UNSTRUCTURED, STRUCTURED, AND REDUCED ORDER MODELING OF BIOLOGICAL SYSTEMS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

M.S.

by

Mason Minot

August 2016

ABSTRACT

The advent of recombinant DNA technology and genetic engineering has provided the fields of biology and engineering with a powerful tool set for discovery, rational design, and manipulation of biological systems and processes. Mathematical modeling of these biological systems has the potential to provide valuable insight, to guide experimental design, and to optimize industrial products and processes. Modeling of cell biology, growth, metabolism, and product formation can assist in strain development and fermentation processes. This thesis seeks to utilize chemical engineering principles to model several biological systems using a variety of approaches. Unstructured and structured models are developed to model an industrially relevant riboflavin-producing strain of *Bacillus subtilis* and to maximize product formation based on fed-batch feeding. A new technique is developed that combines cybernetic control variables and flux balance analysis to model cellular metabolism of *Escherichia coli* capable of capturing dynamic behavior during cell growth while reducing computational overhead. A parameter optimization algorithm is developed for the Julia programming language and applied to biological systems. Finally, a reduced order computational model of the complement system, which plays a significant role in inflammation, host defense as well as many disease processes, is developed and validated in through its ability to predict experimental data trends.

BIOGRAPHICAL SKETCH

Mason Minot was born and raised in Orange County, California. He attended the University of California, Irvine from 2011-2014 and was awarded a B.S. in chemical engineering and a B.S. in chemistry. He has held internships in chemical engineering, process development, and process engineering at Zymo Research, the National Renewable Energy Lab, and DuPont Industrial Biosciences. In August of 2014 he began his M.Eng degree at Cornell University. In the spring of the following year he was awarded an M.Eng degree in chemical engineering, was accepted to the M.S. program in chemical engineering, and began working with Professor Jeffrey Varner. In the summer of 2016 he will be working Genentech in South San Francisco.

To my parents, my uncle, and my sister.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Jeffrey Varner, for his guidance and perspective and for the time he has spent training me in systems biology and biochemical engineering. I would also like to thank Mike Vilkhovoy, Adithya Sagar, and David Dai, whose collaboration directly contributed to a majority of this thesis. Lastly, I want to thank all of the members of the Varner lab and the wonderful people I have met at Cornell University for the fantastic experiences and friendships I have been fortunate enough to be a part of.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Unstructured Models of Cell Growth

In 1949 Monod proposed a basic model for cell growth based on Michaelis-Menten kinetics and empirical data that has become a staple in the field of biochemical engineering [52]. The growth of the cell was postulated to be limited by the availability of one or more substrates and could be described by a hyperbolic equation now referred to as the specific growth rate equation [85]. An additionally notable achievement of the mathematics that Monod proposed is the ability to predict diauxie, or growth on multiple substrates. Monod kinetics can be used to model and optimize bioprocesses with low computational overhead in a relatively simple manner [16] [61]. These models, however, assume the cell to be a 'black-box' type system in which the input is substrate and the output is cell mass and extracellular metabolites. One drawback of the Monod approach to modeling cell growth is this oversimplification of the complex nature of cellular metabolism; as a result, kinetic parameters often differ for the same organism and same strain under different culture conditions.

1.2 Structured and Cybernetic Models of Metabolism

Constraints based models are important tools to understand and ultimately to predict how cells utilize nutrients to produce products. Constraints based methods such as flux balance analysis (FBA) [60] and network decomposition ap-

proaches such as elementary modes (EMs) [83] or extreme pathways (EPs) [81] model intracellular metabolism using the biochemical stoichiometry and other constraints such as thermodynamical feasibility under pseudo-steady state conditions. FBA has been used to efficiently estimate the performance of metabolic networks of arbitrary complexity, including genome scale networks, using linear programming [13]. On the other hand, EMs (or EPs) catalog all possible metabolic behaviors such that any flux distribution predicted by FBA is a convex combination of the EMs (or EPs) [110]. However, the calculation of EMs (or EPs) is computationally expensive and currently infeasible for genome scale networks [42].

Cybernetic models have been developed to approach cell metabolism as a resource optimization problem; control variables are used to direct the allocation of resources through the cell in a manner most favorable to cell survival, growth, and product formation [67]. Cybernetic models have been successful in describing a variety of biological systems including diauxic growth [39], growth in low carbon substrate environments [101], simultaneous uptake patterns [55], and growth on complementary substrates [2]. Young [111] developed an optimality framework based on elementary modes representing flux pathways through a metabolic network to describe the batch culture of *E. coli*. Kim et al. (2008) built on the work of Kompala et al. [39] and Young [111] to incorporate a hybrid cybernetic model with elementary modes (HCM-EM) framework [37]. This methodology combined the pseudo steady state assumption of intracellular metabolism and the flux pathways described by elementary modes with dynamic balances describing substrate consumption, product formation, and cell growth. The cybernetic control variables regulate the allocation of substrate resources through specific combinations of elementary modes by weight-

ing Michaelis-Menten type rate equations describing the substrate uptake rate and enzyme synthesis and degredation for each mode.

1.3 JuPOETs Approach to Estimating Biochemical Model Ensembles

Parameter estimation can often be an obstacle in the modeling of biological systems; as the level of detail increases, the model size and number of parameters increase exponentially. Ensemble modeling is a promising approach for obtaining robust predictions and course grained population behavior in deterministic mathematical models. Ensemble approaches address model uncertainty by using parameter or model families instead of single best-fit parameters or fixed model structures. Parameter ensembles can be selected based upon simulation error, along with other criterion such as diversity or steady-state performance. Simulations using parameter ensembles can estimate confidence intervals on model variables, and robustly constrain model predictions, despite having many poorly constrained parameters.

1.4 Reduced Order Modeling of the Complement System

Complement is an important pathway in innate immunity, inflammation, and many disease processes. However, despite its importance, there have been few validated mathematical models of complement activation. The central challenge of complement model identification is the estimation of model parame-

ters from experimental measurements. Unlike other important cascades, such as coagulation where there are well developed experimental tools and publicly available data sets, the data for complement is relatively sparse. Data sets with missing or incomplete data, and limited dynamic data also make the identification of large mechanistic complement models difficult. Thus, reduced order approaches which describe the biology of complement using a limited number of species and parameters could be important for pharmacokinetic model development, and for our understanding of the varied role of complement in the body.

1.5 This Work

This thesis seeks to build upon the advancements in biochemical engineering and systems biology in an attempt to accurately model the fermentation of *B. subtilis* and *E. coli*. Two approaches were taken towards the modeling of *B. subtilis* fermentation and riboflavin production: a classical Monod model describing batch and fed-batch behavior, and an HCM-EM of a network of *B. subtilis* metabolism consisting of glycolysis, the pentose-phosphate pathway, and the TCA cycle. A new technique was developed to model the fermentation of *E. coli* that builds upon the principles used in HCM-EM and utilizes FBA to drastically reduce the number of modes, and therefore, the number of ODEs, kinetic parameters, and computational overhead. The new technique is termed hybrid cybernetic modeling with flux balance analysis or HCM-FBA.

The *B. subtilis* Monod model was capable of fitting to experimental data and model robustness was assessed through parameter perturbation and a sensitivity analysis. The developed Monod model was then used to predict a feed-

ing profile capable of increasing the yield of riboflavin in a fed-batch culture. It was found that an exponential feed improved the titer three-fold. A structured model using elementary modes to describe intracellular metabolism at a pseudo steady state was then developed. A hybrid cybernetic model framework was used to control dynamic flux of substrate through metabolic pathways and describe growth, substrate uptake, and product formation over the course of the modeled experiment. The structured model was able to describe cell mass growth and acetate formation, however, it over predicted the formation of riboflavin. This overestimation is attributed to the large number of ordinary differential equations (ODEs) and estimated kinetic parameters coupled with the need for multiple objective functions. A method is proposed to better train the model to experimental data using the JuPOETS package in the Julia programming language. The development and future validation of a *B. subtilis* HCM-EM has the potential to predict experimentally determined intracellular fluxes and give the metabolic engineer insight into gene modification targets for strain improvement.

HCM-FBA was developed to assess metabolic networks that are too large or computationally expense to analyze with HCM-EM. First, HCM-FBA performance was shown comparable to HCM-EM using a small proof of concept model and a reduced network model of anaerobic *E. coli*. HCM-FBA was then applied to a larger metabolic network of aerobic *E. coli* metabolism which was infeasible for HCM-EM. HCM-EM produced 153,000 elementary modes compared to 29 modes generated with HCM-FBA. HCM-FBA was able to capture the shift from glucose to acetate consumption observed in experimental data that HCM-EM was not able to model. Global sensitivity analysis further reduced the number of FBA modes required to describe the aerobic *E. coli* data,

while maintaining model fit. Thus, HCM-FBA is a promising alternative to HCM-EM for large networks where the generation of elementary modes is infeasible.

JuPOETs, the Pareto Optimal Ensemble Technique in the Julia programming language, was developed as a multiobjective based technique to estimate parameter or model ensembles. JuPOETs integrates simulated annealing with Pareto optimality to estimate ensembles on or near the optimal tradeoff surface between competing training objectives. We demonstrate JuPOETs on a suite of multiobjective problems, including test functions with parameter bounds and system constraints as well as for the identification of a proof-of-concept biochemical model with four conflicting training objectives. JuPOETs identified optimal or near optimal solutions approximately six-fold faster than a corresponding implementation in Octave for the suite of test functions. For the proof-of-concept biochemical model, JuPOETs produced an ensemble of parameters that gave both the mean of the training data for conflicting data sets, while simultaneously estimating parameter sets that performed well on each of the individual objective functions.

Finally, an ensemble of experimentally validated reduced order complement models was developed. Ordinary differential equations were combined with logical rules to produce a compact yet predictive complement model. The model, which described the lectin and alternative pathways, was an order of magnitude smaller than comparable models in the literature. An ensemble of model parameters was estimated from *in vitro* dynamic measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a measurements not used for model training. Despite its small size, the model was surprisingly predictive. Global sensitivity and robust-

ness analysis were performed to elucidate key reactions and binding events in the system. The model described experimental data, and predicted the need for multiple points of therapeutic intervention to fully disrupt complement activation.

CHAPTER 2

UNSTRUCTURED AND STRUCTURED MODELING OF THE RIBOFLAVIN PRODUCING *BACILLUS SUBTILIS* STRAIN RB50::PRF69 AND ITS QUINOL OXIDASE KNOCKOUT MUTANTS

2.1 Introduction

Bacillus subtilis is a model organism used for the industrial production of chemicals in the food and feed industries [17] [18]. It is a gram-positive bacteria commonly found in soil and the gastrointestinal tract of humans. *B. subtilis* also has a number of resilient properties that lend it to being ideal for industrial processes including the ability to ferment in acidic, neutral, and basic environments and the ability of certain strains to withstand extreme temperatures [80]. Previously, strains of *B. subtilis* have been engineered to produce industrially relevant quantities of ionosine and guanosine (20-40 $\frac{g}{L}$) and the genes responsible for riboflavin biosynthesis pathway were sequenced and shown to be organized in a single operon [65] [93]. GTP is a precursor to riboflavin and the ability of the strain to produce it in high quantities stimulated efforts to engineer a strain capable of producing industrial amounts of riboflavin, or vitamin B_2 [4] [20]. The RB50::pRF69 strain was engineered by Perkins et al. (1999) to increase the production of GTP and was equipped with additional copies of the *rib* operon containing the riboflavin biosynthetic genes resulting in a drastic increase in riboflavin yield [65] [112].

Several metabolic engineering studies have been undertaken to characterize and improve the RB50::pRF69 strain [14] [15] [112] . Zamboni et al. hypothesized that improving energy generation efficiency and reducing maintenance

metabolism would increase the production of riboflavin by redirecting electron flow. A knockout study was conducted of two quinol oxidases in the aerobic electron transport chain, aa_3 oxidase and *bd* oxidase, encoded by *qoxA-D* and *cydAB* respectively [112]. The work demonstrated that the *cyd* knockout mutant was capable of growing to higher culture densities and producing a larger amount of riboflavin in chemostat studies than the unmodified RB50::pRF69. It was concluded that the knockout of the less efficient *cyd* quinol oxidase (1 H^+/e^-) redirected energy flow to a more efficient route, one of which may have been the the aa_3 oxidase (*qox*) (2 H^+/e^-) [105] [113]. Knockout of the *qox* quinol oxidase resulted in reduced TCA cycle fluxes indicating an inefficient transport of energy [112]. In an attempt to further characterize these strains and potentially gain insight into areas for either process improvement or genetic modification targets for strain improvement, this study modeled the fermentation of RB50::pRF69 and its mutants using an unstructured model based on Monod kinetics and a structured model of *B. subtilis* metabolism using a hybrid cybernetic framework combined with elementary modes.

Monod kinetics were used to model the batch culture of RB50::pRF69 and the *cyd* and *qox* mutants and the fed-batch culture of RB50::pRF69. The model was composed of 5 ODEs and 7 kinetic parameters that were fit to describe experimental data. For the batch culture simulations, ensembles of randomly perturbed parameter sets about the best fit were generated to assess model robustness. A global sensitivity analysis using the python package SALib and Sobol methodology [27] was performed to determine the parameters that influenced the model the greatest. It was determined that specific growth rate had the largest overall sensitivity. The developed model was then applied to fed-batch growth of RB50::pRF69 and fit to experimental data. The estimated kinetic

parameters were then fixed and code was generated to optimize riboflavin yield by varying the feeding profile. The model determined an exponential feed rate would yield higher culture density and a three-fold improvement in riboflavin production. Although the Monod model is capable of describing experimental data and fermentation processes, it fails to take into account the *B. subtilis* intracellular metabolism, drastically simplifying the complexity of the organism.

To computationally analyze *B. subtilis* intracellular metabolism a hybrid cybernetic model using elementary modes (HCM-EM) based on the methodology of Kim et al (2008) was developed to describe the batch culture of *B. subtilis* strain RB50::pRF69 [37]. The metabolic network considered was adapted from Dauner & Sauer [15] and consisted of glycolysis, the pentose phosphate pathway, and the TCA cycle (Figure 2.1). The cell mass reaction applied to the network was adapted from Dauner, Bailey, and Sauer (2001) [14].

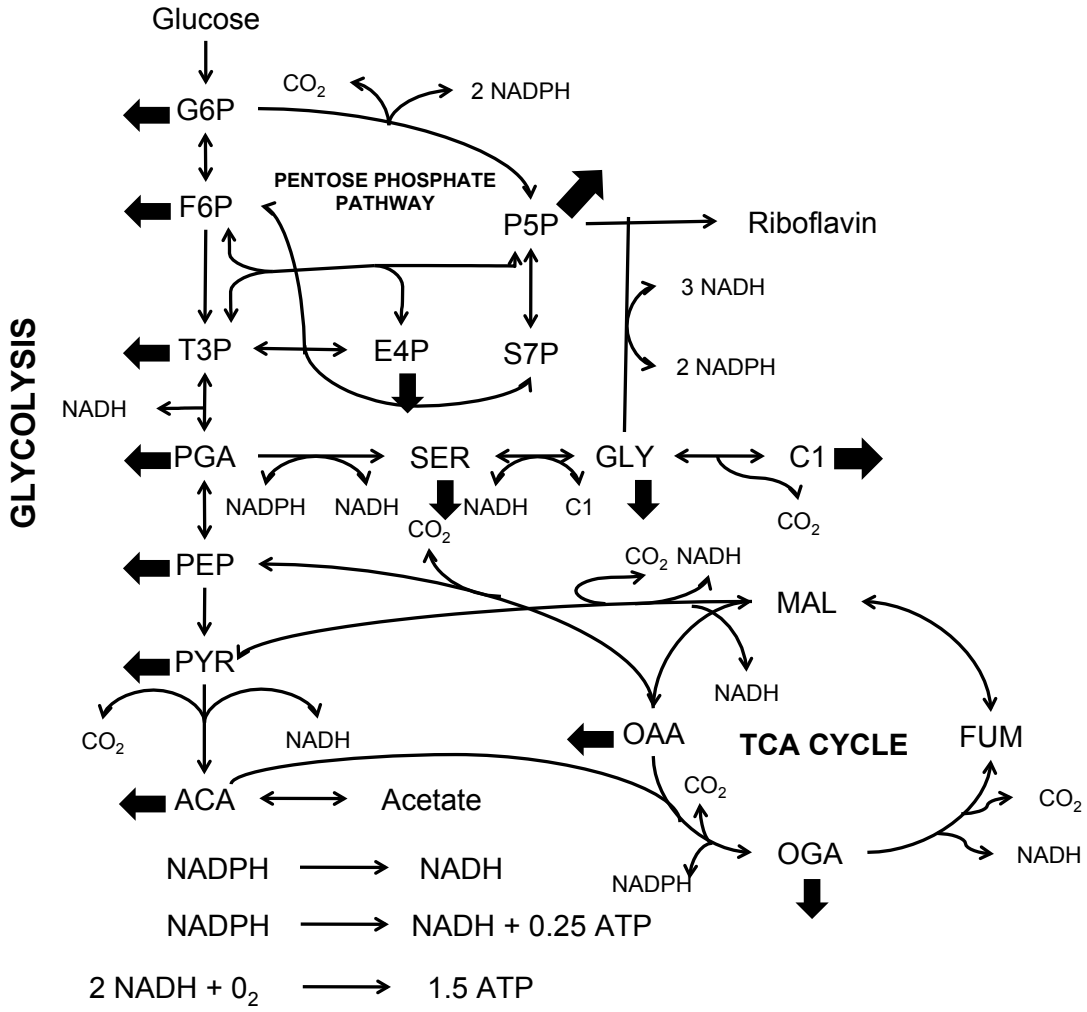

Figure 2.1: *B. subtilis* reaction network adapted from Dauner & Sauer (2001) [15]. Network includes 32 reactions involving glycolysis, pentose-phosphate pathway, and TCA cycle. Boldface arrows indicate that chemical species are reactants for the formation of biomass. Chemical species abbreviations: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; T3P, triose-3-phosphates; PGA, 3-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; ACA; acetyl coenzyme A; P5P, pentose-5-phosphates; S7P, seduheptulose-7-phosphate; E4P, erythrose-4-phosphate; SER, serine; GLY, glycine; C1, methyl group bound to tetrahydrofolate; OAA, oxaloacetate; OGA, oxoglutarate; FUM, fumarate; MAL, malate.

METATOOL 5.1[34] was used to generate 170 elementary modes and an HCM-EM framework was generated consisting of 174 ODEs and 386 kinetic parameters. The developed model is capable of describing batch culture of

RB50::pRF69 with respect to cell growth and acetate formation, however, it overestimates the production of riboflavin. This overestimation is likely due to the large network size, number of estimated kinetic parameters, and ODEs coupled with the need for multiple objective functions. The Julia programming language package JuPOETs is proposed as a way to better fit the model to experimental data in the future.

2.2 Results and Discussion

2.2.1 Monod Kinetics for Batch Culture of RB50::pRF69 and the *cyd* and *qox* Mutants

Monod kinetics were used to model the batch culture of *B. subtilis* strain RB50::pRF69 and its *cyd* and *qox* mutants. The model, composed of 5 ODEs and 7 kinetic parameters, was fit to experimental data from Zamboni's dissertation (2003) [112] by minimizing the sum of the squared residuals between simulated and experimental results. An ensemble of 1,600 parameter variants was generated by randomly perturbing the best fit parameter set for each strain to demonstrate model robustness. Best fit and ensemble results are shown in Figure 2.2 for RB50::pRF69 and its two mutant strains. Best fit parameter values can be found below in Table 2.1.

Parameter	RB50::pRF69	*cyd* Mutant	*qox* Mutant
$\mu\ (hr^{-1})$	0.713	0.838	0.933
$K_s\ (\frac{mmol,glc}{L})$	0.001	0.003	0.084
$k_d\ (hr^{-1})$	0.239	0.367	0.442
$q_{m,ace}\ (\frac{mmol,ace}{OD})$	0.511	1	2.88
$Y_{ace,glc}\ (\frac{mmol,ace}{mmol,clc})$	4.02	2.67	8.28
$Y_{ribo,glc}\ (\frac{mmol,ribo}{mmol,glc})$	8.16	3.64	4.22
$Y_{bio,glc}\ (\frac{OD}{mmol,glc})$	1.70	2.96	2.03

Table 2.1: Monod Best Fit Kinetic Parameters. Estimated using MATLAB function fmincon

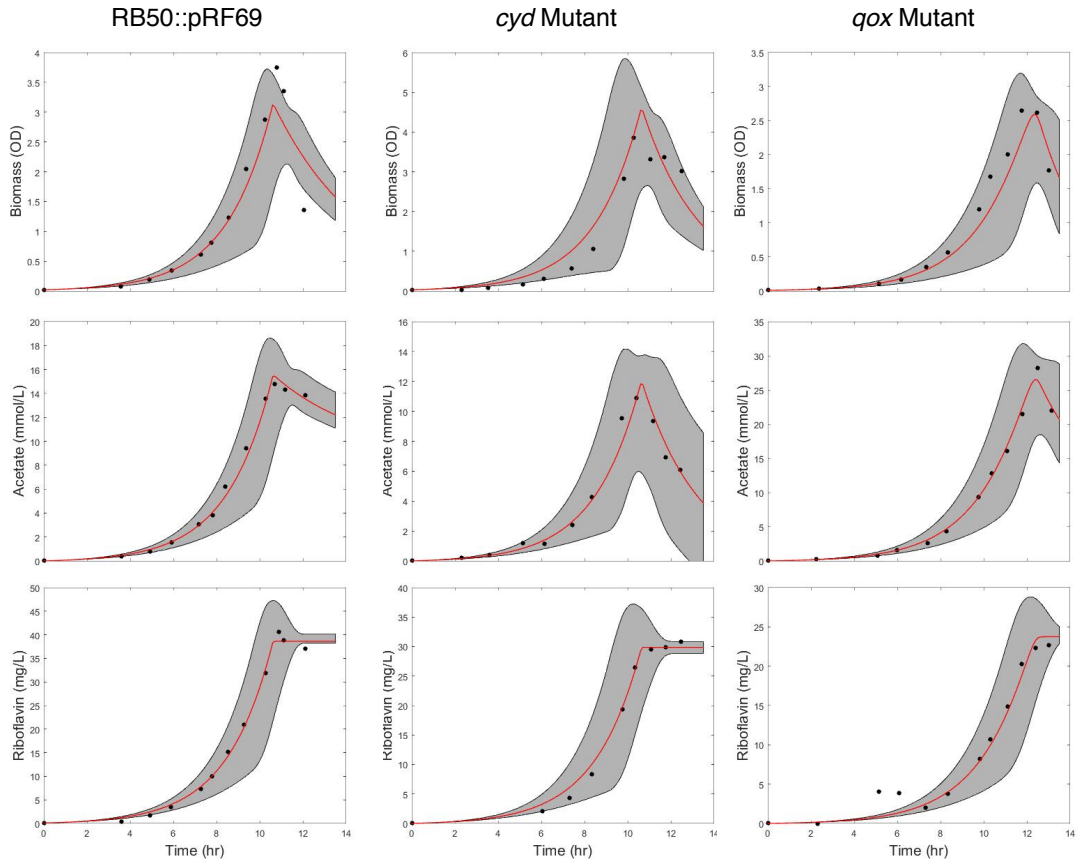

Figure 2.2: Batch culture modeling results of *Bacillus subtilis* strain RB50::pRF69 and its *cyd* and *qox* mutants with Monod kinetics are presented in Figure 2.2. Experimental data (black dots) compared to model best fit (red line) and 99% confidence interval range (grey area) from the model output of 1,600 parameter variant ensemble set generated by perturbing the best fit solution by ±10%.

