A computational framework for predicting CHO cell culture performance in bioreactors

by

Martín Cárcamo Behrens

M.Eng., Bioprocess Engineering, Pontificia Universidad Católica at Chile, 2013

Submitted to the MIT Department of Biological Engineering and the Sloan School of Management in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Biological Engineering
and
Master of Business Administration
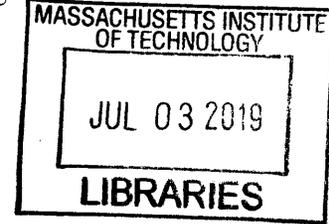
In conjunction with the Leaders for Global Operations Program

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Signature of Author ....
Department of Biological Engineering
MIT Sloan School of Management
May 10th, 2019

Signature redacted

Certified by ....
Douglas Lauffenburger,
Ford Professor of Biological Engineering, Chemical Engineering, and Biology
Head, Department of Biological Engineering | Thesis Supervisor

Signature redacted

Certified by ....
Roy Welsch
Eastman Kodak Leaders for Global Operations | Professor of Statistics and Management Science,
MIT Sloan School of Management | Thesis Supervisor

Signature redacted

Accepted by...
Forest White
Professor of Biological Engineering
Chair of Graduate Program, Department of Biological Engineering

Signature redacted

Accepted by...
Maura Herson, Assistant Dean,
MBA Program MIT Sloan School of Management

This page has been intentionally left blank.

A computational framework for predicting CHO cell culture performance in bioreactors

by

Martín Cárcamo Behrens

Submitted to the MIT Department of Biological Engineering and the MIT Sloan School of Management on May 10, 2019 in Partial Fulfillment of the Requirements for the Degrees of Master of Science in Biological Engineering and Master of Business Administration.

Abstract

Breaking the trade-off between speed and productivity is a key milestone across industries. In particular, in the biopharmaceutical industry this trade-off is exacerbated by a highly regulated environment, which hinders continuous improvement and fixes future manufacturing costs.

Given the complexity of living organisms and the improvement in quality of life offered by the product – which demand agile development – the industry has traditionally taken phenomenological approaches to process development, generally sacrificing costs. Nonetheless, technological developments and lower entry barriers make the biopharmaceutical industry far more competitive than in its origins, demanding efficient and reliable processes. Developing efficient manufacturing processes for new products while being agile to market is a key differentiating capability of Amgen's process development organization.

In collaboration with the process development team at Amgen, a computational framework for *in-silico* upstream bioprocess development has been developed, allowing for faster, more robust, and more optimal process development. Specifically, a mechanistic model of a bioreactor has been designed, implemented, and applied to an Amgen product. The project was divided into three major components: The first was a survey of internal Amgen capabilities and the state of the art in external industrial and academic models to identify the algorithms and design the signal flow required to support the range of expected process engineering applications. The second consisted of implementing a modular, extensible software platform with the architecture and interfaces dictated by the first component. The third part consisted of applying the software to an actual product development problem capturing the primary process variables.

A constraint-based model of a metabolic network consisting of 35 reactions of the main carbon-nitrogen metabolism relevant in energy and redox balance was adapted from literature (Nolan & Lee, 2011). The metabolic network was coupled with glucose, glutamine and asparagine kinetics with temperature, dissolved oxygen, pH and osmolarity dependence. Stress induced by temperature shifts was modeled as a first-order step response coupled to a non-growth associated ATP of maintenance. The cellular model was coupled with a well-mixed bioreactor model consisting of mass balance equations. We solved the model using dynamic Flux Balance Analysis (dFBA). We first calibrated the model with experimental process characterization data for a product in development. We used a Non-dominated Sorted Genetic Algorithm (NSGA-II) to solve the calibration problem, minimizing the error in metabolite concentrations to yield estimates of 13 strain-specific parameters. We then assessed the calibrated model's predictions of biomass growth and metabolite concentrations against a second experiment run with different process settings. Finally, I developed a graphical user interface for subject-matter-experts to simulate experiments and test hypotheses using the model. We applied the tool to three process-relevant case studies, and analyzed the in-silico results. The calibrated model can predict biomass and titer from process settings, potentially reducing experimental time from 20 days to 30 seconds, in addition to reducing the experimental cost.

This page has been intentionally left blank

Acknowledgment

This page has been intentionally left blank

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1. Definitions and Market

Biologics are defined as pharmaceutical therapeutics that are biologically manufactured, most commonly by recombinant DNA technology. Biologics are a subdivision of biotechnology known as red biotechnology that offers new mechanisms of generating novel pharmaceuticals compared to synthetics or botanicals. Biologics range in size from ~6 kDa (small peptides such as insulin) to up to ~350 kDa (antibodies, fragments, vaccines, or even recombinant blood factors). Biologics are most commonly complex three-dimensional molecules which are manufactured in the cells with post-translational modifications.

Since the approval of human biotech insulin by Genentech in 1982, the biotechnology industry has exploded. The biologics market was worth $221 billion in 2017, had 29 new approvals in 2017 and over 1000 biologics under development with growth projections of over 20% in the upcoming years (*Global Biologics Market—Companies-to-Action*, 2017; *Global Biologics Market, Opportunities And Strategies*, 2018). Biologic products include treatment for numerous diseases, such as diabetes, arthritis, cancer, anemia, and HIV, amongst others (Butler, 2005).

Manufacturing is often times divided in upstream process, where a recombinant host cell is cultured and the desired product is manufactured, and a downstream process, consisting of series of purification and formulation steps that generate the final drug product. Manufacturing scale varies according to the product ranging from multi ton scale (*e.g.* Insulin) to 100 kg. (*e.g.* monoclonal antibodies for cancer treatment) (Strube et al., 2018). Host cell lines vary depending on the complexity of the product and the metabolic machinery required to bio-manufacture the product. Approved cell lines range from bacteria such as *E. coli* or mammalian cells such as CHO (Chinese hamster ovary) or BHK (baby hamster kidney). Other types of cell factories have also been used such as yeasts, insects and plant cells.

1.2. Current Challenges

The biopharmaceutical industry faces several challenges: while it has seen explosive growth with ongoing technological development it has also been faced with decreasing internal rate of returns, growing complexity of products targeting progressively more complex diseases and increasing pressure to reduce cost.

A recent study from Deloitte showed that in 2018 biopharmaceuticals hit the highest level by spending over $2 billion for product development while having an historical minimum of 1.9% internal rate of returns on R&D (Neil & Sonal, 2018). The growing capital investment combined with the lower success rate translates in higher financial risk.

Additionally, technological breakthroughs in the field have been multiplying, and so have the complexity of the products. In fact, innovative biologics, such as cell therapeutics, virus-like particles, exosomes, recombinant proteins, and peptides seem likely to be the main therapeutics in the coming years. Consequently, product complexity is predicted to increase in the short- to medium-term, resulting in low success approval rates and manufacturing and operational challenges in scaling the new technologies.

Technological developments and cost reduction on diagnosis tools have enabled population clustering allowing for targeted therapies, with higher success rates. Additionally, regulatory incentives such as the orphan drug disease act (21 C.F.R. 316, 2004) added to lower competition in smaller but better characterized patient populations have pushed biopharmaceuticals towards market stratification and even personalization (such as CART therapy).

Nonetheless, from a healthcare perspective, lower competition has generated increasingly expensive drugs that certainly will not be sustainable. For example, for rheumatoid arthritis, a biologic therapy can cost from $1,000 to $3,000 a month (Metzger, 2012). Therefore, a significant pressure is imposed to drug manufacturers to lower cost of therapeutics.

A good opportunity for cost reduction will come from biosimilars. With several blockbuster biologics soon facing patent expiries, several biopharmaceutical companies have looked at this as a business opportunity (Rader & Langer, 2014). The global biosimilars market is expected to show a 31.5% growth in the next 7-8 years to reach $66.3 billion in 2025 (*Global Biologics Market—Companies-to-Action*, 2017).

1.3. How to Solve these Problems

These challenges can be overcome, from a simplistic perspective in two dimensions: by reducing risk of product development and/or by improving the process efficiencies.

Reducing product development risk can be overcome both from a scientific as well as a financial perspective. From a scientific perspective, the main goal is to increase the success rate of a new therapeutic, for example by having orthogonal analytics to cluster and target successfully a drug to a population target (Neil & Sonal, 2018), and better designing clinical trials (Berry, 2006; Montazerhodjat & Lo, 2015). From a financial perspective reducing the financial risk of investment by diversifying the portfolio, by financial engineering techniques such as securitization, can help incentivize investors and accelerate developments (Fernandez, Stein, & Lo, 2012).

The second dimension can come from having a better understanding of the process itself. This leads to faster product development, scaling and overall reduction of cost by more efficient processes. Nonetheless, because of the strict regulatory environment in biopharmaceutical, continuous improvement is hindered, and this evolution is suboptimal. The focus of this project will be on this dimension.

1.4. Challenges in Biopharmaceutical Process Development

The complexity of the product requires sophisticated and sensitive manufacturing processes, that must assure identical product quality across the culture and downstream processing. This means that due to regulation, the product quality and the manufacturing process are fixed during the supply of the first clinical

trials. Process changes generating differences in more than 1% of side product profiles might require new clinical trials (Strube et al., 2018), making any continuous improvement effort futile.
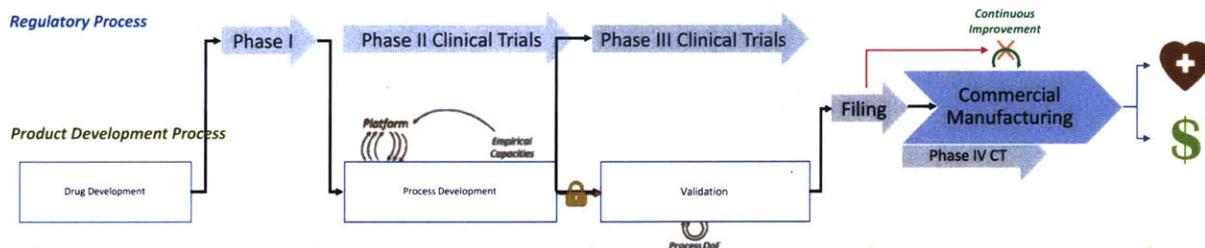
Figure 1. 1. Current Biopharmaceutical Product Development. Product Development tightly interconnected with a regulatory process that after filing and approval hinders continuous improvement thus affecting cost.

Consequently, any process improvement must be defined early in process development and before first amount for clinical trials are generated. This translates in a tradeoff between process efficiency – and potentially all future manufacturing costs – and speed-to-market of the therapeutic, which is crucial both ethically and strategically. Thus process improvement efforts need to be ready according to the regulatory timelines, greatly increasing the cost (Dhanasekharan et al., 2016; Linz, 2012)). For this reason, the biopharmaceutical industry has been reticent to industrialize process intensification, unless proven elsewhere, and have adopted a platform-based approach to process development.

A good example of this tradeoff between process development improvements comes from large biopharmaceuticals. Historically, in the early 80's, bioprocess technology was very limited – mostly inspired from the learnings from the wine and beer industry. Early recombinant cell cultures for antibody productions had low titers compared to current standards (~0.2 vs ~3 grams per liter) (Rader & Langer, 2014). Later, in the late 90's, high growth projections combined with low titers triggered capacity shortages and consequently drove many companies and contract manufacturing organizations (CMOs) to invest heavily in building large production plants – known as 'multi-packs' (plants with multiple bioreactors with 10,000 L or larger). At the same time, process improvements both upstream (e.g. protein expression systems, culture media, control strategies) and downstream (e.g. disposable, flow-through process) showed

impressive productivity gains. Both developments combined led to large biopharmaceutical companies facing excess production capacity often sitting idle. On the contrary, smaller companies were greatly benefited by being able to achieve similar throughputs at low scale with disposable systems, thus greatly reducing capital investment. This lowered the barrier for new entrants in the market, making it much easier to meet production demand for clinical trials or even commercial launch, without depending on large CMOs or pharma partners (Chon & Zarbis-Papastoitsis, 2011).

1.5. Process Intensification in Biotechnology

Biotechnology is in several aspects very similar to the chemical industry, and as such, it can use and learn from the best-practices from chemical processes. Particularly, process intensification has been particularly promising. The concept is defined as "a strategy for making dramatic reductions in the size of a plant so as to reach a given production objective" (Stankiewicz & Moulijn, 2000). To achieve this purpose several tools have been used, such as process modelling and comprehension, miniaturization/ scaled-down equipment for rapid process development, advanced process control, innovative analytics, mass transfer enhancement and process integration (Strube et al., 2018).

Examples of process intensification have been explored in biotechnology, but mostly for industrial products (white biotechnology). These products are mostly commodities - i.e. low-margin, high volume products – that face much higher economic pressure and thus need continuous improvement as (opposite to red biotechnology, high-margin/low-volume and no alternative manufacturing process exists). Interesting examples of process intensification in white biotechnology, arise in different fields, such as bio-fuels, bio-flavors, artemisinin production and biorefinery (Kochergin & Kearney, 2006; Vaghari et al., 2015). In such products, the main goal is maximizing economics and environmental footprint by having smaller, cleaner, safer, and more energy efficient processes. Since these benefits are indispensable to improve the economy of the product and be competitive in the market, these innovations are transferred freely and directly to industrial manufacturing (Dunn, Wells, & Williams, 2010; Whitesides, 2015). Process

enhancement examples are media recycling, in-situ product recovery, batch-continuous manufacturing transformation (Gronemeyer et al., 2014; F. Grote, Ditz, & Strube, 2009; Florian Grote, Ditz, & Strube, 2012; Heinzle, Biwer, & Cooney, 2006)

1.6. Quality by Design

Regulatory authorities acknowledge these limitations and thus started demanding development approaches: quality by design (QbD) approaches, and methods: process analytical technology (PAT). In the QbD approach the first step is to define the quality target product profile (QTPP) that identifies the critical quality attributes (CQA) of a drug product. The main goal is to understand the influence of critical process parameters (CPPs) and critical material attributes (CMAs) and how these affect the CQA (Figure 1. 2).

$$CQAs = f (CPP_1, CPP_2, CPP_3 ...CMA_1, CMA_2, CMA_3...)$$
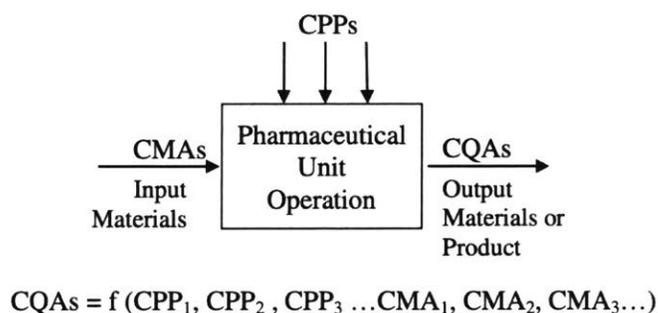
Figure 1. 2. Quality by Design for Pharmaceutical Unit Operations. The Critical Material Attributes (CMAs) together with Critical Process Parameters (CPPs) affect the Critical Quality Attributes (CQAs). Extracted from Yu et al. (2014)

Mechanistic understanding of pharmaceutical unit operations allows for optimal design, faster-targeted experimentation, designing for scale, optimal process parameter identification as well as control strategies that can potentially guarantee quality and consistency. Mechanistic models are important tools that rely on first principles and current knowledge to virtually define the process, and can achieve faster process improvements allowing for the biopharmaceutical companies achieve faster process intensification, and thus reduce cost and achieve faster but optimal process developments.

1.7. This Work

The present work consists on developing a multi-scale mechanistic modeling framework for a CHO cell bioreactor to enable process intensification despite regulatory constraints. This thesis is divided into sections: (1) *Literature Review*, provides a thorough landscaping of the state of the art in the biological modeling space; (2) *Methods*, consists of a detailed description of the three components of the work: the framework architecture, the software architecture and the validation with a case model; (3) *Results*, calibration and validation of the case model applied to an Amgen pipeline molecule; (4) *Discussion*, consists of analyzing three different cases studies to optimize the current process platform and discusses the opportunities of this framework; (5) the final sections, *Conclusion*, closes the document with the main takeaways of this work.

2. LITERATURE REVIEW

Biological systems compared to chemical systems present additional challenges due to biological intrinsic metabolic regulation. Nonetheless, designing, controlling, and optimizing bioreactors - similarly to chemical systems - can be achieved based on macroscopic first principle models. Technological developments and cost reduction in bio-analytical assays have increased bioreactor monitoring capabilities and cell culture knowledge, allowing for further development and validation of mathematical approaches of biology. In the present section a review of current state of mathematical models and their use in bioprocess control. The present section is divided in three: (1) the modeling frameworks in systems biology assuming perfectly mixed bioreactors; (2) how these models can be expanded to heterogeneous environments; and finally (3) a brief description of characteristic mammalian cell culture models.

2.1. Modeling a Homogeneous Bioreactor

To describe and predict the behavior of cell populations, mechanistic models of varying complexity have been developed. In systems biology, two different dimensions have been defined to classify cellular models, as illustrated by Figure 2. 1. The horizontal axis in Figure 2. 1 represents the model's description of the structure of the biological system - *i.e.* the degree of description of cellular physiology. The vertical axis named segregated, describes how much heterogeneity the model is capturing - *i.e.* the variability of the system in terms of biological differences (different types of cells or different states amongst the same types of cells).
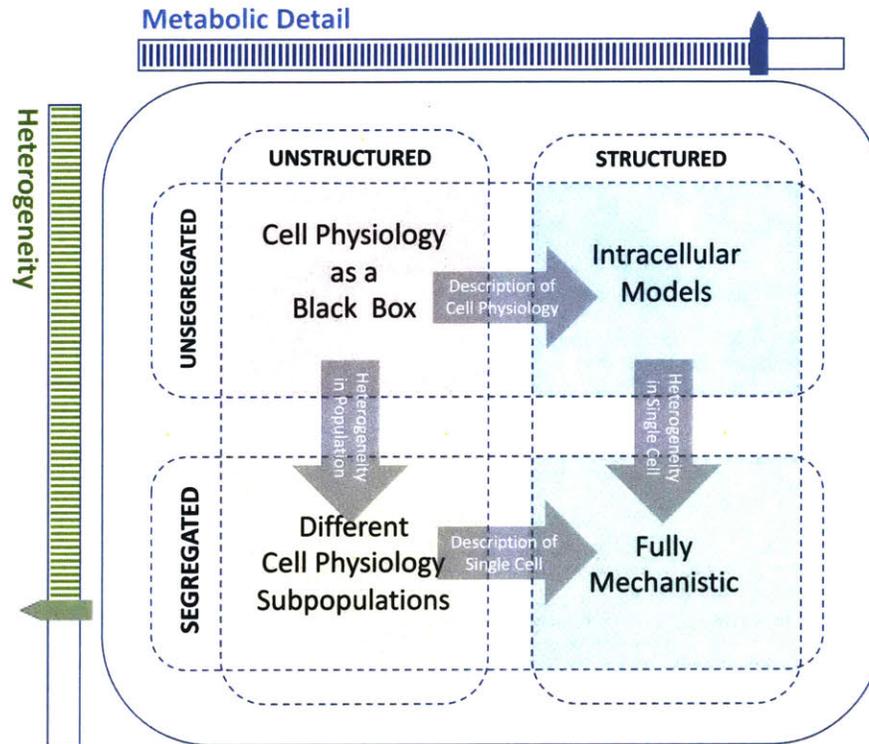
Figure 2. 1. Systems Biology Mechanistic Modeling Classification. Horizontal dimension, structuredness: description of the biomass (metabolism/compartment). Vertical Dimension, Segregatedness: description of heterogeneity (intrinsic physiological difference amongst cells). Adapted from (Lencastre Fernandes et al., 2011)

The simplest model is an unstructured-unsegregated model. Unsegregated models assume all cells behave in the same manner, and the description of physiology is based on an average cell description of the biomass. These models account for the vast majority of models used nowadays (Lencastre Fernandes et al., 2011). An Unstructured model considers physiology as a black box, meaning intracellular activity is not described and the behavior of the cell is based exclusively on extracellular concentration of substrates and products.

2.1.1. Structured Models

The explosion of the omics era and the affordability of analytical techniques have allowed for a much better understanding of cellular physiology, particularly in the context of whole-cell metabolism. This expansion has allowed 'opening the box' and looking inside the cell, allowing for vast development of

19

structured mathematical formulations of metabolism. Different approaches to describe physiology have been explored. Figure 2. 2 describes and compare several of these approaches.

Method	Model systems	Parameterization	Typical prediction type	Advantages	Disadvantages
Stochastic kinetic modelling	Small-scale biological processes	Detailed kinetic parameters	Reaction fluxes, component concentrations and regulatory states	• Mechanistic • Dynamic • Captures biological stochasticity and biophysics	• Computationally intensive • Difficult to parameterize • Challenging to model multiple timescales
Deterministic kinetic modelling	Small-scale biological processes	Detailed kinetic parameters	Reaction fluxes, component concentrations and regulatory states	• Mechanistic • Dynamic	• Computationally intensive • Difficult to parameterize
Constraint-based modelling	Genome-scale metabolism	Network topology, and uptake and secretion rates	Metabolic flux states and gene essentiality	• Mechanistic • Large scale • No kinetic information is required	• No inherent dynamic or regulatory predictions • No explicit representation of metabolic concentrations
Logical, Boolean or rule-based formalisms	Signalling networks and transcriptional regulatory networks	Rule-based interaction network	Global activity states and on–off states of genes	Can model dynamics and regulation	Biological systems are rarely discrete
Bayesian approaches	Gene regulatory networks and signalling networks	High-throughput data sets	Probability distribution score	• Non-biased • Can include disparate and even non-biological data • Takes previous associations into account	• Statistical • Issues of over-fitting • Requires comprehensive training data
Graph and interaction networks	Protein–protein and genetic interaction networks	Interaction network that is based on biological data	Enriched clusters of genes and proteins	• Incorporates prior biological data • Encompasses most cellular processes	• Dynamics are not explicitly represented
Pathway enrichment analysis	Metabolic and signalling networks	Pathway databases (for example, KEGG, Gene Ontology and BioCyc)	Enriched pathways	• Simple and quick • Takes prior knowledge into account	• Biased to human-defined pathways • Non-modelling approach

Figure 2. 2. Structured physiological models in systems biology comparison. Extracted from Bordbar, Monk, King, & Palsson (2014)

Under a fundamental assumption that cell physiology can be described as a network of enzymatic reactions, it can be argued that a fully descriptive network together with the knowledge of reaction kinetics (and their underlying dependency), cell physiology could be abstracted. Fully modeling the cell via a metabolic network relies on the current efforts to describe cellular metabolism as a set of biochemical reactions mediated by enzymes. The increasing availability of high-throughput technologies has allowed the view of the cell as the system of study, allowing for novel and detailed models to describe metabolism. The underlying assumptions and description of the enzyme kinetics is generally what differentiate the different modeling approaches.

