+00	-1.89E+00	-1.88E+00
	GAPD	8.59E-02	2.94E-12	7.17E-12	1.28E+01	1.29E+01	1.28E+01	-5.01E+00	-2.91E+01	-2.82E+01
	PGK	1.68E-01	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-4.34E+00	-2.56E+00	-2.56E+00
	PGM	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	ENO	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	PYK	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	PYRt2m	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	PDH	2.04E-10	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.49E+01	-2.56E+00	-2.56E+00
TCA	CSM	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	ACONT	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	ICDH	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	AKGD	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	SUCOAS	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	SUCD	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	FUM	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	MDH	1.00E+00	1.00E+00	1.00E+00	1.28E+01	1.29E+01	1.28E+01	-2.56E+00	-2.56E+00	-2.56E+00
	GOGAT	1.00E+00	1.00E+00	1.00E+00	-1.10E-16	-1.10E-16	-1.10E-16	1.11E-16	1.11E-16	1.11E-16

Table B.5: Predicted enzyme activities, fluxes, and free energy for glycolysis-TCA pathway from MCA-local, MCA, and RL methods.