Contents

1 Introduction

Bioprocesses, especially microbial bioprocesses, are very important chemical processes. Yet microbial systems in bioprocesses are far from being accurately modelled and analysed, nor can they be controlled as precisely as other standard chemical engineering reaction systems, owning to the complicated cell kinetics that cannot be readily estimated. Traditional bioprocess uses classic kinetic models analyse only a handful of observable biomarkers. In other words, these classic kinetic models can only calculate the extracellular environmental information with limited biological insights, and have difficulties perform extrapolation.

To gain a comprehensive understanding of cellular insights as an integrated system, some metabolic networks have been proposed using systems biology approach. Flux Balance Analysis (FBA) is a genome-scale constraint-based modelling framework that has attracted increased attentions for its ability to calculate the solution of the metabolic network with a pre-defined objective function. In essence, FBA represents the steady state biological systems as a linear program (LP).

However, FBA models only concern with steady state process which limits its industrial application. Dynamic Flux Balance Analysis (DFBA), an extension of FBA approach, helps introduce dynamic variation in the FBA framework. Detailed information of FBA and DFBA models are depicted in the following sections.

Due to the complexity of most DFBA models, many DFBA solving methods and toolkits have been proposed to boost the efficiency of bioprocess simulation. A recent method based on an Interior Point Method (IPM), proposed by Scott et al. (2018), has been highlighted. This new method seems promising to solve large-scale DFBA problem accurately and efficiently, while further research has to be done to explore its reliability and applicability. The can be explored from different aspects Moreover, DFBA are used to model a single strain of microbe in a culture. Unfortunately, in reality, industrial bioprocesses contain more than one species; and therefore, the interactions between microbes should be represented by the DFBA model.

The scope of this project is to improve and extend the methodology proposed by Scott et al. (2018), using IPM to simulate and optimise DFBA models. Specifically, this project will focus on exploring the applicability of IPM on a co-culture systems.

The aims of this works are listed as follows:

- To present the the progress of models developed on the bioprocess industry found in the literature.
- To introduce the appropriate method to solve the models, highlighting the Interior Point Method methodology
- To improve the IPM methodology to design optimise and control bioprocesses

This report has the following structure: in Section 2, a literature survey of the models developed to simulate and control bioprocesses is presented. Section 3 detailed a list of the current solving techniques of the DFBA model, by no mean exhaustive, with their respective advantages and disadvantages. Section 4 explores the theory of the IPM and how it can be used to reformulate the DFBA model. In Section 5, the IPM methodology applied on co-culture microbial system has been detailed. A few examples of the simulation result of this methodology are presented as well. Conclusions and future research of this project are mentioned in Section 6.

2 Bioprocessing Model

Bioprocesses play a paramount industrial role as it is used extensively in the production of high value-added bioproducts, *e.g.* food, nutraceuticals, pharmaceuticals, biofuel, chemicals and polymers *etc.*, from living cells or their components. There has been a predicted growing market share of bio-based chemicals from 2% in 2008 to 22% in 2025 with a market size of 9.7 billion US dollar (Biddy et al. 2016). To successfully model the bioproducts (the desired material) in relation to all the inputs (the nutrients provided for the cells as well as reactor temperature *etc.*) enables better monitoring and subsequently controlling over the process to achieve the optimal outcome.

2.1 The unstructured model

As mentioned in the Introduction, classical kinetic models, such as Monod model, Droop model, Contois model and Luedeking-Piret model were used for modelling industrial bioprocesses despite they overlook the intracellular metabolic reactions. They are mostly constructed phenomenologically, if not empirically, with only the observable extracellular environmental information. These models, treating the cell as a black box, are often termed as unstructured model (Tziampazis & Sambanis 1994).