Fully describing the system with deterministic or stochastic kinetics is currently infeasible due to the large number of metabolites in the cell. Additionally, nonlinear interactions among metabolites (such as feedback loops) make the process very complex to determine and model. In the kinetic approach each biochemical reaction's kinetics is explicitly described, and consequently the models are described by non-linear rate laws that generally involve highly parameterized mathematical expressions difficult to estimate and expensive to compute (for a detailed review see Saa & Nielsen, 2017). For this reason, scaling this approach to genome or metabolome scale – generally required for bioprocess development- has been extremely difficult.

To overcome this challenge, equilibrium models of metabolism have emerged as promising tools in systems biology. In Constraint based modeling (CBM), an exhaustive set of stoichiometric equations – representing reactions in equilibrium – is converted to a mathematically consistent format, known as the stoichiometry matrix. These reactions are generally constructed from annotated Gene-Protein-Reactions (GPR) associations. The GPR annotations come from different sources such as large (genome-scale) and/or compiled small scale annotations. The stoichiometric matrix is commonly underdetermined and thus needs and additional objective function to be solved. Three computational approaches have been used in CBM: (i) Flux Balance Analysis (FBA), consists of analyzing the network based on the pseudo-steady state mass balance around intracellular metabolites and an optimization function (for a detailed review see Lewis, Nagarajan, & Palsson, 2012), (ii) pathway analysis, which seeks to identify feasible flux routes through the network and (iii) random flux sampling, which consists of finding probability distributions of feasible steady state fluxes.

Particularly, in FBA, obtaining biologically accurate flux representations of metabolism depends on the choice of objective function and the scope of the modeled network. Different objective functions have been explored (Figure 2. 3), the most accepted being to maximize biomass growth. Particularly novel is the concept of parsimonious FBA (pFBA), which finds the flux distribution that maximizes biomass while minimizing overall metabolic expenditure (by minimizing total flux) (Lewis et al., 2010). Despite the

21

bold simplification of pseudo-steady state in metabolite concentration, this approach allows for genome-scale mechanistic models with small kinetic information and has shown great promise, particularly in steady-state cultures. Additionally, these CBM models can be compartmentalized and thus used to model subcellular compartments within the cell. Over 85 consensus models have been developed in recent years ranging from simple organisms such as *E.coli* to complex ones such as *Homo sapiens* (retrieved from BiGG database as of March 11th 2019, refer to King et al., 2016). To incorporate regulatory dependencies, recently this approach has been modified to include transcriptomic data either by changing the boundaries of the constraints (Guo & Feng, 2016; Lewis et al., 2010) or by adding the enzymes as part of the network (Sánchez et al., 2017).

The need for dynamic models of metabolism has been explored by generating hybrid models that simulate whole-cell metabolic in a dynamically changing environment. These models, known as dynamic FBA (dFBA) use kinetic rate expressions to modify the boundaries of a FBA (Mahadevan, Edwards, Doyle, & 3rd, 2002; Sánchez, Pérez-Correa, & Agosin, 2014). Mathematically speaking, dFBAs consist of embedded linear optimization problems within a system of ordinary differential equations. Despite the conceptual simplicity there are computational challenges that need to be overcome and different approaches to solve these models have been explored. The classical approach consists of a Static Optimization Approach (SOA), which uses pre-defined time intervals. Then came the Direct Approach in which adaptive time steps were used allowing for more accurate and less computationally expensive solutions. (Höffner, Harwood, & Barton, 2013), later Differential Algebraic Equations (DAE) were used to solve the system by reformulating the dFBA to reutilize solutions. The most exhaustive solution, which is limited to small networks, is the Dynamic Optimization Approach (DOA) in which she system is simultaneously solved over the entire simulation. Recently (Jeong, Yoo, Kim, & Lee, 2016) developed a parametric programing approach, on which precomputing FBA solutions were used to generate a single layer optimization, the authors named this approach xDFBA. An alternative, but promising approach is the combination of FBA with cybernetic models, where cells are modeled as goal seeking individuals, and uptake kinetics can be

cybernetically modeled according to environmental changes (Namjoshi, Hu, & Ramkrishna, 2003; Song & Ramkrishna, 2010).

Objective function[a]	Mathematical definition	Explanation	Rationale
Max biomass[b]	$\max \dfrac{v_{biomass}}{v_{glucose}}$	Maximization of biomass yield	Evolution drives selection for maximal biomass yield ($Y_{X/S}$)
Max ATP	$\max \dfrac{v_{ATP}}{v_{glucose}}$	Maximization of ATP yield	Evolution drives maximal energetic efficiency ($Y_{ATP/S}$)
$\text{Min} \sum v^{2}$ [c]	$\min \displaystyle\sum_{i=1}^{n} v_i^2$	Minimization of the overall intracellular flux	Postulates maximal enzymatic efficiency for cellular growth (analogous to minimization of the Euclidean norm)
Max ATP per flux unit[c]	$\max \dfrac{v_{ATP}}{\sum\limits_{i=1}^{n} v_i^2}$	Maximization of ATP yield per flux unit	Cells operate to maximize ATP yield while minimizing enzyme usage
Max biomass per flux unit[c]	$\max \dfrac{v_{biomass}}{\sum\limits_{i=1}^{n} v_i^2}$	Maximization of biomass yield per flux unit	Cells operate to maximize biomass yield while minimizing enzyme usage
Min glucose	$\min \dfrac{v_{glucose}}{v_{biomass}}$	Minimization of glucose consumption	Evolution drives selection for most efficient usage of substrate
Min reaction steps[c]	$\min \displaystyle\sum_{i=1}^{n} y_i^2, y_i \in \{0, 1\}$	Minimization of reaction steps	Cells minimizes number of reaction steps to produce biomass
Max ATP per reaction step[c]	$\min \dfrac{v_{ATP}}{\sum\limits_{i=1}^{n} y_i^2}, y_i \in \{0, 1\}$	Maximization of ATP yield per reaction step	Cells operate to maximize ATP yield per reaction step
Min redox potential[d,e]	$\min \dfrac{\sum\limits_n v_{NADH}}{v_{glucose}}$	Minimization of redox potential[f]	Cells decrease number of oxidizing reactions thus conserving their energy or using their energy in the most efficient way possible
Min ATP production[d,e]	$\min \dfrac{\sum\limits_n v_{ATP}}{v_{glucose}}$	Minimization of ATP producing fluxes[g]	Cells grow while using the minimal amount of energy, thus conserving energy
Max ATP production[d,e]	$\max \dfrac{\sum\limits_n v_{ATP}}{v_{glucose}}$	Maximization of ATP producing fluxes[h]	Cells produce as much ATP as possible

Figure 2. 3. Different Objective Functions explored in Flux Balance Analysis. Extracted from Schuetz, Kuepfer, & Sauer (2007)

Post-translational modifications are a big differentiator for mammalian cell cultures. Particularly, the Golgi Apparatus is important due to its role in protein glycosylation. Understanding, modeling, and controlling the metabolic machinery of this subcellular organelle is particularly important for glycoproteins such as monoclonal antibodies, and as mentioned previously particularly useful in the development of biosimilars. While CBM can include subcellular compartment and reactions, they are limited to quasi-

steady state assumptions within the compartment which often oversimplifies the process. Detailed kinetic models of glycolysis have been developed by several authors (Jedrzejewski et al., 2014; Jimenez del Val, Nagy, & Kontoravdi, 2011; Sou et al., 2015).

2.1.2. Segregated Models

Segregated models consider the extrinsic cell-to-cell variability in a cellular culture by acknowledging distributed rather than uniform cell properties (Refer to Heins & Weuster-Botz, 2018 for a detail review of sources of heterogeneity). Cell physiology is not uniform and is dependent on the cell cycle and state of cell. In a cell culture, a population of cells proliferates, and cells dynamically change their state asynchronously, depending both on intrinsic as well as extrinsic variables (Figure 2. 4). From a modeling perspective, cells can be differentiated by means of scalar properties, such as age and mass, or specific intracellular metabolites as descriptors of the metabolic cell state. Several studies (Pan, Dalm, Wijffels, & Martens, 2017; Sunley & Butler, 2010) have described the difference in protein productivity in different cell states and thus have developed models that could help maximize specific productivity.
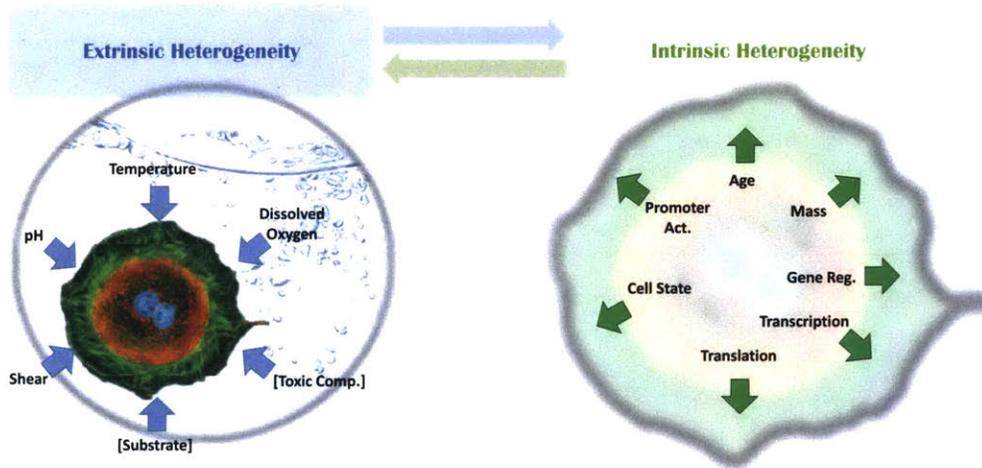
Figure 2. 4. Sources of Biological Heterogeneity. Adapted from Heins & Weuster-Botz (2018)

Two different approaches have been utilized to account for cell-to-cell heterogeneity: Population Balance Models (PBM) and Agent Based Models (ABM).

PBMs provide the most generic approach for modeling distributed properties. Instead of modeling a composition vector of biomass description (grams per gram of dry cell weight), cell heterogeneity is described by certain properties (*e.g.* age, size) and a dynamic balance of the distribution of the cells around the property or properties is modeled.

PBMs consists of population balance equations, boundary, and initial conditions as well as coupled equations that describe cell division probability, partitioning of cell content, stage transition, and cellular kinetics in the case of structured models. Population balance equations can be defined as equations that account for the various processes that change the number of cells in a population. The equations take the form of first-order partial integro-differential equations; the supplementary equations are usually coupled in a non-linear way, and are typically integro-differential equations (Lencastre Fernandes et al., 2011; Mantzaris, Daoutidis, & Srienc, 2001a).

The dimension of the property space can vary, single variable PBMs have included mass-, growth-rate- and age- structured models. Multi-dimensional PBM have also been explored, but are much limited due to both formulation limitations (which are mainly constrained by singe-cell measurements) as well as computational requirements to solve the problem. Refer to (Lencastre Fernandes et al., 2011) for a review on single and multi-dimensional PBMs in microorganisms.

PBMs are complex equations and analytical solutions are often limited. Different numerical schemes have been used to obtain dynamical solutions, including three-step procedure, finite differences, finite elements, orthogonal collocation, spectral methods (Lencastre Fernandes et al., 2011). Other approaches such as Monte Carlo, method of characteristics and method of moments have also been (for a detailed review refer to Mantzaris et al., 2001a; Mantzaris, Daoutidis, & Srienc, 2001b, 2001c; Nielsen, Villadsen, & Lidén, 2002).

Agent-based modeling (ABM), consists of a model that defines the system as a collection of autonomous decision-making entities called agents. In the system, each agent individually assesses its

situation and makes decisions on the basis of a set of rules (Bonabeau, 2002). ABM have been proved advantageous in modeling systems because of their ability to: (1) capture emergent phenomena, *i.e.* predict properties that are decoupled from the properties individual parts, (2) provides a natural description of the system, and (3) its flexibility, *i.e.* has a natural framework for tuning the agent complexity, and level of description and aggregation. ABM has been applied to the social, political and economic sciences amongst others. ABM has also been applied in the field of biology in translational systems biology (An, Mi, Dutta-Moscato, & Vodovotz, 2009), cancer modeling (Zhang, Wang, Sagotsky, & Deisboeck, 2009), biofilm formation by microbial communities (Jayathilake et al., 2017) as well as in bioreactor dynamics (Elif S. Bayrak et al., 2016; Elif Seyma Bayrak, Wang, Cinar, & Undey, 2015).

In the context of cell populations, the term Individual Based Models (IBM) is also used (González-Cabaleiro, Mitchell, Smith, Wipat, & Ofiţeru, 2017). In these models the individual cells or a cluster/population of cells are discrete particles (agents) which interact with each other and with their environment. Physiology is described at an agent level, thus allowing the study of the behavior of the system as a result of the properties, performance, and interaction of the individual components (Railsback & Grimm, 2012). The main drawback of these modeling techniques is computing constraints (González-Cabaleiro et al., 2017). Critical model choices are the number and type of agents used as well as the level of detail for each agent. IBMs have gained popularity in microbiology in the last decades (Ferrer, Prats, & López, 2008; Hellweger, Clegg, Clark, Plugge, & Kreft, 2016; Schuler et al., 2011) mainly due to accelerated advancement in computational power and the development of specialized software (González-Cabaleiro et al., 2017). In fact, nowadays open source platforms are now available (Coakley et al., 2016; Lardon et al., 2011; Rudge, Steiner, Phillips, & Haseloff, 2012; Sklar, 2007). Nonetheless, due to their complex structure ABMs require more computing skills than other modeling approaches, such as PBMs. Recently a Gaussian process emulator for IBM was generated to overcome reduce the computational requirements of this modeling approach (Oyebamiji et al., 2017)

2.2. Spatial Heterogeneity

Cellular physiology and its impact on growth and bioproduction is a complex interaction between the extracellular environment and cellular machinery. Maintaining homogeneity at a large scale is often practically infeasible due to energetic constraints (Hermelink, Brauer, Lasch, & Naumann, 2009; Lencastre Fernandes et al., 2011). As a result, large scale bioreactors often have significant gradients of substrates (*e.g.*, glucose or dissolved oxygen), inhibitory byproducts (such as carbon dioxide and lactate), pH and temperature, amongst others. Oxygen limitation is often a matter of concern and arises when biological consumption rate exceeds the physical transfer rates, which is often present in certain zones of the bioreactor. Additionally, substrates (both gaseous and liquid) are often added in certain regions of the bioreactor making the gradients more pronounced.

These gradients may significantly affect the biological process and thus make the mathematical models inapplicable (Albers, Larsson, Lidén, Niklasson, & Gustafsson, 1996; Lencastre Fernandes et al., 2011). Coupling biological models with extracellular environment models enables the mathematical formulation to capture spatial heterogeneity and thus make them applicable across scales.

Computational Fluid Dynamics (CFD) models consist of three fundamental principles: continuity, momentum and energy conservation; which can be solved -often numerically- under a given set of conditions to represent fluid flow (Lencastre Fernandes et al., 2011). CFD coupling with biological models have gained importance, both in academia and industry for modeling bioprocesses (Bezzo, Macchietto, & Pantelides, 2003; Fang, 2010a, 2010b; Generalis & Cartland Glover, 2005).

Two computational approaches for modeling the phase interactions are commonly used (Kelly, 2008): the Euler-Euler approach (Eulerian) consists of treating the phases mathematically as interpenetrating continua (Barrue, Bertrand, Cristol, & Xuereb, 2001; Micale, Montante, Grisafi, Brucato, & Godfrey, 2000) and the Euler-Lagrange approach (Lagrangian) consists in treating the fluid phase as a

continuum and solving the dispersed phase by following a large number of particles through the calculated flow field (Derksen, 2003).

2.2.1. Eulearian

Elqotbi et al. (2013) used an Euler-Euler multi-phase CFD model with an unstructured model for gluconic acid production by Aspergillus niger. By considering a cellular continuum phase the authors argue a better prediction of bioreactor dynamics due to correct prediction of bioreactor viscosity its effect on oxygen transfer rate and consequently on product formation.

Bezzo et al. (2003) used a steady-state CFD calculations to compartmentalize the bioreactor into discrete isothermal homogeneous zones. With an unstructured-segregated (PBM) model of xanthan gum production in a batch aerobic culture the authors were able to capture the description of imperfect mixing.

2.2.2. Lagrangian

Lapin, Müller, & Reuss (2004) utilized an Euler-Lagrange approach combined with a fractional-step method and simulated three examples: unstructured biological model, a tracer particle and a fairly simple structured-segregated model. The same group later used a similar approach to model the phosphotransferase system in E. coli cells growing in fed-batch cultures in large scale heterogeneous bioreactors (Lapin, Schmid, & Reuss, 2006). The authors emphasize the advantage of the Euler-Lagrange approach over the Euler-Euler approach based on a more realistic description of the biological phase that reflects history of the cells. Additionally, they justify that momentum transfer between the particles and the fluid phase can be neglected because the particles are smaller than the meshing, and thus can be treated as a quasi-single phase. Based on similar assumptions (Morchain, Gabelle, & Cockx, 2014) assumed that the cells behave like tracers of the continuous phase, and thus represented them through their concentration in the liquid phase. Furthermore, they assumed non-identical individuals represented by a population balance model. With a Euler-Lagrange approach the authors coupled an unstructured-segregated biological model to a CFD model. Comparing both lab scale and industrial aerated fermenters the authors used a simple

substrate/oxygen unstructured kinetic model with PBM that differentiates individuals based on the specific growth rate. (Kaul, Cui, & Ventikos, 2013) using a Euler-Lagrange approach used an Agent Based Modeling approach to represent proliferation, migration, chemotaxis and apoptosis of cells in a bioreactor.

Delafosse et al. (2015) through abiotic bioreactor experiments with tracer particles validated an Euler-Lagrange modeling approach with a CFD compartment model. (Haringa, Noorman, & Mudde, 2017) also developed guidelines to determine the tradeoff between modeling resolution and computational cost, particularly by defining the optimal number of particles and time-steps of the system.

To the best of our knowledge no previous experiment has explored the combination of CFD modeling with truly structured-segregated mathematical representations of biological systems. Here we propose a flexible framework that enables different applications and modularly changed the configuration of the system. Its most granular configuration is represented by structured-segregated biological models in heterogeneous bioreactors, with a combination of agents and PBMs.

2.3. Mammalian Cell Models

In the present section a brief description of the different mammalian cell biological models is described according the definition stated in Figure 2. 1. The section is divided in (1) Unsegregated Unstructured models; (2) Structured-Unsegregated; (3) Unstructured-Segregated; and (4) Structured-Segregated models.

2.3.1. Unsegregated unstructured

Unsegregated unstructured models are the simplest models and comprise the original interpretation of biological system derived from chemical principles and contextualized by limited bio-analytical power and physiological knowledge. Refer to (Nielsen et al., 2002) for detailed description of this classification of models.

2.3.2. Structured-Unsegregated

Diverse *Structured-Unsegregated* models of mammalian cells have been developed. Quek *et. al* reconstructed a genome scale mouse model and simplified to a final version of model with 272 reactions and 228 metabolites compartmentalized into the cytoplasm and mitochondria (Quek, Dietmair, Krömer, & Nielsen, 2010). The authors analyzed the network using FBA by minimizing overall fluxes with two examples: a continuous hybridoma systems and batch CHO culture. Nolan *et. al* had a similar approach, starting from a network of reactions collected from the KEGG database for the hamster analog mouse, and after a series of simplifications and reductions reached a model with 34 reactions and 24 metabolites (Nolan & Lee, 2011). The authors used the fully determined network and experimental perfusion data to calibrate 47 kinetic parameters that describe 12 intracellular reaction kinetics. Recently, Hefzi *et. al* developed a consensus genome-scale reconstruction of CHO cell metabolism in a network 6633 reactions and 4456 metabolites, associated with 1766 genes. The model provides the biochemical basis of growth and recombinant protein production (Hefzi et al., 2016).

2.3.3. Unstructured Segregated

Bartonlini *et. al* developed a single-dimension PBM describing mass of the cells, with an unstructured cellular model in a well-mixed bioreactor gas/liquid model (Bartolini et al., 2015). The authors evaluated different initial concentrations of substrate and oxygen, inoculum and feeding rate. The Mantalaris group developed a multi-dimensional PBM for the different phases of the cell cycle determined by intracellular metabolites (cyclins and DNA), with an unstructured cell model and calibrated with batch experiments, to predict cell cycle and viability particularly by capturing cell growth arrest (García Münzer, Kostoglou, Georgiadis, Pistikopoulos, & Mantalaris, 2015; Kostoglou et al., 2016).

Bayrak *et. al* developed an ABM combined with an unstructured cellular model to predict CHO performance in fed-batch bioreactors (Elif Seyma Bayrak et al., 2015). The model accounted for

environmental variables as well as internal cellular states to define the rule-based logic of growth of the unstructured model.

2.3.4. Structured Segregated

A detailed model of glycolysis was coupled with a segregated model to describe the growth of five different Madin-Darby Canine Kidney (MDCK) strains based on cellular diameter (Rehberg, Ritter, & Reichl, 2014).

Sidoli et al. (2006) presented a complex mammalian kinetic model with over 700 parameters by coupling a population balance model based on cell mass with a very detailed physiological model of mammalian cells.

Despite the detailed description of these models, often times they result in a poor fit due to limited experimental data to estimate the large number of parameters they require, particularly for intracellular sampling and spatial location of the cells (Kyriakopoulos et al., 2018).

3. METHODS

The present section describes the methods developed in the present work. It is divided in four sections: (1) the *framework architecture* (FA), capable of implementing all of the established modeling techniques found in literature; (2) *the software architecture* (SA), designed in a modular and expansible fashion allowing for continuous development of any configuration of the FA; (3) a *case-model* of the architecture (FA-01) capable of predicting the impact of process variables on bioreactor performance, that forms the first effort at Amgen of the FA and validates the SA; and (4) the *calibration methods* for FA-01, with the experimental and mathematical methods used to obtain the relevant parameters.

3.1. Framework Architecture (FA)

The first goal of this project is generating the *engineering design* to model cell culture in a bioreactor. For the clarity of the reader, this will be referred as Framework Architecture (FA). The FA design does not consider any current computational limitations. Instead it is a framework capable of integrating all model components necessary to properly define the physical-chemical-biological phenomena of interest, while also allowing those components to be included or excluded from the model for any given application as the tradeoff in speed and precision requires. The main applications that the FA intends to address include process optimization, engineering design, soft sensing, model-predictive control, and process variable sensitivity analysis. For proprietary reasons the present section contains a high-level description of the FA.

Even though the FA design might not be computationally feasible at the moment, this exercise envisions the long-term goal of bioreactor modeling effort at Amgen, and enables a software architecture that allows for a sustainable growth of the project.

32

The FA design is presented in Figure 3. 1. The FA has been divided into 3 domains: the bioreactor, the cell, and subcellular compartments. The bioreactor domain is further understood to comprise liquid and gas phases. The FA is further divided into controlled variables (or inputs) and process variables (physical, chemical & biological variables that represent the system).