In order to determine the parameters with the greatest influence on model performance, a sensitivity analysis (Figure 2.3) of estimated kinetic parameters was conducted using the python package SALib and Sobol sensitivity program [27].

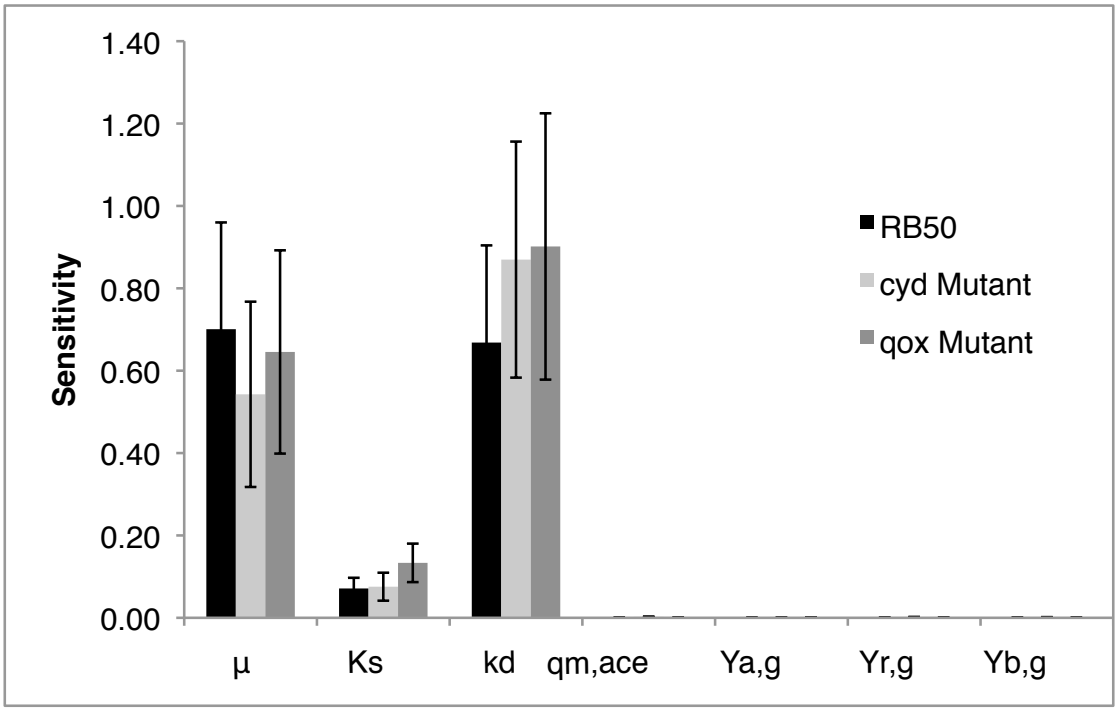

Figure 2.3: Sobol sensitivity analysis results for Monod batch culture model of *B. subtilis* strain RB50::pRF69 and the *cyd* and *qox*. Specific growth rate (μ), and cell death (k_d), exhibited the greatest sensitivity. The saturation constant (K_s) exhibited a smaller, but still meaningful degree of sensitivity. Sobol sensitivity analysis performed in python using the SALib package to generate an ensemble of 1,600 parameter variants perturbed around each simulations best fit values. Error bars represent the 95% confidence intervals of the sensitivity coefficients.

The analysis determined that the values of the specific growth rate of each strain (μ) and cell death rate (k_d) influence the model output the greatest, while the Monod saturation constant (K_s) also has a significant, yet smaller sensitivity. It is an expected result that these parameters yield the highest sensitivity values for the model developed. These three parameters account for the growth and

death of cell mass which directly influences product formation.

2.2.2 Fed-Batch Feeding Optimization

Given the ability of classical Monod kinetics to accurately describe experimental data for the batch culture of strain RB50::pRF69 and its mutants, a fed-batch model of *B. subtilis* growth was developed to assess culture conditions in a more industrially relevant process. One shortcoming of modeling cell growth via Monod kinetics is that kinetic constants are often different for the same organism and same strain under different culture conditions. As a result of this shortcoming, when the model was adapted to fed-batch culture the kinetic parameters had to be refit to describe the data of Ruehl et al. (2010) [75]. The fed-batch model was able to fit experimental data relatively well with the sole addition of cell feeding and a dilution due to growth term on each of the mass balances (Figure 2.4); however, the decrease in acetate concentration over time observed experimentally could not be captured exactly. The addition of the MATLAB *heaviside* function to delay the consumption of acetate due to maintenance facilitated the model to accurately describe the change in acetate observed experimentally (Figure 2.5).

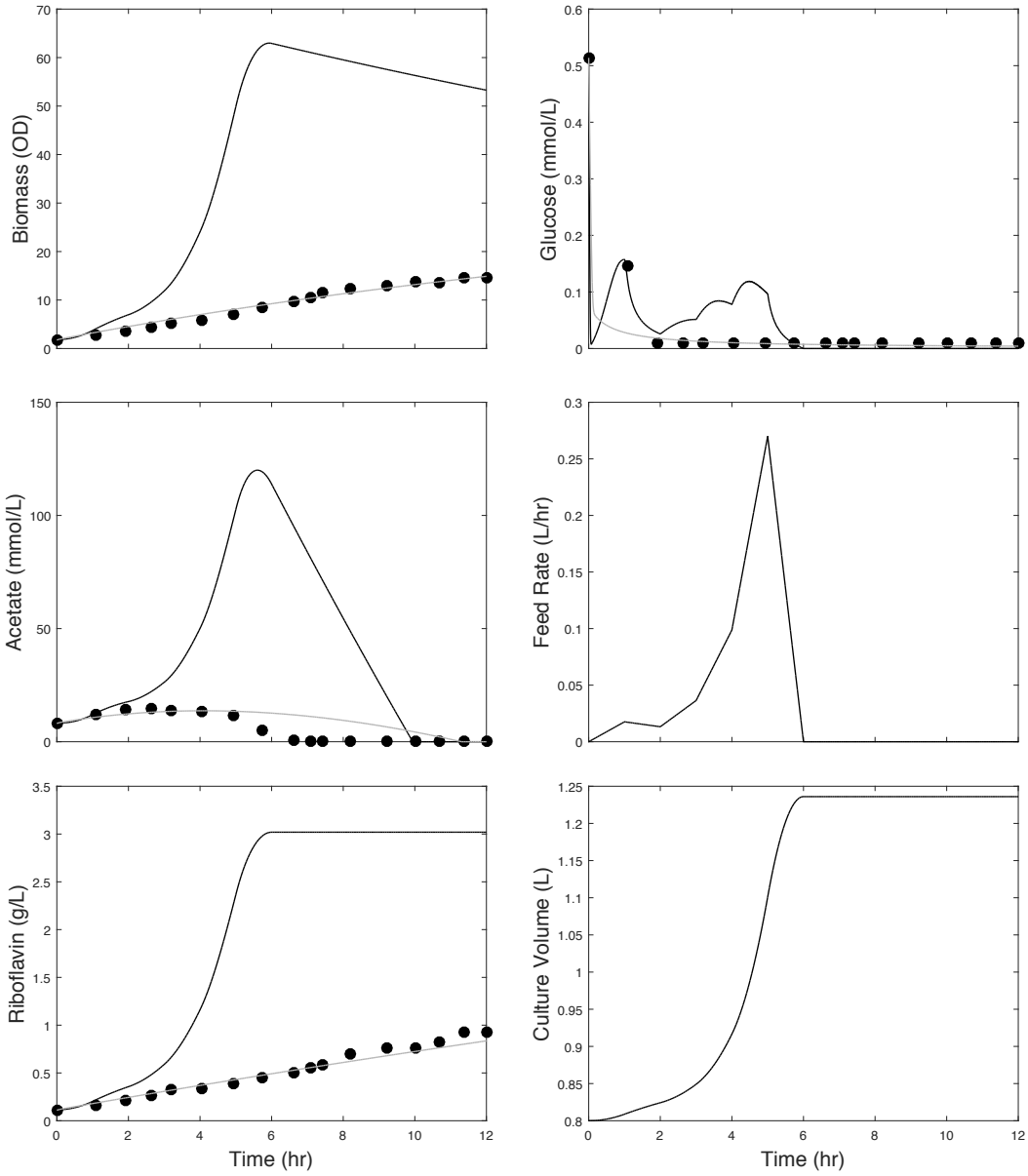

Figure 2.4: Feeding profile optimization of *B. subtilis* RB50::pRF69 based on objective function to increase riboflavin production (black line) per hour yielded an exponential feeding profile. Model was first fit (grey line) to experimental data [75] (black circles) in order to identify kinetic parameters. Glucose concentration in feed was set to $170\frac{g}{L}$ for both parameter estimation and feed optimization cases to coincide with experimental conditions. Parameter estimation model feed rate was fixed at 6.4 $\frac{mL}{hr}$. Monod kinetic parameters identified from experimental data were fixed during feed optimization study.

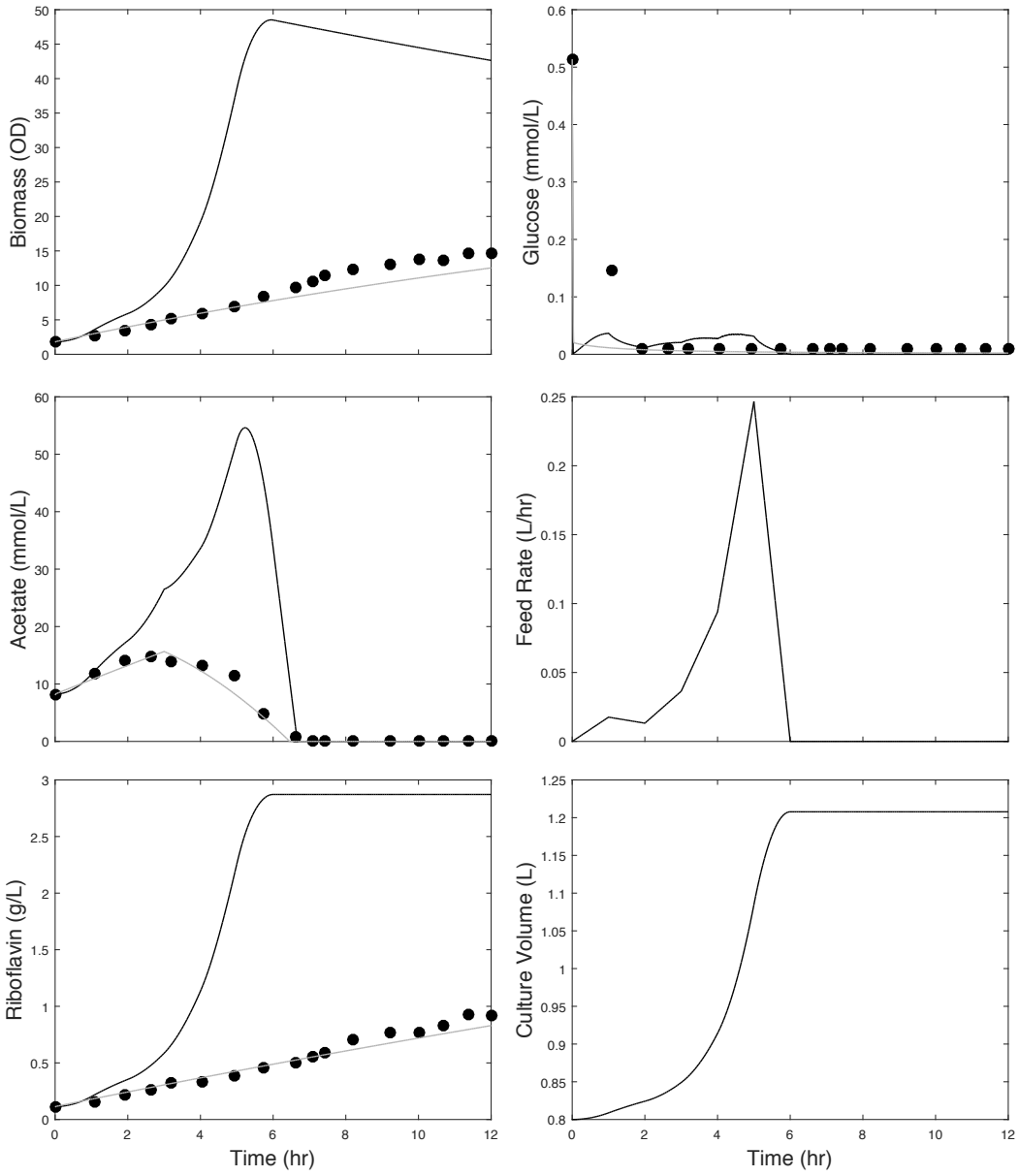

Figure 2.5: Feeding profile optimization of *B. subtilis* RB50::pRF69 with time delay heaviside function placed on acetate maintenance coefficient. Acetate maintenance set to zero until t = 3 hours. Optimized case (black line) based on objective function to increase riboflavin production per hour yielded an exponential feeding profile. Model was first fit (grey line) to experimental data [75] (black circles) in order to identify kinetic parameters. Glucose concentration in feed was set to $170\frac{g}{L}$ for both parameter estimation and feed optimization cases to coincide with experimental conditions. Parameter estimation model feed rate was fixed at 6.4 $\frac{mL}{hr}$. Monod kinetic parameters identified from experimental data were fixed during feed optimization study.

Code was then written to maximize the production of riboflavin by varying the feeding rate subject to reactor volume constraints and limiting the change in feeding to ±1.75 times the previous volume of feed delivered in order to maintain a realistic feeding profile. The optimized model produced an exponential feed rate capable of achieving a higher cell density and a three-fold improvement on riboflavin yield. The result of an exponential feeding profile is unsurprising, as such methods are commonly used in industry to maximize product formation while maintaining a relatively constant specific growth rate [61] [66]. These results, however, are purely mathematical. Given the opportunity, an experimental study should be undertaken to validate the prediction that the optimized feeding profile will indeed produce a higher yield of riboflavin.

Monod kinetics were used to build a model of *B. subtilis* growth and product formation capable of fitting experimental data for batch and fed batch scenarios. The benefits of such a model include the small number of kinetic parameters, mass balances, and computational overhead coupled with the ability to provide insight into experimental design for process improvement (i.e. cell feeding). The Monod approach, however, is limited in the sense that it assumes the complex inner workings of the cell to be a 'black-box' type environment in which inputs (substrates) are fed in and outputs (biomass and extracellular metabolites) are produced. This approach fails to provide any biological insight for the metabolic engineer or biologist to rationally design experiments for strain development and improvement.

2.2.3 Structured Model of *B. subtilis* Metabolism via HCM-EM

The inability of Monod kinetics to describe intracellular metabolism, combined with the variance of Monod parameters for the same strain under different culture conditions has motivated the undertaking of the development of a dynamic model of *B. subtilis* metabolism. A reaction network consisting of 32 reactions describing the *B. subtilis* glycolysis backbone, coupled with the pentose-phosphate pathway and TCA cycles was considered to model cell growth and product formation. METATOOL 5.1 [34] analysis was used to compute the elementary modes for this reaction network and resulted in the generation of 170 elementary modes. Three types of modes were generated depending on the substrate(s) consumed: 125 modes responsible for consuming glucose as the only substrate, 3 modes consuming only acetate, and 42 modes consuming both glucose and acetate simultaneously. An HCM-EM framework was then developed to describe dynamic growth behavior through the incorporation of dynamic balances and cybernetic control variables responsible for the regulation of flux through each elementary mode. The resulting model contained 386 kinetic parameters and 174 ODEs. The preliminary results of this model are presented in Figure 2.6.

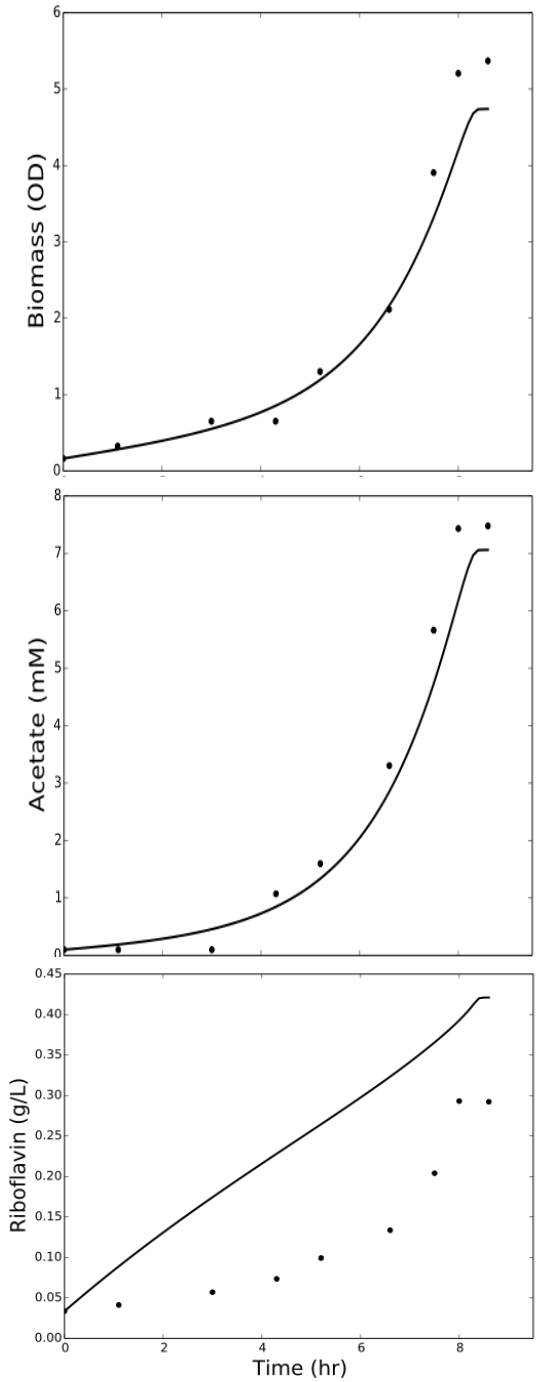

Figure 2.6: HCM-EM *B. subtilis* model (line) with experimental data (circles) from Ruehl et al 2010 [75]. The HCM-EM model decomposed a network of 32 reactions describing glycolysis, the pentose phosphate pathway, and the TCA cycle into 170 elementary modes, 174 ordinary differential equations, and 386 kinetic parameters. The model was build in the Julia programming language and trained to experimental data using simulated annealing.

It should be noted that although the HCM-EM model was able to accurately describe biomass growth and acetate formation, the model overestimated the production of riboflavin in the culture. This is most likely due to the large number of kinetic parameters and ODEs in the model and the use of multiple objective functions to describe several outputs. In order to better describe the experimental data, a study utilizing the JuPOETs code, an algorithm specifically designed for large models and multiple objective functions, will be undertaken to improve the model. The model will then be assessed to determine if it is capable of predicting experimental intracellular flux measurements. If successful, the model has the potential to facilitate the metabolic engineer with a platform to run simulated experiments of the organism under a variety of conditions, to observe the differences in intracellular fluxes, and to identify metabolic bottlenecks and sites for knockouts or gene modifications for rational strain design.

2.3 Materials and Methods

2.3.1 Unstructured Monod Kinetics Model

Traditional Monod Kinetics [85] were used to model culture of riboflavin producing *B subtilis* strain RB50::pRF69 and two quinol oxidase knockout mutants. Five ordinary differential equations (ODEs) were created to model the time evolution of cell biomass, glucose, acetate and riboflavin, and culture volume:

$$\frac{dBio}{dt} = (\mu - k_d)Bio - \frac{F}{V}Bio \tag{2.1}$$

$$\frac{dGlc}{dt} = -\left(\frac{\mu}{Y_{bio/glc}}\right)Bio + \frac{F}{V}(Glc_{in} - Glc) \tag{2.2}$$

$$\frac{dAce}{dt} = \left(\frac{\mu}{Y_{ace/glc}} - q_{m,ace}\right)Bio - \frac{F}{V}Ace \tag{2.3}$$

$$\frac{dRibo}{dt} = \left(\frac{\mu}{Y_{ribo/glc}}\right)Bio - \frac{F}{V}Ribo \tag{2.4}$$

$$\frac{dVol}{dt} = F \tag{2.5}$$

Where *Bio* refers to biomass, *Glc* to glucose, Glc_{in} to the concentration of glucose in the feed, *Ace* to acetate, *Ribo* to riboflavin, *Vol* to culture volume, and F to feed rate. Biomass is presented in units of Optical Density (OD_{600}), glucose and acetate in $\frac{mmol}{L}$, and riboflavin in $\frac{g}{L}$ or $\frac{mg}{L}$. A conversion factor 0.33 g cell dry weight per OD_{600} was experimentally determined for the RB50::pRF69 and mutant strains of *B. subtitlis* [114]. The model results are presented in units of OD_{600}. The yield coefficients $Y_{bio/glc}$, $Y_{ace/glc}$, and $Y_{ribo/glc}$ refer to the production of biomass, acetate, and riboflavin due to glucose consumption respectively. $q_{m,ace}$ accounts for the consumption of acetate for cellular maintenance. μ refers to the specific growth rate on glucose as governed by the equation:

$$\mu = \frac{\mu_{max} Glc}{K_s + Glc} \tag{2.6}$$

The mass balances were subject to the initial conditions of the experiment being modeled. Batch culture experimental data from [112] was used to train the batch mutant models and feed rate F was set to zero. For fed-batch model training, the feed rate F was fixed at 6.4 $\frac{mL}{hr}$ and glucose feed concentration to $170\frac{g}{L}$ to coincide with culture conditions of Ruehl et al 2010 [75]. The model equations were solved via MATLAB *ode45*.