It is known that biological systems sometimes experience changes of phenotype. For example, the diauxic growth of *Escherichia coli* on either glucose or acetate under different conditions (Mahadevan et al. 2002) shows different cell behaviour will happen in simulating these cells. The classical kinetic models have trouble predicting these changes in biological states; therefore, their extrapolating ability severe limits their applicability in industry. Moreover, as reported by Vatcheva et al. (2006), to select one classical kinetic model over the other is not well-adapted due to the various assumptions to use under different circumstances, that further discounts the reliability of the unstructured model.

2.2 The structured model

With the advancement of the computation capability, systems biology emerged to shed light upon metabolic pathways inside the cells that enable the construction of models with more fundamental basis. Systems biology applies advance analytic toolkits to understand cellular interactions. With the knowledge it gains, metabolic networks are constructed by separating biomaterials into compartments with the help of biological knowledge and bridging different compartments with equations on stoichiometry and kinetics (Tziampazis & Sambanis 1994). The structured models created are therefore, incorporated with a holistic metabolic network of the cell.

Hollywood et al. (2006) pointed out the metabolic networks constructed can be directly linked with physiological states of the cell as well as the genomic information; Sonntag et al. (2011) indicated that the metabolic networks open doors to better understanding cell-based activities and precise computation simulation of bioprocesses. Moreover, in the field of biotechnology, to develop a modelling approach that capture the intracellular metabolic variation helps study of effect of genetic mutations (Hiesinger & Hassan 2005). Therefore, it is safe to say that the structured models are a promising tool to investigate cellular interactions.

2.3 Genome-scale reconstruction

A common approach to analyse the metabolic network is by reconstruction of the genome-scale metabolic network in which the different metabolic reactions are connected through their common substrates and products. To date, even though the kinetics for the enzyme inside the cell are unknown, metabolic network is constructed with the well-established reaction stoichiometry of biochemical reactions of the enzyme (Baart & Martens 2012). Orth J (2010) presented a systematic procedure to construct the metabolic network. In essence, genome-scale metabolic network reconstruction provides accurate metabolic pathway analysis. This network represents mathematically the mass (and energy) balances of the metabolites within the cell with linear equations:

$$\mathbf{N}\mathbf{v} = \mathbf{0} \tag{1}$$

$\mathbf{N} \in \mathbb{R}^{\tilde{m}\times n}$ with $\text{rank}(\mathbf{N}) = \tilde{m}$ is an full rank stoichiometric matrix. It represents the metabolic network of n reactions (fluxes) inside the cell and $\tilde{m}$ metabolites. Each column of $\mathbf{N}$ specifies the stoichiometry of the metabolite in a given reaction from the metabolic network, and $\mathbf{v} \in \mathbb{R}^n$ is the flux distribution of metabolites. However, the stoichiometric matrix in Equation (1) is usually underdetermined: it has more unknown fluxes than the number of metabolites ($\tilde{m} < n$). Therefore, it is not possible to find a unique solution of $\mathbf{v}$ just by solving Equation (1). To address this issue, some specific modelling frameworks have been proposed.

2.4 Flux Balance Analysis

FBA is a constraint-based model with the assumption that the metabolites are in a steady state condition. Based on genome-scale reconstruction of metabolic network, FBA adds a pre-defined metabolic objective, *e.g.* maximising biomass or energy reserve (ATP) rates with additional bound constraints (van Gulik & Heijnen 1995, Price et al. 2004):

$$\min_{\mathbf{v}} \quad -\mathbf{c}^{\mathrm{T}}\mathbf{v} \tag{2a}$$

$$\text{s.t.} \quad \mathbf{N}\mathbf{v} = \mathbf{0} \tag{2b}$$

$$\mathbf{v}^{\text{lo}} \leq \mathbf{v} \leq \mathbf{v}^{\text{up}} \tag{2c}$$

where $\mathbf{c} \in \mathbb{R}^n$ is a column vector of nonnegative entries, serving as a weighing factor for the biological objective; $\mathbf{v}^{\text{lo}} \in \mathbb{R}^n$ and $\mathbf{v}^{\text{up}} \in \mathbb{R}^n$ define the fixed lower and upper bounds of the flux distribution. Note, in most literature, a minimisation function is considered. In this work, a minimisation problem with a negated objective is used. Since the objectives and the constraints in the Equation (2) are all linear, FBA turns the problem into a linear programming (LP) problem.