Control variables are:

o   Gaseous inputs: input flow of gas as a function of position and time;

o   Liquid inputs: input flow of solute – generally feeding media or pH control agents– as a function of position and time;

o   Temperature of the bioreactor as a function of time, which is assumed homogeneous; and

o   Agitation: angular velocity of the impellers as a function of time – it is noteworthy to mention that the physical dimensions as well as the location of each impeller is known.

Process variables represent the

o   Gaseous phase of the system in the form of bubbles. Particularly interesting is the partial pressure of oxygen and carbon dioxide in the bubbles that allows proper gaseous exchange rates for the cells to breathe correctly. The gas concentration will dynamically change based on the surface area of the bubble and the solutes which it interacts with.

o   Extracellular liquid components are the concentrations of solutes. Liquid components include dissolved oxygen, carbon dioxide and all the carbonated species, feeding nutrients for the cell to grow, and products of cellular metabolism (either target products, such as an antibody, or byproducts such as lactate or ammonia). Note the spatial dependence of these components allows for extrinsic heterogeneity in the system to be represented in the model.

o   Biomass, representing individual cells or groups of cells. In the event biomass represents an individual cell, it comprises a position and state as a function of time (these states account for the

33

intrinsic heterogeneity of the system, and the transitioning of states result from the interaction with the environment and cell physiology).

o One or more of a number of candidate subcellular compartment models, for example using the plug flow model developed by Jimenez del Val et al. (2011) the system can be characterized by a flow velocity rate and the concentration of glycosylated profile within the compartment as a function of time. Analogously, other subcellular compartments can be modelled and coupled to the cellular model, such as the ribosome.

For clarity of the reader the FA is described in the following section as a causal system. To update the state variables a series of sequential steps are performed in the different phases of the system in an interconnected fashion. Here it will be described in a sequential manner.
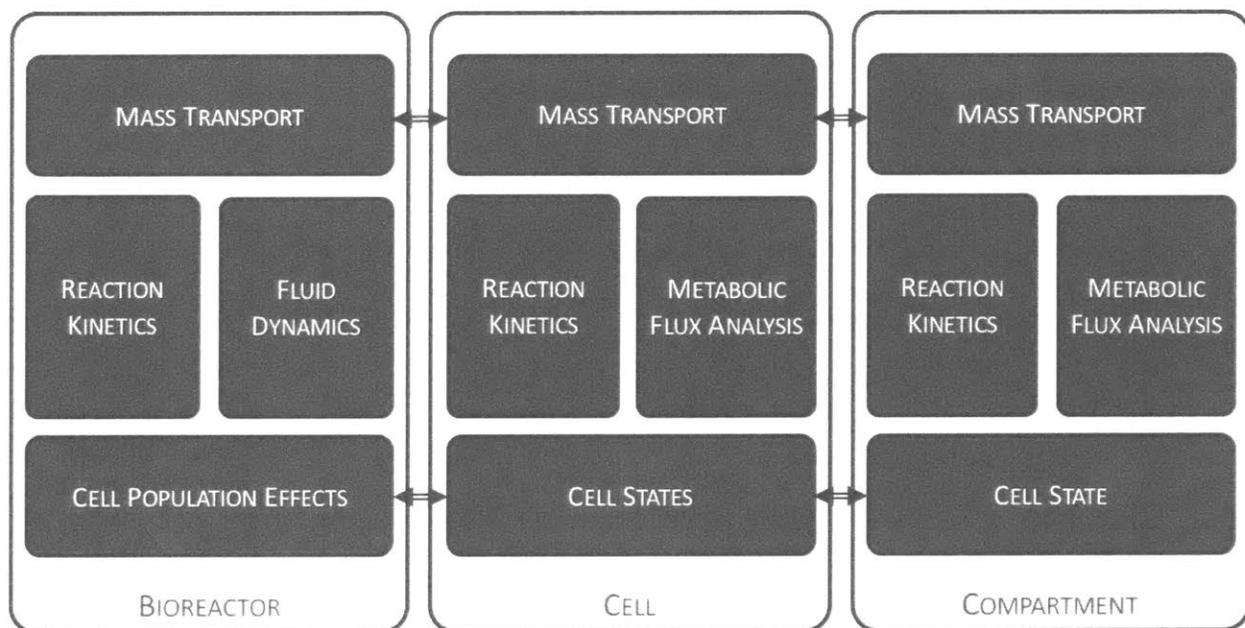
Figure 3. 1. Engineering design of the cell culture model in a bioreactor.

3.1.1. Bioreactor Process Modeling

In the bioreactor domain, conservation equations are solved in both the liquid as well as the gaseous phase. Mass, momentum, or energy conservation may be implemented as required by a particular application. In some cases, solution of partial differential equations will be necessary, which may be performed using finite volume methods (viz. computational fluid dynamics – CFD). The inputs of the system are temperature, agitation, boundary conditions of the gas and liquid domains, viscosity, and density of the system. The outputs are the vector velocity fields in the gas and liquid domains necessary for the convection diffusion step, the shear stress necessary to account for stress-related biological conditions (*e.g.* lysis), and quantities accounting for the gaseous bubble dynamics.

Chemical species (gas bubbles and solutes) are transferred within the bioreactor due to diffusion and convection. The inputs for mass conservation during convection and diffusion are concentrations of the liquid and gas components as well as the velocity field obtained from solving the momentum conservation equation. A similar transport process is solved for the cells.

Mass transfer between the liquid and gas domains may account for the size of bubbles and the solubility of gasses. This step is crucial given the scaling difficulty of good gas-liquid transfer rates at high biomass levels essential for healthy cell metabolism, *i.e.* constant supply of oxygen without toxic accumulation of carbon dioxide.

In addition to the bioreactor-scale physical phenomena, a biological cascade of processes is simulated at the cell scale. Given that cellular concentrations are on the order of 10^6 cells per mL, modeling each cell at each position would be practically impossible and not particularly necessary. For this reason, a proprietary discretization methodology may be performed to approximate the heterogeneity in the cell population.

Each discretized group of cells needs to react to their physical environment and produce a response. Coupling of the physical and biological systems is accomplished by equality of a source term in the physical domain with a boundary condition in the biological system. The cell module uses the extrinsic variables

35

(transformed according to the discretization scheme) to mechanistically determine key boundary conditions of the selected biological model such as reaction rates or metabolic fluxes.

Given the bioprocess goals of the overall model, the use of metabolic network models to account for the biological activity of the cell factories is proposed. Using a flux balance analysis approach, given a specific objective function, the solution of the system can be obtained. The solution of the system consists of all the fluxes that define the solution for that particular discretization, including intracellular fluxes, production/consumption rates and growth rate. It is important to mention that for each specific state different objective functions and or network compositions can be defined (for example different biomass composition) to account for different biological goals of each state.
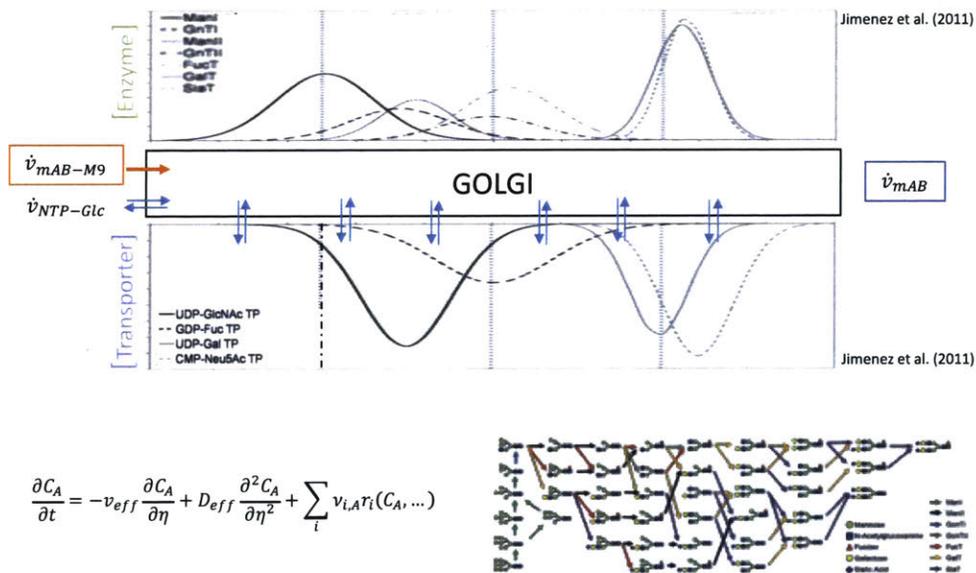
Figure 3. 2. Golgi Kinetic Model. Developed based on the model from (Jedrzejewski et al., 2014; Jimenez del Val et al., 2011)

If a subcellular model – for a specific subcellular process – is included in the FA, the state of the cell model can serve as the input for the subcellular model. To exemplify this in a commercial relevant product monoclonal antibodies (mAb) are selected as an illustrative case study due to the importance of glycosylation in the biosimilar market. For mAb formation, the glycosylation process occurring in the Golgi

apparatus is crucial for product profile. In this case model the input fluxes to the subcellular compartment are the mAb-M9 (antibody precursor) and the activated nucleotide sugar donors (NSD) as well as an effective 'processing' velocity of the plug flow bioreactor model (Figure 3. 2). The former two variables can be obtained from the FBA, while the latter is determined by process variables (such as temperature).

The biological solution is then projected back to bioreactor domain using the inverse of the discretization transform and mass, momentum, or energy conservation resolved as required. Finally, the viscosity and density of the system may be updated as a function of biomass growth.

3.1.2. *Process Control Modeling*

In addition to the spatiotemporal update of the process variables, the process control is simulated. In the physical system, the process variables are estimated or sensed via on-line instruments or off-line sampling. These measurements serve as inputs for control schemes to manipulate the input variables to the system and control the biomanufacturing process.
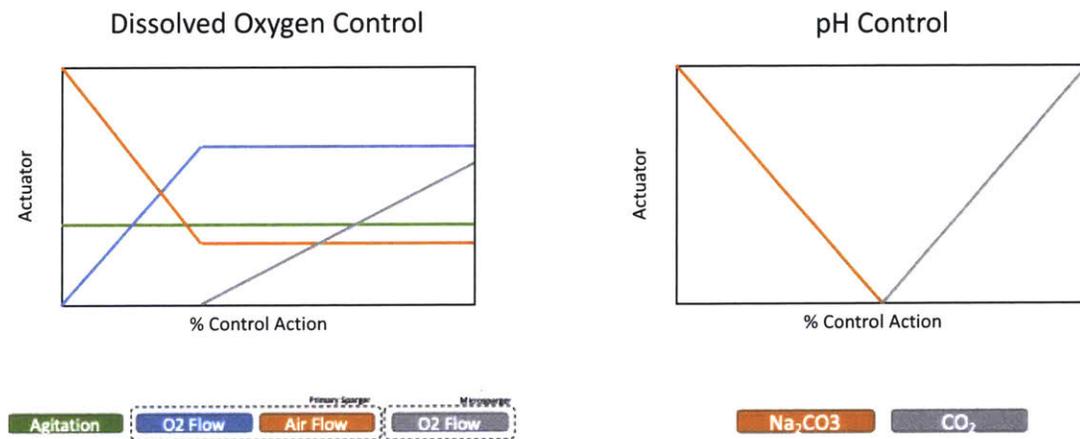
Figure 3. 3. Typical Control scheme for Oxygen and pH.

Online sampling includes temperature, dissolved oxygen and pH – the control logic of these variables is standard, and the control logic is incorporated in the FA. Other online measurements can be included in the model as needed, such as on-line biomass sensors or metabolite sensors (e.g., glucose or lactate).

Off-line sampling includes biomass concentration, metabolite profile, product concentration and quality (*e.g.*, glycoform composition in the case of antibody production). Usually off-line measurements are manually sampled and have considerable delays due to slow analytical processing of the samples (e.g. microscopy, chromatography). Due to these limitations, off-line sampling is not typically included in bioprocess control systems; the modeling framework nonetheless allows its incorporation.

Both off-line and on-line measurements are the input for a process control strategy that can manipulate the feeding strategy of the bioreactor (including different feeding media, inflow rates, outflow rates) as well as the set-points for process variables such as temperature, pH, and dissolved oxygen. Much of the current strategy in bioprocess development is pre-defined open-loop strategies, that manipulate these variables on a predefined schedule or when the culture achieves a desired biomass. The proposed FA includes the possibility to generate novel control strategies.

It is important to mention that the proposed FA enables many different instantiations of individual models that incorporate different assumptions and levels of structure (*i.e.*, metabolic detail) as well as segregation (*i.e.*, both intrinsic and extrinsic heterogeneity). The level of structure can be changed by modulating the underlying biological model of biomass in the system, *i.e.* the metabolic detail of the model and the number of different states that can be included. The level of segregation or heterogeneity is modulated by changing the discretization scheme of the system. The simplest case being that of a well-mixed average-cell bioreactor. Another important aspect of solving the system is the characteristic times of the different elements of the system. That is, how do the speeds of physico-chemical, cellular, and subcellular dynamics compare, and what are the appropriate approximations that can be performed in the system? For example, if metabolism is slow compared to the fluid dynamics, an average trajectory of agents can be used, and coarser time steps solved for the cell model. Figure 3. 4 shows the characteristic time steps of the different processes.

Mixing time scale (turbulent stirred tank):

$$t_m = \frac{5.9}{N}\left(\frac{1}{Np_0^{\frac{1}{3}}}\right)\left(\frac{T}{D}\right)^2$$

N: Impeller rotation speed
p_0: ungased power consumption;
T, D geometric config.

The characteristic time associated with oxygen mass transfer:

$$t_{tr} = (k_L a)^{-1}$$

The time scale of the substrate assimilation:

$$t_s = \frac{\langle S \rangle}{\mu_{max} Y_{SX} \langle X \rangle}$$

The time scale for biomass growth:

$$t_X = (\mu_{max})^{-1}$$

The time scale for biological adaptation:

$$t_a = \begin{cases} > T_{circulation} \\ \sim T_{circulation} \end{cases}$$

The characteristic time for Golgi

$$\tau_{char} = max\left(\frac{L}{V}, \frac{L^2}{D}\right) \sim 15 \text{ min}$$

Figure 3. 4. Characteristic Time steps of Bio-, Chemical- and Physical Processes in FA. Adapted from Morchain et al. (2014)

3.2. Software Architecture

With the model vision defined, the second stage of this research project consisted of developing a software architecture that is capable of supporting continuous and sustainable development of this modeling initiative. The software needed to be capable of instantiating multiple models based on different requirements, and to be flexible enough to persist in time and evolve with the different applications of the FA. Ideally this platform should enable integration of previously independent models developed at Amgen to centralize and unify effort, thereby avoiding duplication and non-value-added efforts.

Python was selected as the language of choice due to its versatility, accessibility, and the large community of developers dedicated to open-source release of extensions. The system needed to be modular, allowing customizable applications and development. It needed to be interoperable, meaning that it needs to be able to communicate with different software if required (e.g., proprietary Computational Fluid Dynamics solvers). And it needed to be extensible, allowing it to evolve in time without hindering prior development.

For proprietary reasons a brief overview of the description of the software architecture is presented in this section. At a high-level, the software architecture is dictated by the concept of a computational study consisting of a model, a problem to which the model is applied, and a solver which computes the solution to the problem. By incorporating the concept of nested, modular studies and including some additional objects for design expediency, the software architecture may be illustrated as in Figure 3. 5.
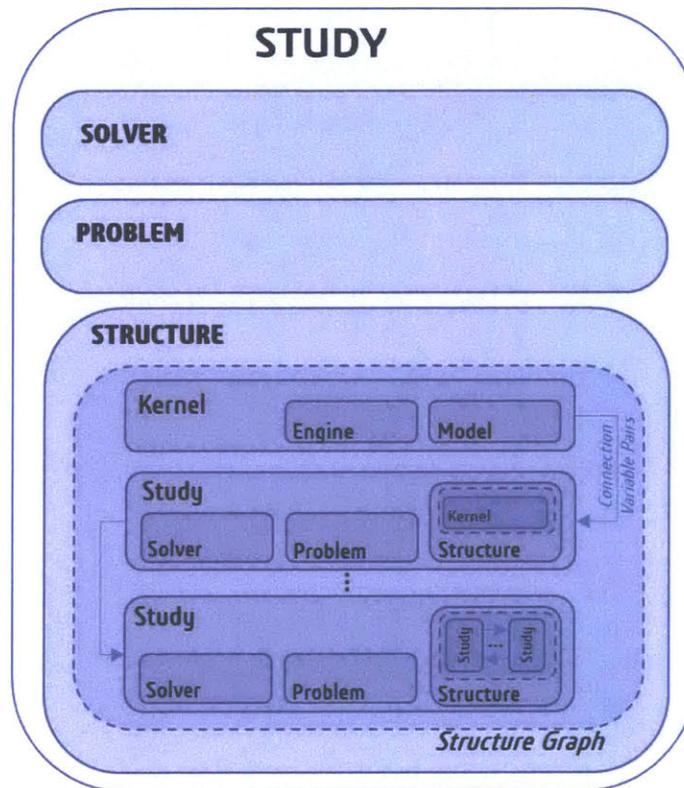
Figure 3. 5. Overview of Software Architecture.

The main objects and relationships are:

- **Model**: Object that represents the mathematical equations of the system to be solved. The model cannot be solved per se, it requires the definition of a specific problem to be solved (specified by the problem object), an algorithm that can solve the problem (solver object) and a software capable of finding the numerical solutions (provided by the engine object).

- **Problem**: Object that specifies what variables of the model are known – specified by problem data – and which variables are unknown and need to be solved for. The problem in combination

with the model determines the mathematical objective. The problem has the capabilities to apply itself to the engine, and by doing so execute the model.

- **Solver**: Object that defines the algorithm by which the model can be executed to solve a specific problem. The solver interacts with the problem and determines a procedure to solve the specific problem.

- **Engine**: Object that acts as the interface with the software that provides the modeling and numerical solution capabilities. The engine is called by the problem, and together with the solver and the model objects comprise a fully determined mathematical problem. The engine is responsible of interfacing this representation to a different software if necessary and by doing so obtaining the solution of the problem.

- **Kernel**: The combination of an Engine and a Model, the minimal configuration of a structure element.

- **Structure**: Multiple elements (either studies or kernels) represented in a Structure Graph. The variables of the connected elements (Variable Pairs) are connected via the Connections Class.

- **Study**: is the combination of a structure, a problem and a solver. It is the minimum class that can instantiate an application.

- **Application**: Is the interface with the user, that generates a study, solves it and retrieves the user-requested solutions.

Note that a study's structure may be composed of other studies or an atomic kernel. This property of the platform allows it to be extensible and modular: studies may be combined hierarchically in any combination in which their input and output quantities agree.

41

3.3. Case Model: FA-01

In the present chapter a dynamic, process-dependent model of a generic CHO cell bioreactor is introduced. The model consists of three different studies: Process Variable and Kinetic Study, a Flux Balance Analysis Study, and a Mass Balance Study. This case model is the first application of the FA, created using the developed software architecture (described in section 3.2) and will be referred to as FA-01.

FA-01 consists of a constraint-based flux balance model of a simplified network of CHO metabolism, constrained by kinetics representing the interaction of the cell with the bioreactor environment. It considers the main energetic components as well as inhibitory by-products; a variable maintenance energy and a stress variable to account for induction-related process perturbations (temperature shift). Temperature, pH, osmolarity and stress are inputs to the mechanistic kinetic equations of metabolism by means of changing the boundaries of the network. A mass balance system accounts for the time and the user-defined inputs.
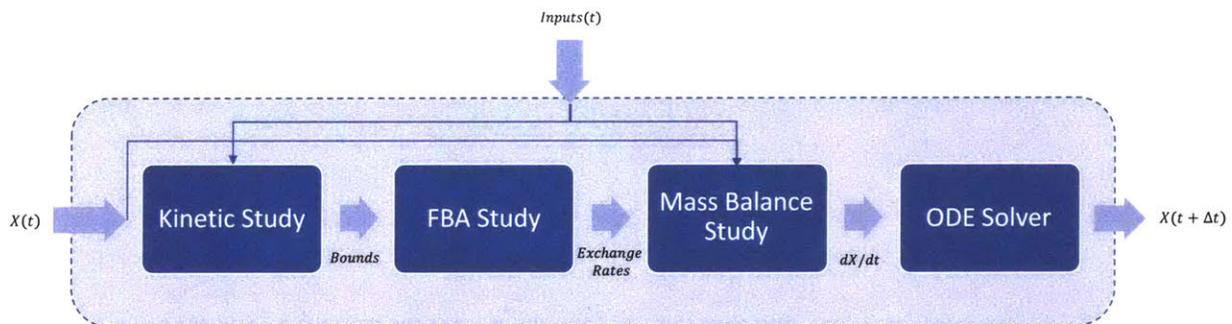
Figure 3. 6. Logic of the Composition of FA-01.

The main **assumptions** of FA-01 are:

o   Non-segregation or Biological homogeneity, *i.e.*, average biomass. Regardless of the difference amongst cells in the culture, all cells are assumed to be the same, and the model characterizes the average physiology of all cells in the culture. (see literature review section for more detail).

o Physical homogeneity, *i.e.*, well-mixed bioreactor. Physical gradients are neglected, lab-scale experimental data are used (2-liter bioreactors) with proper mixing system. Thus, regardless of point addition of media inputs (i.e., oxygen and feeding media) it can be assumed that they distribute homogeneously inside the bioreactor.

o Oxygen is assumed to be in excess due to scale of the experiments. Oxygen needs to be dissolved in solution prior to be metabolized by the cell. As biomass in the culture increases, so does the oxygen demand, and thus a higher liquid-gas transfer capacity is required. Here it is assumed that the control system, mainly due to scale, is sufficient to properly oxygenate the culture at the cellular densities modeled, and thus can be modeled as an excess nutrient.

o Cell physiology is modeled with a simplified version of carbon/nitrogen network. If feasible, goal of the cell is to maximize reproduction while minimizing energetic expenditure. Intracellular metabolites are assumed at pseudo-steady state and flux distribution is determined by a pFBA objective function (see literature review for further details)

o Biomass formula is constant across the duration of the experiment. Normally, the conditions of the physical-chemical conditions of the culture (including cell concentration itself) change with time and so does the cellular composition. In the present work it is assumed that the elemental composition per unit mass remains the same across the culture. Mass is determined by the multiplication of cell diameter and cell count.

o Process variables are assumed to mechanistically modulate metabolism. pH, Temperature and Osmolarity measurements are assumed to affect enzyme kinetics of certain equations. The effect of varying pH, temperature and osmolarity on physiology is not fully understood, in this work it is assumed that deviation from ideal conditions lower enzyme kinetics and thus modulate physiological response.

o Stress is accounted by the perturbations to the system. Abrupt perturbations to the system have an impact on metabolism. In the present work it is assumed that fast changes to the environment affect non-growth associated energy, and thus account for the stress that cells are subject to.