2.3.2 Estimation of Model Parameters

Model parameters were estimated by minimizing the difference between simulations and experimental measurements (squared residual):

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \tag{2.7}$$

where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ, $x_j(\tau, \mathbf{k})$ denotes the simulated value for species j at time τ, and $\omega_j(\tau)$ denotes the experimental measurement variance for species j at time τ. The outer summation is with respect to time, while the inner summation is with respect to state. The model residual was minimized using the MATLAB function *fmincon*.

2.3.3 Feeding Profile Optimization

To first estimate the kinetic parameters for the fed-batch culture conditions, Equation 2.7 was used and culture feed rate was set to $6.4 \frac{mL}{hr}$ and feed concentration of glucose to $170 \frac{g}{L}$ to coincide with the experimental conditions of [75]. The model timestep was set to 1 hour and the area under the riboflavin curve was approximated using the MATLAB function *trapz*. The MATLAB function *fmincon* was then used to maximize the area under the curve for riboflavin for each hour during culture by varying the feeding rate with a constant glucose inlet concentration of $170 \frac{g}{L}$. Several constraints were placed on the optimization routine to keep the results meaningful: each timestep feed rate was prevented from increasing more than ±1.75 times the amount fed during the previous timestep and the maximum culture volume was restricted to 1.2 L to mimic the size of

the bioreactor used in the experiment. The fed-batch model acetate fit was improved by multiplying the acetate maintenance coefficient by a *heaviside* function to delay the consumption of acetate by 3 hours.

2.3.4 Ensemble Generation

Parameter ensembles (N = 1,600) were generated by randomly perturbing parameters around their best fit values. Plots of the mean model performance, 99% confidence interval, and best fits were generated to demonstrate model robustness under variation.

2.3.5 Global Sensitivity Analysis

Variance based sensitivity analysis was used to estimate which parameters were critical to model performance. The performance function used in this study was previously mentioned squared residual. Candidate parameter sets (N = 1,600) were generated using Sobol sampling by perturbing the best fit parameter set for each mutant [27]. Model performance, calculated for each of these parameter sets, was then used to estimate the total-order sensitivity coefficient and 95% confidence interval for each model parameter.

2.3.6 Hybrid Cybernetic Model using Elementary Modes

The HCM-EM approach, adapted from [37] is a modification of the Flux Balance Analysis (FBA) technique in which intracellular reactions are assumed to be at a

pseudo steady state. HCM-EM modifies this approach by including a set of differential equations to describe the dynamic behavior of extracellular variables. In this study, the abundance of extracellular species i (x_i), the pseudo enzyme e_l (catalyzes flux through mode l), and cell mass are governed by:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sum_{l=1}^{\mathcal{L}} \sigma_{ij} z_{jl} q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})\, c \qquad i = 1, ..., \mathcal{M} \tag{2.8}$$

$$\frac{de_l}{dt} = \alpha_l + r_{El}(\mathbf{k}, \mathbf{x})\, u_l - (\beta_l + r_G)\, e_l \qquad l = 1, ..., \mathcal{L} \tag{2.9}$$

$$\frac{dc}{dt} = r_G c \tag{2.10}$$

where $\mathcal{R}$ and $\mathcal{M}$ denote the number of reactions and extracellular species in the model and $\mathcal{L}$ denotes the number of elementary modes. The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j and z_{jl} denotes the normalized flux for reaction j in mode l. If $\sigma_{ij} > 0$, species i is produced by reaction j; if $\sigma_{ij} < 0$, species i is consumed by reaction j; if $\sigma_{ij} = 0$, species i is not connected with reaction j. Extracellular species balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$ determined from experimental data. All pseudo enzymes were set to the initial value of 0.8. The term $q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$ denotes the specific uptake/secretion rate for mode l where $\mathbf{e}$ denotes the pseudo enzyme vector, $\mathbf{k}$ denotes the unknown kinetic parameter vector, and $\mathbf{x}$ denotes the extracellular species vector; $q_l(\mathbf{e}, \mathbf{k}, \mathbf{x})$ is the product of a kinetic term ($\bar{q}_l$) and a control variable governing enzyme activity. Flux through each mode was catalyzed by a pseudo enzyme e_l, synthesized at the regulated specific rate $r_{E,l}(\mathbf{k}, \mathbf{x})$, and constitutively at the rate α_l. The term u_l denotes the cybernetic variable controlling the synthesis of enzyme l. The term β_l denotes the rate constant governing nonspecific enzyme degradation, and r_G denotes the specific growth rate through all modes. The specific uptake/secretion rates and the specific rate of enzyme

synthesis were modeled using saturation kinetics. The specific growth rate was given by:

$$r_G = \sum_{l=1}^{\mathcal{L}} z_{\mu l} q_l (\mathbf{e}, \mathbf{k}, \mathbf{x}) \tag{2.11}$$

where $z_{\mu l}$ denotes the growth flux μ through mode l. The control variables u_l and v_l, which control the synthesis and activity of each enzyme respectively, were given by:

$$u_l = \frac{z_{sl}\bar{q}_l}{\sum_{l=1}^{\mathcal{L}} z_{sl}\bar{q}_l} \qquad v_l = \frac{z_{sl}\bar{q}_l}{\max_{l=1,\ldots,\mathcal{L}} z_{sl}\bar{q}_l} \tag{2.12}$$

where z_{sl} denotes the uptake flux of substrate s through mode l.

Elementary Modes

The model developed contained 386 kinetic parameters and 174 ODEs for the reaction network of 32 reactions consisting of *B. subtilis* glycolysis, pentose-phosphate pathway, and TCA cycle (Figure 2.1) adapted from [15]. The biomass formation reaction used was modeled after Dauner, Bailey, and Sauer (2001) [14]. 170 elementary modes were calculated for the network using METATOOL 5.1 [34]. The model equations were implemented in Julia (v.0.4.2) [8] and solved using SUNDIALS [28]. The model parameters were estimated by minimizing the difference between simulated and experimental measurements (Eqn 2.7).

CHAPTER 3

EFFECTIVE DYNAMIC MODELS OF METABOLIC NETWORKS

Submitted to IEEE Life Sciences Letters for publication. Michael Vilkhovoy, Mason Minot, Jeffrey D. Varner. doi: http://dx.doi.org/10.1101/047316.

3.1 Introduction

Cybernetic models are an alternative to the constraints based approach which hypothesize that metabolic control is the output of an optimal decision. Cybernetic models have predicted mutant behavior [103, 90], steady-state multiplicity [38], strain specific metabolism [89], and have been used in bioprocess control applications [24]. Hybrid cybernetic models (HCM) have addressed earlier shortcomings of the approach by integrating cybernetic optimality concepts with EMs. HCMs dynamically choose combinations of biochemical modes (each catalyzed by a pseudo enzyme whose expression is controlled by an optimal decision) to achieve a physiological objective (Fig. 3.1A). HCMs generate intracellular flux distributions consistent with other approaches such as metabolic flux analysis (MFA), and also describe dynamic extracellular measurements [37]. However, HCMs are restricted to networks which can be decomposed into EMs (or EPs).

In this study, we developed the hybrid cybernetic modeling with flux balance analysis (HCM-FBA) technique. HCM-FBA is a modification of the hybrid cybernetic approach of Ramkrishna and coworkers [37] which uses FBA solutions (instead of EMs) in conjunction with cybernetic control variables to dynamically simulate metabolism. We compared the performance of HCM-

FBA to HCM-EM for a prototypical metabolic network along with two *E. coli* networks. HCM-FBA performed comparably to HCM-EM for the prototypical network and a reduced anaerobic *E. coli* network, despite having fewer parameters in each case. Next, HCM-FBA was applied to an aerobic *E. coli* metabolic network that was infeasible for HCM-EM. HCM-FBA described cell mass growth and the shift from glucose to acetate consumption with only a few modes. Global sensitivity analysis allowed us to further reduce the aerobic *E. coli* HCM-FBA model to the minimal model required to describe the data. Thus, HCM-FBA is a promising approach for the development of reduced order dynamic metabolic models and a viable alternative to HCM-EM, especially for large networks where the generation of EMs is infeasible.

3.2 Results

HCM-FBA was equivalent to HCM-EM for a prototypical metabolic network (Fig. 3.1). The proof of concept network, consisting of 6 metabolites and 7 reactions (Fig. 3.1B), generated 3 FBA modes and 6 EMs. Using the EMs and synthetic parameters, we generated test data from which we estimated the HCM-FBA model parameters. The best fit HCM-FBA model replicated the synthetic data (Fig. 3.1C). The HCM-EM and HCM-FBA kinetic parameters were not quantitatively identical, but had similar orders of magnitude; the FBA approach had 3 fewer modes, thus identical parameter values were not expected. Taken together, the HCM-FBA approach replicated synthetic data generated by HCM-EM, despite having 3 fewer modes. Next, we tested the ability of HCM-FBA to replicate experimental data.

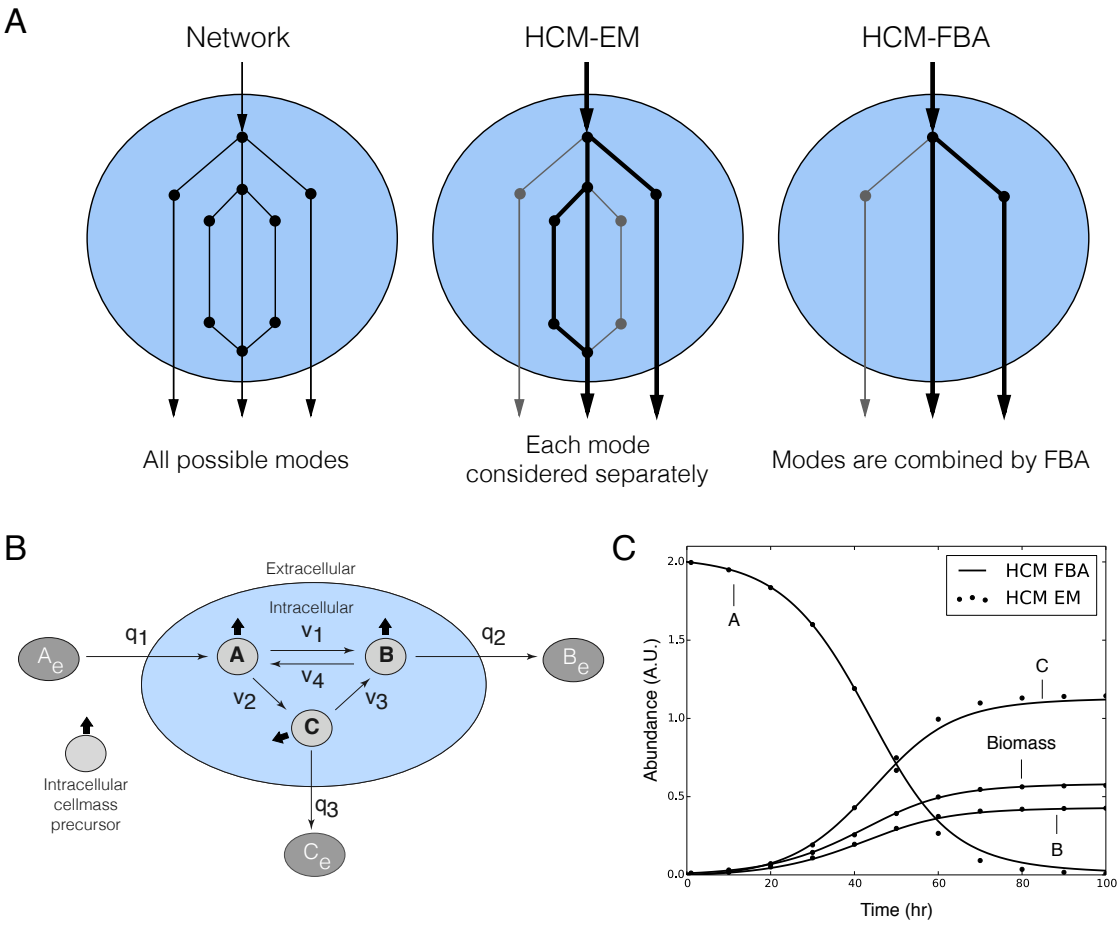

Figure 3.1: HCM proof of concept metabolic study. A: HCMs distribute uptake and secretion fluxes amongst different pathways. For HCM-EM, these pathways are elementary modes; for HCM-FBA these are flux balance analysis solutions. HCM-EM combines all possible modes within a network; whereas HCM-FBA combines only steady-state paths estimated by flux balance analysis. B: Prototypical network with six metabolites and seven reactions. Intracellular cell mass precursors A, B, and C are balanced (no accumulation) while the extracellular metabolites (A_e, B_e, and C_e) are not balanced (can accumulate). The oval denotes the cell boundary, q_j is the jth flux across the boundary, and v_k denotes the kth intracellular flux. C: Simulation of extracellular metabolite trajectories using HCM-FBA (solid line) versus HCM-EM (points) for the prototypical network.

The performance of HCM-FBA was equivalent to HCM-EM for anaerobic *E. coli* metabolism (Fig. 3.2A). We constructed an anaerobic *E. coli* network [37], consisting of 12 reactions and 19 metabolites, which generated 7 FBA modes and 9 EMs. HCM-EM reproduced cell mass, glucose, and byproduct trajectories

using the kinetic parameters reported by Kim et al. [37] (Fig. 3.2A, points versus dashed). HCM-FBA model parameters were estimated in this study from the Kim et al. data set using simulated annealing. Overall, HCM-FBA performed within 5% of HCM-EM (on a residual standard error basis) for the anaerobic *E. coli* data (Fig. 3.2A, solid), despite having 2 fewer modes and 4 fewer parameters (17 versus 21 parameters). Thus, while both HCM-EM and HCM-FBA described the experimental data, HCM-FBA did so with fewer modes and parameters.

HCM-FBA captured the shift from glucose to acetate consumption for a model of aerobic *E. coli* metabolism that was infeasible for HCM-EM (Fig. 3.2B). An *E. coli* metabolic network (60 metabolites and 105 reactions) was constructed from literature [82][63]. Elementary mode decomposition of this network (and thus HCM-EM) was not feasible; 153,000 elementary modes were generated before the calculation became infeasible. Conversely, flux balance analysis generated only 29 modes for the same network. HCM-FBA model parameters were estimated from cell mass, glucose, and acetate measurements [102] using simulated annealing (Fig. 3.2B, solid). HCM-FBA captured glucose consumption, cell mass formation, and the switch to acetate consumption following glucose exhaustion. HCM-FBA described the dynamics of a network that was infeasible for HCM-EM, thereby demonstrating the power of the approach for large networks. Next, we demonstrated a systematic strategy to identify the critical subset of FBA modes required for model performance.

Global sensitivity analysis identified the FBA modes essential to model performance (Fig. 3.3). Total order sensitivity coefficients were calculated for all kinetic parameters and enzyme initial conditions in the aerobic *E. coli* model. Five of the 29 FBA modes were significant; removal of the most significant of

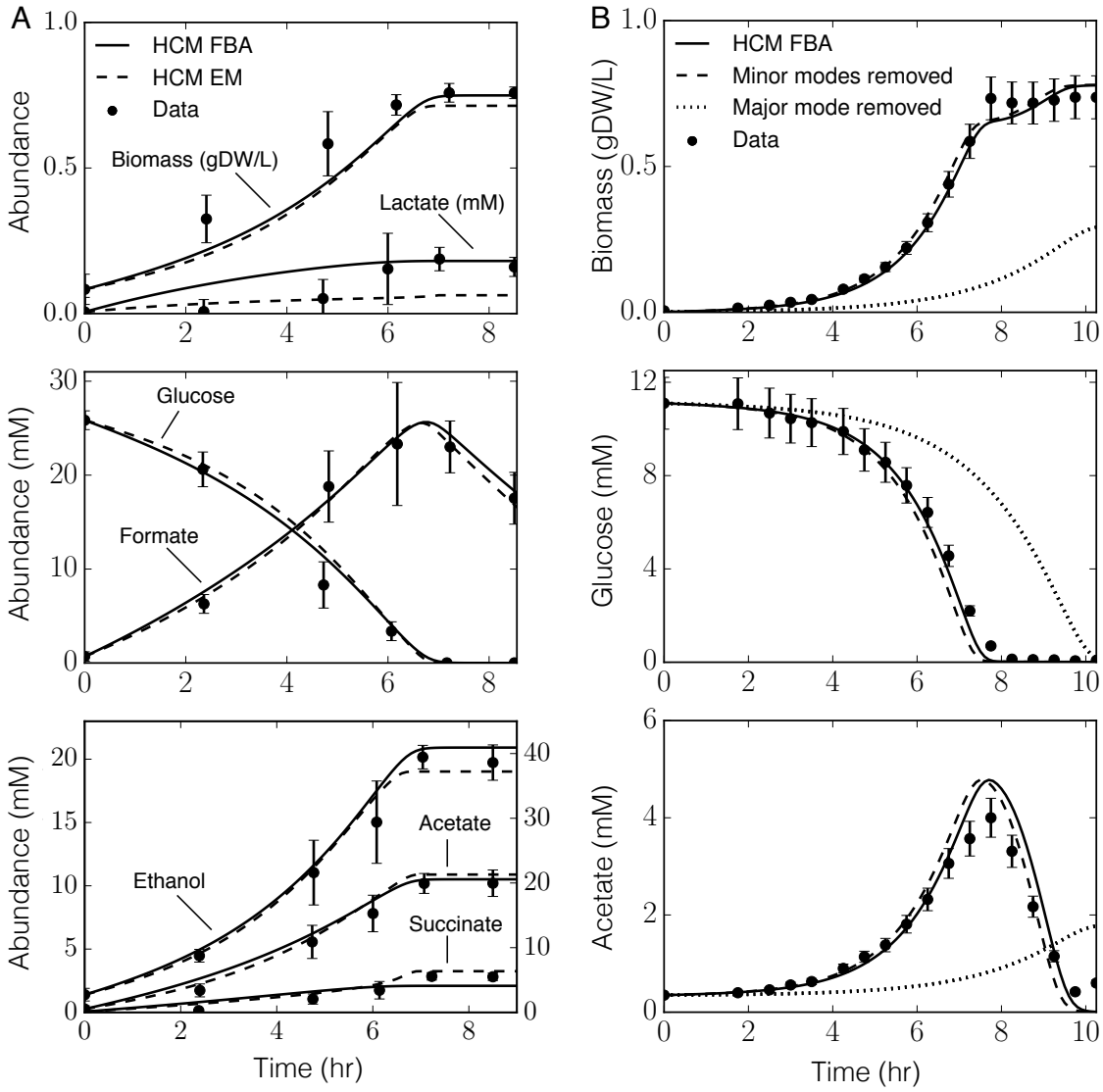

Figure 3.2: HCM-FBA versus HCM-EM performance for small and large metabolic networks. A: Batch anaerobic *E. coli* fermentation data versus HCM-FBA (solid) and HCM-EM (dashed). The experimental data was reproduced from Kim et al.[37]. Error bars represent the 90% confidence interval. B: Batch aerobic *E. coli* fermentation data versus HCM-FBA (solid). Model performance is also shown when minor modes (dashed) and major modes (dotted) were removed from the HCM-FBA model. The experimental data was reproduced from Varma & Palsson [102]. Error bars denote a 10% coefficient of variation.

these modes (encoding aerobic growth on glucose) destroyed model performance (Fig. 3.2B, dotted). Conversely, removing the remaining 24 modes had a negligible effect upon model performance (Fig. 3.2B, dashed). The sensitivity

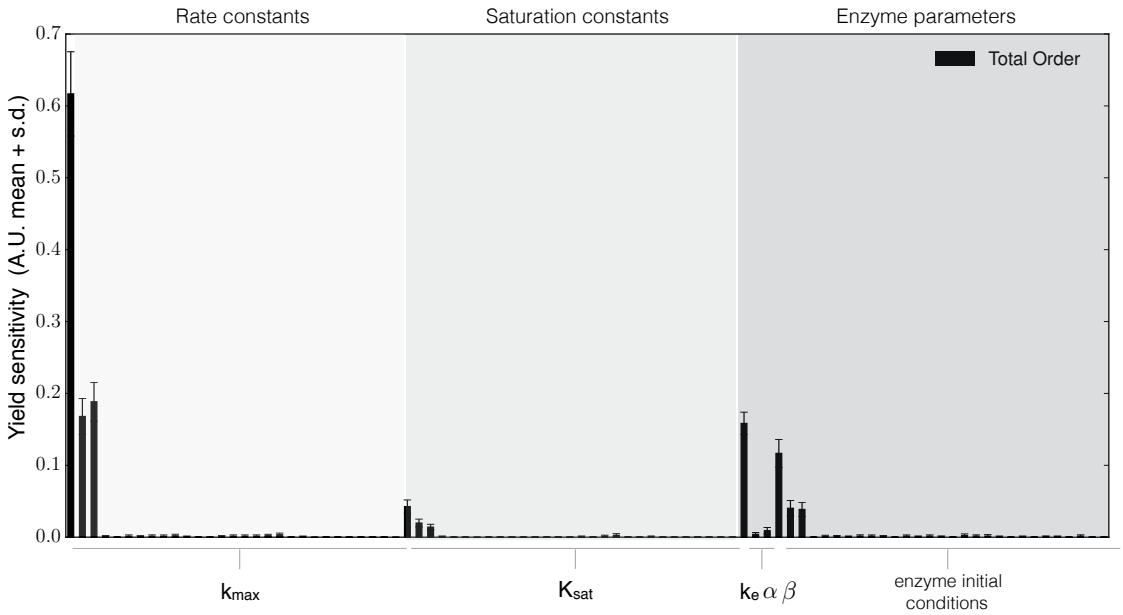

Figure 3.3: Global sensitivity analysis of the aerobic *E. coli* model. Total order variance based sensitivity coefficients were calculated for the biomass yield on glucose and acetate. Sensitivity coefficients were computed for kinetic parameters and enzyme initial conditions (N = 183,000). Error bars represent the 95% confidence intervals of the sensitivity coefficients.

analysis identified the minimal model structure required to explain the experimental data.

3.3 Discussion

In this study, we developed HCM-FBA, an effective modeling technique to simulate metabolic dynamics. HCM-FBA uses flux balance analysis solutions (instead of elementary modes) in conjunction with cybernetic control variables to dynamically simulate metabolism. We studied the performance of HCM-FBA on a prototypical metabolic network, along with two *E. coli* networks. First, we showed that the performance of HCM-FBA and HCM-EM were comparable for the prototypical network and a small model of anaerobic *E. coli* metabolism. For

the anaerobic case, both approaches described the experimental data. However, HCM-FBA (which was within 5% of HCM-EM and slightly better than HCM-EM for lactate secretion) had fewer modes and parameters. Next, HCM-FBA was applied to an aerobic *E. coli* metabolic network that was not feasible for HCM-EM. Elementary mode decomposition of the aerobic network generated over 153,000 elementary modes. Conversely, the HCM-FBA approach described cell mass growth and the shift from glucose to acetate consumption with only 29 FBA modes. Global sensitivity analysis further showed that only 5 of the 29 FBA modes were critical to model performance. Removal of these modes crippled the model, but removal of the remaining 24 modes had a negligible impact. Thus, HCM-FBA is an alternative approach to HCM-EM, especially for large networks where the generation of elementary modes is infeasible.