FBA is a powerful tool to incorporate biochemical information in bioprocess simulation. However, it has two shortcomings.

The first shortcoming of FBA models is that they don't always guarantee the uniqueness of the solution, even though the objective function guides the search for the optimum in the polytope of feasible region formed by constraints. Given that the FBA model is an LP problem, the optimal solutions can always find itself located either on the vertex or the facet of the feasible polytope: in case when the optimal solution is at the vertex, a unique solution is found; in the other case where the objective function hyperplane is parallel to

a facet of the polytope, all points along the facet are "optimal solutions"–nonuniqueness issue appears. The development of the solving method for the FBA modelling approach is the focus of this work and it is detailed in the Section 3.

The second shortcoming of FBA is the subjectively selected objective function. The choice of objective function greatly impacts the model prediction accuracy (García Sánchez & Torres Sáez 2014). For example, there is no way to be sure if the cells trying to maximise its biomass or ATP rates during the whole processing time. Consequently, the reaction pathways the model highlighted might not be active in reality. Therefore, there is no a single objective function that can be used universally (Toya et al. 2011). There are different methods have been proposed in the pass to select the most appropriate objective function, such as ObjFind (Schuetz et al. 2007), Biological Objective Solution Search (BOSS) (Mahadevan et al. 2002), and Bayesian-based approaches (Knorr et al. 2006). Karlsen (2017) proposed a systematic data-based approach for finding both biological objective function and a minimum set of active constraints to limit the solution space. However, the detail of finding the most appropriate objective function for the FBA is not the main focus of this work. Hence the maximisation of the biomass is assumed to be the objective function in the following discussion with non-linear objectives to be a work to carry out in the future (see Section 6.2).

2.5 Dynamic Flux Balance Analysis

As mentioned above, FBA assumes steady state conditions. Although it may be true for continuous processes, most industrial bioprocesses are either batch or fed-batch which allow cells to grow through their life circle. Hence, a model to include cellular dynamics is needed.

Dynamic Flux Balance Analysis (DFBA), an extension of FBA, is introduced to add the dynamic variation of the metabolites in the modelling formalism of FBA, allowing the model to precisely capture the time evolution of the key extracellular species, given the microbial system is in (intracellular) quasi-steady state (QSS). QSS assumes the intracellular metabolic system quickly reaches equilibrium (steady state) compare to the extracellular environmental change. This has been proven to be a mild assumption which does not detract from the predictive ability of the methodology (Mashego et al. 2006, Kresnowati et al. 2008).

A classic DFBA model contains two parts, a dynamic part at the upper level (extracellular environment) and an optimisation part at the lower level (intracellular environment). The upper level has a system of equations the mass (and energy) balance of all species transporting through the cell membrane, while the lower level is similar to the LP problem of the FBA (see Equation (2)), except for the metabolic network in Equation (1) is expanded to include the species uptake rates.

However, a slight modification to the DFBA model has been proposed by Scott et al. (2018) that, instead of the assuming the uptake bounds of the model to be fixed, a new model with varied bounds that constrain the values for the uptake rates by the state variables is suggested:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{v}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{3a}$$