The following section of the work describes each of the different blocks in Figure 3. 6 separately and then will be integrated to form FA-01.

Table 3. 1. Variables and Inputs of the Model.

Variable/Input Symbol	Variable Description
t	Time [day]
[GLC]	Bioreactor Glucose Concentration [mM]
[LAC]	Bioreactor Lactate Concentration [mM]
[O_2]	Bioreactor Dissolved Oxygen Concentration [mM]
[GLN]	Bioreactor Glutamine Concentration [mM]
[GLU]	Bioreactor Glutamate Concentration [mM]
[CC]	Bioreactor Cystine Concentration [mM]
[CYS]	Bioreactor Cysteine Concentration [mM]
[H]	Bioreactor Hydrogen Concentration [mM]
[GLY]	Bioreactor Glycine Concentration [mM]
[SER]	Bioreactor Serine Concentration [mM]
[ASP]	Bioreactor Aspartate Concentration [mM]
[ASN]	Bioreactor Asparagine Concentration [mM]
[NH3]	Bioreactor Ammonia Concentration [mM]
[ALA]	Bioreactor Alanine Concentration [mM]
[X]	Bioreactor Biomass Concentration [10^6 cells/mL]
[X_{dead}]	Bioreactor Dead Biomass Concentration [10^6 cells/mL]
[P]	Bioreactor Product Concentration [g/L]
T_{mem}	Effective Cellular Temperature [C]
T_{σ}	Stress Variable [a.u.]
V_{Harv}	Cumulative Harvest Volume [L]
P_{Harv}	Cumulative Harvest Product Mass [g]
V	Bioreactor Volume [L]
T	Bioreactor Temperature [C]
pH	Bioreactor pH Function [a.u.]
Π	Bioreactor Osmolarity Function []
F_i	Flow rate of Feed Solution i [L/day]

3.3.1. Kinetic Study

The kinetic study is divided in two: (1) the description of mechanistic models that captures the effect of **Process Variables** in physiological responses; and (2) the description of process-dependent mechanistic models of **Reaction Kinetics** on a set of enzymatic reactions.

To account for the cellular sensing of their environment, the effect of Temperature, pH and Osmolarity were incorporated by affecting the enzyme kinetics of reactions of the system. The main assumption is that reaction kinetics can be modeled mechanistically for an ideal state and for different conditions there's a loss in efficiency due to un-ideal conditions. The term happiness ($\mathcal{H}$) was coined to refer to the fractional deviation from optimality, where the optimal condition has happiness of 1, and the rest is a fraction of that.

Temperature Effect

Temperature (T) dependence was modeled according to the mechanistic relationship described by (Ross, Ratkowsky, Mellefont, & McMeekin, 2003). The ideal temperature is 37C, the average blood temperature of a Chinese Hamster.

Equation 3. 1 $$\mathcal{H}_{T}(T) = \frac{\left((T - T_{min})\cdot\left(1 - e^{(d_{temp}\cdot(T - T_{max}))}\right)\right)}{max(\mathcal{H}_T)}$$

pH Effect

The effect of pH was considered according the mechanistic relationship developed by Ross et al. (2003). Note that his function is symmetric on an optimal pH value.

Equation 3. 2 $$\mathcal{H}_{pH}(pH) = \frac{|(1 - 10^{pH_{min} - pH})\cdot(1 - 10^{pH - pH_{max}})|}{max(\mathcal{H}_{pH})}$$

Osmolarity Effect

Osmolarity (Π) has a dual effect on cellular physiology: first it lowers water availability and therefore affects overall metabolism, and secondly high ionic concentrations affect protein folding and thus

functioning. To account for the effect of osmolarity on cell growth the mechanistic relationship developed

by (Zhu et al., 2008) was used.

Equation 3. 3
$$\mathcal{H}_\Pi(\Pi) = \frac{1 - \frac{1}{e^{(\frac{-1}{\pi_{slope}}) \cdot (\pi - \pi_{k_{0.5}}))} + 1}}{\max(\mathcal{H}_\pi)}$$

Stress Effect

An important factor of the process is cellular stress. Stress can come from environmental factors

such as osmolarity, high cellular density, concentration of byproducts, or even purposefully induced stress

– such as temperature shift for target product induction. In this work a time varying stress variable was

created to account for the latter effect. The stress variable was hypothesized to have memory and account

for process perturbations with first order delay kinetics. Figure 3. 7 illustrates the effect of temperature shift

on the stress variable.
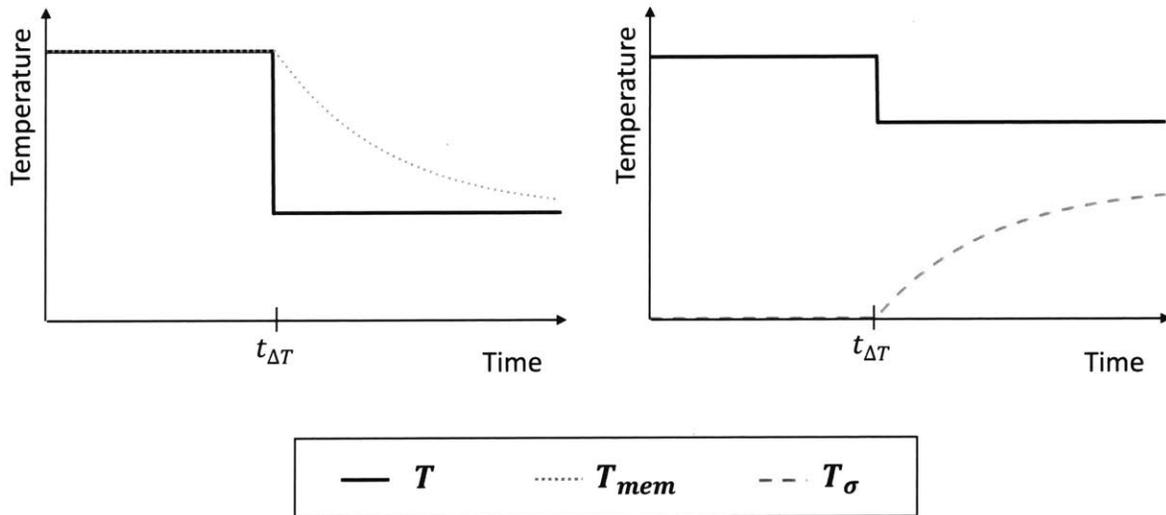
Figure 3. 7. Biological stress caused by temperature perturbance at time $t_{\Delta T}$. Left panel: An auxiliary variable termed T_{mem} was modeled with a first order delay kinetics with respect to the physical bioreactor temperature, T. Right Panel: The time integral of the difference between T_{mem} and T was used to model biological stress (T_σ).

Equation 3. 4 $\frac{dT_{mem}}{dt} = \tau_{temp} \cdot (T_{mem} - \mathbf{T})$

Equation 3. 5 $T_\sigma(t) = \int_0^t abs(T_{mem} - \mathbf{T}) dt$

Equation 3. 4 represents the generation of a memory temperature, that is, the temperature that the cells sense – and thus determines metabolism – and can be different to extracellular temperature. In addition to this, the integral of the difference between memory temperature and extracellular temperature was defined as the stress variable (Equation 3. 5). This term accounts for the memory term for perturbations in the system. The model has one parameter, τ_{temp}, which was calibrated (see calibration section).

Combined Effect

To account for the combined effect of temperature, pH and osmolarity a combined happiness function for each enzyme reaction was generated:

Equation 3. 6 $\mathcal{H}_j = \prod_p \left(1 - k_j^p \cdot (1 - \mathcal{H}_p) \right)$

where j stands for the specific enzyme reaction (i.e. GLN, ASN, GLC, O_2 or $ANTI$), p stands for the process variable (*i.e.* T, pH or Π), and k_j^p is a process variable-reaction specific parameter. Note that if k_j^p is 0, process variable p has no effect on reaction j, but sill allows for other process variables to affect the reaction.

3.3.1.2.   *Reaction Kinetics*

Mechanistic   reaction   kinetics   for   the   controlling   reactions   (*i.e.* $GLN, ASN, GLC, O_2, Maintainance$ and $ANTI$) were obtained from literature (Nolan & Lee, 2011). The effect of process variables was added to each reaction kinetic, modifying either the upper or lower bound:

Upper bound

- Glucose Uptake Rate:

Equation 3. 7

$$v_{GLC} = \mathcal{H}_{GLC}(T, pH, \Pi) \cdot \frac{\frac{v_{max_{GLC}} \cdot [GLC]}{Km_{GLC}}}{\left(1 + \frac{[LAC]}{K_{LAC/GLC}}\right)\left(1 + \frac{[GLC]}{Km_{GLC}}\right)}$$

- Glutamine Uptake Rate:

Equation 3. 8

$$v_{GLN} = \mathcal{H}_{GLN}(T, pH, \Pi) \cdot \frac{\left(v_{max_{GLN_f}} \cdot \left(\frac{[GLN]}{Km_{GLN}}\right) - v_{max_{GLN_r}} \cdot \left(\frac{[GLU]}{Km_{GLN}}\right) \cdot \left(\frac{[NH_3]}{Km_{GLN}}\right)\right)}{1 + \left(\frac{[GLN]}{Km_{GLN}}\right) + \left(\frac{[GLU]}{Km_{GLN}}\right) + \left(\frac{[NH_3]}{Km_{GLN}}\right) + \left(\frac{[GLU]}{Km_{GLN}}\right) \cdot \left(\frac{[NH_3]}{Km_{GLN}}\right)}$$

- Asparagine Uptake Rate:

Equation 3. 9

$$v_{ASN} = \mathcal{H}_{ASN}(T, pH, \Pi) \cdot \frac{\left(v_{max_{ASN_f}} \cdot \left(\frac{[ASN]}{Km_{ASN}}\right) - v_{max_{GLN_r}} \cdot \left(\frac{[ASP]}{Km_{ASN}}\right) \cdot \left(\frac{[NH_3]}{Km_{ASN}}\right)\right)}{1 + \left(\frac{[ASN]}{Km_{ASN}}\right) + \left(\frac{[ASP]}{Km_{ASN}}\right) + \left(\frac{[NH_3]}{Km_{ASN}}\right) + \left(\frac{[ASP]}{Km_{ASN}}\right) \cdot \left(\frac{[NH_3]}{Km_{ASN}}\right)}$$

- Oxygen Uptake Rate

Equation 3. 10

$$v_{O_2} = \mathcal{H}_{O_2}(T, pH, \Pi) \cdot \frac{q_{O_{2max}} \cdot [O2]}{Km_{O_2} + [O2]}$$

It is noteworthy to mention that oxygen was assumed in excess, thus Equation 3. 10 simplifies to:

Equation 3. 11 $\quad DO \gg K_{m_{O_2}} \rightarrow v_{O_2} = \mathcal{H}_{O_2}(T, pH, \Pi) \cdot q_{O_{2max}}$

Lower bound

- Maintenance

A term for non-growth associated energy consumption (maintenance energy) was added to the model, it was modeled to account its dependency on stress and inhibitory by-product concentration:

Equation 3. 12 $$v_{atp} = m_{atp} \cdot \left(1 + \left(\frac{STRESS}{k_{stress_{maint}}}\right)\right) + k_{maint_0} \cdot \left(\frac{[NH_3]}{[NH_3] + K_{maint}^{NH_3}}\right) \cdot \left(\frac{[LAC]}{[LAC] + K_{maint}^{LAC}}\right)$$

- Antibody production

Product Formation was modeled proportional to cellular maintenance, and inhibited by lactate concentration, as follows:

Equation 3. 13 $$v_{ANTI} = \mathcal{H}_{ANTI}(T, pH, \Pi) \cdot \beta \cdot \left(\frac{1}{1 + \frac{LAC}{K_{i_{LAC}}^{ANTI}}}\right) \cdot q_{atp}$$

Therefore, the increase in stress generates higher maintenance requirements, which in turn are associated with higher specific product formation rates, which has been observed empirically, and is in fact the objective of the temperature shift.

Table 3. 2. Model Parameters of Kinetic Study.

MODELING PROCESS	PARAMETERS
Temperature Dependence	$d_{temp}; T_{max}; T_{min}$
Ph Dependence	$pH_{min}; pH_{max};$
Osmolarity Dependence	$\pi_{k_{0.5}}; \pi_{slope}$
Process Kinetic Dependency (Happiness)	$k^T_{gln}; k^{pH}_{gln}; k^{\Pi}_{gln}; k^T_{asn}; k^{pH}_{gln}; k^{\Pi}_{gln}; k^T_{glc}; k^{pH}_{glc}; k^{\Pi}_{glc}; k^T_{O_2}; k^{pH}_{O_2}; k^{\Pi}_{O_2}; k^T_{ANTI}; k^{pH}_{ANTI}; k^{\Pi}_{ANTI}$
Glucose Kinetics	$v_{max_{GLC}}; K_{m_{GLC}}; K_{LAC/GLC}$
Glutamine Kinetics	$v_{max_{GLN_f}}; v_{max_{GLN_r}}; K_{m_{GLN}}_{\overline{GLN}}; K_{m_{GLN}}_{\overline{GLU}}; K_{m_{GLN}}_{\overline{NH3}}$
Asparagine Kinetics	$v_{max_{ASN_f}}; v_{max_{GLN_r}}; K_{m_{ASN}}_{\overline{ASN}}; K_{m_{ASN}}_{\overline{ASP}}; K_{m_{ASN}}_{\overline{NH3}}$
Oxygen Kinetics	$q_{O_{2max}}$
Maintenance	$m_{atp}; k_{stress_{maint}}; k_{maint_0}; K^{NH_3}_{maint}; K^{LAC}_{maint}$
Product Formation	$\beta; K^{ANTI}_{i_{LAC}}$
Stress	k_{temp}

3.3.2. Constraint-Based Modeling Study

The selected network was adapted from (Nolan & Lee, 2011). The network consists of the main energetic carbon reactions as well as selected amino acids that are relevant in terms of redox balance (Figure 3. 8). The reaction stoichiometrics were determined and lumped accordingly to have a simplified energetic network (Nolan & Lee, 2011). The network consisted of 55 reactions and 47 metabolites (Table 3. 3).

This network was selected as the first step approach for several reasons:

- Simplicity. If properly adapted the network can incorporate central energetic metabolism and include the Warburg effect.

- Main sources of energy: Glucose, Glutamine and Asparagine are the main substrates that determine the metabolic state of metabolism.

- Ammonia, Lactate and CO2 are byproducts of the system and can be further incorporated in the inhibition kinetics as needed.

Figure 3. 8. Metabolic Network of the system (Adapted from Nolan & Lee (2011)).

Differently from the approach executed by Nolan of a fully determined network (rank 0 problem), the present work proposes network optimization, particularly a parsimonious flux balance analysis (Lewis et al., 2010) The fluxes are determined such that the overall networks maximize reproduction while reducing overall energy expenditure (Lewis et al., 2010), according to the following equation:

Equation 3. 14 $Min \sum q$

$$st: max\ \mu$$

$$st: S \cdot q = 0$$

where q is the velocity vector of the reactions in the system, μ is the biomass reaction (reproduction) and S is the stoichiometric matrix of the system. Note that v refers to the calculated rate for the boundary condition while q refers to the variable rate, whose optimal value is obtained after solving the optimization problem in 3.14.The boundaries of the system were set according to the maximum and minimum velocities determined by the kinetics developed by (Nolan & Lee, 2011).

Additionally, the biomass reaction was divided into two reactions: a purely growth associated reaction and a non-growth associated energy consumption, or maintenance reaction.

14 boundaries were dynamically changed at each iteration of the FBA. Of the 14 reactions, 9 reaction rates correspond to the intra/extracellular transmembrane reactions: their upper bounds were modified to account for the availability of the substrate in the system. This was done by updating their boundaries with the instantaneous concentration in the bioreactor, or the concentration in the feed media. The remaining 5 boundaries of the network were dynamically controlled by mechanistic enzyme kinetics as explained in the previous section (in red in Table 3. 4). These mechanistic reactions consider the effect of environmental factors, thus allowing the cells to "sense" their environment. The controlled enzyme kinetics are the following: glutamine, glucose, asparagine, maintenance ATP, and product formation. The first three reactions consist of energy consumption and thus the kinetic equations affect the upper bound of the reaction. The latter two modify the lower bound of the reaction, given that they account for energetic costs of metabolism. Note that at each step of the simulation, these 14 boundaries will be dynamically changed based on the inputs and state variables of the system. This will affect the consumption/production rates, which will in turn change the mass balance equation.

To avoid unfeasible solutions, a three-level optimization function was developed. In the top layer, the objective is to maximize biomass while minimizing overall energy expenditure (Equation 3. 14). The addition of a stress-dependent variable maintenance energy can make the system not have enough resources for growth (i.e., positive μ), thus if the first solution was unfeasible a second level's objective was to

maximizes maintenance energy. This maintenance energy will be less than the required energy of maintenance, so that difference was used to calculate an additional death rate due to starvation ($q_{X_{d_{starv}}}$):

Equation 3. 15 $\qquad q_{X_{d_{starv}}} = |min\{0, \ v_{maint} - q_{maint}\}|$

Finally, if this second objective was not feasible, the solution was halted as infeasible and all the fluxes were set equal to zero.

The main parameters of the metabolic model include the network, the biomass and product reaction formulae, the boundary conditions of the system, and the objective function. The network is a simplified metabolic network that includes the main energetic components, that can model the main components of process development. The product reaction and stoichiometric was calculated from the aminoacidic sequence (proprietary data not disclosed). The biomass formula as well as the boundaries of the system that are not dynamically modified were extracted from (Nolan & Lee, 2011). The objective function, as previously explained, was set according to the procedure developed by (Lewis et al., 2010).

Table 3. 3. Metabolite ID, names and empirical formula for the network.

METABOLITE ID	COMPARTMENT	METABOLITE NAME	METABOLITE FORMULA
P_{I_c}	Cytosol	Phosphate	HO_4P
H_e	Extracellular	H+	H
H_2O_e	Extracellular	H2O	H_2O
CYS_{L_e}	Extracellular	L-Cysteine	$C_3H_7NO_2S$
CC_e	Extracellular	C-C	$C_6H_{13}N_2O_4S_2$
GLU_{L_e}	Extracellular	L-Glutamate	$C_5H_8NO_4$
GLN_{L_e}	Extracellular	L-Glutamine	$C_5H_{10}N_2O_3$
CO_{2_e}	Extracellular	CO2	CO_2
LAC_{L_e}	Extracellular	L-Lactate	$C_3H_5O_3$
GLC_{D_e}	Extracellular	D-Glucose	$C_6H_{12}O_6$
ALA_{L_e}	Extracellular	L-Alanine	$C_3H_7NO_2$
NH_{3_e}	Extracellular	Ammonia	H_3N
ASN_{L_e}	Extracellular	L-Asparagine	$C_4H_8N_2O_3$
ASP_{L_e}	Extracellular	L-Aspartate	$C_4H_6NO_4$
SER_{L_e}	Extracellular	L-Serine	$C_3H_7NO_3$
GLY_e	Extracellular	Glycine	$C_2H_5NO_2$
$ANTI_e$	Extracellular	Antibody	$C_{4.81}N_{1.28}O_{2.51}H_{9.44}$
$BIOM_e$	Extracellular	Biomass	$C_{5.68}N_{1.40}O_{2.92}H_{11.49}P_{0.09}$
NAD_M	Mithocondria	Nicotinamide adenine dinucleotide	$C_{21}H_{26}N_7O_{14}P_2$
NAD_c	Cytosol	Nicotinamide adenine dinucleotide	$C_{21}H_{26}N_7O_{14}P_2$
H_c	Cytosol	H+	H
H_2O_c	Cytosol	H2O	H_2O
ATP_c	Cytosol	ATP	$C_{10}H_{12}N_5O_{13}P_3$
ADP_c	Cytosol	ADP	$C_{10}H_{12}N_5O_{10}P_2$
$NADH_c$	Cytosol	Nicotinamide adenine dinucleotide - reduced	$C_{21}H_{27}N_7O_{14}P_2$
FAD_c	Cytosol	Flavin adenine dinucleotide oxidized	$C_{27}H_{31}N_9O_{15}P_2$
$NADH_c$	Mithocondria	Nicotinamide adenine dinucleotide - reduced	$C_{21}H_{27}N_7O_{14}P_2$
$FADH_{2_c}$	Cytosol	Flavin adenine dinucleotide reduced	$C_{27}H_{33}N_9O_{15}P_2$
GLU_{L_c}	Cytosol	L-Glutamate	$C_5H_8NO_4$
O_{2_c}	Cytosol	O2	O_2
NH_{3_c}	Cytosol	Ammonia	H_3N
SER_{L_c}	Cytosol	L-Serine	$C_3H_7NO_3$
GLY_c	Cytosol	Glycine	$C_2H_5NO_2$
GLN_{L_c}	Cytosol	L-Glutamine	$C_5H_{10}N_2O_3$
CC_c	Cytosol	C-C	$C_6H_{13}N_2O_4S_2$
ASP_{L_c}	Cytosol	L-Aspartate	$C_4H_6NO_4$
ASN_c	Cytosol	L-Asparagine	$C_4H_8N_2O_3$
ALA_c	Cytosol	L-Alanine	$C_3H_7NO_2$
LAC_{L_c}	Cytosol	L-Lactate	$C_3H_5O_3$
$G6P_c$	Cytosol	D-Glucose 6-phosphate	$C_6H_{11}O_9P$
$ANTI_c$	Cytosol	Antibody	$C_{4.81}N_{1.28}O_{2.51}H_{9.44}$
$BIOM_c$	Cytosol	Biomass	$C_{5.68}N_{1.40}O_{2.92}H_{11.49}P_{0.09}$
CYS_{L_c}	Cytosol	L-Cysteine	$C_3H_7NO_2S$
OAA_c	Cytosol	Oxaloacetate	$C_4H_2O_5$
MAL_{L_c}	Cytosol	L-Malate	$C_4H_4O_5$
PYR_c	Cytosol	Pyruvate	$C_3H_3O_3$
AKG_c	Cytosol	2-Oxoglutarate	$C_5H_4O_5$

Table 3. 4. Reaction ID, stoichiometrics and bounds of the Metabolic Network. Note that excess reactions are included in the table to facilitate the optimization procedure (these reactions are not illustrated in Figure 3. 8).