HCM-FBA is a promising approach to model large metabolic networks where elementary modes calculations are infeasible. However, there are additional studies that should be performed. First, the intracellular flux distribution predicted by HCM-FBA should be compared to HCM-EM and to flux measurements calculated using MFA or FBA in combination with carbon labeling. HCM-EM predicted intracellular fluxes that were similar to MFA results [37]; however, the fluxes predicted by HCM-FBA have not yet been validated. Next, the performance of HCM-FBA should be compared to lumped hybrid cybernetic models (L-HCM). L-HCMs, which combine elementary modes into mode families based upon metabolic function [88, 89], have been applied to an *E. coli* network with 67 reactions and a *Saccharomyces cerevisiae* network with 70 reactions; both cases had satisfactory fits to extracellular experimental data. However, while L-HCM reduces the dimension of possible alternative modes that must be considered, it still requires the calculation of an initial set of modes. For

metabolic networks of even moderate size, EM (or EP) decomposition may not be possible. On the other hand, the generation of flux balance solutions (convex combinations of the elementary modes or extreme pathways) is trivial, even for genome scale metabolic networks. Thus, HCM-FBA opens up the possibility for dynamic genome scale models of bacterial and perhaps even of mammalian metabolism.

3.4 Materials and Methods

3.4.1 Elementary mode and flux balance analysis

The HCM-FBA approach is a modification of HCM-EM, where elementary modes are replaced with flux balance analysis solutions. The dynamic equations describing cybernetic variables, the formation of extracellular metabolites, and specific growth have been presented in Section 2.3.6. Elementary modes were calculated using METATOOL 5.1 [34]. FBA modes were defined as the solution flux vector through the network connecting substrate uptake to cell mass and extracellular product formation. The FBA problem was formulated as:

$$\begin{gathered} \max_{\boldsymbol{w}} \left(w_{obj} = \boldsymbol{\theta}^T \boldsymbol{w} \right) \\ \text{Subject to : } \mathbf{Sw} = \mathbf{0} \\ \alpha_i \leq w_i \leq \beta_i \qquad i = 1, 2, ..., \mathcal{R} \end{gathered} \tag{3.1}$$

where **S** denotes the stoichiometric matrix, **w** denotes the unknown flux vector, $\boldsymbol{\theta}$ denotes the objective selection vector and α_i and β_i denote the lower and upper bounds on flux w_i, respectively. The flux balance analysis problem was solved using the GNU Linear Programming Kit (v4.52) [1]. For each FBA mode,

the objective w_{obj} was to maximize either the specific growth rate or the specific rate of byproduct formation. Multiple FBA modes were calculated for each objective by allowing the oxygen and nitrate uptake rates to vary. For aerobic metabolism, the specific oxygen and nitrate uptake rates were constrained to allow a maximum flux of 10 mM/gDW·hr and 0.05 mM/gDW·hr, respectively. Each FBA mode was normalized by the specified objective flux.

3.4.2 Global sensitivity analysis

Variance based sensitivity analysis was used to estimate which FBA modes were critical to model performance. The performance function used in this study was the biomass yield on substrate. Candidate parameter sets (N = 182,000) were generated using Sobol sampling by perturbing the best fit parameter set ±50% [27]. Model performance, calculated for each of these parameter sets, was then used to estimate the total-order sensitivity coefficient for each model parameter.

3.4.3 Estimation of model parameters

Model parameters were estimated by minimizing the difference between simulations and experimental measurements (Eqn 2.7). The model residual was minimized using simulated annealing implemented in the Julia programming language.

CHAPTER 4

JUPOETS: A CONSTRAINED MULTIOBJECTIVE OPTIMIZATION APPROACH TO ESTIMATE BIOCHEMICAL MODEL ENSEMBLES IN THE JULIA PROGRAMMING LANGUAGE

Submitted for publication to BMC Systems Biology. David Bassen, Michael Vilkhovoy, Mason Minot, Jonathan T Butcher, and Jeffrey D. Varner.

doi: http://dx.doi.org/10.1101/056044

Author contributions: J.V. developed the software presented in this study. M.M and M.V developed the proof-of-concept biochemical model. The manuscript was prepared and edited for publication by D.B, J.B, and J.V.

4.1 Introduction

Ensemble modeling is a promising approach for obtaining robust predictions and course grained population behavior in deterministic mathematical models. Ensemble approaches address model uncertainty by using parameter or model families instead of single best-fit parameters or fixed model structures. Parameter ensembles can be selected based upon simulation error, along with other criterion such as diversity or steady-state performance. Simulations using parameter ensembles can estimate confidence intervals on model variables, and robustly constrain model predictions, despite having many poorly constrained parameters.

In this study, we present a multiobjective based technique to estimate parameter or models ensembles, the Pareto Optimal Ensemble Technique in the

Julia programming language (JuPOETs). JuPOETs integrates simulated annealing with Pareto optimality to estimate ensembles on or near the optimal trade-off surface between competing training objectives. We demonstrate JuPOETs on a suite of multiobjective problems, including test functions with parameter bounds and system constraints as well as for the identification of a proof-of-concept biochemical model with four conflicting training objectives. JuPOETs identified optimal or near optimal solutions approximately six-fold faster than a corresponding implementation in Octave for the suite of test functions. For the proof-of-concept biochemical model, JuPOETs produced an ensemble of parameters that gave both the mean of the training data for conflicting data sets, while simultaneously estimating parameter sets that performed well on each of the individual objective functions.

JuPOETs is a promising approach for the estimation of parameter and model ensembles using multiobjective optimization. JuPOETs can be adapted to solve many problem types, including mixed binary and continuous variable types, bilevel optimization problems and constrained problems without altering the base algorithm. JuPOETs is open source available under an MIT license, and can be installed using the Julia package manager from the JuPOETs GitHub repository

4.2 Background

Ensemble modeling is a promising approach for obtaining robust predictions and course grained population behavior in deterministic mathematical models. It is often not possible to uniquely identify all the parameters in biochemical

models, even when given extensive training data [23]. Thus, despite significant advances in standardizing biochemical model identification [25], the problem of estimating model parameters from experimental data remains challenging. Ensemble approaches address parameter uncertainty in systems biology and other fields like weather prediction [7, 41, 32, 62] by using parameter families instead of single best-fit parameter sets. Parameter families can be selected based upon simulation error, along with other criterion such as diversity or steady-state performance. Simulations using parameter ensembles can estimate confidence intervals on model variables, and robustly constrain model predictions, despite having many poorly constrained parameters [26, 91]. There are many techniques to generate parameter ensembles. Battogtokh et al., Brown et al., and later Tasseff et al. generated experimentally constrained parameter ensembles using a Metropolis-type random walk [7, 32, 97, 98]. Liao and coworkers developed methods that generate ensembles that all approach the same steady-state, for example one determined by fluxomics measurements [100]. They have used this approach for model reduction [96], strain engineering [12, 95] and to study the robustness of non-native pathways and network failure [43]. Maranas and coworkers have also applied this method to develop a comprehensive kinetic model of bacterial central carbon metabolism, including mutant data [36]. We and others have used ensemble approaches, generated using both sampling and optimization techniques, to robustly simulate a wide variety of signal transduction processes [48, 91, 97, 98, 56], neutrophil trafficking in sepsis [90], patient specific coagulation behavior [47], uncertainty quantification in metabolic kinetic models [3] and to capture cell to cell variation [44]. Thus, ensemble approaches are widely used to robustly simulate a variety of biochemical systems.

Identification of biochemical models with hundreds or even thousands of

states and parameters may not be tractable as a single objective optimization problem. Further, large models require significant training data perhaps taken from diverse sources, for example different laboratories or cell-lines. These data are often heterogenous, and contain intrinsic conflicts that complicate parameter estimation. Parameter ensembles which optimally balance tradeoffs between submodels and conflicts in training data can lead to robust model performance. Multiobjective optimization is an ensemble generation technique that naturally balances conflicting training data. Previously, we developed the Pareto Optimal Ensemble Technique (POETs) algorithm to address the challenge of competing or conflicting objectives. POETs, which integrates simulated annealing (SA) and multiobjective optimization through the notion of Pareto rank, estimates parameter ensembles which optimally trade-off between competing (and potentially conflicting) experimental objectives [92]. However, the previous implementation of POETs, in the Octave programming language [19], suffered from poor performance and was not configurable. For example, Octave-POETs does not accommodate user definable objective functions, bounds and problem constraints, cooling schedules, different variable types e.g., a mixture of binary and continuous design variables or custom diversity generation routines. Octave-POETs was also not well integrated into a package or source code management (SCM) system. Thus, upgrades to the approach containing new features, or bug fixes were not centrally managed.

4.3 Implementation

In this study, we present an open-source implementation of the Pareto optimal ensemble technique in the Julia programming language (JuPOETs). JuPOETs

offers many advantages and improvements compared to Octave-POETs. JuPOETs takes advantage of the unique features of Julia. Julia is a cross-platform, high-performance programming language for technical computing that has performance comparable to C but with syntax similar to MATLAB/Octave and Python [8]. Julia also offers a sophisticated compiler, distributed parallel execution, numerical accuracy, and an extensive function library. Further, the architecture of JuPOETs takes advantage of the first-class function type in Julia allowing user definable behavior for all key aspects of the algorithm, including objective functions, custom diversity generation logic, linear/non-linear parameter constraints (and parameter bounds constraints) as well as custom cooling schedules. Julia's ability to naturally call other languages such as Python or C also allows JuPOETS to be used with models implemented in a variety of languages across many platforms. Additionally, Julia offers a built-in package manager which is directly integrated with GitHub, a popular web-based Git repository hosting service offering distributed revision control and source code management. Thus, JuPOETs can be adapted to many problem types, including mixed binary and continuous variable types, bilevel problems and constrained problems without altering the base algorithm, as was required in the previous POETs implementation.

4.3.1 JuPOETs optimization problem formulation.

JuPOETs solves the $\mathcal{K}-$dimensional constrained multiobjective optimization problem:

$$\min_{\mathbf{p}} \{ O_1(\mathbf{x}(t,\mathbf{p}),\mathbf{p}) \ldots O_{\mathcal{K}}(\mathbf{x}(t,\mathbf{p}),\mathbf{p}) \tag{4.1}$$

subject to:

$$\begin{aligned} \mathbf{f}(t,\mathbf{x}(t,\mathbf{p}),\dot{\mathbf{x}}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}) &= \mathbf{0} \\ g_1(t,\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}) &\geq 0 \\ &\vdots \\ g_C(t,\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}) &\geq 0 \end{aligned}$$

and parameter bound constraints:

$$\mathcal{L} \leq \mathbf{p} \leq \mathcal{U} \tag{4.2}$$

using a modified simulated annealing approach. The quantity t denotes time, $\mathbf{x}(t,\mathbf{p})$ denotes the model state (with an initial state $\mathbf{x}_0$), and $\mathbf{u}(t)$ denotes an input vector. The terms $\mathbf{f}(t,\mathbf{x}(t,\mathbf{p}),\dot{\mathbf{x}}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p})$ denote the system of model equations (e.g., differential equations, differential algebraic equations or linear/non-linear algebraic equations) where $\mathbf{p}$ denotes the unknown parameter vector ($\mathcal{D} \times 1$). The parameter search can be subject to parameter bound constraints, where $\mathcal{L}$ and $\mathcal{U}$ denote the lower and upper parameter bounds, respectively as well as C problem specific constraints $g_i(t,\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), i = 1,\ldots,C$.

JuPOETs integrates simulated annealing with Pareto optimality to estimate parameter sets on or near the optimal tradeoff surface between competing training objectives (Fig. 4.1 and Fig. 4.2).

```
input  : User specified neighbor, objective, acceptance and cooling functions. Initial parameter
         guess (D × 1)
Output: Rank archive R, parameter solution archive S and objective archive O

initialize: R, S and O using initial guess;
initialize: T ←1.0;
initialize: T_min ←1/10000;
initialize: Maximum number of steps per temperature I;

while T > T_min do
    i ← 1;
    while i< I do

        // Generate a new parameter solution using user neighbor function
        p_{i+1} ← user-function::neighbor(p*);

        // Evaluate p_{i+1} using user objective function
        o_{i+1} ← user-function::objective(p_{i+1});

        Add p_{i+1} to solution archive S;
        Add o_{i+1} to objective archive O;

        // Calculate Pareto rank of solutions in O using builtin rank function
        R ← builtin-function::rank(O);

        // Accept p_{i+1} into the archive with user defined probability
        P ← user-function::acceptance(R,T);
        if P >rand then
            // Update the best solution with p_{i+1}
            p* ← p_{i+1};
            prune S, R and O of all solutions above a rank threshold;
        else
            Remove p_{i+1} from solution archive S;
            Remove o_{i+1} from error archive O;
        end

        i ← i + 1;
    end

    // Update T using the user cooling function
    T ← user-function::cooling(T);
end
```

Figure 4.1: Pseudo-code for the main run-loop of JuPOETs. The user specifies the `neighbor`, `acceptance`, `cooling` and `objective` functions along with an initial parameter guess. The rank archive $\mathcal{R}$, solution archive $\mathcal{S}$ and objective archive $\mathcal{O}$ are initialized from the initial guess. The initial guess is perturbed in the `neighbor` function, which generates a new solution whose performance is evaluated using the user supplied `objective` function. The new solution and objective values are then added to the respective archives and ranked using the builtin `rank` function. If the new solution is accepted (based upon a probability calculated with the user supplied `acceptance` function) it is added to the solution and objective archive. This solution is then perturbed during the next iteration of the algorithm. However, if the solution is not accepted, it is removed from the archive and discarded. The computational temperature is adjusted using the user supplied `cooling` function after each $\mathcal{I}$ iterations.

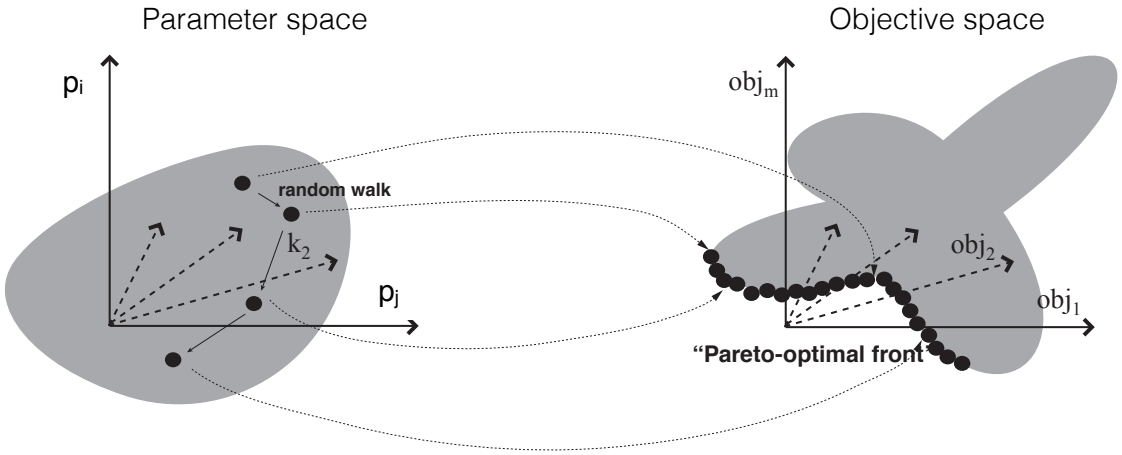

Figure 4.2: Schematic of multiobjective parameter mapping. The performance of any given parameter set is mapped into an objective space using a ranking function which quantifies the quality of the parameters. The distance away from the optimal tradeoff surface is quantified using the Pareto ranking scheme of Fonseca and Fleming in JuPOETs.

The central idea of POETs is a mapping between the value of the objective vector evaluated at $\mathbf{p}_{i+1}$ (parameter guess at iteration $i+1$) and Pareto rank. JuPOETs calculates the performance of a candidate parameter set $\mathbf{p}_{i+1}$ by calling the user defined `objective` function; `objective` takes a parameter set as an input and returns a $\mathcal{K} \times 1$ objective vector. Candidate parameter sets are generated by the user supplied `neighbor` function. The error vector associated with $\mathbf{p}_{i+1}$ is ranked using the builtin Pareto `rank` function, by comparing the current error at iteration $i+1$ to the error archive O_i (all error vectors up to iteration $i-1$ meeting a ranking criteria). Pareto rank is a measure of distance from the trade-off surface; parameter sets on or near the optimal trade-off surface between the objectives have a rank equal to 0 (no other current parameter sets are better). Sets with increasing non-zero rank are progressively further away from the optimal trade-off surface. Thus, a parameter set with a rank = 0 is *better* in a trade-off sense than rank > 0. We implemented the Fonseca and Fleming ranking scheme in the builtin `rank` function [22]:

$$\texttt{rank}\left(O_{i+1}\left(\mathbf{p}_{i+1}\right) \mid O_i\right) = r \tag{4.3}$$

where rank r is the number of parameter sets that dominate (are better than) parameter set $\mathbf{p}_{i+1}$, and $O_{i+1}\left(\mathbf{p}_{i+1}\right)$ denotes the objective vector evaluated at $\mathbf{p}_{i+1}$. We used the Pareto rank to inform the SA calculation. The parameter set $\mathbf{p}_{i+1}$ was accepted or rejected by the SA, by calculating an `acceptance` probability $\mathcal{P}\left(\mathbf{p}_{i+1}\right)$:

$$\mathcal{P}(\mathbf{p}_{i+1}) \equiv \exp\left\{-\texttt{rank}\left(O_{i+1}\left(\mathbf{p}_{i+1}\right) \mid O_i\right)/T\right\} \tag{4.4}$$

where T is the computational annealing temperature. As $\texttt{rank}\left(O_{i+1}\left(\mathbf{p}_{i+1}\right) \mid O_i\right) \rightarrow 0$, the acceptance probability moves toward one, ensuring that we explore parameter sets along the Pareto surface. Occasionally, (depending upon T) a parameter set with a high Pareto rank was accepted by the SA allowing a more

diverse search of the parameter space. However, as T is reduced, the probability of accepting a high-rank set occurring decreases. Parameter sets could also be accepted by the SA but *not* permanently archived in $\mathcal{S}_i$. Only parameter sets with rank less than or equal to threshold (rank ≤4 by default) were included in $\mathcal{S}_i$, where the archive was re-ranked and filtered after every new parameter set was accepted. Parameter bounds were implemented in the `neighbor` function as box constraints, while problem specific constraints were implemented in `objective` using a penalty method:

$$O_i + \lambda \sum_{j=1}^{C} \min\left\{0, g_j\left(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}\right)\right\} \qquad i = 1, ...\mathcal{K} \tag{4.5}$$

where λ denotes the penalty parameter (λ = 100 by default). However, because both the neighbor and objective functions are user defined, different constraint implementations are easily defined. JuPOETs can be installed using the Julia package manager from the JuPOETs repository at https://github.com/varnerlab/POETs.jl. Sample code is included in the `sample/biochemical` subdirectory of the JuPOETs repository to help users get started using JuPOETs in their projects.

4.4 Results and Discussion

JuPOETs identified optimal or nearly optimal solutions sigdfnificantly faster than Octave-POETs for a suite of multiobjective test problems (Table 4.1). The wall-clock time for JuPOETs and Octave-POETs was measured for 10 independent trials for each of the test problems.

The same `cooling`, `neighbor`, `acceptance`, and `objective` logic was employed between the implementations, and all other parameters were held

Name	Dimension	Function	Domain	Constraints
Schaffer function	1	$O_1(x) = x^2$ $O_2(x) = (x-2)^2$	$-10 \leq x \leq 10$	
Binh and Korn function	2	$O_1(x,y) = 4x^2 + 4y^2$ $O_2(x,y) = (x-5)^2 + (y-5)^2$	$0 \leq x \leq 5$ $0 \leq y \leq 3$	$g_1(x,y) = (x-5)^2 + y^2 \leq 25$ $g_2(x,y) = (x-8)^2 + (y+3)^2 \geq 7.7$
Fonseca and Fleming function	3	$O_1(x_i) = 1 - \exp\left(-\sum_{i=1}^{N}\left(x_i - \frac{1}{\sqrt{N}}\right)^2\right)$ $O_2(x_i) = 1 - \exp\left(-\sum_{i=1}^{N}\left(x_i + \frac{1}{\sqrt{N}}\right)^2\right)$	$-4 \leq x_i \leq 4$	

Table 4.1: Multi-objective optimization test problems. We tested the JuPOETs implementation on three two-dimensional test problems, with one-, two- and three-dimensional parameter vectors. Each problem had parameter bounds constraints, however, on the Binh and Korn function had additional non-linear problem constraints. For the Fonesca and Fleming problem, N = 3.

constant. For each test function, the search domain was partitioned into 10 segments, where an initial parameter guess was drawn from each partition. The number of search steps for each temperate was $\mathcal{I}$ = 10 for all cases, and the cooling parameter was α = 0.9. On average, JuPOETs identified optimal or near optimal solutions for the suite of test problems six-fold faster (60s versus 400s) than Octave-POETs (Fig. 4.3).

JuPOETs produced the characteristic tradeoff curves for each test problem, given both parameter bound and problem constraints (Fig. 4.4).

Thus, JuPOETs estimated an ensemble of solutions to constrained multiobjective optimization test problems significantly faster than the current Octave implementation. Next, we tested JuPOETs on a proof-of-concept biochemical model identification problem.