$$\mathbf{v}^{\text{up}} = \mathbf{g}(\mathbf{x}(t)), \tag{3b}$$

$$\mathbf{v}^{\text{lo}} = \mathbf{g}'(\mathbf{x}(t)) \tag{3c}$$

where $\mathbf{x} \in \mathbb{R}^d$ represents the time dependent extracellular concentration whose evolution is modelled by a continuous function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^d$. The upper and lower bounds of the specific uptake rates of n species of the cellular system are represented by the $\mathbf{v}^{\text{up}}, \mathbf{v}^{\text{lo}} \in \mathbb{R}^n$ in Equation (3b) and (3c) respectively. $\{\mathbf{g}, \mathbf{g}'\} : \mathbb{R}^d \to \mathbb{R}^n$ are continuous C^1 functions denoting the uptake rates of species. This modification has an important physical meaning that has been omitted by classic DFBA model. For example, in a case where two species (A and B) taken by a cell for metabolism, if species A has a low uptake rates, intuitively the metabolism reaction will be transfer limited which will in turn slow down the uptake of the other species B, therefore, species B will not meet its maximal uptake flux even though the objective function will maximise the metabolism rate. This phenomenon is shown by Acevedo et al. (2014), a trade-off exists between fluxes of species uptake. This modification improves the accuracy of the model, especially when the systems have more than one species that the uptake rate of one might be limited by that of the others.

Moreover, the FBA models usually include a subset $e \in n$ of fluxes with equal lower and upper bounds ($\mathbf{v}_{\text{e}}^{\text{up}} = \mathbf{v}_{\text{e}}^{\text{lo}}$). When simulating, these bound values will create problem. Therefore, these bounds are relaxed (by usually adding and subtracting an arbitrary value to the upper and lower bound), at the same time represents these bounds as equality constraints by appending rows to the matrix $\mathbf{N}$ in Equation (1) and expanding the metabolic network of the DFBA model, as formulated below:

$$\mathbf{A}\mathbf{v} = \begin{bmatrix} \mathbf{N} \\ \mathbf{E} \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_{\text{e}}^{\text{up}} \end{bmatrix} = \mathbf{b} \tag{4}$$

where $\mathbf{E} \in \mathbb{R}^{e \times n}$ has a single one in the positions of the eth-flux in subset e and zeros in the remaining entries. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a modified stoichiometric matrix after converting the fluxes with equal lower and upper bounds as equality constraints, where

$$m = \tilde{m} + e$$

is the total number of rows in the stoichiometric matrix $\mathbf{A}$ also represents the number of the equality constraints the DFBA model has. Since N is full rank, $\mathbf{A}$ will be full rank with the addition of independent equality constraints.

Hence the lower part of the DFBA model can be expressed as follow:

$$\min_{\mathbf{v}} \quad -\mathbf{c}^{\text{T}}\mathbf{v} \tag{5a}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{v} = \mathbf{b} \tag{5b}$$

$$\mathbf{v}^{\text{lo}}(\mathbf{x}) \leq \mathbf{v} \leq \mathbf{v}^{\text{up}}(\mathbf{x}) \tag{5c}$$

Combining the upper and lower part of the DFBA model, it is essentially a system of ordinary differential equations (ODEs) with an embedded LP problem:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{v}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{6a}$$

$$\mathbf{v}^{\text{up}} = \mathbf{g}(\mathbf{x}(t)) \tag{6b}$$

$$\mathbf{v}^{\text{lo}} = \mathbf{g}'(\mathbf{x}(t)) \tag{6c}$$

$$\mathbf{v}(t) \in \underset{\mathbf{w}}{\text{argmin}} \left\{ -\mathbf{c}^{\text{T}}\mathbf{w} | \mathbf{A}\mathbf{w} = \mathbf{b}(t), \quad \mathbf{v}^{\text{lo}}(\mathbf{x}) \leq \mathbf{w} \leq \mathbf{v}^{\text{up}}(\mathbf{x}) \right\} \tag{6d}$$

Note that the new variable $\mathbf{w}$ introduced here is to distinguish from $\mathbf{v}$, where the latter is the argument of the minimum obtaining from the LP problem.

Every time the lower level LP problem–Equation (6d)–is calculated, the optimal flux distribution obtained from the LP will alter the extracellular concentration via the system of ODEs (Equation (6a)), which in turn will change the uptake rates $\mathbf{g}(\mathbf{x}(t))$ and $\mathbf{g}'(\mathbf{x}(t))$; the upper and lower uptake rates of the species is then altered, whose impact subsequently propagate back in the lower level via the constraints of the LP problem. The key concept of DFBA is illustrated in Figure 1 below. To efficiently as well as accurately solve the DFBA model remains a heated topic, and will be detailed in the following sections.