ID	REACTION	BOUNDS
q_{O2}	$\rightarrow O_{2_c}$	$(0, v_{O_2})$
q_{NH3}	$NH_{3_c} \rightarrow NH_{3_e}$	$(0, 100000)$
q_{SER}	$SER_e \rightarrow SER_c$	$(0, v_{av})$
q_{GLY}	$GLY_e \leftrightarrow GLY_c$	$(-100000, v_{av})$
q_{GLN}	$GLN_e \leftrightarrow GLN_c$	$(-100000, v_{av})$
q_{CC}	$CC_e + GLU_c \rightarrow CC_c + GLU_e$	$(0, v_{av})$
q_{ASP}	$ASP_e \leftrightarrow ASP_c$	$(-100000, v_{av})$
q_{ASN}	$ASN_e \rightarrow ASN_c$	$(0, v_{av})$
q_{GLU}	$GLU_c \rightarrow GLU_e$	$(0, 308.75)$
q_{ALA}	$ALA_e \leftrightarrow ALA_c$	$(-100000, v_{av})$
q_{LAC}	$H_e + LAC_e \leftrightarrow H_c + LAC_c$	$(-100000, v_{av})$
q_{GLC}	$ATP_c + GLC_{D_e} \rightarrow ADP_c + G6P_c + H_c$	$(0, v_{av})$
q_P	$ANTI_c \rightarrow ANTI_e$	$(0, 100000)$
q_X	$BIOM_c \rightarrow BIOM_e$	$(0, 100000)$
q_{maint}	$ATP_c + H_2O \rightarrow ADP_c + H_c + P_{I_c}$	$(v_{maint}, 1000.0)$
q_{19}	$0.045\,ALA_c + 0.038\,ASN_c + 0.038\,ASP_c + 4.032\,ATP_c + 0.027\,CYS_c + 0.045\,GLN_c + 0.048\,GLU_c + 0.128\,GLY_c$ $+ 10.43\,H_2O + 0.101\,SER_c \rightarrow ANTI_c + 4.032\,ADP_c + 5.3\,H_c + 4.032\,P_{I_c}$	$(v_{ANTI}, 525)$
q_{18}	$ASN_c + H_2O \leftrightarrow ASP_c + H_c + NH_{3_c}$	$(-20, v_{ASN})$
q_{17}	$GLN_c + H_2O \leftrightarrow GLU_c + H_c + NH_{3_c}$	$(-200, v_{GLU})$
q_{16}	$3.0\,ADP_c + G6P_c + 2.0\,NAD_c + 2.0\,P_{I_c} \rightarrow 3.0\,ATP_c + 2.0\,H_2O + H_c + 2.0\,NADH_c + 2.0\,PYR_c$	$(0.0, v_{GLC})$
q_{15}	$0.5\,FAD_c + 0.5\,H_c + NADH_c \rightarrow 0.5\,FADH_{2_c} + 0.5\,NAD_c + 0.5\,NADH_m$	$(0, 100000)$
q_{14}	$2.0\,CYS_e + 3.0\,H_e + O_{2_c} \rightarrow CC_e + 2.0\,H_2O$	$(0, 100000)$
q_{13}	$0.084\,ALA_c + 0.041\,ASN_c + 0.08\,ASP_c + 3.78\,ATP_c + 0.026\,CYS_c + 0.004\,FAD_c + 0.452\,G6P_c + 0.087\,GLN_c$ $+ 0.056\,GLY_c + 3.78\,H_2O + 0.639\,NAD_c + 0.427\,OAA_c + 0.096\,SER_c$ $\rightarrow BIOM_c + 3.78\,ADP_c + 0.004\,FADH_{2_c} + 0.008\,GLU_c + 3.705\,H_c + 0.445\,MAL_c + 0.639\,NADH_c$ $+ 0.452\,P_{I_c} + 0.209\,PYR_c$	$(0, 2500)$
q_{12}	$1.5\,ADP_c + FADH_{2_c} + 1.5\,H_c + 0.5\,O_{2_c} + 1.5\,P_{I_c} \rightarrow 1.5\,ATP_c + FAD_c + 2.5\,H_2O$	$(0, 100000)$
q_{11}	$2.5\,ADP_c + 3.5\,H_c + NADH_m + 0.5\,O_{2_c} + 2.5\,P_{I_c} \rightarrow 2.5\,ATP_c + 3.5\,H_2O + NAD_m$	$(0, 100000)$
q_{10}	$CC_c + NADH_c \rightarrow 2.0\,CYS_c + NAD_c$	$(0, 150)$
q_9	$CO_{2_c} + H_c + NADH_c + NH_{3_c} + SER_c \leftrightarrow 2.0\,GLY_c + H_2O + NAD_c$	$(-25, 225.0)$
q_8	$AKG_c + ASP_c \rightarrow GLU_c + OAA_c$	$(0, 261.25)$
q_7	$AKG_c + 2.0\,H_c + NADH_m + NH_{3_c} \leftrightarrow GLU_c + H_2O + NAD_m$	$(-261.25, 350)$
q_6	$MAL_c \rightarrow CO_{2_c} + H_c + PYR_c$	$(0, 100000)$
q_5	$MAL_c + NAD_m \rightarrow H_c + NADH_m + OAA_c$	$(0, 100000)$
q_4	$ADP_c + AKG_c + FAD_c + H_2O + NAD_m + P_{I_c} \rightarrow ATP_c + CO_{2_c} + FADH_{2_c} + MAL_c + NADH_m$	$(0, 100000)$
q_3	$H_2O + 2.0\,NAD_m + OAA_c + PYR_c \rightarrow AKG_c + 2.0\,CO_{2_c} + H_c + 2.0\,NADH_m$	$(0, 100000)$
q_2	$GLU_c + PYR_c \leftrightarrow AKG_c + ALA_c$	$(-150, 350)$
q_1	$H_c + NADH_c + PYR_c \leftrightarrow LAC_c + NAD_c$	$(-11000, 13200)$

3.3.3. Mass Balance Study

The solution obtained from the constraint-based study explained previously was used to update a set of mass balance differential equations describing the bioreactor dynamics. The mass balance study captures the variables that a user can manipulate to modify the response of the biological process. Such variables are temperature, media flows, filter, batch concentration, initial seed density, amongst others.

The extracellular exchange rates (q_{EX}) obtained from the constraint-based study, serve as inputs for the Mass Balance Study. For each substrate/product M, the mass balance for the system is:

Equation 3. 16 $\quad\quad \frac{d[M](t)}{dt} = q_M \cdot [X] + \frac{1}{V} \cdot \left(\sum_i F_{IN_i} \cdot [M]_{F_i} \right) - \Phi_M(t) \cdot \frac{1}{V} \left(\sum_i F_{IN_i} \cdot [M](t) \right)$

where $[M]$ is the instantaneous concentration of the substrate or product of interest, q_M is the net specific consumption/production rate of product/substrate M (obtained from the solution of the FBA), $[X]$ is the instantaneous biomass concentration). F_{IN_i} represent the ith inflow to the bioreactor in units of volume per time. V is the bioreactor liquid volume (note that for the derivation of this formula, the volume needs to be constant, and consequently $\sum F_{in} = \sum F_{out}$). M_{F_i} is the concentration of substrate/product M in the media solution i. $\Phi_M(t)$ is the instantaneous permeability of metabolite M through the outlet filter.

In FA-01 a 2-inflow system was defined and media solution 1 was defined as the feeding media, while media solution 2 was defined as a concentrated solution of glucose added as needed to the system.

Equation 3. 17 $\quad\quad \Phi_M(t) = \begin{cases} 0 & if\ M = X \\ \Phi_P & if\ M = P \\ 1 & elsewise \end{cases}$

The case where $\Phi_M(t) = 1$ in Equation 3. 17 applies for all the modeled substrate/products (referred to as Λ) including glucose (GLC), lactate (LAC), carbon dioxide (CO2), glutamine (GLN),

glutamate (GLU), Cystine (CC), Cysteine (CYS), glycine (GLY), serine (SER), aspartic acid (ASP), asparagine (ASN), ammonia (NH3) and alanine (ALA). For all of these cases, Equation 3. 16 reduces to:

Equation 3. 18 $\quad \frac{d[\Lambda](t)}{dt} = q_\Lambda \cdot [X] + \frac{F_{IN_1}}{V} \cdot \left([\Lambda]_{F_1} - [\Lambda]\right) + \frac{F_{IN_2}}{V} \cdot \left([\Lambda]_{F_2} - [\Lambda]\right)$

Note that F_{IN_2} is a concentrated solution of glucose, thus except for glucose all $[\Lambda]_{F_2}$ are null.

For the biomass balance, in addition to cellular growth, a constitutive death rate (k_{death}) and a starvation death rate (due to unfeasible pFBA solution) were incorporated into the net production rate, μ, as follows:

Equation 3. 19 $\quad \mu = q_X - k_{death} - q_{X_{d_{starv}}}$

Given that neither feeding media contain live cells, plugging Equation 3. 17 and Equation 3. 19 into Equation 3. 16 the biomass balance reduces to:

Equation 3. 20 $\quad \frac{d[X](t)}{dt} = \mu \cdot [X]$

The dead biomass balance was calculated according to the following equation:

Equation 3. 21 $\quad \frac{dx_{dead}}{dt} = \left(k_{death} - min\left\{0, q_X - q_{X_{d_{starv}}}\right\}\right) \cdot [X]$

where x_{dead} is the dead cell concentration. Note that the dead cell concentration increases with a constant k_d rate and an additional term to account for an infeasible solution of the pFBA.

For the product mass balance a more complex model is required. Generally, in perfusion bioreactors to concentrate and selectively recover the product the outlet media filter is changed towards the end of the

culture (at t= $t_{\Delta F}$). Initially, the outlet filter has low porosity, retaining both biomass and product. The second filter has a slightly higher porosity and retains biomass but allows the product to flow through. This step in the process facilitates the concentration of the product and reduces cost in downstream purification by removing the biomass. Nonetheless, it is empirically observed (data not shown) that there is a sieving effect of this second filter, and that the concentration of the product in the bioreactor is different from the concentration of the harvested volume. To account for this effect a sieving variable was accounted for in the model. The formula assumes that the filter efficiency begins at a certain efficiency and reduces in time with the cumulative permeated volume. Thus, the overall permeability of P in the culture is defined by:

Equation 3. 22
$$\Phi_P = \begin{cases} 0 & for\ t < t_{\Delta F} \\ \Phi_{P_0} \cdot 10^{-v_P \cdot V_{Harv}(t)} & for\ t \geq t_{\Delta F} \end{cases}$$

where Φ_P is the filter efficiency for the product, Φ_{P_0} is the initial filter efficiency, and v_P is the sieving parameter. $V_{Harv}(t)$ is the instantaneous harvested volume, defined as the cumulative outflow volume after the filter change:

Equation 3. 23
$$V_{Harv} = \int_{t_{\Delta F}}^{t} (F_{IN_1} + F_{IN_2}) dt$$

By creating a dummy variable $\chi(t)$ (which equals 1 when $t \geq t_{\Delta F}$ and 0 elsewise) Equation 3. 23 can be written in differential terms as:

Equation 3. 24
$$\frac{dV_{Harv}}{dt} = (F_{IN_1} + F_{IN_2}) \cdot \chi(t)$$

Plugging in Equation 3. 22 into Equation 3. 16, and considering null addition of product in feeding media, the mass balance for the product reduces to:

Equation 3. 25
$$\frac{d[P]}{dt} = q_P \cdot [X] - \chi(t) \cdot \Phi_{P_0} \cdot \frac{F_{IN_1} + F_{IN_2}}{V} \cdot [P] \cdot 10^{(-v_P \cdot V_{Harv})}$$

Given that the product is recovered from the harvested volume, a separate mass balance of the product effluent was added to the model:

Equation 3. 26 $\frac{dP_{Harv}}{dt} = (F_{IN_1} + F_{IN_2}) \cdot \Phi_P \cdot [P]$

where P_{Harv} is the cumulative mass of the harvested product and the harvest concentration of product is defined by:

Equation 3. 27 $[P]_{Harv} = \frac{P_{Harv}}{V_{Harv}}$

Finally, the biologically sensed temperature T_{mem}, necessary for the kinetic study, is calculated according to:

Equation 3. 28 $\frac{dT_{mem}(t)}{dt} = \tau_T \cdot (T(t) - T_{mem}(t))$

where τ_T is the adaptation parameter and $T(t)$ is the physical temperature in the bioreactor. And the stress variable (T_σ) is calculated according to:

Equation 3. 29 $\frac{dT_\sigma}{dt} = |T - T_{mem}|$

Table 3. 5. Input Variables for the Mass Balance Study.

Input Variable	Type	Variable
Initial Concentrations	Variable	$BIOM_0; GLC_0; LAC_0; GLN_0; GLU_0; CC_0; CYS_0;$ $GLY_0; SER_0; ASP_0; ASN_0; NH3_0; ALA_0$
Feeding Media	Variable	$BIOM_{IN}; GLC_{IN}; LAC_{IN}; GLN_{IN}; GLU_{IN}; CC_{IN}; CYS_{IN};$ $GLY_{IN}; SER_{IN}; ASP_{IN}; ASN_{IN}; NH3_{IN}; ALA_{IN}$
Feed Flow	Variable	$F_{IN,1}$
Glucose Shot	Variable	GLC_{IN}
Shot Flow	Variable	$F_{IN,2}$
Temperature	Variable	$T_1; T_2; T_{\Delta T}$
Filter Change	Variable	$t_{\Delta F}$
Ph	Variable	pH
Total Time	Variable	T_{Harv}
Sieving Effect	Parameter	v_P

3.4. FA-01 Calibration

Kinetic Study: Given the complexity of the system and non-convex nature of the minimization problem, a bottom-up approach was selected to identify the model. This means that different datasets were used to identify subunits of the model to collectively integrate them into the larger model. For this reason, the parameters for the happiness function parameters were calibrated from experimental data obtained from the literature. Figure 3. 9 illustrates the fractional happiness function calibrated from three different data sets obtained from the literature (Furukawa & Ohsuye, 1998a; Kim & Lee, 2007; Zhu et al., 2008).
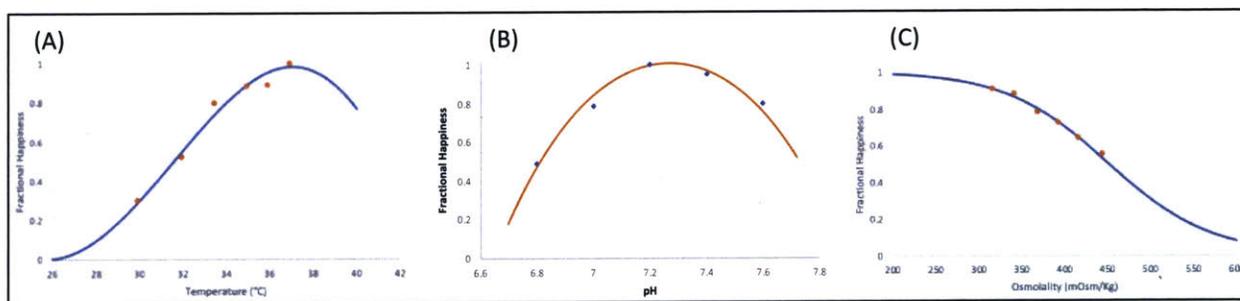

Figure 3. 9. Process Variable Effect on 'Happiness'. [A] Temperature effect on 'happiness'. Parameters $d_{temp} = 0.09$; $T_{max} = 44.8C$; $T_{min} = 25.7C$; $\max(happy_T) = 33.5$. Data extracted from (Furukawa & Ohsuye, 1998b). [B] pH effect on 'happiness'. Parameters $pH_{min} = 6.65$; $pH_{max} = 7.89$; $\max(happy_{pH}) = 0.578$. Data calibrated from (Kim & Lee, 2007). [C] Osmolarity effect on 'happiness'. Parameters $\pi_{k_{0.5}} = 450$; $\pi_{slope} = 60$; $\max(happy_\pi) = 0.6$. Data extracted from (Zhu et al., 2008)

Constraint-Based Study: The Constraint-Based Model Study was solved using pFBA solver in CobraPy (Ebrahim, Lerman, Palsson, & Hyduke, 2013) with the glpk software ("GLPK - GNU Project," 2012) to solve the linear problem.

Mass Balance Study: The SciPy integrate package (Oliphant, 2007) was used to solve ODE in the *Mass Balance Study*, *i.e.* Equations 3. 18, 20-21, 24-29. The Real-valued Variable-coefficient Ordinary Differential Equation solver, with fixed-leading-coefficient implementation ('*vode*') integrator was selected with the Backward Differentiation Formulas ('*bdf*') method, relative tolerance of 1E-3, max step size equal to 0.7 days, order 15 and maximum number of steps to 500.

Experimental Data: Perfusion bioreactors with tangential flow filtration experiments for the Amgen pipeline molecule were used for calibration and validation purposes in commercial 3 Liter bioreactors. An Amgen proprietary batch media was sterilized into the bioreactor. pH and Temperature were adjusted to proprietary settings (pH_{set}, T_{set_1}). The culture was started by inoculating an initial seed density (X_0) of an AMGEN proprietary strain genetically modified to produce a therapeutic antibody. This is hereafter referred to as time 0. After a certain time of culture (at t_{perf}) a proprietary perfusion media was added to the bioreactor at time-varying flowrate ($F_{1_{in}}$) in a feedforward fashion. To control the volume of the bioreactor, a scale was used to estimate the volume and control the action of a pump to modulate the outflow media. The outflow media has a tangential flow filter that selectively permeates cells and/or the antibody, at $t_{\Delta f}$ the filter was changed from F_1 to F_2. F_1 retains both cells and product and F_2 retains cells and perfusion of the antibody. At $t_{\Delta T}$ the bioreactor temperature was changed from T_1 to T_2, to purposefully induce a stress response and shift metabolism towards product formation. A concentrated solution of glucose ($[GLC]_{F_2}$) was added as-needed to the bioreactor at a fixed flow (F_2). After the filter change perfused media was stored in a sterile container for further downstream purification. The experiment was ended at t_{harv}.

Analytical samples were taken daily from the bioreactor. Harvest samples were taken at $t_{\Delta f}$, $t_{\Delta T}$ and t_{harv}. Additionally, at t_{harv} a perfusion outflow sample was taken to calculate filter sieving. Glucose, Asparagine, Glutamine, Lactate, ammonia and antibody concentrations were estimated using liquid chromatography. Four different experimental data-sets were used from a design of experiment data-set in process development. Two of these correspond to the central point, and are different batches of the exact process set-up, the third corresponds to an early temperature shift experiment and the fourth corresponds to a data-set of a delayed temperature shift experiment. The central point data-set was used for calibration purposes, while the remaining three were used for validation purposes.

Cost Function and Optimization Problem: With initial estimates of the parameters, initial process conditions and defined operational set-up FA-01 was used to obtain a simulated trajectory $C_k^{sim}(t)$. This trajectory was compared with experimental data by a normalized residual squared error cost function. A set of parameters (ρ) was estimated in order to minimize the cost function:

Equation 3. 30
$$\arg\min_{\rho} \overrightarrow{f_{(\rho)}} := \sum_t \left(\frac{C_k^{sim}(t) - C_k^{exp}(t)}{C_k^{exp}(t)} \right)^2$$

where $C_k^{sim}(t)$ is the instantaneous simulated concentration of metabolite k. and $C_k^{exp}(t)$ is the instantaneous experimental concentration of substrate/product k. The variables used for calibration purposes are glucose, glutamine, glutamate, product in bioreactor, product in permeate, product in the harvest, ammonia and lactate. The set of calibrated parameters (ρ) is explained in the Optimization Procedure section.

Three different optimization methods were used to calibrate the parameters. Feasible candidate optima where explored by two global optimization techniques: Differential Evolution ((Oliphant, 2007)) and a Non Dominated Sorted Genetic Algorithm (NSGA-II) (Biscani & Izzo, 2018). The former is a univariate optimization technique, while the latter is a global optimization technique that has the added

benefit of parallelization and has a multi-objective cost function. Given the intensive memory requirements, once feasible candidates where found the local, multivariate Levenberg-Marquardt (SciPy: Least Square Method - (Oliphant, 2007)) was used to find the optimum.

Table 3. 6. Optimization Methods.

	Local Optimization	Global Optimization
Univariate ($\sum_k \vec{f_k}$)	-	Differential Evolution
Multivariate ($\vec{f_k}$)	Levenberg Marquardt	NSGA-II

Optimization Procedure: The commercial nature of the data set, and the dynamic nature of the process, adds an extreme challenge for calibration purposes. For this reason, the experimental data was divided into two pieces, a pre-shift data set (Calibration Part I), and a post-shift data set (Calibration Part II); as illustrated in Figure 3. 10. Calibration Part I, consisted of an initial batch phase plus a perfusion phase and the calibrated parameters were related to the kinetics of glucose, oxygen and non-growth-related energy (maintenance). In this phase the Warburg effect was correctly characterized as seen by the decrease in lactate concentration. Calibration Part II consisted in the characterization of the effect of temperature on the controlled kinetics, the effect of stress on the energy of maintenance, and the harvest and sieving effect. It is important to mention that the input variables are not disclosed due to proprietary reasons.
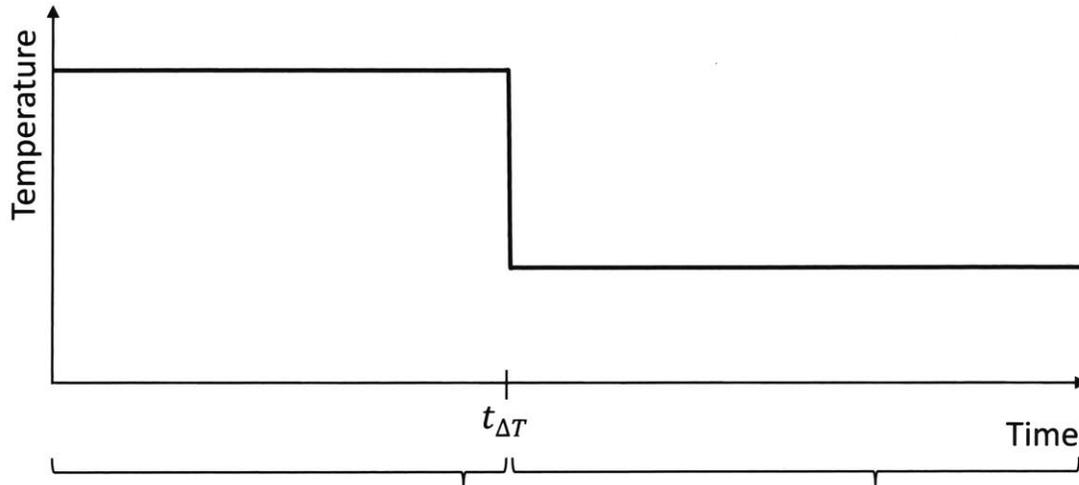
Calibration Part	Part I	Part II
Physiological/Physical Effect	Warburg, Maintenance, Uptake	Stress, Sieving
Calibrated Parameter	$v_{max_{GLC}}$; $K_{m_{ASN}}$; $q_{o_{2max}}$; m_{atp}; k_{maint_0}	k_{gln}^T; k_{asn}^T; k_{glc}^T; $k_{O_2}^T$; k_{ANTI}^T; $k_{stress_{maint}}$; k_{temp}; $K_{sieving}$

Figure 3. 10. Calibration Methodology. Experimental set up division at temperature shift time ($t_{\Delta T}$) into Part I and II. Description of physiological and physical effects captured in each part and calibrated parameters.