JuPOETs estimated an ensemble of biochemical models that was consistent with the mean of synthetic training data. Four synthetic training data sets were

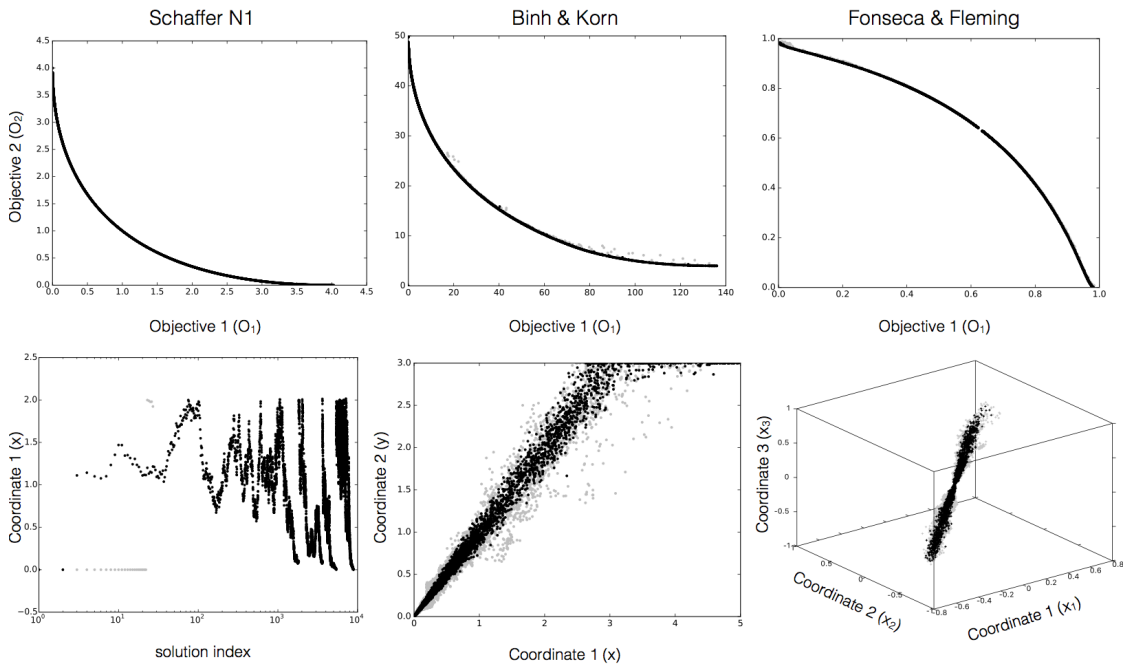

Figure 4.3: The performance of JuPOETs on the multi-objective test suite. The execution time (wall-clock) for JuPOETs and POETs implemented in Octave was measured for 10 independent trials for the suite of test problems. The number of steps per temperature $\mathcal{I} = 10$, and the cooling parameter $\alpha = 0.9$ for all cases. The problem domain was partitioned into 10 equal segments, an initial guess was drawn from each segment. For each of the test functions, JuPOETs estimated solutions on (rank zero solutions, black) or near (gray) the optimal tradeoff surface, subject to bounds and problem constraints.

generated from a prototypical biochemical network consisting of 6 metabolites and 7 reactions (Fig. 4.5).

We considered a common case in which the same measurements were made on four hypothetical cell types, each having the same biological connectivity but different performance. Network dynamics were modeled using the hybrid cybernetic model with elementary modes (HCM-EM) approach of Ramkrishna and coworkers [37]. In the HCM-EM approach, metabolic networks are first decomposed into a set of elementary modes (EMs) (chemically balanced steady-state pathways, see [83]). Dynamic combinations of elementary modes are then used to characterize network behavior. Each elementary mode is catalyzed by

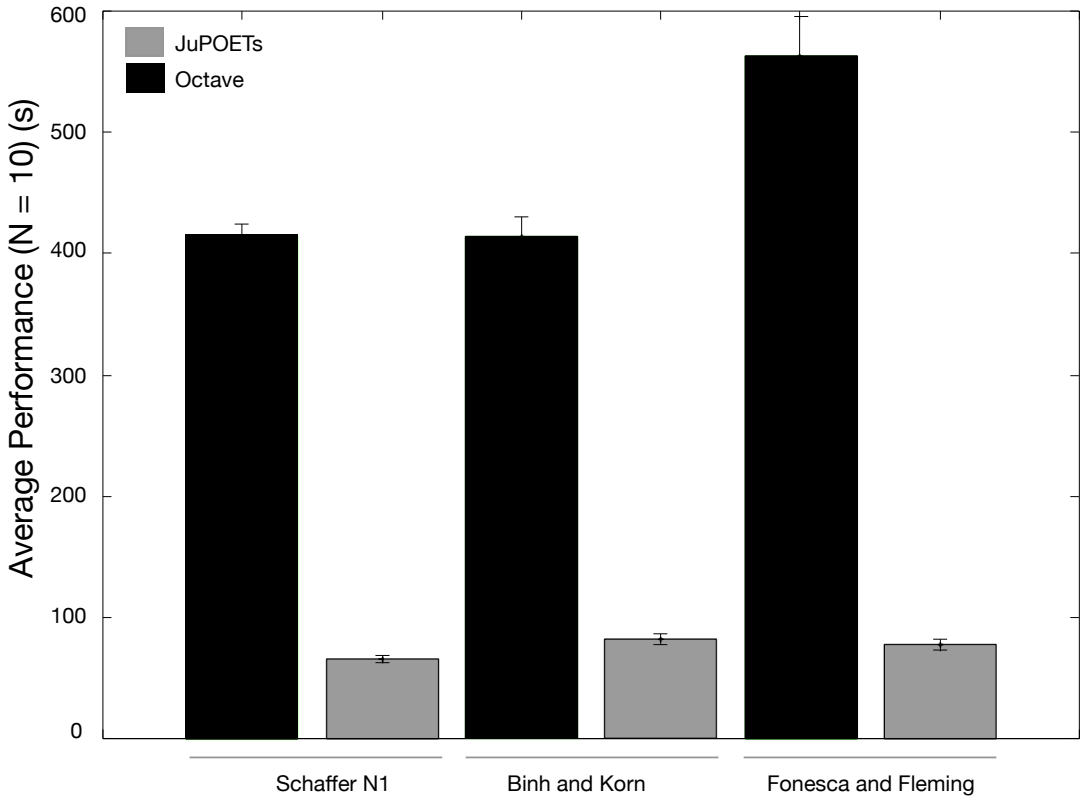

Figure 4.4: Representative JuPOETs solutions for problems in the multi-objective test suite. The number of steps per temperature $\mathcal{I} = 10$, and the cooling parameter $\alpha = 0.9$ for all cases. The problem domain was partitioned into 10 equal segments, an initial guess was drawn from each segment. For each of the test functions, JuPOETs estimated solutions on (rank zero solutions, black) or near (gray) the optimal tradeoff surface, subject to bounds and problem constraints.

a pseudo enzyme; thus, each mode has both kinetic and enzyme synthesis parameters. The proof of concept network generated 6 EMs, resulting in 13 model parameters. The synthetic data was generated by randomly varying these parameters. JuPOETs produced an ensemble of approximately $\dim \mathcal{S} \simeq 13{,}000$ parameters that captured the mean of the measured data sets for extracellular metabolites and cellmass (Fig. 4.5A and B). JuPOETs minimized the difference between the simulated and measured values for A_e, B_e, C_e and cellmass, where the residual for each data set was treated as a single objective (leading to four

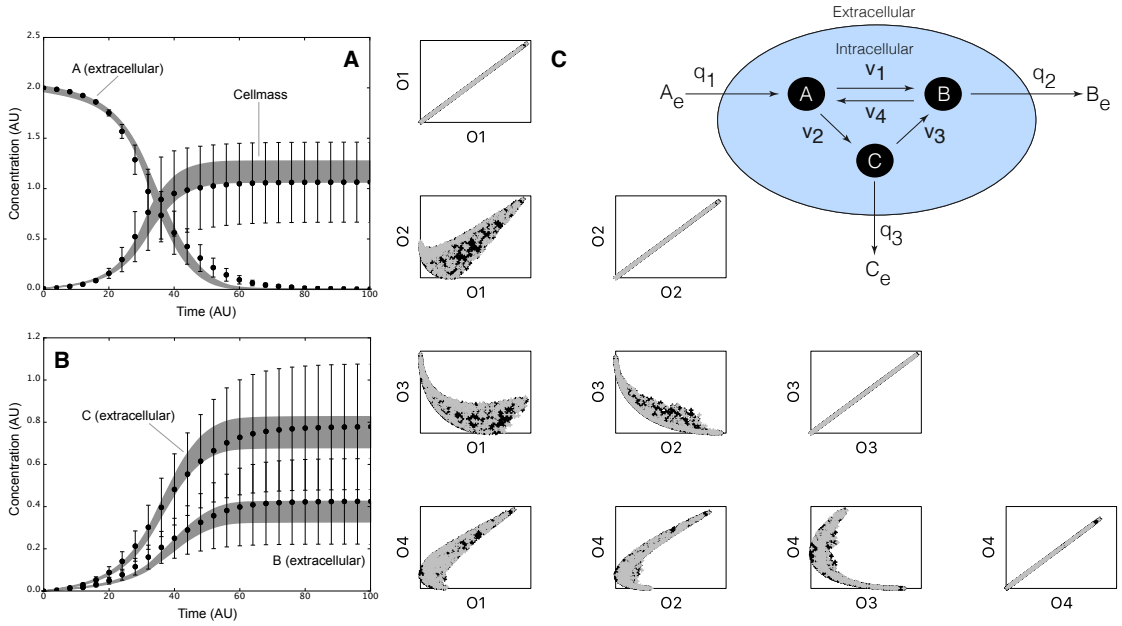

Figure 4.5: Proof of concept biochemical network study. Inset right: Prototypical biochemical network with six metabolites and seven reactions modeled using the hybrid cybernetic approach (HCM). Intracellular cellmass precursors A, B, and C are balanced (no accumulation) while the extracellular metabolites A_e, B_e, and C_e are dynamic. The oval denotes the cell boundary, q_j is the jth flux across the boundary, and v_k denotes the kth intracellular flux. Four data sets (each with A_e, B_e,C_e and cellmass measurements) were generated by varying the kinetic constants for each biochemical mode. Each data set was a single objective in the JuPOETs procedure. A: Ensemble simulation of extracellular substrate A_e and cellmass versus time. B: Ensemble simulation of extracellular substrate B_e and C_e versus time. The gray region denotes the 95% confidence estimate of the mean ensemble simulation. The data points denote mean synthetic measurements, while the error bars denote the 95% confidence estimate of the measurement computed over the four training data sets. C: Trade-off plots between the four training objectives. The quantity O_j denotes the jth training objective. Each point represents a member of the parameter ensemble, where black denotes a rank 0 set, while gray denotes rank 1 set.

objectives). The 95% confidence estimate produced by the ensemble was consistent with the mean of the measured data, despite having significant uncertainty in the training data. JuPOETs produced a consensus estimate of the synthetic data by calculating optimal trade-offs between the training data sets (Fig. 4.5C). Multiple trade-off fronts were visible in the objective plots, for example between

data set 3 (O_3) and data set 2 (O_2). Thus, without a multiobjective approach, it would be challenging to capture these data sets as fitting one leads to decreased performance on the other. However, the ensemble contained parameter sets that described each data set independently (Fig. 4.6).

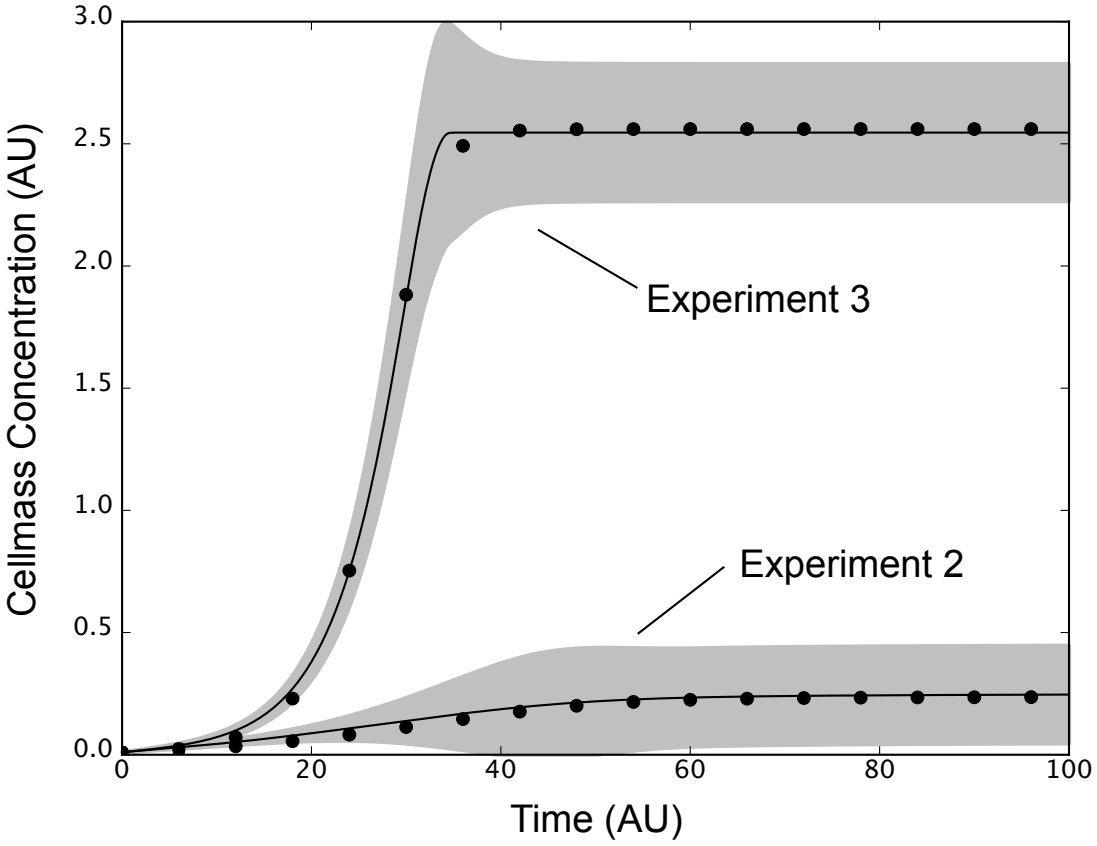

Figure 4.6: Experiment to experiment variation is captured by a single ensemble. Cellmass measurements (points) versus time for experiment 2 and 3 were compared with ensemble simulations. The full ensemble was sorted by simultaneously selecting the top 25% of solutions for each objective with rank ≤ 1. The best fit solution for each objective (line) ± 1-standard deviation (gray region) for experiment 2 and 3 brackets the training data despite significant differences the training values between the two data sets.

Thus, JuPOETs produced an ensemble of parameters that gave the mean of the training data for conflicting data sets, while simultaneously estimating parameter sets that performed well on each individual objective function.

4.5 Conclusions

JuPOETs is a significant advance over the previous POETs implementation. It offers improved performance and is highly adaptable to different problem types. We demonstrated JuPOETs on a suite of test problems, and a proof-of-concept biochemical model. However, there are several areas that could be explored further to improve JuPOETs. First, JuPOETs should be compared with other multiobjective evolutionary algorithms (MOEAs) to determine its relative performance on test and real world problems. Many evolutionary approaches e.g., the nondominated sorting genetic algorithm (NSGA) family of algorithms, have been adapted to solve multiobjective optimization problems [33, 31]. It is unclear if JuPOETs will perform as well as these other approaches; one potential advantage that JuPOETs may have is the local refinement step which temporarily reduces the problem to a single objective formulation. Previously, this hybrid approach led to better convergence on a proof-of-concept signal transduction model [92]. For many real world parameter estimation problems, the bulk of the execution time is spent evaluating the objective functions. One strategy to improve performance could be to optimize surrogates [10], while another would be parallel execution of the objective functions. Currently, JuPOETs serially evaluates the objective function vector. However, parallel evaluation of the objective functions could be easily implemented using a variety of techniques without changing the main run loop of JuPOETs. Because of the flexible function pointer architecture of JuPOETs, the only changes required are in the user defined objective function. Taken together, JuPOETs has demonstrated improved flexibility, and performance over POETs in parameter identification and ensemble generation for multiple objectives. JuPOETs has the potential for widespread

use due to the flexibility of the implementation, and the high level syntax and distribution tools native to Julia.

CHAPTER 5

REDUCED ORDER MODELING AND ANALYSIS OF THE HUMAN COMPLEMENT SYSTEM

Submitted for publication to Scientific Reports. Adithya Sagar, Wei Dai, Mason Minot, and Jeffrey D. Varner.

doi: http://dx.doi.org/10.1101/059386

Author contributions: J.V directed the study. A.S developed the reduced order complement model and the parameter ensemble. A.S, W.D. and M.M analyzed the model ensemble and generated figures for the manuscript. The manuscript was prepared and edited for publication by A.S, W.D, M.M, and J.V.

5.1 Introduction

Complement is an important pathway in innate immunity. It plays a significant role in inflammation, host defense as well as many disease processes. Complement was discovered in the late 1880s where it was found to 'complement' the bactericidal activity of natural antibodies [59]. However, research over the past decade has shown the importance of complement extends beyond innate immunity. For example, complement contributes to tissue homeostasis [69], and has been linked with several diseases including Alzheimers, Parkinson's, multiple sclerosis, schizophrenia, rheumatoid arthritis and sepsis [70, 73]. Complement also plays positive and negative roles in cancer; attacking tumor cells with altered surface proteins in some cases, while potentially contributing to tumor growth in others [79, 71]. Lastly, several other important biochemical systems are integrated with complement including the coagulation cascade, the

autonomous nervous system and inflammation [71]. Thus, complement is important in a variety of beneficial and potentially harmful functions in the body. However, despite its importance, there have been relatively few approved complement specific therapeutics, largely because of safety concerns and challenging pharmacokinetic constraints.

The complement cascade involves many soluble and cell surface proteins, receptors and regulators [107, 108]. The outputs of complement are the Membrane Attack Complex (MAC), and the inflammatory mediator proteins C3a and C5a. The membrane attack complex, generated during the terminal phase of the response, forms transmembrane channels which disrupt the membrane integrity of targeted cells, leading to cell lysis and death. On the other hand, the C3a and C5a proteins act as a bridge between innate and adaptive immunity, and play an important role in regulating inflammation [79]. Complement activation takes places through three pathways: the classical, the lectin and the alternate pathways. The classical pathway is triggered by antibody recognition of foreign antigens or other pathogens. A multimeric protein complex C1 binds antibody-antigen complexes and undergoes a conformational change, leading to an activated form with proteolytic activity. The activated complex cleaves soluble complement proteins C4 and C2 into C4a, C4b, C2a and C2b, respectively. The C4a and C2b fragments bind to form the C4bC2a protease, also known as the classical pathway C3 convertase (CP C3 convertase). The lectin pathway is initiated through the binding of L-ficolin or Mannose Binding Lectin (MBL) to carbohydrates on the surfaces of bacterial pathogens. These complexes, in combination mannose-associated serine proteases 1 and 2 (MASP-1/2), also cleave C4 and C2, leading to additional CP C3 convertase. Thus, the classical and lectin pathways, initiated by different cues on foreign surfaces, converge at the CP C3

convertase. However, the alternate pathway works differently. It is activated by a 'tickover' mechanism in which complement protein C3 is spontaneously hydrolyzed to form an activated intermediate C3w; C3w recruits factor B and factor D, leading to the formation of C3wBb. C3wBb cleaves C3 into C3a and C3b, where the C3b fragment further recruits additional factor B and factor D to form C3bBb, the alternate C3 convertase (AP C3 convertase) [64]. The role of classical and alternate C3 convertases is varied. First, AP C3 convertases mediate signal amplification. AP C3 convertases cleave C3 into C3a and C3b; the C3b fragment is then free to form additional alternate C3 convertases, thereby forming a positive feedback loop. Next, AP/CP C3 convertases link complement initiation with the terminal phase of the cascade through the formation of C5 convertases. Both classical and alternate C3 convertases can recruit C3b subunits to form the classical pathway C5 convertase (C4bC2aC3b, CP C5 convertase), and the alternate pathway C5 convertase (C3bBbC3b, AP C5 convertase), respectively. Both C5 convertases cleave C5 into the C5a and C5b fragments. The C5b fragment, along with the complement proteins C6, C7, C8 and multiple C9s, form the membrane attack complex. On the other hand, both C3a and C5a are important inflammatory signals involved in several responses [107, 108]. Thus, the complement cascade attacks invading pathogens, while acting as a beacon for adaptive immunity.

The complement cascade is regulated by plasma and host cell proteins which balance host safety with effectiveness. The initiation of the classical pathway via complement protein C1 is controlled by the C1 Inhibitor (C1-Inh); C1-Inh irreversibly binds to and deactivates the active subunits of C1, preventing chronic complement activation [106]. Regulation of upstream processes in the lectin and alternate pathways also occurs through the interaction of the C4 binding

protein (C4BP) with C4b, and factor H with C3b [9]. Interestingly, both factor H and C4BP are capable of binding their respective targets while in convertase complexes as well. At the host cell surface, membrane cofactor protein (MCP or CD46) can interact with C4b and C3b, which protects the host cell from complement self-activation [72]. Delay accelerating factor (DAF or CD55) also recognizes and dissociates both C3 and C5 convertases on host cell surfaces [49]. More generally the well known inflammation regulator Carboxypeptidase-N has broad activity against the complement proteins C3a, C4a, and C5a, rendering them inactive by cleavage of carboxyl-terminal arginine and lysine residues [45]. Although Carboxypeptidase-N does not directly influence complement activation, it silences the important inflammatory signals produced by complement. Lastly, assembly of the MAC complex itself can be inhibited by vitronectin and clusterin in the plasma, and CD59 at the host surface [11, 115]. Thus, there are many points of control which influence complement across the three activation pathways.

Developing quantitative mathematical models of complement will be crucial to fully understanding its role in the body. Traditionally, complement models have been formulated as systems of linear or non-linear ordinary differential equations (ODEs). For example, Hirayama et al., modeled the classical complement pathway as a system of linear ODEs [29], while Korotaevskiy and co-workers modeled the classical, lectin and alternate pathways as a system of non-linear ODEs [40]. More recently, large mechanistic models of sections of complement have also been proposed. For example, Liu et al., analyzed the formation of the classical and lectin C3 convertases, and the regulatory role of C4BP using a system of 45 non-linear ODEs with 85 parameters [46]. Zewde and co-workers constructed a detailed mechanistic model of the alternative pathway

which consisted of 107 ODEs and 74 kinetic parameters and delineated between the fluid, host and pathogen surfaces [115]. However, these previous modeling studies involved large models with little experimental validation. Thus, while these models are undoubtably important theoretical tools, it is unclear if they can describe or quantitatively predict complement measurements. The central challenge of complement model identification is the estimation of model parameters from experimental measurements. Unlike other important cascades, such as coagulation where there are well developed experimental tools and publicly available data sets, the data for complement is relatively sparse. Data sets with missing or incomplete data, and limited dynamic data also make the identification of large mechanistic complement models difficult. Thus, reduced order approaches which describe the biology of complement using a limited number of species and parameters could be important for pharmacokinetic model development, and for our understanding of the varied role of complement in the body.