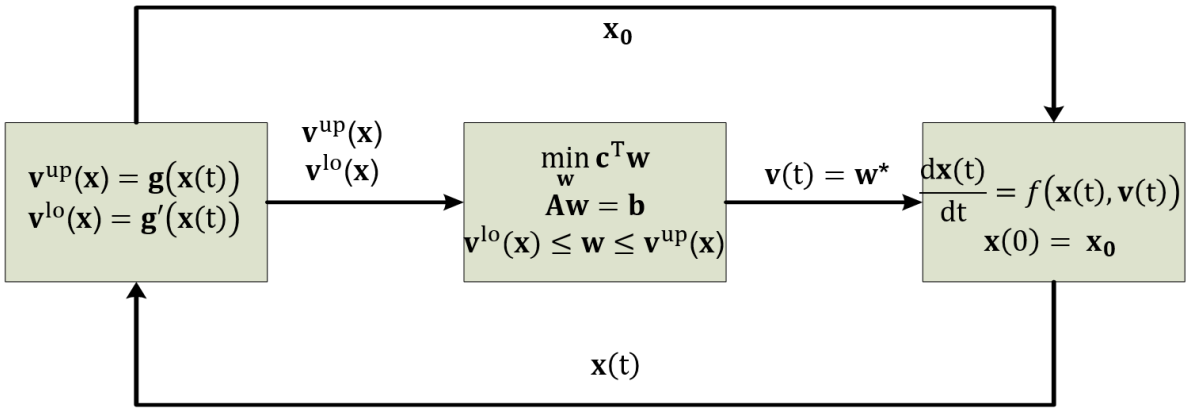

Figure 1: The schematic diagram of DFBA

2.6 Other models

There are alternative dynamic modelling approaches, such as structured kinetic models (Steinmeyer & Shuler 1989), log linear models (Hatzimanikatis et al. 1998) and cybernetic models (Jones & Kompala 1999) are limited by the need for *in vivo* enzyme kinetics (Henson & Hanly 2014), which as mentioned above, might not be readily estimated.

Apart from the above mentioned models, there is still strong candidates competing with DFBA model. As mentioned in Section 2.4, the FBA model (as well as DFBA) has a shortcoming that it requires a pre-set objective function. An alternative structured model, the Metabolic Flux Analysis (MFA) is proposed. MFA tries to solve the underdetermined metabolic system by leaving out insignificant or unobservable reactions to make the system exactly determined (R Antoniewicz 2013). Though inaccurate the MFA method is, several researches have built on the MFA modelling approach recently.

Leighty & Antoniewicz (2011), Fernandes et al. (2016) have proposed the Dynamic Metabolic Flux Analysis (DMFA) where no objective function nor constraint is needed. DMFA coupled the online measurement of data during the bioprocesses, and create DMFA time points accordingly. These DMFA time points then used to create inflection points on the prediction of flux changes, between which, the flux is assumed to be point-wise linear. However, this method create non-smooth function, implementation of the system-wise control base on this method is difficult.

Fernandes et al. (2016) managed to apply DMFA on batch process, while the microbial system they tested are relatively small size (60 metabolites and 72 reactions, *cf.* Section 3).

Toya et al. (2011) demonstrated a novel modelling approach, ^{13}C-MFA, is also free from a pre-set objective function. In essence, the author proposed to use ^{13}C-labelled

carbon source to fathom whether a reaction pathway is active or not. The resulting simulation yields a more realistic intracellular flux distribution with an extra effort spent *in vitro* to determine the cellular parameters.

However, as suggested by Henson & Hanly (2014), one of the important advantage of DFBA model is the direct incorporation of the metabolic reconstruction model in the dynamic system, this convenience is not found in other modelling approaches.