4. RESULTS

The present section describes the main results regarding FA-01. It is important to mention that the main results of this work come from the development of the FA and the software architecture that enables it. Nonetheless to validate them, FA-01 was selected as a case-model, and this section describes the: (1) calibration of FA-01 with experimental data; and (2) validating the optimized model different data-sets.

4.1. Calibration

Figure 4. 1 shows the calibration results for the control experiment, following the methodology described in section 3.4, the parameters are summarized in Table 4. 1. Note that the experiment is able to reproduce the production and perfusion of the inhibitory byproducts (lactate and ammonia), the biomass effect (both before and after the temperature shift) as well as the glucose consumption. The model also predicts the product formation both in the bioreactor as well as in the harvest.
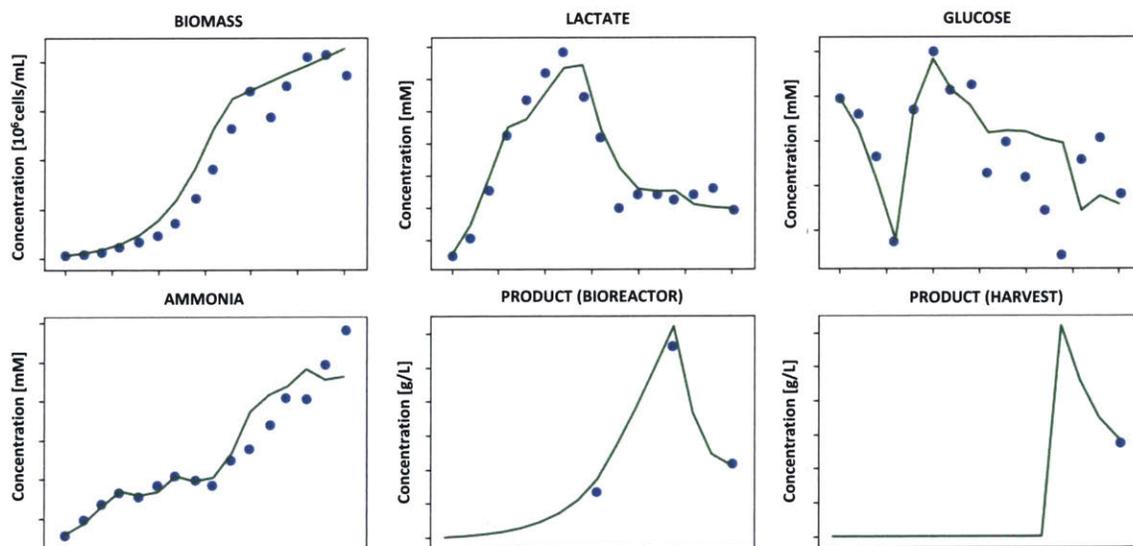

Figure 4. 1. Calibration results for control experiment. Experimental data shown by blue points and simulated data by FA-01 in green lines. Simulation and experimental data are in the same scale, with x- and y- axis deleted to protect proprietary information.

It is important to discuss that the results of the parameters are in accordance to those observed in literature. The temperature dependence on the kinetics updates are consistent with those observed by Nolan & Lee (2011). The calibrated strain showed a slower glucose consumption rate and more tolerance to lactate levels than those observed by the same authors.

It is noteworthy to mention that $q_{O_{2max}}$ and m_{atp} structurally have a high degree of correlation. $q_{O_{2max}}$ describes the physiological maximum oxygen uptake rate and thus determines the internal energetic bottleneck. m_{atp} describes the non-growth associated energy, or the energy required for maintenance. Thus, energy conversion efficiency is correlated with energy requirements, and thus these parameters are structurally correlated and in a certain degree non-identifiable currently in FA-01. Using different experiments to measure the maximum oxygen uptake rate, or adding oxygen kinetics to FA-01 would allow to fix one of these parameters and properly estimate the second. Despite this, the maximum oxygen uptake rate was calibrated to approximately 3 pmol/cell/day, similar lower than that observed by Goudar (2011) of ~5 pmol/cell/day.

Figure 4. 1 shows that the model slightly overestimates biomass concentration as compared to experimental data. This can be explained by the underlying logic of the model used, and generally of FBAs, that perform metabolic mass balances and thus describe fluxes in terms of specific production/consumption rates, i.e., per unit mass of biomass. Given that the biomass measurements are in cells per volume, and usually cells increase in diameter in perfusion experiments, this principle is violated. For this reason, it is advisable to convert cellular concentration measurements to mass measurements. If measurements of the cell diameter are available together with cell-count, biomass can be estimated by coupling the measurements together, and thus account for the effect of varying cell diameter.

Table 4. 1. FA-01 Calibrated Parameters.

PARAMETER	VALUE	UNITS	SOURCE
d_{temp}	0.09	$[°C^{-1}]$	Calibrated from Furukawa & Ohsuye (1998b)
T_{max}	44.8	$[°C]$	Calibrated from furukawa & ohsuye (1998b)
T_{min}	25.7	$[°C]$	Calibrated from (furukawa & ohsuye, 1998b)
pH_{min}	6.65	$[-]$	Calibrated from Kim & Lee (2007)
pH_{max}	7.89	$[-]$	Calibrated from Kim & Lee (2007)
$\pi_{k_{0.5}}$	450	$[mOsm/Kg]$	Calibrated from Zhu et al. (2008)
π_{slope}	60	$[mOsm/Kg]$	Calibrated from Zhu et al. (2008)
k_{gln}^{T}	1.00	$[-]$	Calibration part ii
k_{gln}^{pH}	0	$[-]$	Omitted in study
k_{asn}^{T}	0.442	$[-]$	Calibration part ii
k_{asn}^{pH}	0	$[-]$	Omitted in study
k_{glc}^{T}	1.54	$[-]$	Calibration part ii
k_{glc}^{pH}	0	$[-]$	Omitted in study
$k_{O_2}^{T}$	-1.04	$[-]$	Calibration part ii
$k_{O_2}^{pH}$	0	$[-]$	Omitted in study
k_{ANTI}^{T}	1.61	$[-]$	Calibration part ii
k_{ANTI}^{pH}	0	$[-]$	Omitted in study
$v_{max_{GLC}}$	3112	$[nmol/10^6 cells \cdot day]$	Calibration part i
$K_{m_{GLC}}$	10	$[mM]$	Obtained from Nolan & Lee (2011)
$K_{LAC/GLC}$	28	$[mM]$	Obtained from Nolan & Lee (2011)
$v_{max_{GLN_f}}$	2200	$[nmol/10^6 cells \cdot day]$	Obtained from Nolan & Lee (2011)
$v_{max_{GLN_r}}$	200	$[nmol/10^6 cells \cdot day]$	Obtained from Nolan & Lee (2011)
$K_{m_{\frac{GLN}{GLN}}}$	2.5	$[mM]$	Obtained from Nolan & Lee (2011)
$K_{m_{\frac{GLN}{GLU}}}$	1	$[mM]$	Obtained from Nolan & Lee (2011)
$K_{m_{\frac{GLN}{NH3}}}$	1	$[mM]$	Obtained from Nolan & Lee (2011)
$v_{max_{ASN_f}}$	475	$[nmol/10^6 cells \cdot day]$	Obtained from Nolan & Lee (2011)
$v_{max_{GLN_r}}$	20	$[nmol/10^6 cells \cdot day]$	Obtained from Nolan & Lee (2011)
$K_{m_{\frac{ASN}{ASN}}}$	2.5	$[mM]$	Calibration part i
$K_{m_{\frac{ASN}{ASP}}}$	1	$[mM]$	Obtained from Nolan & Lee (2011)
$K_{m_{\frac{ASN}{NH3}}}$	2	$[mM]$	Obtained from Nolan & Lee (2011)
$q_{O_2 max}$	2861	$[nmol/10^6 cells \cdot day]$	Calibration part i
m_{atp}	8780	$[nmol/10^6 cells \cdot day]$	Calibration part i
$k_{stress_{maint}}$	0.43	$[-]$	Calibration part ii
k_{maint_0}	1.5E4	$[nmol/10^6 cells \cdot day]$	Calibration part i
$K_{maint}^{NH_3}$	14.39	$[mM]$	Adapted from (batt & kompala, 1989; pörtner & schäfer, 1996)
K_{maint}^{LAC}	310.8	$[-]$	Adapted from Batt & Kompala (1989) & Pörtner & Schäfer, (1996)
β	2.94E-02	$[-]$	Calibration part i
$K_{i_{LAC}}^{ANTI}$	30	$[-]$	Obtained from Nolan & Lee (2011)
τ_T	7.4	$[day^{-1}]$	Calibration part ii
v_P	0.0298	$[L^{-1}]$	Calibration part ii
Φ_{P_0}	1	$[-]$	Assumption

In terms of the stress variable, the calibration results showed that the biological temperature adaptation time (τ_{Temp}) was relatively fast, and reached the physical temperature in about 12 hours. This is similar to the observations by Nolan & Lee (2011), that based on flux data, estimated that the population of cells exhibit a reduction of enzymatic activity on the order of 10 hours (while bioreactor temperature dropped in 1 hour). Consequently, the stress variable reached an amount of 0.5 °C in the same time. The stress effect on the energy of maintenance (explained by the $k_{stress_{maint}}$ parameter) showed that post-temperature shifts the maintenance energy was doubled. This is consistent with the experimental data, seen by high nutrient consumption rate with limited growth rate after temperature shift. Whether the increase in energy consumption is due to a stress related variable or if it is associated with a higher energy expenditure due to lower temperature needs to be further investigated. Nonetheless, the model correctly captures this effect and is able to predict this change in a lumped manner.

It is important to mention, that normally a growth rate maximizing FBA approach would not solve for product formation (because it is not beneficial for growth). For this reason, fixing the lower bound with a mechanistic relation was used. Three different factors determine a higher specific product formation rate post-temperature shift: higher maintenance energy due to stress, higher productivities due to temperature effect ($\mathcal{H}_{ANTI}$); and lower lactate formation rates due to a higher $q_{O_{2max}}$ (see equation 3.13). The first two effects have some correlation in the parameters $k_{stress_{maint}}$ and k_{ANTI}^{T}, respectively, and thus further identification of these parameters is advisable.

The temperature effect on oxygen consumption rate (represented by $k_{O_2}^{T}$) showed a negative value, representing a higher oxygen availability with lower temperatures. This is in concordance with empirical observations (and the same principle applied for temperature limited fed batch cultures Jahic, Wallberg, Bollok, Garcia, & Enfors (2003)) of overflow metabolism, and in particular with CHO dynamics.

Finally, the sieving effect parameter was obtained from the bioreactor data, the final concentration in the harvest and the instantaneous concentration of the permeate at the time of harvest (data not shown).

Intuitively this would seem as it would correspond to one parameter and one degree of freedom, nonetheless given the dynamic specific production rate, the sieving parameter needs to be included in the calibration. Additional dependence of the sieving parameter with other variables, such as biomass concentration, antifoam addition, cell death rate is something that also needs to be correctly validated. Additionally, it is assumed in the model that sieving only affect product concentration and that the rest of the products/substrates freely permeate the filter, regardless of the sieving effect.

4.2. Validation

To validate the results three different experimental data sets were selected:

-        Validation Experiment 1: Same experimental set-up (different batch).

-        Validation Experiment 2: Early Temperature Shift

-        Validation Experiment 3: Delayed Temperature Shift

For all the experiments, all the input variables were modified to reflect the experimental set-up. Note that even though Validation Experiment 1 has the same design, the input variables are different due to different observed variables (*e.g.*, different glucose additions).
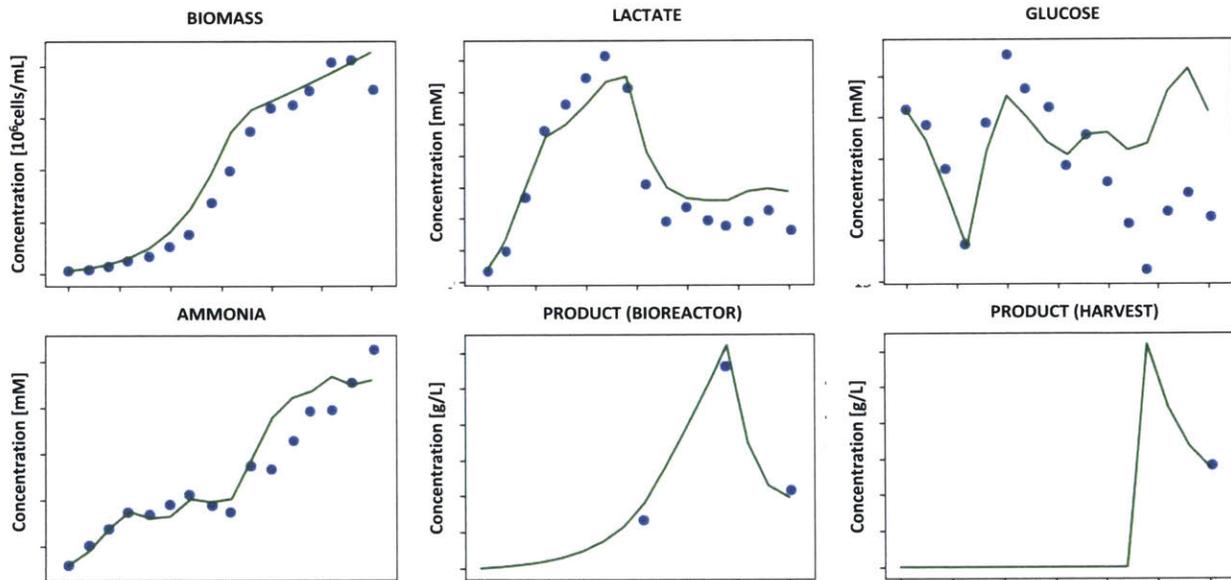
Figure 4. 2. Validation Experiment 1. Same Experimental set-up as Calibration Experiment.
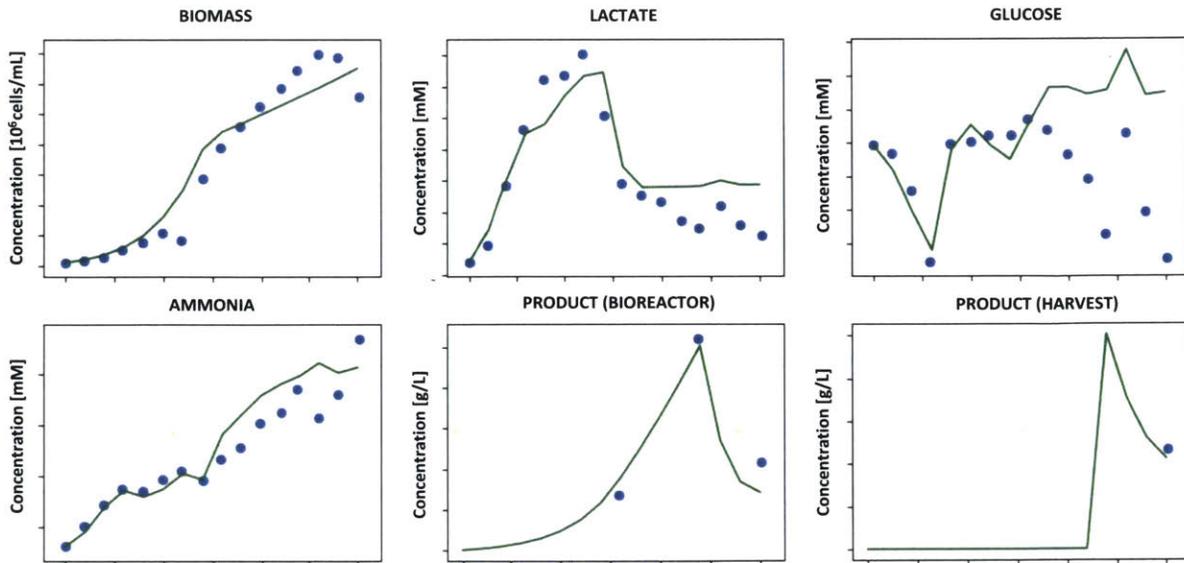
Figure 4. 3. Validation Experiment 2. Temperature Shift at $t - \Delta t$.
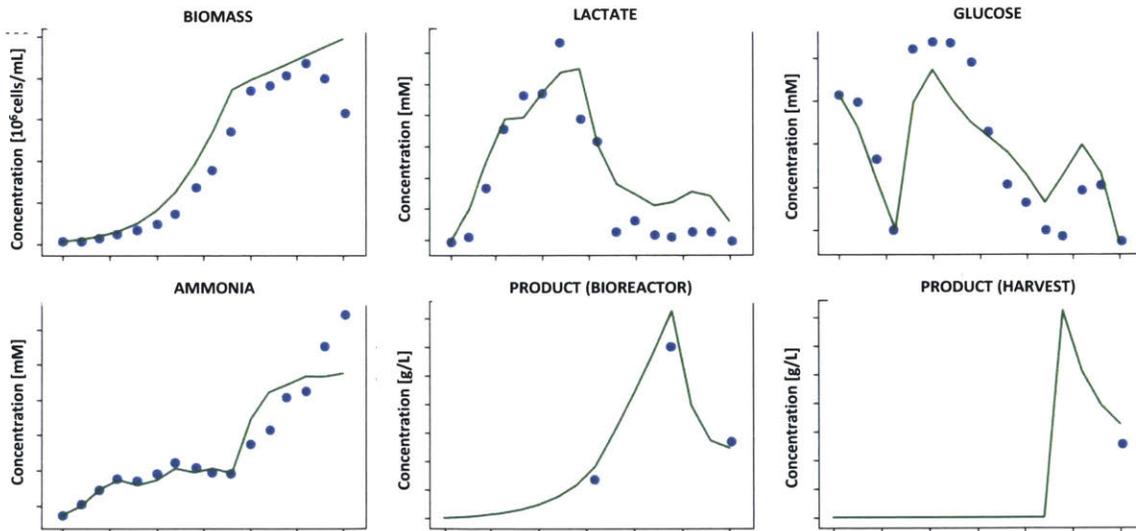

Figure 4. 4. Validation Experiment 3. Temperature Shift at $t + \Delta t$

The validation experiments show that despite the differences in the process conditions the model can reproduce with great accuracy the behavior of the cell culture. For example, for validation experiment 3 there is a lower concentration of lactate, due to the fact that the temperature shift was performed at a later

time when lactate was already perfused from the bioreactor. Additionally, the biomass effect and change in slope was accurately predicted with the model. Product concentration was accurately predicted in all cases.

Glucose concentration is the variable predicted with the lowest accuracy. One of the reasons for this could be the violation of the homogeneous bioreactor assumption by having concentrated feed solutions (from feed 2). In validation cases 1 & 2, the model shows higher glucose concentration compared to the experimental data, particularly at the end of the experiments where the glucose shots were added (data not shown). This could be caused by several factors, first the concentrated glucose solution as well low flow-rates account for great sources of error and could potentially overestimate glucose addition to the system. Additionally, the model assumes constant glucose feeding, while in fact the pump works in an on-and-off fashion, which could affect the cellular activity. Lastly, there could be a physiological effect that is not accounted for in the model, that increases glucose consumption despite the low metabolic activity.

5. DISCUSSION

In the present section the results of this work will be discussed. For this purpose the section is divided into three: (1) the development of a *User Interface*, to exemplify through FA-01 a tool to facilitate the adoption of the model and highlights the application-oriented goal of this work; (2) through the user interface the study of 3 commercially-relevant *Case Studies*; and (3) a closing *Applications and Opportunities* section of this work, that describes what will be the benefits and relevant next steps to successfully continue the present development.

5.1. User Interface

To facilitate a continuous use and development of the tool, a user interface for the model was created so that the subject matter experts at Amgen can take advantage of this platform and generate in-silico experimentation. Figure 5. 1 shows the graphical user interface, where the main inputs of the model can be adjusted, and a summary of the metrics and graphs can be accessed.

5.2. Case Studies

Through the user interface, three different scenarios of commercial interest were tested: decreasing overall production time, reducing feeding volume and delaying filter change.

Define your Experiment

Experimental Conditions

Process Conditions

Initial Seed Density
[UPDATE]
10^5 cells/mL

Initial Viability
[UPDATE]
%

Working Volume
[UPDATE]
Liters

pH
[UPDATE]
.

Filter Change
[UPDATE]
Days

Harvest Time
[UPDATE]
Days

Process Conditions

Initial Temperature
[UPDATE]
°C

Production Temperature
[UPDATE]
°C

Temperature Shift Time
[UPDATE]
Days

Dilution Rates
[UPDATE]
Volumes per Day

Dilution Times
[UPDATE]
Days

Glucose Shots Flow
[UPDATE]
mL/hr

Glucose Shots Time
[UPDATE]
Days

Media Composition

Batch Media

Glucose Concentration [GLC]
[UPDATE]
mM

Glutamine Concentration [GLN]
[UPDATE]
mM

Glutamic Acid Concentration [GLU]
[UPDATE]
mM

Cystine Concentration [CC]
[UPDATE]
mM

Glycine Concentration [GLY]
[UPDATE]
mM

Serine Concentration [SER]
[UPDATE]
mM

Aspartic Acid Concentration [ASP]
[UPDATE]
mM

Asparagine Concentration [ASN]
[UPDATE]
mM

Alanine Concentration [ALA]
[UPDATE]
mM

Perfusion Media

Glucose Concentration [GLC]
[UPDATE]
mM

Glutamine Concentration [GLN]
[UPDATE]
mM

Glutamic Acid Concentration [GLU]
[UPDATE]
mM

Cystine Concentration [CC]
[UPDATE]
mM

Glycine Concentration [GLY]
[UPDATE]
mM

Serine Concentration [SER]
[UPDATE]
mM

Aspartic Acid Concentration [ASP]
[UPDATE]
mM

Asparagine Concentration [ASN]
[UPDATE]
mM

Alanine Concentration [ALA]
[UPDATE]
mM

Figure 5. 1. User Interface for the Model.