5.2 Results

In this study, we developed an ensemble of experimentally validated reduced order complement models using multiobjective optimization. The modeling approach combined ordinary differential equations with logical rules to produce a complement model with a limited number of equations and parameters. The reduced order model, which described the lectin and alternative pathways, consisted of 18 differential equations with 28 parameters. Thus, the model was an order of magnitude smaller and included more pathways than comparable models in the literature. We estimated an ensemble of model parameters from

in vitro time series measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a measurements that were not used for model training. Despite its small size, the model was surprisingly predictive. After validation, we performed global sensitivity and robustness analysis to estimate which parameters and species controlled model performance. Sensitivity analysis suggested CP C3 and C5 convertase parameters were critical, while robustness analyses suggested complement was robust to any single therapeutic intervention; only the knockdown of both C3 and C5 consistently reduced C3a and C5a formation for all cases. Taken together, we developed a reduced order complement model that was computationally inexpensive, and could easily be incorporated into pre-existing or new pharmacokinetic models of immune system function. The model described experimental data, and predicted the need for multiple points of intervention to disrupt complement activation.

5.2.1 Reduced order complement network.

The complement model described the alternate and lectin pathways (Fig. 5.1).

A trigger event initiated the lectin pathway (encoded as a logical rule), which activated the cleavage of C2 and C4 into C2a, C2b, C4a and C4b respectively. Classical Pathway (CP) C3 convertase (C4aC2b) then catalyzed the cleavage of C3 into C3a and C3b. The alternate pathway was initiated through the spontaneous hydrolysis of C3 into C3a and C3b (not C3w). The C3b fragment generated by hydrolysis (or by CP C3 convertase) could then form the alternate pathway (AP) C3 convertase (C3bBb). We did not consider C3w, nor the forma-

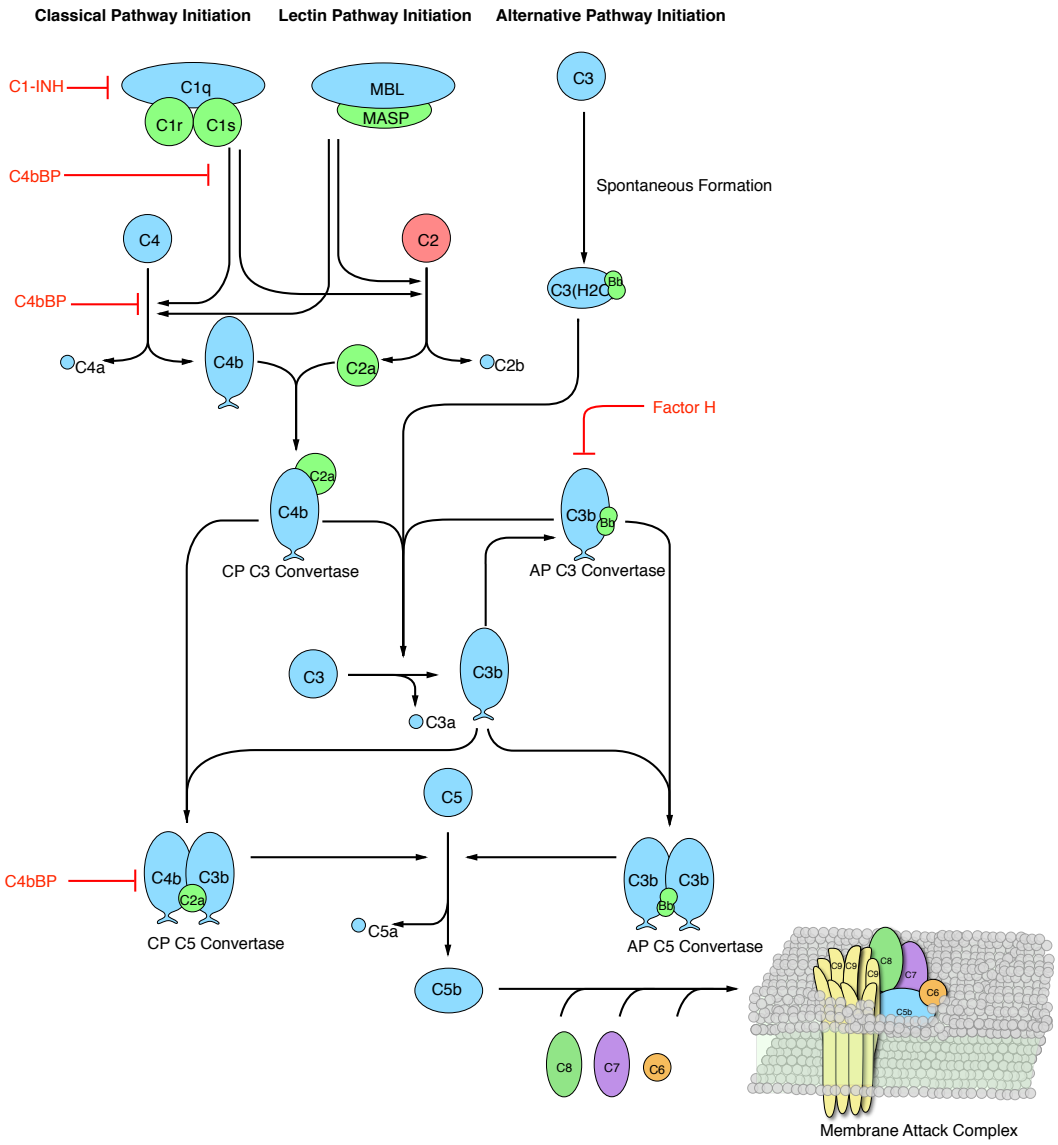

Figure 5.1: Simplified schematic of the human complement system. The complement cascade is activated through three pathways: the classical, the lectin, and the alternate pathways. Complement initiation results in the formation of classical or alternative C3 convertases, which amplify the initial complement response and signal to the adaptive immune system by cleaving C3 into C3a and C3b. C3 convertases further react to form C5 convertases which catalyze the cleavage of the C5 complement protein to C5a and C5b. C5b is critical to the formation of the membrane attack complex (MAC), while C5a recruits an adaptive immune response.

tion of the initial alternate C3 convertase (C3wBb). Rather, we assumed C3w was equivalent to C3b and only modeled the formation of the main AP C3 convertase. Both the CP and AP C3 convertases catalyzed the cleavage of C3 into C3a and C3b. A second C3b fragment could then bind with either the CP or AP C3 convertase to form the CP or AP C5 convertase (C4bC2aC3b or C3bBbC3b). Both C5 convertases catalyzed the cleavage of C5 into the C5a and C5b fragments. In this initial study, we simplified the model by assuming both factor B and factor D were in excess. However, we did explicitly account for two control proteins, factor H and C4BP. Lastly, we did not consider MAC formation, instead we stopped at C5a and C5b. Lectin pathway activation, and C3/C5 convertase activity was modeled using a combination of saturation kinetics and non-linear transfer functions, which facilitated a significant reduction in the size of the model while maintaining performance. Binding interactions were modeled using mass-action kinetics, where we assumed all binding was irreversible. Thus, while the reduced order complement model encoded significant biology, it was highly compact consisting of only 18 differential equations and 28 model parameters. Next, we estimated an ensemble of model parameters from time series measurements of the C3a and C5a complement proteins.

5.2.2 Estimating an ensemble of reduced order complement models.

A critical challenge for the development of any dynamic model is the estimation of model parameters. We estimated an ensemble of complement model parameters using *in vitro* time-series data sets generated with and without zymosan, a

lectin pathway activator [53]. The residual between model simulations and experimental measurements was minimized using the dynamic optimization with particle swarms (DOPS) routine, starting from a random parameter guess. The best fit parameter set estimated by DOPS was then used to generate a parameter ensemble using multiobjective optimization. Unless otherwise specified, all initial conditions were assumed to be their mean physiological values. While we had significant training data, the parameter estimation problem was underdetermined (we were not able to uniquely determine model parameters). Thus, instead of using the best-fit yet uncertain parameter set generated by DOPS, we estimated an ensemble of probable parameter sets to quantify model uncertainty (N = 2100, see materials and methods). The complement model ensemble captured the behavior of both the alternate and lectin pathways (Fig. 5.2). For the alternate pathway, we used C3a and C5a measurements in the absence of zymosan (Fig. 5.2A and B). On the other hand, lectin pathway parameters were estimated from C3a and C5a measurements in the presence of 1mg/ml zymosan (Fig. 5.2C and D). The reduced order model reproduced a panel of alternate and lectin pathway data sets in the neighborhood of physiological factor and inhibitor concentrations. However, it was unclear whether the reduced order model could predict new data, without updating the model parameters. To address this question, we fixed the model parameters and simulated data sets not used for model training.

We tested the predictive power of the reduced order complement model with data not used during model training (Fig. 5.3). Six validation cases were considered, three for C3a and C5a, respectively. All model parameters and initial conditions were fixed for the validation simulations (with the exception of zymosan, and other experimentally mandated changes). The ensemble of re-

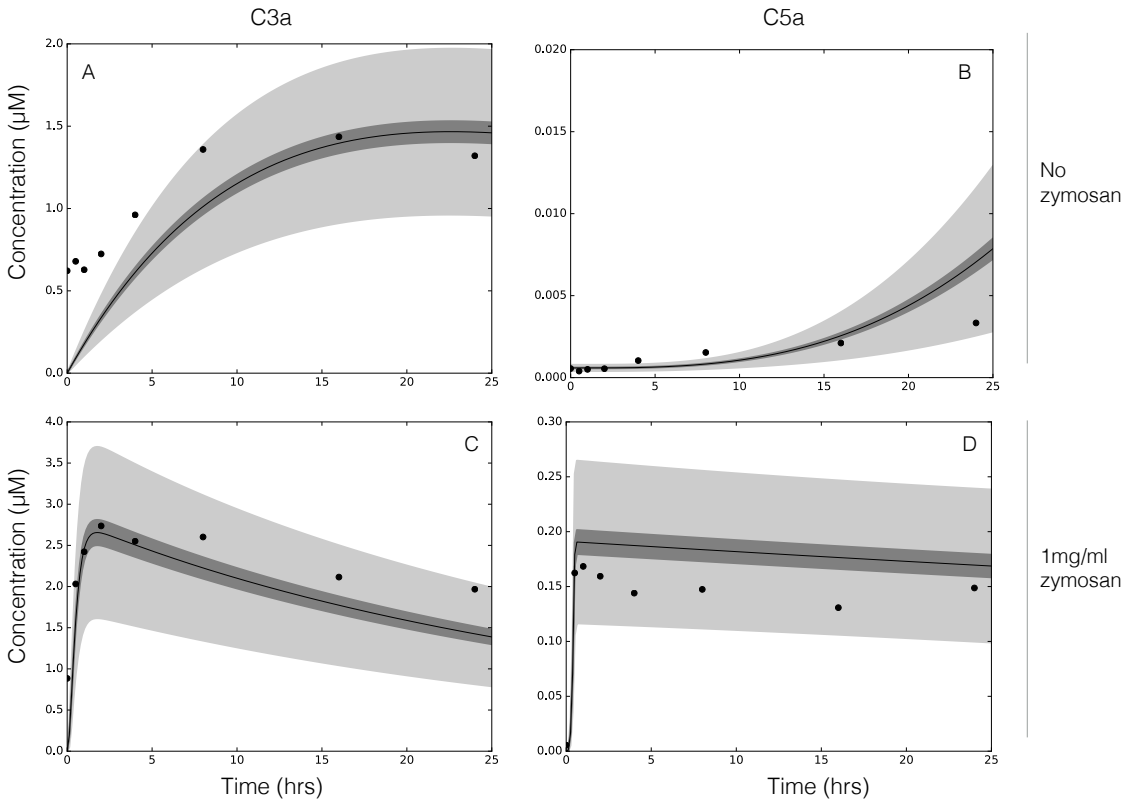

Figure 5.2: Reduced order complement model training. An ensemble of model parameters were estimated using multiobjective optimization from C3a and C5a measurements with and without zymosan [53]. The model was trained using C3a and C5a data generated from the alternative pathway (**A**–**B**) and lectin pathway initiated with 1 mg/ml zymosan (**C**–**D**). The solid black lines show the simulated mean value of C3a or C5a for the ensemble, while the dark shaded region denotes the 99% confidence interval of mean. The light shaded region denotes the 99% confidence interval of the simulated C3a and C5a concentration. All initial conditions were assumed to be at their physiological serum levels unless otherwise noted.

duced order models predicted the qualitative dynamics of C3a formation (Fig. 5.3, top), and C5a formation (Fig. 5.3, bottom) at three inducer concentrations. The rate of C3a formation and C3a peak time were directly proportional to initiator dose. Similarly, the C5a plateau and rate of formation were also directly proportional to initiator dose, with the lag time being indirectly proportional to initiator exposure for both C3a and C5a. However, there were shortcomings with model performance. First, while the overall C3a trend was captured

(within the 99% confidence interval), the C3a dynamics were too fast with the exception of the low dose case. We believe the C3a time scale was related to our choice of training data, how we modeled the tickover mechanism, and factor B and D limitation. We trained the model using either no or 1 mg/ml zymosan, but predicted cases in a different initiator range; comparing training to prediction, the model performance e.g., the shape of the C3a trajectory was biased towards either high or very low initiator doses. Next, tickover was modeled as a first-order generation processes where C3wBb formation and activity was lumped into the AP C3 convertase. Thus, we skipped an important upstream step which could strongly influence AP C3 convertase formation by slowing down the rate C3 cleavage into C3a and C3b. We also assumed both factor B and factor D were not limiting, thereby artificially accelerating the rate of AP C3 convertase formation. The C5a predictions followed a similar trend as C3a; we captured the long-time C5a behavior but over predicted the time scale of C5 cleavage. However, because the C5a time scale depends strongly upon C3 convertase formation, we can likely correct the C5 issues by fixing the rate of C3 cleavage. Despite these shortcomings, we qualitatively predicted unseen experimental measurements typically within the 99% confidence of the ensemble, for three inducer levels. Next, we used global sensitivity and robustness analysis to determine which parameters and species controlled the performance of the complement model.

5.2.3 Global analysis of the reduced order complement model.

We conducted sensitivity analysis to estimate which parameters controlled the performance of the reduced order complement model. We calculated the total

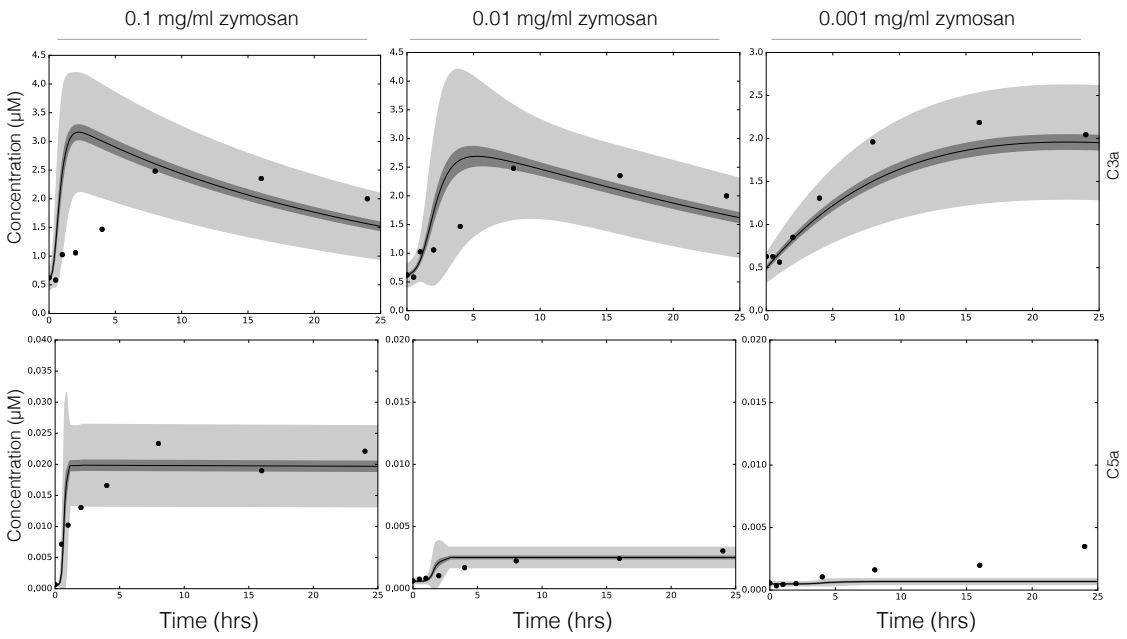

Figure 5.3: Reduced order complement model predictions. Simulations of C3a and C5a generated in the lectin pathway using 0.1 mg/ml, 0.01 mg/ml, and 0.001 mg/ml zymosan were compared with the corresponding experimental measurements. The solid black lines show the simulated mean value of C3a or C5a for the ensemble, while the dark shaded region denotes the 99% confidence interval of mean. The light shaded region denotes the 99% confidence interval of the simulated C3a and C5a concentration. All initial conditions were assumed to be at their physiological serum levels unless otherwise noted.

sensitivity of the C3a and C5a residual to changes in model parameters with and without zymosan (Fig. 5.4). In the absence of zymosan (where only the alternative pathway is active), the most sensitive parameter was the rate constant governing the assembly of the AP C3 convertase, as well as the rate constant controlling basal C3b formation. The C5a trajectory was sensitive to the AP C5 convertase kinetic parameters (Fig. 5.4A). Interestingly, neither the rate nor the saturation constant governing AP C3 convertase activity were sensitive in the absence of zymosan. Thus, C3a formation in the alternative pathway was more heavily influenced by the spontaneous hydrolysis of C3, rather than AP C3 convertase activity, in the absence of zymosan. In the presence of zymosan, the C3a residual was controlled by the formation and activity of the CP C3 con-

vertase, as well as tickover and degradation parameters. On the other hand, the C5a residual was controlled by the formation and activity of CP C5 convertase, and tickover C3b formation in the presence of zymosan (Fig. 5.4B). The lectin initiation parameters were sensitive, but to a lesser extent than CP convertase kinetic parameters and tickover C3b formation. Thus, sensitivity analysis suggested that CP C3/C5 convertase formation and activity dominated in the presence of zymosan, but tickover parameters and AP C5 convertase were more important without initiator. AP C3 convertase assembly was important, but its activity was not. Next, we compared the sensitivity results to current therapeutic approaches; pathways involving sensitive parameters have been targeted for clinical intervention (Fig. 5.4C). In particular, the sensitivity analysis suggested AP/CP C5 convertase inhibitors, or interventions aimed at attenuating C3 or C5 would most strongly influence complement performance. Thus, there was at least a qualitative overlap between sensitivity and the potential of biochemical efficacy. However, sensitivity coefficients are only a local measure of how small changes in parameters affect model performance. To more closely simulate a clinical intervention e.g., administration of an anti-complement inhibitor, we performed robustness analysis.

Robustness analysis suggested there was no single intervention that inhibited complement activation in the presence of both initiation pathways (Fig. 5.5). Robustness coefficients quantify the response of a protein to a macroscopic structural or operational perturbation to a biochemical network. Here, we computed how the C3a and C5a trajectories responded to a decrease in the initial abundance of C3 and/or C5 with and without lectin initiator. We simulated the addition of different doses of anti-complement inhibitor cocktails by decreasing the initial concentration of C3, C5 or the combination of C3 and C5 by 50%, 90%

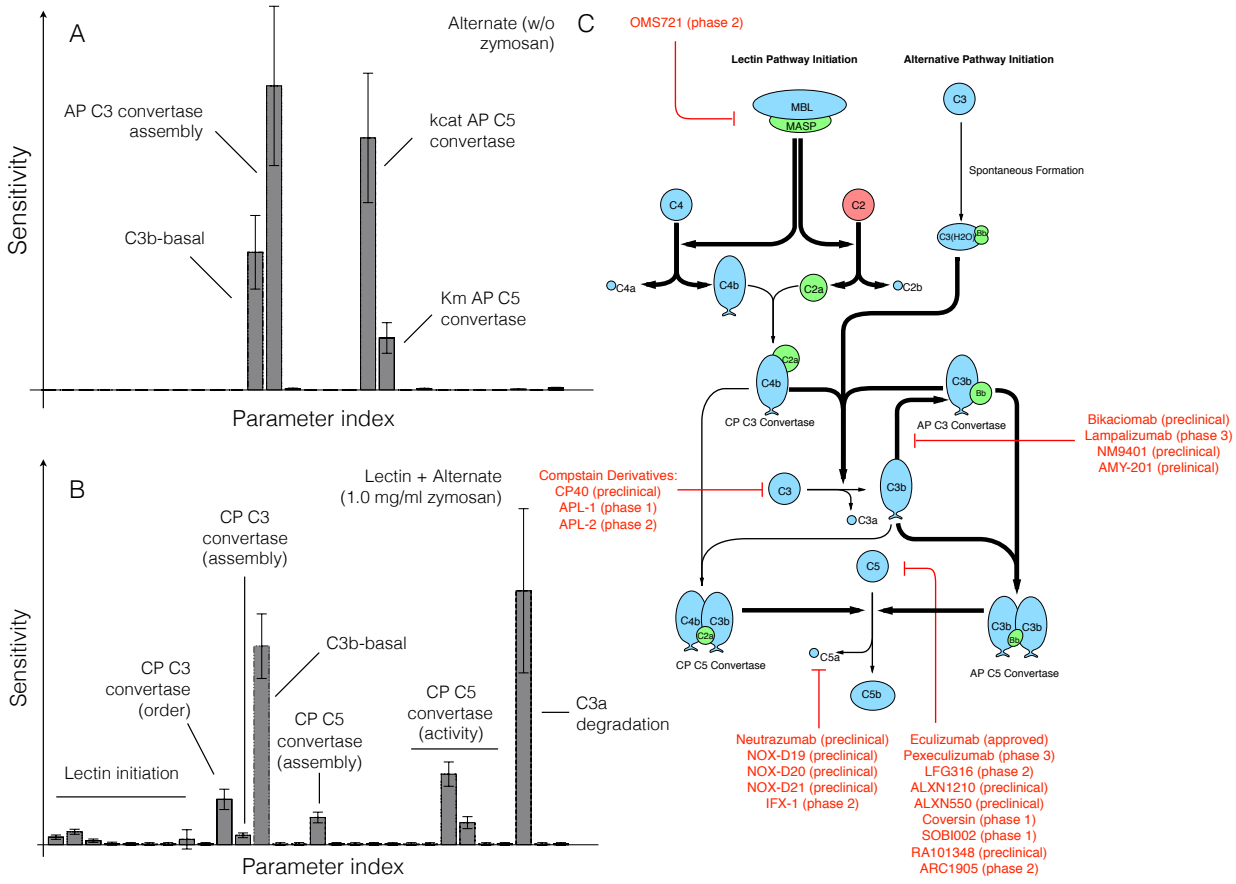

Figure 5.4: Global sensitivity analysis of the reduced order complement model. Sensitivity analysis was conducted on the two objectives used for model training. **A**: Sensitivity of the C3a and C5a residual w/o zymosan. **B**: Sensitivity of the C3a and C5a residual with 1 mg/ml zymosan. The bars denote the mean total sensitivity index for each parameter, while the error bars denote the 95% confidence interval. **C**: Pathways controlled by the sensitivity parameters. Bold black lines indicate the pathway involves one or more sensitive parameters, while the red lines show current therapeutics targets. Current complement therapeutics were taken from the review of Morgan and Harris [54].

and 99%. This would be conceptually analogous to the administration of a C3 inhibitor e.g., Compstatin alone or combination with Eculizumab (Fig. 5.4C). The response of the complement model to different knock-down magnitudes was non-linear; a 90% knock-down had an order of magnitude more impact than a 50% knock-down. As expected, a C5 knockdown had no effect on C3a formation for either the alternate (Fig. 5.5A) or lectin pathways (Fig. 5.5B).