More importantly, some researches have applied FBA (or DFBA) model on industrial-scale processes and have validated against industrial process. Meadows et al. (2010) reported the FBA model captured the important trend despite a simplified model was used on an industrial fermentation process.

R Antoniewicz (2013) also acknowledged that FBA (or DFBA) modelling approach is useful for exploring the biotechnological potential of organisms and improving the bio-process product yield as well as identifying genetic manipulations that could improve cell-specific productivity.

Recent development in flux balance analysis has also seen the development of dynamic internally constrained FBA (dircFBA), which combines DFBA with internally constrained FBA (ircFBA) for enhanced performance (Karlsen 2017). Further development of the DFBA model can go into future study and therefore, only to be briefly touch at this point.

Looking through the literature, DFBA provides a practical alternative to incorporate intracellular structure effectively. Combined with its solving techniques, as detailed in the next section, DFBA can be a promising systems biology approach for comprehensive *in silico* simulate large-scale biological systems.

3 Current DFBA solution techniques

As mentioned in Section 2.5, the DFBA model, a system of ODEs with an embedded LP problem, is difficult to solve. The nonuniqueness of the embedded LP problem adds complexity to the model. Novel solving methods and toolkits have been proposed in recent years to address this issue.

Mahadevan et al. (2002) pioneered and presented two key approaches to solve the DFBA problem that have led to a proliferation of studies and application of DFBA. One is static optimisation approach (SOA) and the other called dynamic optimisation approach (DOA).

3.1 Static optimisation approach

SOA solves the DFBA problem by applying the forward Euler's method following the schematic outlined in Figure 1. it divides the operating time into several intervals, solves the LP problem at each time step followed by integration. Convenient and straightforward as this method is, SOA is sensitive to numerical instability. Due to most DFBA model are stiff, small time steps and large number of LP problems are needed to ensure convergence (Gomez et al. 2014). For example, Feng et al. (2012) applied SOA on DFBA problem with 774 reactions and 634 metabolites, they needed to solve around 400 LP problems, one every five minutes for the 35 hours simulation time. Therefore, SOA is computationally expensive and its capability to solve large-scale DFBA problem has been restricted.

3.2 Dynamic optimisation approach

DOA solves the DFBA problem with collocation method. It discretises the time horizon and the system of ODEs is transformed into a non-linear programming (NLP) problem. Detailed step to solve the DOA can be seen from Mahadevan et al. (2002), Biegler (2007) Compare to SOA, this method only need to solve the NLP once over the whole operating time. However, as evaluated by Mahadevan et al. (2002), solving a large size problem for DOA method is limited due to large number of variables and constraints being introduced in the collocation method.

3.3 Direct approach and DFBAlab

Another approach to solve the DFBA model is direct approach (DA).It is in essence including the LP solver in the right-hand side evaluator for the ODEs, and solve the system with implicit ODE integrator with adaptive step size for error control, thus it requires less step size as compared with SOA method (Gomez et al. 2014). However, as mentioned in Section 2.4, one shortcoming the FBA (or DFBA) model is that the LP problem may not always guarantee a unique solution, and hence will impede the integration of the implicit ODEs. In this regard, Gomez et al. (2014) presented a DA implementation, namely DFBAlab, to incorporate DA approach with lexicographic linear optimisation problem, which provides further objectives to the LP problem in cases where non-uniqueness is reached. Though this approach is efficient and elegant, as recognised by Scott et al. (2018), the lexicographic optimisation problem is inherently discrete and non-differentiable (hybrid system), which renders optimal control problem less desirable.