5.2.1. Reducing Production Time

Biologic growth for current products as well as new products is projected to growth. Consequently, having more efficient processes is highly desirable, not only because it improves the margins, but also because it avoids capacity building or contract manufacturing. By increasing the initial seed density and shifting the process perturbations a day earlier, an in-silico experiment was tested and compared to the control process. Figure 5. 2 illustrates the results.
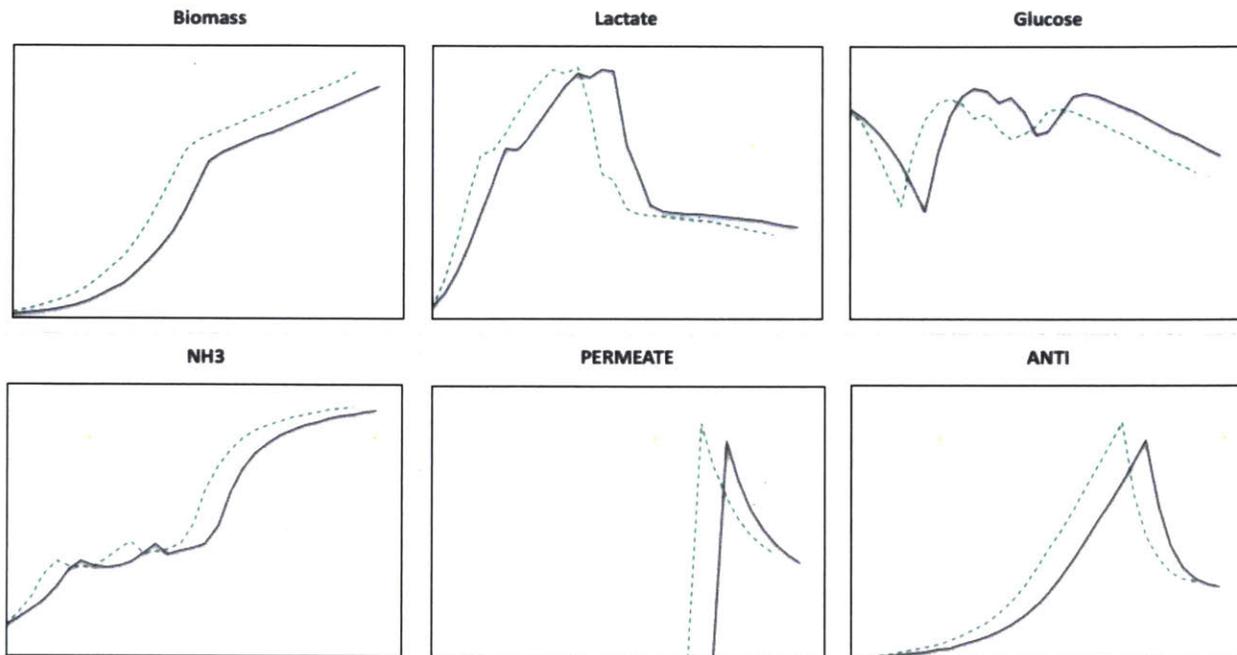
Figure 5. 2. Caste Study 1: Reducing total time by a day. Dashed Line time reduced experiment. Solid Line represent control experiment (or normal operational set-up).

Table 5. 1 summarizes the productivity metrics obtained with this scenario. We observe a 25% increase in the harvest productivity (i.e., final concentration in the harvest divided by the total production time). And additionally, a 9% increase in the harvest titer (grams per liter of harvest) and a 6% increase in the media yield (grams of product per gram of media used), presumably because of the different feeding dynamics. This translates in an 8% increase in the mass recovered in the harvest with the same media yield, harvested volume and filter yield. It is important to note that the lactate peak and the final ammonia concentration were comparable to the control experiment, meaning that potentially no detrimental effect on the cell culture would be observed. These results show that potentially similar productivity metrics can be obtained while reducing the production time in a day by increasing the initial seed density and increasing the biomass generation capacity up-stream of the commercial bioreactor.

Table 5. 1. Percent Change of Productivity Metrics of Case Study 1.

	Percent Change
Harvest Productivity	25%
Harvest Titer	9%
Media Yield	6%
Used Media	0%
Harvested Volume	0%
Mass Of Product Recovered	8%
Equivalent Titer	9%
Filter Yield	0%

5.2.2. *Reducing Feeding Volumes*

Perfusion experiments require great amount of feeding media, to both feed the culture as well as eliminate harmful byproducts. Having large volumes of sterile media is expensive and also a source of contamination. Reducing the feed volume while maintaining a correct elimination of waste is a critical aspect in culture performance. To test whether it was possible to reduce the media volume by doubling the substrate concentration we tested the results via simulation (Figure 5. 3)

From the simulation it was observed that the biomass levels are notably reduced, potentially due to the higher levels of inhibitory by-products (both lactate and ammonia) due to a lower waste extraction rate leading to a higher glucose concentration. Additionally, the increased inhibition generates an excess of glucose on the media that is not consumed. The product level on the bioreactor is lowered, but due to a lower perfusion rate the final harvest concentration is 25% higher (Table 5. 2). Nonetheless, note that the final mass of the product in the harvest is reduced by 39%. Additionally, there is a 44% reduction in the media yield and a 20% reduction in the filter yield. These results show that such a feeding strategy is not optimal due to inhibitory effects, but further simulations to optimize the feeding strategy can be explored to achieve the desired goal.
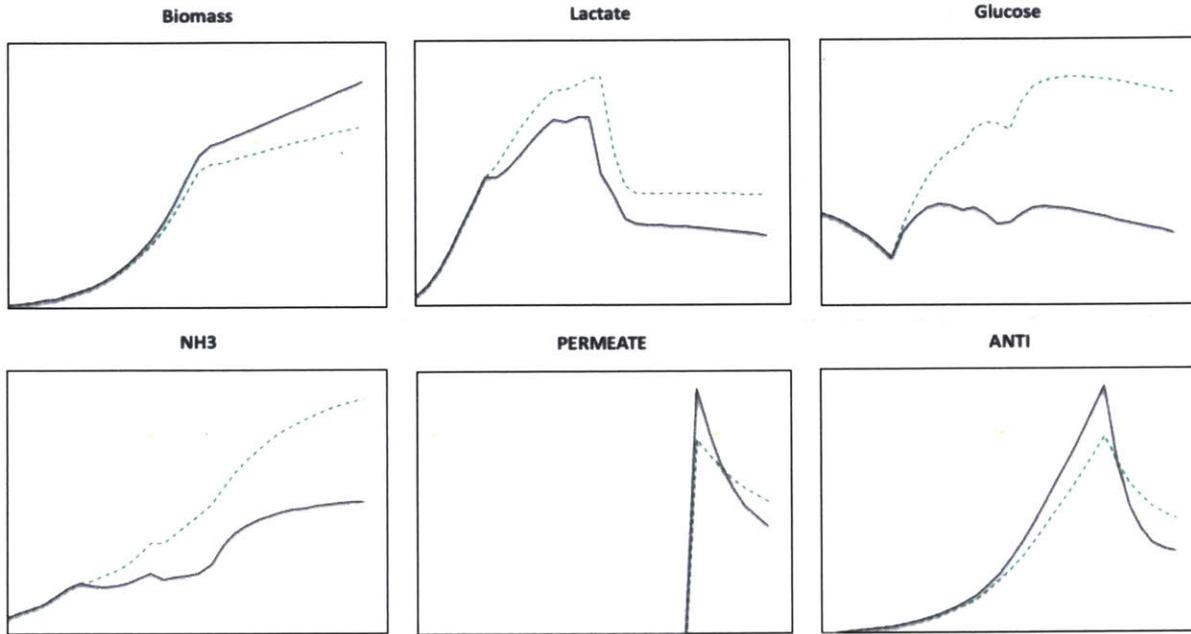
Figure 5. 3. Caste Study 2: Reducing media feed. Dashed Line media reduced experiment. Solid Line control.

Table 5. 2. Productivity Metrics for Case Study 2.

	Percent Change
Harvest Productivity	25%
Harvest Titer	23%
Media Yield	-44%
Used Media	-45%
Harvested Volume	-50%
Recovered Product (mass)	-39%
Equivalent Titer	-23%
Filter Yield	-20%

5.2.3. Delaying Harvesting Time

Final harvest concentration is a key productivity metric. Higher concentrations allow for less intensive and expensive down-stream processing. Nonetheless, there is a trade-off between final concentration and total recovered mass of the product (or harvest volume). Correctly designing a process

to optimize this trade-off is key from a process perspective. Figure 5. 4 shows the in-silico results of postponing harvest filtration by a day.
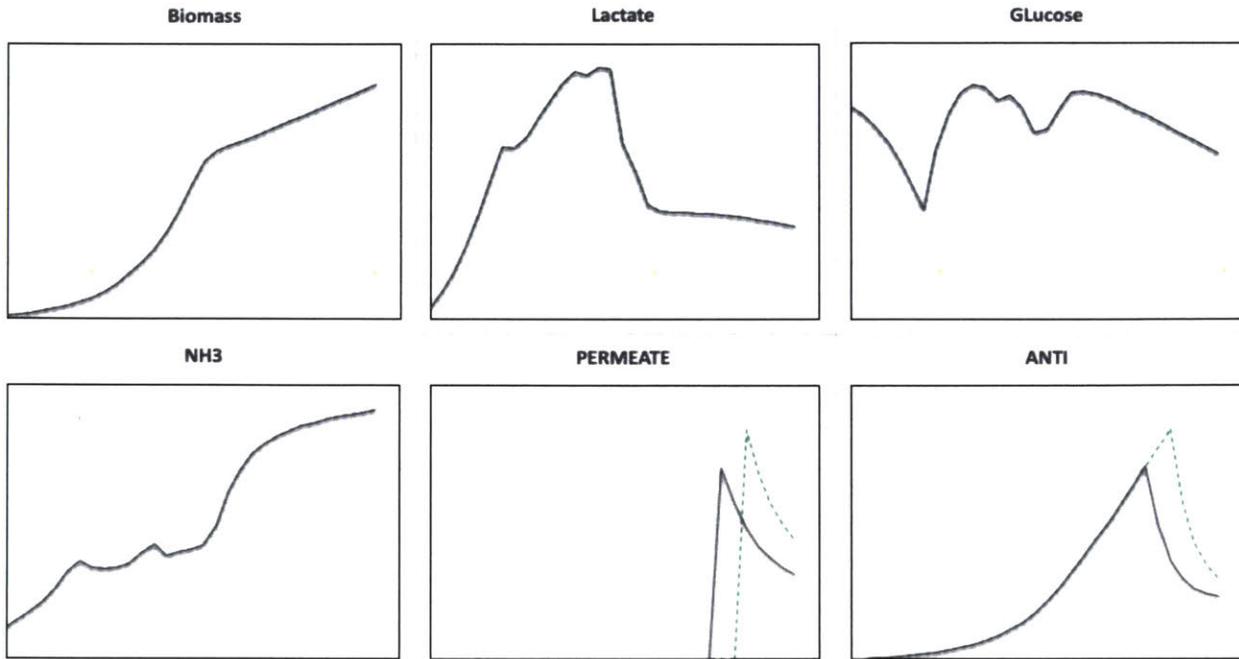
Figure 5. 4. Caste Study 2: Delaying harvest filtration by a day. Dashed Line delayed harvest experiment. Solid Line control.

It was observed from the simulation results, that as expected only the permeate and product dynamics are changed. As expected, only the permeate and antibody variables show a change as compared to the control. Additionally, when we observe the productivity metrics we observe a 50% increase in harvest productivity with only a 6% loss in mass of recovered product (and filter yield). Whether these results are commercially interesting depend on the down-stream processing steps and associated costs. Nonetheless this is an interesting tool to explore scenarios. Additionally, one could design a process which combines both a reduction in time (like in case study 1) and a shift in the harvest filter time (such as in this case study) and find optimal strategies in-silico that can be further validated experimentally.

Table 5. 3. Productivity Metrics for Case Study 3.

	Percent Change
Harvest Productivity	50%
Harvest Titer	41%
Media Yield	-6%
Used Media	0%
Harvested Volume	-33%
Mass Of Product Recovered	-6%
Equivalent Titer	1%
Filter Yield	-6%

5.3. Applications and Opportunities

Process development agility and speed to market is often prioritized over manufacturing cost (or productive capacity). Adding this factor to a highly regulated process hinders continuous improvement, fixing future manufacturing costs. Combining wet experimentation with in-silico modeling will allow for more efficient processes without sacrificing speed to market, forming the foundations for process intensification efforts and also enabling the quality-by-design framework.

In the present work a biological modeling framework and software architecture have been developed and validated with a case-model application (FA-01). This case model was designed to understand and predict the impact of process variables on productivity metrics, but most importantly to validate the framework foundations. FA-01 has successfully defined and validated the first step of the development of upstream biomanufacturing modeling effort. The underlying framework and software architecture, together with the modular development allow for a growing effort that can be easily expansible to broader models with different applications.

This modeling framework envisions several opportunities in the biopharmaceutical space:

Agile Process Development: Each approximately 30-second solution of the model presented here has the potential to eliminate bench-, lab-scale experiments with costs up to $500-$2,500 and take 15-25 days. The savings translated to commercial-scale account for savings of several order magnitudes greater. This allows

for faster and better resource allocation in process development and most importantly having a streamlined process development strategy. Combining wet-experimentation with *in-silico* modeling will allow for faster and optimal process design, faster and cheaper transfers across scales, and full exploration and characterization of the process space.

Increased Productivity: Modeling tools form one of the foundations of process intensification principles. By characterizing the relevant unit operations, the process can be further optimized in terms of scale or process parameters that increase the yields, productivity and final titer. This will allow for overall unit cost reduction. Examples of increased productivity can be obtained from applying the framework with trajectory optimization to define optimal feeding strategies, or optimize media composition to reduce overfeeding nutrients. Furthermore, it can help integrate early stage R&D data to project efficiencies and better estimate capacity allocation. Or it can in fact help design new manufacturing plants if needed.

Improved Process Efficiencies: Additionally, this tool will allow for further improvement in process automation that can further increase productivity or reduce human dependencies in production of commercial batches. The tool can also help design and define novel platforms such as continuous manufacturing, or new platforms required for innovative biologics where no platform processes are available.

Towards Quality by Design: Future generations and development of the model can incorporate relevant metrics such as quality attributes (*e.g.*, glycosylation profiles in the case of biosimilars or monoclonal antibodies), oxygen transfer dynamics relevant when scaling and transferring processes, and pH effects on osmolarity in the cell culture. This will enable applications where the process set-up will be determined by the model to achieve quality targets and advance in the Quality by Design framework.

Despite the value and hypothesized results, to be successful this needs to be implemented and adapted within the corporate culture at Amgen – and for this several changes need to occur:

Continuing the virtualization journey in bioprocess development requires a paradigm change. The modeling tool and hypothesis needs to be accessible and easy to use for the end users. This will allow for the experimenters to accurately test and compile their current knowledge in a systematic and reproducible manner across products. Additionally, this will allow for the modeling tools to persist and evolve with the end-users varying needs.

The paradigm shift also demands a change in the way that experiments and hypothesis testing are performed. The present project consisted of calibrating a model to existing commercial and dynamic processes. To correctly define modeling tools, targeted experimental designs must be defined to calibrate individual components of the model representing well-conserved physical and biological behavior. This will only occur after the models have been validated and sufficient buy-in across end-users as well as management is achieved, for this it is essential that end-users together with engineers join efforts in co-develop and iterate this modeling framework

It is advisable to start from current knowledge, and a good start is to start from the lessons learned in the chemical industry, such as developing in-house database of physical properties of the unit operations and having streamlined information systems for accurate and reliable data flow, amongst others (Zobel-Roos et al., 2019). Additionally, it is recommended that a white biotechnology mindset of bioprocess engineering be adopted – that due to economic pressure have the continuous need to become more efficient.

From a sustainability perspective, there needs to be clear connections between models, their underlying assumptions, and the application space. This needs to be clearly informed to end-users as well as management, as to have clear expectations of what are the outcomes.

Next steps:

It is not difficult to envision process development where small-scale experiments are executed to calibrate a model, an in-silico process optimization is performed, and then a limited subset of the simulated processes are selected for experimental validation. This change can come together with novel culture

technologies that allow for miniature and high throughput experimentation and will greatly help improve process development in the biopharmaceuticals.

From a technical perspective, the model needs to be further refined to incorporate oxygen kinetics and physical capacities of the manufacturing plants to properly transfer processes across scales. It should incorporate relevant reaction kinetics related to carbonate equilibria, pH control and integrate the osmolarity as a predictive variable. It should also consider incorporating the quality attributes of the product, the most important aspect of the biopharmaceutical attributes.

It is advisable also to incorporate modern statistical or Machine Learning tools, particularly in the definition of the kinetic reaction rates. Usually these formulas have been obtained empirically in perfectly controlled experiments, and when transferred to whole-scale models inaccurate behaviors are observed, possibly due to the bottom-up approach. Novel ML tools are very promising in helping to overcome these problems and can help accurately generate more robust models if sufficient process data exists a priori. Additionally, model order reduction methods can be used to simplify the computational complexity of the system and work in faster times that can interact in real-time with the process.

Aside from the technical challenges, the most important dimension for the success of the proposed framework is the adoption of the tool within the organization. A gradual introduction is recommended for two reasons: (1) to safeguard the technical development and readiness of the tool and (2) to guarantee the work force is ready and embraces the change.

Developing this tool will require investment in time and capital, and collaboration across different functional areas. It is fundamental for the success to clearly define expectations, achievable timelines and cost. Starting with low-impact applications, that can be developed in parallel experiments will reduce the risk of unwanted failures. Considering separate dedicated experimentation capabilities for calibration allows for more robust models. As the tool gradually matures and becomes validated it can be introduced in riskier processes and become part of the process development pipeline.

People will also see the change, both a cultural change as well as a skill-set change will be required to move from platform-based/empirical-exploration to model-driven process development workflow. This change requires buy-in across the organization and it is recommended to co-develop this tool with the end-users. This is beneficial in two ways: (1) it will accelerate the end-user learning rate and thus adoption (2) will also accelerate engineering learning rate and thus having better defined and useful models.

6. CONCLUSION & PERSPECTIVES

Society needs to innovative yet affordable medicines. Manufacturing efficiency is one of the main challenges that the biopharmaceutical industry will need to address in the upcoming years. This is challenge both from a capacity constraint (~20% growth rate) as well as for a cost and user-adoption perspective. In this work a model-driven approach towards process development has been covered. This approach could potentially break the trade-off between speed and process efficiency, enabling process intensification despite regulatory restrictions and therefore allowing to have fast, economical and sustainable manufacturing processes.

In the present study the core of an in-silico modeling framework for a CHO cell bioreactor was successfully designed. Through a thorough landscaping of industrial and academic efforts a framework architecture capable of implementing all of the established modeling techniques was defined (FA). To develop this vision a sustainable, modular and expansible software architecture supporting these requirements was designed and constructed. This software architecture was designed so that the user can easily tune parameters to target specific applications. Finally, these modeling foundations were validated with a novel dynamic process-dependent flux balance analysis with a stress-related variable affecting the energy of maintenance (FA-01) specifically designed for the end-user's requirements.

A subset of the model parameters was estimated using a two-stage calibration procedure with a single perfusion experiment of a pipeline molecule. The calibrated model was able to correctly reproduce the growth dynamics, product formation and inhibitory by-product accumulation. Thirteen parameters were predicted in accordance with literature values and the model was further validated with three different experimental set ups showing close agreement.

Additionally, the calibrated model was used to execute three different commercially-relevant case studies, demonstrating the potential to develop fast hypothesis testing to re-examine the current process development platform at Amgen. The modeling platform was shown to enable optimized resource

allocation, replacement of unnecessary experimentation with in-silico tools, and optimization of process design.

Proceeding with the bioprocess virtualization journey enabled by this modeling framework will support agile process development, increased productivity, improved process efficiencies, and advance towards a quality by design mindset. This framework will help differentiate Amgen in terms of process development and be more efficient in the increasingly competitive industry while allowing for innovative yet affordable medicines for patients.

7. BIBLIOGRAPHY

21 C.F.R. 316. (2004). Orphan Drugs. Retrieved from https://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&SID=51cf70689d51f0ea4147c0a8ac649321&rgn=div5&view=text&node=21:5.0.1.1.6&idno=21

Albers, E., Larsson, C., Lidén, G., Niklasson, C., & Gustafsson, L. (1996). Influence of the nitrogen source on Saccharomyces cerevisiae anaerobic growth and product formation. *Applied and Environmental Microbiology*, *62*(9), 3187–3195. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=168115&tool=pmcentrez&rendertype=abstract

An, G., Mi, Q., Dutta-Moscato, J., & Vodovotz, Y. (2009). Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *1*(2), 159–171. https://doi.org/10.1002/wsbm.45

Barrue, H., Bertrand, J., Cristol, B., & Xuereb, C. (2001). Eulerian Simulation of Dense Solid-Liquid Suspension in Multi-Stage Stirred Vessel. *JOURNAL OF CHEMICAL ENGINEERING OF JAPAN*, *34*(5), 585–594. https://doi.org/10.1252/jcej.34.585

Bartolini, E., Manoli, H., Costamagna, E., Jeyaseelan, H. A., Hamad, M., Irhimeh, M. R., ... Abbas, A. (2015). Population balance modelling of stem cell culture in 3D suspension bioreactors. *Chemical Engineering Research and Design*, *101*, 125–134. https://doi.org/10.1016/J.CHERD.2015.07.014

Batt, B. C., & Kompala, D. S. (1989). A structured kinetic modeling framework for the dynamics of hybridoma growth and monoclonal antibody production in continuous suspension cultures. *Biotechnology and Bioengineering*, *34*(4), 515–531. https://doi.org/10.1002/bit.260340412

Bayrak, E. S., Wang, T., Cinar, A., & Undey, C. (2015). Computational Modeling of Fed-Batch Cell Culture Bioreactor: Hybrid Agent-Based Approach. *IFAC-PapersOnLine*, *48*(8), 1252–1257. https://doi.org/10.1016/J.IFACOL.2015.09.140

Bayrak, E. S., Wang, T., Jerums, M., Coufal, M., Goudar, C., Cinar, A., & Undey, C. (2016). In Silico

    Cell Cycle Predictor for Mammalian Cell Culture Bioreactor Using Agent-Based Modeling

    Approach. *IFAC-PapersOnLine*, *49*(7), 200–205. https://doi.org/10.1016/J.IFACOL.2016.07.249

Berry, D. A. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery*, *5*(1), 27–36.

    https://doi.org/10.1038/nrd1927

Bezzo, F., Macchietto, S., & Pantelides, C. C. (2003). General hybrid multizonal/CFD approach for

    bioreactor modeling. *AIChE Journal*, *49*(8), 2133–2148. https://doi.org/10.1002/aic.690490821

Biscani, F., & Izzo, D. (2018). esa/pagmo2: pagmo 2.9. https://doi.org/10.5281/ZENODO.1406840

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems.

    Proceedings of the National Academy of Sciences of the United States of America, *99 Suppl 3*(Suppl

    3), 7280–7287. https://doi.org/10.1073/pnas.082080899

Bordbar, A., Monk, J. M., King, Z. A., & Palsson, B. O. (2014). Constraint-based models predict

    metabolic and associated cellular functions. *Nature Reviews Genetics*, *15*(2), 107–120.

    https://doi.org/10.1038/nrg3643

Butler, M. (2005). Animal cell cultures: recent achievements and perspectives in the production of

    biopharmaceuticals. *Applied Microbiology and Biotechnology*, *68*(3), 283–291.

    https://doi.org/10.1007/s00253-005-1980-8

Chon, J. H., & Zarbis-Papastoitsis, G. (2011). Advances in the production and downstream processing of

    antibodies. *New Biotechnology*, *28*(5), 458–463. https://doi.org/10.1016/J.NBT.2011.03.015

Coakley, S., Richmond, P., Gheorghe, M., Chin, S., Worth, D., Holcombe, M., & Greenough, C. (2016).