However, C3a and to a greater extent C5a abundance decreased with decreasing C3 concentration in the alternate pathway (Fig. 5.5A). This agreed with the sensitivity results; changes in AP C3-convertase formation affected the downstream dynamics of C5a formation. Thus, if we only considered the alternate pathway, C3 alone could be a reasonable target, especially given that C5a formation was surprisingly robust to C5 levels in the alternate pathway. Yet, when both pathways were activated, C5a levels were robust to the initial C3 concentration (Fig. 5.5B); even 1% of the nominal C3 was able to generate enough AP/CP C5 convertase to maintain C5a formation. Thus, the only reliable intervention that consistently reduced both C3a and C5a formation for all cases was a knockdown of both C3 and C5. For example, a 90% decrease of both C3 and C5 reduced the formation of C5a by an order of magnitude, while C3a was reduced to a lesser extent (Fig. 5.5B).

5.3 Discussion

In this study, we developed an ensemble of experimentally validated reduced order complement models using multiobjective optimization. The modeling approach combined ordinary differential equations with logical rules to produce a complement model with a limited number of equations and parameters. The reduced order model, which described the lectin and alternative pathways, consisted of 18 differential equations with 28 parameters. Thus, the model was an order of magnitude smaller and included more pathways than comparable mathematical models in the literature. We estimated an ensemble of model parameters from *in vitro* time series measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a

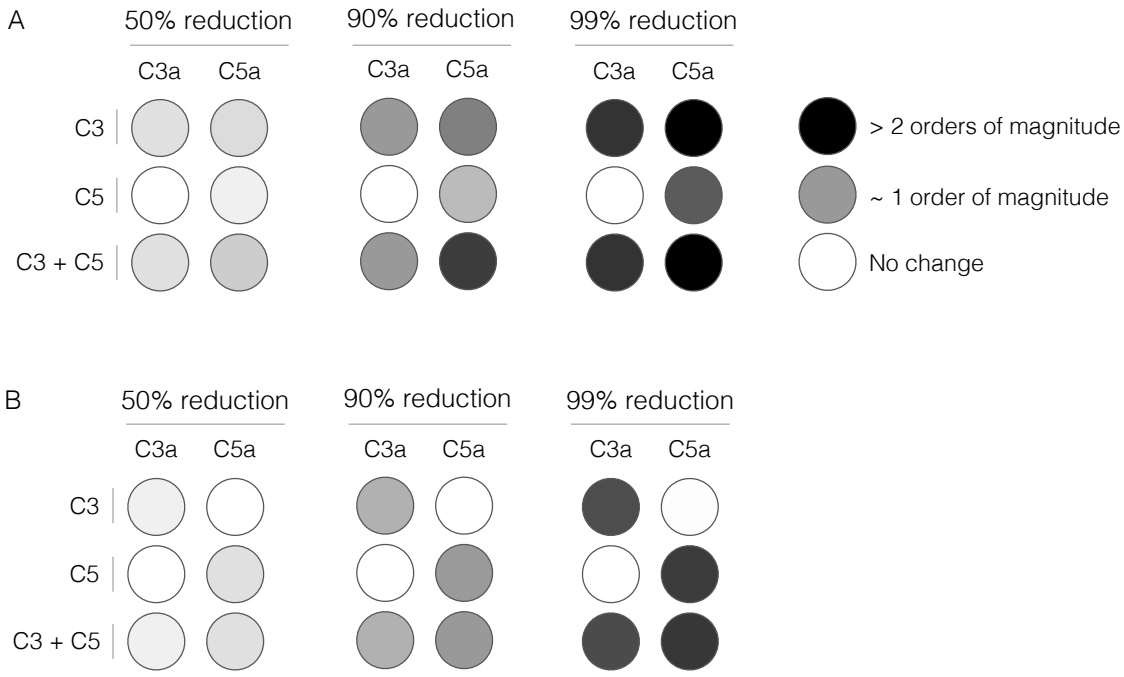

Figure 5.5: Robustness analysis of the complement model. Robustness coefficients were calculated for a 50, 90, and 99 percent reduction in C3, C5, or C3 and C5 initial conditions. **A**: Mean robustness index for C3a and C5a generated from the alternate pathway (w/o zymosan). **B**: Mean robustness index for C3a and C5a generated from the lectin and alternate pathway (1 mg/ml zymosan). The color describes the degree of reduction of C3a or C5a following the network perturbation. Robustness coefficients were calculated using all parameter sets with Pareto rank less than five (N = 65).

measurements that were not used for model training. Despite its small size, the model was surprisingly predictive. After validation, we performed global sensitivity and robustness analysis to estimate which parameters and species controlled model performance. These analyses suggested complement was robust to any single therapeutic intervention. The only intervention that consistently reduced C3a and C5a formation for all cases was a knockdown of both C3 and C5. Taken together, we developed a reduced order complement model that was computationally inexpensive, and could easily be incorporated into pre-existing or new pharmacokinetic models of immune system function. The model described experimental data, and predicted the need for multiple points

of intervention to disrupt complement activation.

Despite its importance, there has been a paucity of validated mathematical models of complement pathway activation. To our knowledge, this study is one of the first complement models that combined multiple initiation pathways with experimental validation of important complement products like C5a. However, there have been several theoretical models of components of the cascade in the literature. Liu and co-workers modeled the formation of C3a through the classical pathway using 45 non-linear ODEs [46]. In contrast, in this study we modeled lectin mediated C3a formation using only five ODEs. Though we did not model all the initiation interactions in detail, especially the cross-talk between the lectin and classical pathways, we successfully captured C3a dynamics with respect to different concentrations of lectin initiators. The model also captured the dynamics of C3a and C5a formed from the alternate pathway using only seven ODEs. The reduced order model predictions of C5a were qualitatively similar to the theoretical complement model of Zewde et al., which involved over 100 ODEs [115]. However, we found that the C3a produced in the alternate pathway was nearly three orders of magnitude greater than the C5a generated. While this was in agreement with the experimental data [53], it differed from the theoretical predictions made by Zewde et al., who showed C3a was eight orders of magnitude greater than the C5a concentration [115]. In our model, the time profile of both C3a and C5a generated changed with respect to the quantity of zymosan (the lectin pathway initiator). In particular, the C3a peak time was directly proportional to initiator, while the lag phase for generation was inversely proportional to the initiator concentration. Korotaevskiy et al. showed a similar trend using a theoretical model of complement, albeit for much shorter time scales [40]. Thus, the reduced order complement

model performed at least as well as existing larger mechanistic models, despite being significantly smaller.

Global analysis of the complement model suggested potentially important therapeutic targets. Complement malfunctions are implicated in a spectrum of diseases, however the development of complement specific therapeutics has been challenging [70, 54]. Previously, we have shown that mathematical modeling and analysis can be useful tools to estimate therapeutically important mechanisms [48, 57, 97, 68]. In this study, we analyzed a validated ensemble of reduced order complement models to better understand the strengths and weaknesses of the cascade. In the presence of an initiator, C3a and C5a formation was sensitive to CP C3/C5 convertase assembly and activity, and to a lesser extent lectin initiation parameters. Formation of the CP convertases can be inhibited by targeting upstream protease complexes like MASP-1,2 from the lectin pathway (or C1r, C1s from classical pathway). For example, Omeros, a protease inhibitor that targets the MASP-2 complex, has been shown to inhibit the formation of downstream convertases [84]. Lampalizumab and Bikaciomab, which target factor B and factor D respectively, or naturally occurring proteins such as Cobra Venom Factor (CVF), an analogue of C3b, could also attenuate AP convertase formation [104, 35, 30]. Removing supporting molecules could also destabilize the convertases. For example, Novelmed Therapeutics developed the antibody, NM9401 against propedin, a small protein that stabilizes alternate C3 convertase [5]. Lastly, convertase catalytic activity could be attenuated using small molecule protease inhibitors. All of these approaches are consistent with the results of the sensitivity analysis. On the other hand, robustness analysis suggested C3a and C5a generation could only be significantly attenuated by modulating the free levels of C3 and C5. The most commonly used

anti-complement drug Eculizumab, targets the C5 protein [54]. Several other antibodies targeting C5 are also being developed; for example, LFG316 targets C5 in Age-Related Macular Degeneration [74], while Mubodina is used to treat Atypical Hemolytic-Uremic Syndrome (aHUS) [51]. Other agents such as Coversin [109] or the aptamer Zimura [21] could also be used to knockdown C5. The peptide inhibitor Compstatin and its derivatives are promising approaches for the inhibition of C3 [50]. However, while the knockdown of C3 and C5 affect C3a and C5a levels downstream, the abundance, turnover rate and population variation of these proteins make them difficult targets [86, 94]. For example, the eculizumab dosage must be significantly adjusted during the course of treatment for aHUS [58]. A validated complement model, in combination with personalized pharmacokinetic models of immune system function, could be an important development for the field.

The performance of the complement model was impressive given its limited size. However, there are several questions that should be explored further. A logical progression for this work would be to expand the network to include the classical pathway and the formation of the membrane attack complex (MAC). However, time course measurements of MAC abundance (and MAC formation dynamics) are scarce, making the inclusion of MAC challenging. On the other hand, inclusion of classical pathway activation is straightforward. Liu et al., have shown cross-talk between the activation of the classical and lectin pathways through C reactive proteins (CRP) and L-ficolin (LF) under inflammation conditions [46]. Thus, inclusion of these species, in addition to a lumped activation term for the classical pathway should allow us to capture classical activation. Next, we should address the C3a time scale issue. We believe the C3a time scale was related to our choice of training data, how we modeled

the tickover mechanism, and factor B and D limitation. Tickover was modeled as a first-order generation processes where C3wBb formation and activity was lumped into the AP C3 convertase. Thus, we skipped an important step which could strongly influence AP C3 convertase formation by slowing down the rate C3 cleavage into C3a and C3b. The model should be expanded to include the C3wBb intermediate, where C3wBb catalyzes C3 cleavage at a slow rate compared to normal AP or CP C3 convertases. We also assumed both factor B and factor D were not limiting, thereby artificially accelerating the rate of AP C3 convertase formation. This shortcoming could be addressed by including balances around factor B and D, and including these species in the appropriate kinetic rates. The C5a predictions also had an accelerated time scale. However, because the C5a time scale depended strongly upon C3 convertase formation, we can likely correct the C5 issues by fixing the rate of C3 cleavage. Lastly, we should also consider including the C2-bypass pathway, which was not included in the model. The C2-bypass mediates lectin pathway activation, without the involvement of MASP-1/2. Thus, this pathway could be important for understanding the role of MASP-1/2 inhibitors on complement activation.

5.4 Materials and Methods

5.4.1 Formulation and solution of the complement model equations.

We used ordinary differential equations (ODEs) to model the time evolution of complement proteins (x_i) in the reduced order model:

$$\frac{1}{\tau_i}\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, ..., \mathcal{M} \tag{5.1}$$

where $\mathcal{R}$ denotes the number of reactions and $\mathcal{M}$ denotes the number of proteins in the model. The quantity τ_i denotes a time scale parameter for species i which captures unmodeled effects. For the current study, τ scaled with the level of initiator (z) for C5a and C5b; $\tau_i = z/z^*$ for i = C5a, C5b where z^* was 1mg/ml, $\tau_i = 1$ for all other species. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of biochemical and enzyme species abundance, as well as unknown model parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, species i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, species i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates species i is not connected with reaction j. Species balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$.

Rate processes were written as the product of a kinetic term ($\bar{r}_j$) and a control term (v_j) in the complement model. The kinetic term for the formation of C4a, C4b, C2a and C2b, lectin pathway activation, and C3 and C5 convertase activity was given by:

$$\bar{r}_j = k_j^{max} \epsilon_i \left(\frac{x_s^\eta}{K_{js}^\eta + x_s^\eta} \right) \tag{5.2}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the abundance of the enzyme catalyzing reaction j, η denotes a cooperativity parameter, and K_{js} denotes the saturation constant for species s in reaction j. We used mass action kinetics to model protein-protein binding interactions within the network:

$$\bar{r}_j = k_j^{max} \prod_{s \in m_j^-} x_s^{-\sigma_{sj}} \tag{5.3}$$

where k_j^{max} denotes the maximum rate for reaction j, σ_{sj} denotes the stoichiometric coefficient for species s in reaction j, and $s \in m_j$ denotes the set of *reactants* for reaction j. We assumed all binding interactions were irreversible.

The control terms $0 \leq v_j \leq 1$ depended upon the combination of factors which influenced rate process j. For each rate, we used a rule-based approach to select from competing control factors. If rate j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $v_j = \mathcal{I}_j\left(f_{1j}(\cdot), ..., f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on rate j. The function $\mathcal{I}_j(\cdot)$ is an integration rule which maps the output of regulatory transfer functions into a control variable. Each regulatory transfer function took the form:

$$f_{ij}\left(\mathcal{Z}_i, k_{ij}, \eta_{ij}\right) = k_{ij}^{\eta_{ij}} \mathcal{Z}_i^{\eta_{ij}} / \left(1 + k_{ij}^{\eta_{ij}} \mathcal{Z}_i^{\eta_{ij}}\right) \tag{5.4}$$

where $\mathcal{Z}_i$ denotes the abundance of factor i, k_{ij} denotes a gain parameter, and η_{ij} denotes a cooperativity parameter. In this study, we used $\mathcal{I}_j \in \{min, max\}$ [77]. If a process has no modifying factors, $v_j = 1$. The model equations were implemented in Julia and solved using the `CVODE` routine of the Sundials package [8, 28]. The model code and parameter ensemble is freely available under an MIT software license and can be downloaded from http://www.varnerlab.org.

5.4.2 Estimating an ensemble of complement model parameters.

We estimated a single initial parameter set using the Dynamic Optimization with Particle Swarms (DOPS) technique [76]. DOPS is a novel hybrid metaheuristic which combines a multi-swarm particle swarm method with the dynamically dimensioned search approach of Shoemaker and colleagues [99]. DOPS minimized the squared residual between simulated and C3a and C5a measurements with and without zymosan as a single objective. The best fit set estimated by DOPS served as the starting point for multiobjective ensemble generation using Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) [6]. JuPOETs is a multiobjective approach which integrates simulated annealing with Pareto optimality to estimate model ensembles on or near the optimal tradeoff surface between competing training objectives. JuPOETs minimized training objectives of the form:

$$O_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j}\left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k})\right)^2 + \left(\frac{\mathcal{M}'_{ij} - \max y_{ij}}{\mathcal{M}'_{ij}}\right)^2 \tag{5.5}$$

subject to the model equations, initial conditions and parameter bounds $\mathcal{L} \leq \mathbf{k} \leq \mathcal{U}$. The first term in the objective function measured the shape difference between the simulations and measurements. The symbol $\hat{\mathcal{M}}_{ij}$ denotes a scaled experimental observation (from training set j) while the symbol $\hat{y}_{ij}$ denotes the scaled simulation output (from training set j). The quantity i denotes the sampled time-index and $\mathcal{T}_j$ denotes the number of time points for experiment j. The scaled measurement is given by:

$$\hat{\mathcal{M}}_{ij} = \frac{\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}}{\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}} \tag{5.6}$$

Under this scaling, the lowest measured concentration become zero while the highest equaled one, where a similar scaling was defined for the simulation output. The second-term in the objective function quantified the absolute error in the estimated concentration scale, where the absolute measured concentration (denoted by $\mathcal{M}'_{ij}$) was compared with the largest simulated value. In this study, we minimized two training objectives, the total C3a and C5a residual w/o zymosan (O_1) and the total C3a and C5a residual for 1 mg/ml zymosan (O_2). JuPOETs identified an ensemble of N $\simeq$ 2100 parameter sets which were used for model simulations and uncertainty quantification subsequently. JuPOETs is open source, available under an MIT software license. The JuPOETs source code is freely available from the JuPOETs GitHub repository at https://github.com/varnerlab/POETs.jl. The objective functions used in this study are available in the GitHub model repository available from http://varnerlab.org.

5.4.3 Sensitivity and robustness analysis of complement model performance.

We conducted global sensitivity and robustness analysis to estimate which parameters and species controlled the performance of the reduced order model. We computed the total variance-based sensitivity index of each parameter relative to the training residual for the C3a/C5a alternate and C3a/C5a lectin objectives using the Sobol method [87]. The sampling bounds for each parameter were established from the minimum and maximum value for that parameter in the parameter ensemble. We used the sampling method of Saltelli *et al.* to com-

pute a family of $N(2d+2)$ parameter sets which obeyed our parameter ranges, where N was the number of trials per parameters, and d was the number of parameters in the model [78]. In our case, $N = 400$ and $d = 28$, so the total sensitivity indices were computed using 23,200 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language[27].

Robustness coefficients quantify the response of a marker to a structural or operational perturbation to the network architecture. Robustness coefficients were calculated as shown previously [98]. Log-transformed robustness coefficients denoted by $\hat{\alpha}\left(i, j, t_o, t_f\right)$ are defined as:

$$\hat{\alpha}\left(i, j, t_o, t_f\right) = \log_{10}\left[\left(\int_{t_o}^{t_f} x_i(t)\,dt\right)^{-1}\left(\int_{t_o}^{t_f} x_i^{(j)}(t)\,dt\right)\right] \tag{5.7}$$

Here t_o and t_f denote the initial and final simulation time, while i and j denote the indices for the marker and the perturbation, respectively. A value of $\hat{\alpha}\left(i, j, t_o, t_f\right) > 0$, indicates increased marker abundance, while $\hat{\alpha}\left(i, j, t_o, t_f\right) < 0$ indicates decreased marker abundance following perturbation j. If $\hat{\alpha}\left(i, j, t_o, t_f\right) \sim 0$, perturbation j did not influence the abundance of marker i. In this study, we perturbed the initial condition of C3 or C5 or a combination of C3 and C5 by 50%, 90% and 99% and measured the area under the curve (AUC) of C3a or C5a with and without lectin initiator. We computed the robustness coefficients for a subset of the parameter ensemble.

CHAPTER 6

CONCLUSION

This thesis seeks to add to the fields of biochemical engineering and systems biology in an attempt to accurately model the fermentation of *B. subtilis* and *E. coli*. The *B. subtilis* fermentation models developed include batch and fed-batch Monod kinetics and an HCM-EM model of glycolysis, the pentose phosphate pathway, and the TCA cycle. HCM-FBA was developed to model *E. coli* fermentation and proposed as an alternative to HCM-EM due to its ability to reduce model size and computational overhead. JuPOETs was developed to estimate model parameters and ensembles and applied to a variety of test problems. Finally, a reduced order model of the complement system was developed and capable of describing experimental data and predicting targets for therapeutic intervention.

The Monod kinetic models were trained with experimental data and capable of describing the fed-batch fermentation of the RB50::pRF69 strain and batch fermentation of RB50::pRF69 and the *cyd* and *qox* knockout mutants. The Monod models consisted of 5 ODEs and 7 kinetic parameters determined by minimizing the difference in model and experimental results. The fed-batch model was then modified to maximize the increase in riboflavin production by varying the feeding profile. An exponential feeding profile was determined to increase the riboflavin yield 3-fold. However, as these results are purely mathematical, they must be validated experimentally.

The HCM-EM model of RB50::pRF69 incorporated cybernetic control variables that dynamically directed flux through the modeled metabolism. The model consisted of 174 ODEs and 386 kinetic parameters and was capable of de-

scribing biomass and acetate formation when trained with experimental data. The model overestimated the production of riboflavin, however, and it was proposed that the Julia programming language package JuPOETs could be a potential method to retrain the model to describe the experimental data completely. If the model can be improved it could potentially be used to predict intracellular flux data found in literature. If validated via flux prediction, the model could be used to rapidly run a large number of simulated experiments with the organism to identify metabolic bottlenecks and knockout targets for strain improvement.

HCM-FBA was shown to have comparable performance to HCM-EM for a proof of concept metabolic network and a reduced network of anaerobic *E. coli*. HCM-FB A was then applied to a larger aerobic *E. coli* network that was computationally infeasible by HCM-EM due to the large number of elementary modes. HCM-EM generated 153,000 elementary modes while HCM-FBA generated 29 modes. Sobol sensitivity analysis was used to further reduce the number of FBA modes from 29 to 5 while maintaining model fit. HCM-FBA was also able to capture the experimentally observed switch from glucose to acetate consumption that HCM-EM could not. HCM-FBA is a promising alternative to HCM-EM for large networks where the generation of elementary modes is infeasible, but still requires further validation. HCM-FBA could potentially be applied to a wide variety of biological systems and organisms for the production of proteins and other valuable products in more complex cells in which networks are too large to be considered by other methods.

JuPOETs is a promising approach for the estimation of parameter and model ensembles using multiobjective optimization. JuPOETs integrates simulated annealing with Pareto optimality to estimate ensembles on or near the opti-

mal tradeoff surface between competing training objectives. JuPOETs can be adapted to solve many problem types, including mixed binary and continuous variable types, bilevel optimization problems and constrained problems without altering the base algorithm. JuPOETs is open source available under an MIT license, and can be installed using the Julia package manager from the JuPOETs GitHub repository.

Finally, a reduced order model of the complement system that was computationally inexpensive and could easily be incorporated into pharmacokinetic models of immune system function was developed. The model described experimental data, and predicted the need for multiple points of therapeutic intervention to fully disrupt complement activation. ODEs, saturation and Michaelis-Menten type rate laws, and, logical rules were combined to produce a computationally inexpensive model without sacrificing performance and predictive capability. The model described the lectin and alternative pathways and was an order of magnitude smaller than comparable models in the literature. An ensemble of model parameters was estimated from *in vitro* dynamic measurements of the C3a and C5a complement proteins. The model was capable of predicting unseen C3a and C5a experimental data trends. Global sensitivity and robustness analysis suggested complement was robust to any single therapeutic intervention, however, the knockdown of both C3 and C5 consistently reduced C3a and C5a formation from all pathways.