3.4 Karush-Kuhn-Tucker condition-based method

What is more, Zhao et al. (2017) demonstrated the use of a Karush-Kuhn-Tucker (KKT) condition-based solution approach for DFBA models with non-linear objective function in the lower level–the embedded LP problem. The KKT based solution approach transforms a DFBA model into a quasi-differential algebraic equation (DAE) system, which is then solved by classical DAE integrators. Regularity of the derived quasi-DAE system is ensured through an extreme-ray-based transformation in order to avoid changes in the active set. However, application of this method to genome-scale metabolic networks is a challenging task due to the computational demand of computing extreme rays. The proposed method cannot guarantee global optimality of the inner optimisation problem as KKT conditions are local optimality conditions. Solving DFBA models with guaranteed global optimality remains a challenging yet highly interesting research field (Zhao et al. 2017).

4 Theoretical background

With the information gathered from the Section 3, one can conclude that a method that is smooth, differentiable, accurate while at the same time computational less expensive is desired. In this section, a novel method proposed by Scott et al. (2018) using Interior Point Method (IPM) is presented following the theory behind using IPM.

4.1 Interior Point Method

Interior point methods are a class of algorithms used to solve constrained optimisation problems, linear or nonlinear. The key idea is to introduce a logarithmic barrier function in the objective function which takes up the inequality consrtaints as appended weighting factor. By following a path of optimising (either minimising or maximising depending on the original objective function) the logarithmic barrier family of problems, the so called "central path", the solution can be found. The theories underpinning the interior point method presented below are based on the understanding of Monteiro & Adler (1989), Biegler (2010). Note should be taken that, three assumptions have been made in order for the IPM to work, and they are assumed to be automatically fulfilled in the discussion below: the set of variable is non-empty throughout the integration time-span; the set of the dual variables as mentioned below is non-empty; and the equality constraint matrix $\mathbf{A}$ is full rank (the assumption has been stressed in Section 2.5).

Consider a Linear programming minimisation problem with constraints

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 \quad i = 1, 2 \ldots, k \\ & h_j(\mathbf{x}) = 0 \quad j = 1, 2 \ldots, m \end{aligned} \tag{7}$$

where $\mathbf{x} \in \mathbb{R}^n$; the objective function $f_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, the inequality constraints $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$, and the equality constraints $h_1, \ldots, h_m : \mathbb{R}^n \to \mathbb{R}$ are all linear functions. The equality constraints are linearly independent in order to meet the assumption with Monteiro & Adler (1989).

Given that the optimal solution is denoted by $\mathbf{x}^*$ and the optimal objective function value is denoted as $f_0(\mathbf{x}^*)$.This LP problem in Equation (7) can be solved by finding its first-order necessary conditions (KKT):

$$\begin{aligned} h_j(\mathbf{x}^*) = 0, \quad & j = 1, \cdots, m \\ f_i(\mathbf{x}^*) \leq 0, \quad & i = 1, \cdots, k \\ \nu_i^* \geq 0, \quad & i = 1, \cdots, k \\ \nu_i^* f_i(\mathbf{x}^*) = 0, \quad & i = 1, \cdots, k \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{k} \nu_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(\mathbf{x}^*) = 0 & \end{aligned} \tag{8}$$

which can be demanding in large-scale problem. Therefore, the Equation (7) is expressed alternatively by including the inequality constraints in the objective function $\min f_0(\mathbf{x})$, where

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) + \sum_{i=1}^{k} I\left(f_i(\mathbf{x})\right) \\ \text{s.t.} \quad & h_j(\mathbf{x}) = 0 \quad j = 1, 2 \ldots, m \end{aligned} \tag{9}$$

In this case, the original problem can be solved by minimising the new objective without introducing unknown multiplier (as the case for the inequality constraint). The indicator function I in the above Equation (9) is defined as

$$I(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \infty & \text{otherwise} \end{cases} \tag{10}$$

Considering $k = 1$ for simplicity (only one inequality constraint case), one can substitute the indicator into Equation (9), and compare it with the original problem: when $u \leq 0$ (*i.e.* $f_i(\mathbf{x}) \leq 0$, and the inequality constraint is met), $I(u)$ equals to zero, thus Equation (9) returns the original problem; in case when $u > 0$, $I(u)$ becomes infinity, which $\min f(\mathbf{v})$ will try to avoid. Similarly, if more than one inequality constraints are in place, the summation of the positive indicator function will always assure every I_i value to be zero, *i.e.* all inequality constraints are met.