    Large-Scale Simulations with FLAME (pp. 123–142). Springer, Cham. https://doi.org/10.1007/978-

    3-319-23742-8_6

Delafosse, A., Calvo, S., Collignon, M.-L., Delvigne, F., Crine, M., & Toye, D. (2015). Euler–Lagrange

approach to model heterogeneities in stirred tank bioreactors – Comparison to experimental flow

characterization and particle tracking. *Chemical Engineering Science, 134*, 457–466.

https://doi.org/10.1016/J.CES.2015.05.045

Derksen, J. J. (2003). Numerical simulation of solids suspension in a stirred tank. *AIChE Journal, 49*(11),

2700–2714. https://doi.org/10.1002/aic.690491104

Dhanasekharan, K., Humbard, K., Kester, B., Choudhari, S., Li, Y., & Vinci, V. (2016). Emerging

technology trends in biologics development: a contract development and manufacturing perspective.

BioProcess Int., 0(3), 32–37. https://doi.org/10.1002/jps.24268

Dunn, P. J. (Peter J., Wells, A. S., & Williams, M. T. (Michael T. (2010). *Green chemistry in the*

pharmaceutical industry. Wiley-VCH. Retrieved from https://www.wiley.com/en-

us/Green+Chemistry+in+the+Pharmaceutical+Industry-p-9783527324187

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based

Reconstruction and Analysis for Python. *BMC Systems Biology, 7*(1), 74.

https://doi.org/10.1186/1752-0509-7-74

Elqotbi, M., Vlaev, S. D., Montastruc, L., & Nikov, I. (2013). CFD modelling of two-phase stirred

bioreaction systems by segregated solution of the Euler–Euler model. *Computers & Chemical*

Engineering, 48, 113–120. https://doi.org/10.1016/j.compchemeng.2012.08.005

Fang, Z. (David). (2010a). Applying Computational Fluid Dynamics Technology in Bioprocesses-Part 1.

BioPharm International, 23(4). Retrieved from http://www.biopharminternational.com/applying-

computational-fluid-dynamics-technology-bioprocesses-part-1

Fang, Z. (David). (2010b). Applying Computational Fluid Dynamics Technology in Bioprocesses-Part 2:

Computational fluid dynamics can resolve performance problems. *BioPharm International, 23*(5).

Retrieved from http://www.biopharminternational.com/applying-computational-fluid-dynamics-

technology-bioprocesses-part-2

Fernandez, J.-M., Stein, R. M., & Lo, A. W. (2012). Commercializing biomedical research through securitization techniques. *Nature Biotechnology*, *30*(10), 964–975. https://doi.org/10.1038/nbt.2374

Ferrer, J., Prats, C., & López, D. (2008). Individual-based Modelling: An Essential Tool for Microbiology. *Journal of Biological Physics*, *34*(1–2), 19–37. https://doi.org/10.1007/s10867-008-9082-3

Furukawa, K., & Ohsuye, K. (1998a). *Effect of culture temperature on a recombinant CHO cell line producing a C-terminal-amidating enzyme. Cytotechnology* (Vol. 26). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466681/pdf/10616_2004_Article_151973.pdf

Furukawa, K., & Ohsuye, K. (1998b). Effect of culture temperature on a recombinant CHO cell line producing a C-terminal α-amidating enzyme. *Cytotechnology*, *26*(2), 153–164. https://doi.org/10.1023/A:1007934216507

García Münzer, D. G., Kostoglou, M., Georgiadis, M. C., Pistikopoulos, E. N., & Mantalaris, A. (2015). Cyclin and DNA Distributed Cell Cycle Model for GS-NS0 Cells. *PLOS Computational Biology*, *11*(2), e1004062. https://doi.org/10.1371/journal.pcbi.1004062

Generalis, S. C., & Cartland Glover, G. M. (2005). Modelling a Biochemical Reaction with Computational Fluid Dynamics. *International Journal of Chemical Reactor Engineering*, *3*(1). https://doi.org/10.2202/1542-6580.1288

Global Biologics Market—Companies-to-Action. (2017). Retrieved from https://cds-frost-com.libproxy.mit.edu/p/58319/#!/ppt/c?id=MD83-01-00-00-00&hq=biologics market

Global Biologics Market, Opportunities And Strategies. (2018). Retrieved from https://www.thebusinessresearchcompany.com/report/biologics-global-market-opportunities-and-strategies-to-2021

GLPK - GNU Project. (2012). Retrieved March 24, 2019, from https://www.gnu.org/software/glpk/

González-Cabaleiro, R., Mitchell, A. M., Smith, W., Wipat, A., & Ofiţeru, I. D. (2017). Heterogeneity in

Pure Microbial Systems: Experimental Measurements and Modeling. *Frontiers in Microbiology, 8*,

1813. https://doi.org/10.3389/fmicb.2017.01813

Gronemeyer, P., Ditz, R., Strube, J., Gronemeyer, P., Ditz, R., & Strube, J. (2014). Trends in Upstream

and Downstream Process Development for Antibody Manufacturing. *Bioengineering, 1*(4), 188–

212. https://doi.org/10.3390/bioengineering1040188

Grote, F., Ditz, R., & Strube, J. (2009). Downstream of downstream processing - Integrated bioprocess

development from upstream to downstream. *Chemie Ingenieur Technik, 81*(8), 1276–1277.

https://doi.org/10.1002/cite.200950116

Grote, F., Ditz, R., & Strube, J. (2012). Downstream of downstream processing: development of recycling

strategies for biopharmaceutical processes. *Journal of Chemical Technology & Biotechnology,

87*(4), 481–497. https://doi.org/10.1002/jctb.2727

Guo, W., & Feng, X. (2016). OM-FBA: Integrate Transcriptomics Data with Flux Balance Analysis to

Decipher the Cell Metabolism. *PLOS ONE, 11*(4), e0154188.

https://doi.org/10.1371/journal.pone.0154188

Haringa, C., Noorman, H. J., & Mudde, R. F. (2017). Lagrangian modeling of hydrodynamic–kinetic

interactions in (bio)chemical reactors: Practical implementation and setup guidelines. *Chemical

Engineering Science, 157*, 159–168. https://doi.org/10.1016/J.CES.2016.07.031

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., … Lewis, N. E.

(2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism.

Cell Systems, 3(5), 434–443.e8. https://doi.org/10.1016/j.cels.2016.10.020

Heins, A.-L., & Weuster-Botz, D. (2018). Population heterogeneity in microbial bioprocesses: origin,

analysis, mechanisms, and future perspectives. *Bioprocess and Biosystems Engineering, 41*(7), 889–

916. https://doi.org/10.1007/s00449-018-1922-3

Heinzle, E., Biwer, A. P., & Cooney, C. L. (2006). *Development of Sustainable Bioprocesses*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470058916

Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., & Kreft, J.-U. (2016). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*, *14*(7), 461–471. https://doi.org/10.1038/nrmicro.2016.62

Hermelink, A., Brauer, A., Lasch, P., & Naumann, D. (2009). Phenotypic heterogeneity within microbial populations at the single-cell level investigated by confocal Raman microspectroscopy. *The Analyst*, *134*(6), 1149. https://doi.org/10.1039/b822574e

Höffner, K., Harwood, S. M., & Barton, P. I. (2013). A reliable simulator for dynamic flux balance analysis. *Biotechnology and Bioengineering*, *110*(3), 792–802. https://doi.org/10.1002/bit.24748

Jahic, M., Wallberg, F., Bollok, M., Garcia, P., & Enfors, S.-O. (2003). Temperature limited fed-batch technique for control of proteolysis in Pichia pastoris bioreactor cultures. *Microbial Cell Factories*, *2*(1), 6. https://doi.org/10.1186/1475-2859-2-6

Jayathilake, P. G., Gupta, P., Li, B., Madsen, C., Oyebamiji, O., González-Cabaleiro, R., … Curtis, T. (2017). A mechanistic Individual-based Model of microbial communities. *PLOS ONE*, *12*(8), e0181965. https://doi.org/10.1371/journal.pone.0181965

Jedrzejewski, P. M., del Val, I. J., Constantinou, A., Dell, A., Haslam, S. M., Polizzi, K. M., & Kontoravdi, C. (2014). Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation. *International Journal of Molecular Sciences*, *15*(3), 4492–4522. https://doi.org/10.3390/ijms15034492

Jeong, D. H., Yoo, S. J., Kim, J. H., & Lee, J. M. (2016). Computationally efficient dynamic simulation of cellular kinetics via explicit solution of flux balance analysis: xDFBA modelling and its biochemical process applications. *Chemical Engineering Research and Design*, *113*, 85–95. https://doi.org/10.1016/J.CHERD.2016.07.002

Jimenez del Val, I., Nagy, J. M., & Kontoravdi, C. (2011). A dynamic mathematical model for

    monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing

    Golgi apparatus. *Biotechnology Progress*, *27*(6), 1730–1743. https://doi.org/10.1002/btpr.688

Kaul, H., Cui, Z., & Ventikos, Y. (2013). A Multi-Paradigm Modeling Framework to Simulate Dynamic

    Reciprocity in a Bioreactor. *PLoS ONE*, *8*(3), e59671. https://doi.org/10.1371/journal.pone.0059671

Kelly, W. J. (2008). Using computational fluid dynamics to characterize and improve bioreactor

    performance. *Biotechnology and Applied Biochemistry*, *49*(4), 225.

    https://doi.org/10.1042/BA20070177

Kim, H. S., & Lee, G. M. (2007). Differences in optimal pH and temperature for cell growth and antibody

    production between two Chinese hamster ovary clones derived from the same parental clone.

    Journal of Microbiology and Biotechnology, *17*(5), 712–720. Retrieved from

    http://www.ncbi.nlm.nih.gov/pubmed/18051290

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., … Lewis, N. E. (2016). BiGG

    Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids

    Research*, *44*(D1), D515–D522. https://doi.org/10.1093/nar/gkv1049

Kochergin, V., & Kearney, M. (2006). Existing Biorefinery Operations That Benefit From Fractal-Based

    Process Intensification. *Applied Biochemistry and Biotechnology*, *130*(1–3), 349–360.

    https://doi.org/10.1385/ABAB:130:1:349

Kostoglou, M., Fuentes-Garí, M., García-Münzer, D., Georgiadis, M. C., Panoskaltsis, N., Pistikopoulos,

    E. N., & Mantalaris, A. (2016). A comprehensive mathematical analysis of a novel multistage

    population balance model for cell proliferation. *Computers & Chemical Engineering*, *91*, 157–166.

    https://doi.org/10.1016/J.COMPCHEMENG.2016.02.012

Kyriakopoulos, S., Ang, K. S., Lakshmanan, M., Huang, Z., Yoon, S., Gunawan, R., & Lee, D.-Y. (2018).

    Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance

Biomanufacturing. *Biotechnology Journal, 13*(3), 1700229. https://doi.org/10.1002/biot.201700229

Lapin, A., Müller, D., & Reuss, M. (2004). Dynamic Behavior of Microbial Populations in Stirred

Bioreactors Simulated with Euler–Lagrange Methods: Traveling along the Lifelines of Single

Cells†. https://doi.org/10.1021/IE030786K

Lapin, A., Schmid, J., & Reuss, M. (2006). Modeling the dynamics of E. coli populations in the three-

dimensional turbulent field of a stirred-tank bioreactor—A structured–segregated approach.

Chemical Engineering Science, 61(14), 4783–4797. https://doi.org/10.1016/J.CES.2006.03.003

Lardon, L. A., Merkey, B. V., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., & Smets, B. F. (2011).

iDynoMiCS: next-generation individual-based modelling of biofilms. *Environmental Microbiology,

13*(9), 2416–2434. https://doi.org/10.1111/j.1462-2920.2011.02414.x

Lencastre Fernandes, R., Nierychlo, M., Lundin, L., Pedersen, A. E., Puentes Tellez, P. E., Dutta, A., …

Gernaey, K. V. (2011). Experimental methods and modeling techniques for description of cell

population heterogeneity. *Biotechnology Advances, 29*(6), 575–599.

https://doi.org/10.1016/J.BIOTECHADV.2011.03.007

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., … Palsson, B.

Ø. (2010). Omic data from evolved E. coli are consistent with computed optimal growth from

genome-scale models. *Molecular Systems Biology, 6*(1), 390. https://doi.org/10.1038/msb.2010.47

Lewis, N. E., Nagarajan, H., & Palsson, B. O. (2012). Constraining the metabolic genotype–phenotype

relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology, 10*(4), 291–305.

https://doi.org/10.1038/nrmicro2737

Linz, F. (2012). Answering Pharma's Need for Outsourcing to Process 'Intensification' Technology for

the Future of Biopharmaceutical Production. *Int. Pharm. Ind., 4*(3), 88–90. Retrieved from

https://www.ipimediaworld.com/answering-pharmas-need-for-outsourcing-to-process-

intensification-technology-for-the-future-of-biopharmaceutical-production/

Mahadevan, R., Edwards, J. S., Doyle, F. J., & 3rd. (2002). Dynamic flux balance analysis of diauxic

growth in Escherichia coli. *Biophysical Journal, 83*(3), 1331–1340. https://doi.org/10.1016/S0006-

3495(02)73903-9

Mantzaris, N. V., Daoutidis, P., & Srienc, F. (2001a). Numerical solution of multi-variable cell

population balance models: I. Finite difference methods. *Computers & Chemical Engineering,*

25(11–12), 1411–1440. https://doi.org/10.1016/S0098-1354(01)00709-8

Mantzaris, N. V., Daoutidis, P., & Srienc, F. (2001b). Numerical solution of multi-variable cell

population balance models. II. Spectral methods. *Computers & Chemical Engineering, 25*(11–12),

1441–1462. https://doi.org/10.1016/S0098-1354(01)00710-4

Mantzaris, N. V., Daoutidis, P., & Srienc, F. (2001c). Numerical solution of multi-variable cell

population balance models. III. Finite element methods. *Computers & Chemical Engineering,*

25(11–12), 1463–1481. https://doi.org/10.1016/S0098-1354(01)00711-6

Metzger, G. K. (2012). Biologics for RA: Costs and Insurance. Retrieved March 19, 2019, from

https://www.webmd.com/rheumatoid-arthritis/features/ra-biologics-cost#1

Micale, G., Montante, G., Grisafi, F., Brucato, A., & Godfrey, J. (2000). CFD Simulation of Particle

Distribution in Stirred Vessels. *Chemical Engineering Research and Design, 78*(3), 435–444.

https://doi.org/10.1205/026387600527338

Montazerhodjat, V., & Lo, A. (2015). *Is the FDA Too Conservative or Too Aggressive?: A Bayesian*

Decision Analysis of Clinical Trial Design. Cambridge, MA. https://doi.org/10.3386/w21499

Morchain, J., Gabelle, J.-C., & Cockx, A. (2014). A coupled population balance model and CFD

approach for the simulation of mixing issues in lab-scale and industrial bioreactors. *AIChE Journal,*

60(1), 27–40. https://doi.org/10.1002/aic.14238

Namjoshi, A. A., Hu, W.-S., & Ramkrishna, D. (2003). Unveiling steady-state multiplicity in hybridoma

cultures: The cybernetic approach. *Biotechnology and Bioengineering, 81*(1), 80–91.

https://doi.org/10.1002/bit.10447

Neil, L., & Sonal, S. (2018). *Unlocking R&D Productivity: Measuring the Return from Pharmaceutical Innovation 2018*. United Kingdom. Retrieved from https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/measuring-return-from-pharmaceutical-innovation.html

Nielsen, J., Villadsen, J., & Lidén, G. (2002). *Bioreaction Engineering Principles* (Second). Kluwe Academic/Plenum Publishers.

Nolan, R. P., & Lee, K. (2011). Dynamic model of CHO cell metabolism. *Metabolic Engineering, 13*(1), 108–124. https://doi.org/10.1016/j.ymben.2010.09.003

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering, 9*(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Oyebamiji, O. K., Wilkinson, D. J., Jayathilake, P. G., Curtis, T. P., Rushton, S. P., Li, B., & Gupta, P. (2017). Gaussian process emulation of an individual-based model simulation of microbial communities. *Journal of Computational Science, 22*, 69–84. https://doi.org/10.1016/J.JOCS.2017.08.006

Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology, 101*(22), 8101–8113. https://doi.org/10.1007/s00253-017-8531-y

Pörtner, R., & Schäfer, T. (1996). Modelling hybridoma cell growth and metabolism — a comparison of selected models and data. *Journal of Biotechnology, 49*(1–3), 119–135. https://doi.org/10.1016/0168-1656(96)01535-0

Quek, L.-E., Dietmair, S., Krömer, J. O., & Nielsen, L. K. (2010). Metabolic flux analysis in mammalian cell culture. *Metabolic Engineering, 12*(2), 161–171. https://doi.org/10.1016/j.ymben.2009.09.002

Rader, R. A., & Langer, E. S. (2014). Biopharmaceutical Manufacturing: Historical and Future Trends in

Titers, Yields, and Efficiency in Commercial-Scale Bioprocessing. *BioProcessing Journal, 13*(4), 47–54. Retrieved from http://libproxy.mit.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN =112426595&site=ehost-live&scope=site

Railsback, S. F. (Steven F., & Grimm, V. (2012). *Agent-based and individual-based modeling : a practical introduction*. Princeton University Press. Retrieved from https://press.princeton.edu/titles/9639.html

Rehberg, M., Ritter, J. B., & Reichl, U. (2014). Glycolysis Is Governed by Growth Regime and Simple Enzyme Regulation in Adherent MDCK Cells. *PLoS Computational Biology, 10*(10), e1003885. https://doi.org/10.1371/journal.pcbi.1003885

Ross, T., Ratkowsky, D. A., Mellefont, L. A., & McMeekin, T. A. (2003). Modelling the effects of temperature, water activity, pH and lactic acid concentration on the growth rate of Escherichia coli. *International Journal of Food Microbiology, 82*(1), 33–43. https://doi.org/10.1016/S0168-1605(02)00252-0

Rudge, T. J., Steiner, P. J., Phillips, A., & Haseloff, J. (2012). Computational Modeling of Synthetic Microbial Biofilms. *ACS Synthetic Biology, 1*(8), 345–352. https://doi.org/10.1021/sb300031n

Saa, P. A., & Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances, 35*(8), 981–1003. https://doi.org/10.1016/J.BIOTECHADV.2017.09.005

Sánchez, B. J., Pérez-Correa, J. R., & Agosin, E. (2014). Construction of robust dynamic genome-scale metabolic model structures of Saccharomyces cerevisiae through iterative re-parameterization. *Metabolic Engineering, 25*, 159–173. https://doi.org/10.1016/J.YMBEN.2014.07.004

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P.-J., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic

constraints. *Molecular Systems Biology*, *13*(8), 935. https://doi.org/10.15252/msb.20167411

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting

intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, *3*, 119.

https://doi.org/10.1038/msb4100162

Schuler, A. J., Majed, N., Bucci, V., Hellweger, F. L., Tu, Y., & Gu, A. Z. (2011). Is the whole the sum of

its parts? Agent-based modelling of wastewater treatment systems. *Water Science and Technology :*

A Journal of the International Association on Water Pollution Research, *63*(8), 1590–1598.

Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21866756

Sidoli, F. R., Asprey, S. P., & Mantalaris, A. (2006). A Coupled Single Cell-Population-Balance Model

for Mammalian Cell Cultures. https://doi.org/10.1021/IE0511581

Sklar, E. (2007). NetLogo, a Multi-agent Simulation Environment. *Artificial Life*, *13*(3), 303–311.

https://doi.org/10.1162/artl.2007.13.3.303

Song, H.-S., & Ramkrishna, D. (2010). Prediction of metabolic function from limited data: Lumped

hybrid cybernetic modeling (L-HCM). *Biotechnology and Bioengineering*, *106*(2), n/a-n/a.

https://doi.org/10.1002/bit.22692

Sou, S. N., Sellick, C., Lee, K., Mason, A., Kyriakopoulos, S., Polizzi, K. M., & Kontoravdi, C. (2015).

How does mild hypothermia affect monoclonal antibody glycosylation? *Biotechnology and*

Bioengineering, *112*(6), 1165–1176. https://doi.org/10.1002/bit.25524

Stankiewicz, A. I., & Moulijn, J. A. (2000). Process Intensification: Transforming Chemical Engineering.

Chemical Engineering Progress, *96*(1). Retrieved from

http://owens.mit.edu/sfx_local?sid=google&auinit=AI&aulast=Stankiewicz&atitle=Process+intensif

ication:+transforming+chemical+engineering&title=Chemical+Engineering+Progress&volume=96

&issue=1&date=2000&spage=22&issn=0360-7275

Strube, J., Ditz, R., Kornecki, M., Huter, M., Schmidt, A., Thiess, H., & Zobel-Roos, S. (2018). Process

intensification in biologics manufacturing. *Chemical Engineering and Processing - Process Intensification, 133*, 278–293. https://doi.org/10.1016/J.CEP.2018.09.022

Sunley, K., & Butler, M. (2010). Strategies for the enhancement of recombinant protein production from mammalian cells by growth arrest. *Biotechnology Advances, 28*(3), 385–394. https://doi.org/10.1016/j.biotechadv.2010.02.003

Vaghari, H., Eskandari, M., Sobhani, V., Berenjian, A., Song, Y., Jafarizadeh-Malmiri, H., … Jafarizadeh-Malmiri, H. (2015). Process Intensification for Production and Recovery of Biological Products. *American Journal of Biochemistry and Biotechnology, 11*(1), 37–43. https://doi.org/10.3844/ajbbsp.2015.37.43

Whitesides, G. M. (2015). Reinventing Chemistry. *Angewandte Chemie International Edition, 54*(11), 3196–3209. https://doi.org/10.1002/anie.201410884

Yu, L. X., Amidon, G., Khan, M. A., Hoag, S. W., Polli, J., Raju, G. K., & Woodcock, J. (2014). Understanding pharmaceutical quality by design. *The AAPS Journal, 16*(4), 771–783. https://doi.org/10.1208/s12248-014-9598-3

Zhang, L., Wang, Z., Sagotsky, J. A., & Deisboeck, T. S. (2009). Multiscale agent-based cancer modeling. *Journal of Mathematical Biology, 58*(4–5), 545–559. https://doi.org/10.1007/s00285-008-0211-1

Zhu, M. M., Goyal, A., Rank, D. L., Gupta, S. K., Boom, T. Vanden, & Lee, S. S. (2008). Effects of Elevated pCO2 and Osmolality on Growth of CHO Cells and Production of Antibody-Fusion Protein B1: A Case Study. *Biotechnology Progress, 21*(1), 70–77. https://doi.org/10.1021/bp049815s

Zobel-Roos, S., Schmidt, A., Mestmäcker, F., Mouellef, M., Huter, M., Uhlenbrock, L., … Strube, J. (2019). Accelerating Biologics Manufacturing by Modeling or: Is Approval under the QbD and PAT Approaches Demanded by Authorities Acceptable Without a Digital-Twin? *Processes, 7*(2), 94.