BIBLIOGRAPHY

[1] GNU Linear Programming Kit, Version 4.52, March 2016.

[2] Matthew L Alexander and D Ramkrishna. Cybernetic modeling of iron-limited growth and siderophore production. *Biotechnology and bioengineering*, 38(6):637–652, 1991.

[3] Stefano Andreozzi, Ljubisa Miskovic, and Vassily Hatzimanikatis. ischrunk–in silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab Eng*, 33:158–68, Jan 2016.

[4] A Bacher and B Mailänder. Biosynthesis of riboflavin in bacillus subtilis: function and genetic control of the riboflavin synthase complex. *Journal of bacteriology*, 134(2):476–482, 1978.

[5] Rekha Bansal. Humanized and chimeric anti-properdin antibodies. US Patent 8,664,362, March 2014.

[6] David Bassen, Michael Vilkhovoy, Mason Minot, Jonathan T Butcher, and Jeffrey D Varner. JuPOETs: A Constrained Multiobjective Optimization Approach to Estimate Biochemical Model Ensembles in the Julia Programming Language. *bioRxiv*, 10.1101/056044, 2016.

[7] D Battogtokh, D.K Asch, M.E Case, J Arnold, and H.B Shüttler. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa. *Proc Natl Acad Sci U S A*, 99(26):16904–16909, December 2002.

[8] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. November 2014.

[9] Anna M Blom, Lena Kask, and Björn Dahlbäck. Structural requirements for the complement regulatory activities of c4bp. *J Biol Chem.*, 276(29):27136–27144, 2001.

[10] A.J Booker, J.E Dennis, P.D Frank, D.B Serafini, V Torczon, and M.W Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Struct Optim*, 17:1 – 13, 1999.

[11] AK Chauhan and TL Moore. Presence of plasma complement regulatory proteins clusterin (apo j) and vitronectin (s40) on circulating immune complexes (cic). *Clin Exp Immunol.*, 145(3):398–406, 2006.

[12] Carolina A. Contador, Matthew L. Rizk, Juan A. Asenjo, and James C. Liao. Ensemble modeling for strain development of l-lysine-producing escherichia coli. *Metabolic Engineering*, 11(4–5):221 – 233, 2009.

[13] Markus W Covert, Eric M Knight, Jennifer L Reed, Markus J Herrgard, and Bernhard O Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–6, May 2004.

[14] Michael Dauner, James E Bailey, and Uwe Sauer. Metabolic flux analysis with a comprehensive isotopomer model in bacillus subtilis. *Biotechnology and bioengineering*, 76(2):144–156, 2001.

[15] Michael Dauner and Uwe Sauer. Stoichiometric growth model for riboflavin-producing bacillus subtilis. *Biotechnology and bioengineering*, 76(2):132–143, 2001.

[16] M De Tremblay, M Perrier, C Chavarie, and J Archambault. Optimization of fed-batch culture of hybridoma cells using dynamic programming: single and multi feed cases. *Bioprocess Engineering*, 7(5):229–234, 1992.

[17] Miriam Dormeyer, Richard Egelkamp, Martin J Thiele, Elke Hammer, Katrin Gunka, Lorena Stannek, Uwe Völker, and Fabian M Commichau. A novel engineering tool in the bacillus subtilis toolbox: inducer-free activation of gene expression by selection-driven promoter decryptification. *Microbiology*, 161(2):354–361, 2015.

[18] David Dubnau. *The molecular biology of the bacilli*. Elsevier, 2012.

[19] John W. Eaton, David Bateman, and Soren Hauberg. *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*. CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 2009.

[20] Hitoshi Enei, Katsuaki Sato, Yasuo Anzai, and Hiroshi Okada. Fermentative production of riboflavine, August 19 1975. US Patent 3,900,368.

[21] David Epstein and Jeffrey C Kurz. Complement binding aptamers and

anti-c5 agents useful in the treatment of ocular disorders. US Patent App. 12/224,708, March 2007.

[22] C.M. Fonseca and P. J. Fleming. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 416 – 423, 1993.

[23] K. G. Gadkar, J Varner, and F J Doyle. Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol (Stevenage)*, 2(1):17–30, Mar 2005.

[24] Kapil G Gadkar, Francis J Doyle, 3rd, Timothy J Crowley, and Jeffrey D Varner. Cybernetic model predictive control of a continuous bioreactor with cell recycle. *Biotechnol Prog*, 19(5):1487–97, 2003.

[25] Peter Gennemark and Dag Wedelin. Benchmarks for identification of ordinary differential equations from time series data. *Bioinformatics*, 25(6):780–6, Mar 2009.

[26] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):1871–1878, Oct 2007.

[27] JD Herman. Salib. available online: https://github.com/jdherman/salib.

[28] Alan C. Hindmarsh, Peter N. Brown, Keith E. Grant, Steven L. Lee, Radu Serban, Dan E. Shumaker, and Carol S. Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.*, 31(3):363–396, September 2005.

[29] Hirohumi Hirayama, Kiyono Yoshii, Hidetomo Ojima, Norikazu Kawai, Shintaro Gotoh, and Yuzo Fukuyama. Linear systems analysis of activating processes of complement system as a defense mechanism. *Biosystems*, 39(3):173–185, 1996.

[30] Xianzhen Hu, V Michael Holers, Joshua M Thurman, Trent R Schoeb, Theresa N Ramos, and Scott R Barnum. Therapeutic inhibition of the alternative complement pathway attenuates chronic EAE. *Mol Immunol.*, 54:302–308, 2013.

[31] S Huband, P Hingston, L Barone, and L While. A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit. *IEEE Trans. Evol. Comp.*, 10:477 – 506, 2006.

[32] J P Sethna K S Brown. Statistical mechanial approaches to models with many poorly known parameters. *Phys Rev E*, 68:021904:1–9, 2003.

[33] D Kalyanmoy, A Pratap, S Agarwal, and T. Meyarivan. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comp.*, 6:182 – 197, 2002.

[34] A Kamp and S Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931, 2006.

[35] Kenneth J Katschke, Ping Wu, Rajkumar Ganesan, Robert F Kelley, Mary A Mathieu, Philip E Hass, Jeremy Murray, Daniel Kirchhofer, Christian Wiesmann, and Menno van Lookeren Campagne. Inhibiting alternative pathway complement activation by targeting the factor d exosite. *J Biol Chem.*, 287:12886–12892, 2012.

[36] Ali Khodayari, Ali R Zomorrodi, James C Liao, and Costas D Maranas. A kinetic model of escherichia coli core metabolism satisfying multiple sets of mutant flux data. *Metab Eng*, 25:50–62, Sep 2014.

[37] JI Kim, JD Varner, and D Ramkrishna. A hybrid model of anaerobic e. coli gjt001: Combination of elementary flux modes and cybernetic variables. *Biotechnol. Prog.*, 24(5):993–1006, 2008.

[38] Jin Il Kim, Hyun-Seob Song, Sunil R Sunkara, Arvind Lali, and Doraiswami Ramkrishna. Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. *Biotechnol Prog*, 28(5):1160–6, 2012.

[39] Dhinakar S Kompala, Doraiswami Ramkrishna, Norman B Jansen, and George T Tsao. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnology and Bioengineering*, 28(7):1044–1055, 1986.

[40] Andrey A Korotaevskiy, Leonid G Hanin, and Mikhail A Khanin. Nonlinear dynamics of the complement system activation. *Math Biosci.*, 222(2):127–143, 2009.

[41] U. Saucer J. Stelling L. Kuepfer, M. Peter. Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotech*, 25(9):1001–1006, 2007.

[42] L Lee, JD Varner, and K Ko. Parallel extreme pathway computation for metabolic networks. *Comput Syst Bioinformatics Conf, Int IEEE CS*, 0:636–639, 2004.

[43] Yun Lee, Jimmy G Lafontaine Rivera, and James C Liao. Ensemble modeling for robustness analysis in engineering non-native metabolic pathways. *Metab Eng*, 25:63–71, Sep 2014.

[44] Joshua Lequieu, Anirikh Chakrabarti, Satyaprakash Nayak, and Jeffrey D Varner. Computational modeling and analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol*, 7(11):e1002263, Nov 2011.

[45] M Kathryn Liszewski, Timothy C Farries, Douglas M Lublin, Isabelle A Rooney, and John P Atkinson. Control of the complement system. *Adv Immunol.*, 61:201–283, 1995.

[46] Bing Liu, Jing Zhang, Pei Yi Tan, David Hsu, Anna M Blom, Benjamin Leong, Sunil Sethi, Bow Ho, Jeak Ling Ding, and PS Thiagarajan. A computational and experimental study of the regulatory mechanisms of the complement system. *PLoS Comput Biol*, 7(1):e1001059, 2011.

[47] Deyan Luan, Fania Szlam, Kenichi A Tanaka, Philip S Barie, and Jeffrey D Varner. Ensembles of uncertain mathematical models can identify network response to therapeutic interventions. *Mol Biosyst*, 6(11):2272–86, Nov 2010.

[48] Deyan Luan, Michael Zai, and Jeffrey D Varner. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol*, 3(7):e142, Jul 2007.

[49] P Lukacik, P Roversi, J White, D Esser, GP Smith, J Billington, PA Williams, PM Rudd, MR Wormald, DJ Harvey, et al. Complement regulation at the molecular level: the structure of decay-accelerating factor. *Proc Natl Acad Sci USA.*, 101(5):1279–1284, 2004.

[50] Dimitrios C Mastellos, Despina Yancopoulou, Petros Kokkinos, Markus Huber-Lang, George Hajishengallis, Ali R Biglarnia, Florea Lupu, Bo Nilsson, Antonio M Risitano, Daniel Ricklin, et al. Compstatin: a c3-targeted complement inhibitor reaching its prime for bedside intervention. *European journal of clinical investigation*, 45(4):423–440, 2015.

[51] Joost PM Melis, Kristin Strumane, Sigrid R Ruuls, Frank J Beurskens, Janine Schuurman, and Paul WHI Parren. Complement in therapy and disease: Regulating the complement system with antibody-based therapeutics. *Mol Immunol.*, 67:117–130, 2015.

[52] Jacques Monod. The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394, 1949.

[53] Hassan OJ Morad, Samuel C Belete, Thomas Read, and Andrew M Shaw. Time-course analysis of c3a and c5a quantifies the coupling between the upper and terminal complement pathways in vitro. *J Immunol Methods*, 427:13–18, 2015.

[54] B Paul Morgan and Claire L Harris. Complement, a target for therapy in inflammatory and degenerative diseases. *Nat Rev Drug Discov*, 14:857–877, 2015.

[55] Atul Narang, Allan Konopka, and D Ramkrishna. Escherichia coli k12. *Biotechnology and bioengineering*, 55(5):747–757, 1997.

[56] S Nayak, J K Siddiqui, and J D Varner. Modelling and analysis of an ensemble of eukaryotic translation initiation models. *IET Syst Biol*, 5(1):2, Jan 2011.

[57] Satyaprakash Nayak, Saniya Salim, Deyan Luan, Michael Zai, and Jeffrey D Varner. A test of highly optimized tolerance reveals fragile cell-cycle mechanisms are molecular targets in clinical cancer trials. *PLoS One*, 3(4):e2016, 2008.

[58] Marina Noris, Miriam Galbusera, Sara Gastoldi, Paolo Macor, Federica Banterla, Elena Bresin, Claudio Tripodo, Serena Bettoni, Roberta Donadelli, Elisabetta Valoti, et al. Dynamics of complement activation in ahus and how to monitor eculizumab therapy. *Blood*, pages 1715–1726, 2014.

[59] G Nuttall. Experimente über die bacterienfeindlichen Einflüsse des thierischen Körpers. *Z. Hyg. Infektionskr.*, 4:353–394, 1888.

[60] JD Orth, I Thiele, and BØ Palsson. What is flux balance analysis? *Nat. Biotechnol.*, 28(3):245–248, 2010.

[61] L Paciello, C Landi, E De Alteriisb, and P Parascandola. Mathematical modeling as a tool to describe and optimize heterologous protein produc-

tion by yeast cells in aerated fed-batch reactor. *Chem Eng Trans*, 27:79–84, 2012.

[62] T.N Palmer, G.J Shutts, R Hagedorn, F.J Doblas-Reyes, T Jung, and M Leutbecher. Representing model uncertainty in weather and climate prediction. *Ann Rev Earth and Planetary Sci*, 33:163–193, 2005.

[63] BØ Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, New York, NY, USA, 2006.

[64] Michael K Pangburn and Hans J Müller-Eberhard. The alternative pathway of complement. *Springer Semin Immunopathol*, 7:163–192, 1984.

[65] JB Perkins, A Sloma, T Hermann, K Theriault, E Zachgo, T Erdenberger, N Hannett, NP Chatterjee, V Williams II, GA Rufo Jr, et al. Genetic engineering of bacillus subtilis for the commercial production of riboflavin. *Journal of Industrial Microbiology and Biotechnology*, 22(1):8–18, 1999.

[66] OT Ramírez, R Zamora, R Quintero, and A López-Munguía. Exponentially fed-batch cultures as an alternative to chemostats: the case of penicillin acylase production by recombinant e. coli. *Enzyme and microbial technology*, 16(10):895–903, 1994.

[67] Doraiswami Ramkrishna and Hyun-Seob Song. Dynamic models of metabolism: Review of the cybernetic approach. *AIChE Journal*, 58(4):986–997, 2012.

[68] Nicklaus T Rice, Fania Szlam, Jeffrey D Varner, Peter S Bernstein, Arthur D Szlam, and Kenichi A Tanaka. Differential contributions of intrinsic and extrinsic pathways to thrombin generation in adult, maternal and cord plasma samples. *PLoS One*, 11(5):e0154127, 2016.

[69] Daniel Ricklin, George Hajishengallis, Kun Yang, and John D Lambris. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol*, 11(9):785–797, 2010.

[70] Daniel Ricklin and John D Lambris. Complement-targeted therapeutics. *Nat Biotechnol*, 25:1265–1275, 2007.

[71] Daniel Ricklin and John D Lambris. Complement in immune and inflammatory disorders: pathophysiological mechanisms. *J Immunol*, 190(8):3831–3838, 2013.

[72] Rebecca C Riley-Vargas, Darcy B Gill, Claudia Kemper, M Kathryn Liszewski, and John P Atkinson. Cd46: expanding beyond complement regulation. *Trends Immunol*, 25(9):496–503, 2004.

[73] Daniel Rittirsch, Michael A Flierl, and Peter A Ward. Harmful molecular mechanisms in sepsis. *Nat Rev Immunol*, 8(10):776–787, 2008.

[74] Michael Roguska, Igor Splawski, Beate Diefenbach-Streiber, Elizabeth Dolan, Bijan Etemad-Gilbertson, Jean-Michel Rondeau, and Mark Keating. Generation and Characterization of LFG316, A Fully-Human Anti-C5 Antibody for the Treatment of Age-Related Macular Degeneration. *IOVS*, 55:3433–3433, 2014.

[75] Martin Rühl, Nicola Zamboni, and Uwe Sauer. Dynamic flux responses in riboflavin overproducing bacillus subtilis to increasing glucose limitation in fed-batch culture. *Biotechnology and bioengineering*, 105(4):795–804, 2010.

[76] Adithya Sagar, Christine A. Shoemaker, and J. Varner. Dynamic Optimization with Particle Swarms (DOPS): A meta- heuristic for parameter estimation in biochemical models. *Biotechnol. J*, submitted., 2016.

[77] Adithya Sagar and Jeffrey D. Varner. Dynamic modeling of the human coagulation cascade using reduced order effective kinetic models. *Processes*, 3(1):178, 2015.

[78] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Comput Phys Commun*, 181(2):259–270, 2010.

[79] J Vidya Sarma and Peter A Ward. The complement system. *Cell Tissue Res*, 343(1):227–235, 2011.

[80] Marcus Schallmey, Ajay Singh, and Owen P Ward. Developments in the use of bacillus species for industrial production. *Canadian journal of microbiology*, 50(1):1–17, 2004.

[81] C H Schilling, D Letscher, and B O Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3):229–48, Apr 2000.

[82] R Schuetz, L Kuepfer, and U Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Mol. Syst. Biol.*, 3(1), 2007.

[83] S Schuster, D A Fell, and T Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–32, Mar 2000.

[84] Hans-Wilhelm Schwaeble, Cordula Margaret Stover, Clark E Tedford, James B Parent, and Teizo Fujita. Methods for treating conditions associated with masp-2 dependent complement activation. US Patent 7,919,094, April 2011.

[85] Michael L Shuler and Fikret Kargi. *Bioprocess engineering*. Prentice Hall New York, 2002.

[86] JG Sissons, J Liebowitch, N Amos, and DK Peters. Metabolism of the fifth component of complement, and its relation to metabolism of the third component, in patients with complement activation. *J Clin Invest.*, 59(4):704, 1977.

[87] I.M Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simulat*, 55:271 – 280, 2001.

[88] Hyun-Seob Song and Doraiswami Ramkrishna. Prediction of metabolic function from limited data: Lumped hybrid cybernetic modeling (l-hcm). *Biotechnology and Bioengineering*, 106(2):271–284, 2010.

[89] Hyun-Seob Song and Doraiswami Ramkrishna. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol Bioeng*, 108(1):127–40, Jan 2011.

[90] Hyun-Seob Song and Doraiswami Ramkrishna. Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metab Eng*, 14(2):69–80, Mar 2012.

[91] S Song and J Varner. Modeling and Analysis of the Molecular Basis of Pain in Sensory Neurons. *PLoS ONE*, 4:e6758 – e6772, 2009.

[92] Sang Ok Song, Anirikh Chakrabarti, and Jeffrey D Varner. Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). *Biotechnol J*, 5(7):768–80, Jul 2010.

[93] Alexei Sorokin, Emmanuelle Zumstein, Vasco Azevedo, S Dusko Ehrlich, and Pascale Serror. The organization of the bacillus subtilis 168 chromosome region between the spova and sera genetic loci, based on sequence data. *Molecular microbiology*, 10(2):385–395, 1993.

[94] AJG Swaak, A Hannema, C Vogelaar, FA Boom, L van Es, R van Aalst, and LW Statius van Eps. Determination of the half-life of c3 in patients and its relation to the presence of c3-breakdown products and/or circulating immune complexes. *Rheumatol Int.*, pages 161–166, 1982.

[95] Yikun Tan and James C Liao. Metabolic ensemble modeling for strain engineers. *Biotechnol J*, 7(3):343–53, Mar 2012.

[96] Yikun Tan, Jimmy G Lafontaine Rivera, Carolina A Contador, Juan A Asenjo, and James C Liao. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab Eng*, 13(1):60–75, Jan 2011.

[97] Ryan Tasseff, Satyaprakash Nayak, Saniya Salim, Poorvi Kaushik, Noreen Rizvi, and Jeffrey D Varner. Analysis of the molecular networks in androgen dependent and independent prostate cancer revealed fragile and robust subsystems. *PLoS One*, 5(1):e8864, 2010.

[98] Ryan Tasseff, Satyaprakash Nayak, Sang Ok Song, Andrew Yen, and Jeffrey D Varner. Modeling and analysis of retinoic acid induced differentiation of uncommitted precursor cells. *Integr Biol (Camb)*, 3(5):578–91, May 2011.

[99] Bryan A. Tolson and Christine A. Shoemaker. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Res Research*, 43(1):W01413, 2007. W01413.

[100] Linh M Tran, Matthew L Rizk, and James C Liao. Ensemble modeling of metabolic networks. *Biophys J*, 95(12):5606–17, Dec 2008.

[101] Brian G Turner and Doraiswami Ramkrishna. Revised enzyme synthesis rate expression in cybernetic models of bacterial growth. *Biotechnology and bioengineering*, 31(1):41–43, 1988.

[102] A Varma and BØ Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Appl. Environ. Microbiol.*, 60(10):3724–3731, 1994.

[103] J Varner and D Ramkrishna. Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. *Metab Eng*, 1(1):88–116, Jan 1999.

[104] Carl-Wilhelm Vogel, David C Fritzinger, Brian E Hew, Mike Thorne, and Holger Bammert. Recombinant cobra venom factor. *Molecular immunology*, 41:191–199, 2004.

[105] Claes von Wachenfeldt and Lars Hederstedt. Respiratory cytochromes, other heme proteins, and heme biosynthesis. *Bacillus subtilis and its Closest Relatives: from Genes to Cells*, pages 163–179, 2002.

[106] DG Walker, O Yasuhara, PA Patston, EG McGeer, and PL McGeer. Complement c1 inhibitor is produced by brain tissue and is cleaved in alzheimer disease. *Brain Res.*, 675(1):75–82, 1995.

[107] M J Walport. Complement. first of two parts. *N Engl J Med*, 344(14):1058–66, Apr 2001.

[108] M J Walport. Complement. second of two parts. *N Engl J Med*, 344(15):1140–4, Apr 2001.

[109] Wynne H Weston-Davies, Miles A Nunn, Fernando O Pinto, Ian J Mackie, Stephen John Richards, Samuel J Machin, Raymond Prudo, and Peter Hillmen. Clinical and immunological characterisation of coversin, a novel small protein inhibitor of complement C5 with potential as a therapeutic agent in PNH and other complement mediated disorders. *Blood*, 124:4280–4280, 2014.

[110] Sharon J Wiback, Radhakrishnan Mahadevan, and Bernhard Ø Palsson. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. *J Theor Biol*, 224(3):313–24, Oct 2003.

[111] Jamey Dale Young. A system-level mathematical description of metabolic regulation combining aspects of elementary mode analysis with cybernetic control laws. 2005.

[112] Nicola Zamboni. *Metabolic engineering of respiration for improved riboflavin production and elucidation of NADPH metabolism in Bacillus subtilis*. PhD thesis, Diss., Naturwissenschaften ETH Zürich, Nr. 15118, 2003, 2003.

[113] Nicola Zamboni, Nigel Mouncey, Hans-Peter Hohmann, and Uwe Sauer.

Reducing maintenance metabolism by metabolic engineering of respiration improves riboflavin production by bacillus subtilis. *Metabolic engineering*, 5(1):49–55, 2003.

[114] Nicola Zamboni and Uwe Sauer. Knockout of the high-coupling cytochrome aa3 oxidase reduces tca cycle fluxes in bacillus subtilis. *FEMS microbiology letters*, 226(1):121–126, 2003.

[115] Nehemiah Zewde, Ronald D Gorham Jr, Angel Dorado, and Dimitrios Morikis. Quantitative modeling of the alternative pathway of the complement system. *PloS One*, 11(3):e0152337, 2016.