However, this indicator function is not differentiable, hence, it is difficult to find the optimal solution by differentiation. To address this issue, a smooth function that approximate the indicator function in Equation (10) is necessary. An alternative choice would be a logarithmic function

$$I_\mu(u) = -\mu \ln(-u) \tag{11}$$

As such, it is only when $u \leq 0$, $I_\mu(u)$ has a meaning, while if $u > 0$, this indicator $I_\mu(u)$ shoot to positive infinity, which again, the minimisation problem trys to avoid. Here, μ is called the barrier (penalty) parameter, the smaller it is, the closer $I_\mu(u)$ is to the indicator $I(u)$ in Equation (10).

By varying the value of the barrier parameter, the closeness of the approximation can be adjusted. Therefore, the logarithmic barrier function is defined as

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) - \mu \sum_{i=1}^{k} \ln(-f_i(\mathbf{x})) \\ \text{s.t.} \quad & h_j(\mathbf{x}) = 0 \quad j = 1, 2 \ldots, m \end{aligned} \tag{12}$$

One can see that the objective function is convex–as both the original objective $f_0(\mathbf{x})$ as well as the negation of a logarithmic function are convex–the minimisation problem can be solved by KKT optimality conditions in Equation (8). The Jacobian $\mathbf{J}(\mathbf{x})$ of the objective function in Equation (12)

$$\mathbf{J}(\mathbf{x}) = \nabla f_0(\mathbf{x}) + \mu \sum_{i=1}^{k} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \nabla h_j(\mathbf{x}) \tag{13}$$

where λ is the Lagrange multiplier of the equality constraint for the primal problem. The Hessian matrix $\mathbf{H}(\mathbf{x})$ is

$$\mathbf{H}(\mathbf{x}) = \nabla^2 f_0(\mathbf{x}) + \sum_{i=1}^{k} \frac{\mu}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}) + \sum_{i=1}^{k} \frac{\mu}{f_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^{\mathrm{T}} + \sum_{j=1}^{m} \lambda_j \nabla^2 h_j(\mathbf{x}) \tag{14}$$

4.2 Central path

In order to solve the logarithmic barrier function as presented in Equation (12), the idea of "central path" is used. Given that the assumptions proposed in (Monteiro & Adler 1989) hold, for every value of $\mu > 0$, the Equation (12) has a unique solution $\mathbf{x}^*(\mu)$ that is feasible. In other words, it satisfies the constraints

$$h_j(\mathbf{x}^*) = 0 \quad j = 1, 2 \ldots, m \tag{15a}$$
$$f_i(\mathbf{x}^*) \leq 0 \quad i = 1, 2 \ldots, k \tag{15b}$$

as well as $\lambda^* \in \mathbb{R}^m$ such that the Jacobian function Equation (13) becomes

$$\mathbf{J}(\mathbf{x}) = 0 = \nabla f_0(\mathbf{x}^*) + \mu \sum_{i=1}^{k} \frac{1}{-f_i(\mathbf{x}^*)} \nabla f_i(\mathbf{x}^*) + \nabla \mathbf{h}\left(\mathbf{x}^*\right)^{\mathrm{T}} \lambda^* \tag{16}$$

From the above equation, one can see that each central point has a corresponding dual feasible solution that defines the lower bound of the primal problem. More specifically, combining the knowledge of Equation (15b) and (16), as $f_i(\mathbf{x}^*) \leq 0$ to define $\nu^*(\mu) > 0$ as

$$\nu^*(\mu) f_i\left(\mathbf{x}^*\right) = -\mu \quad i = 1, 2 \ldots, k \tag{17}$$

where ν is the Lagrange multiplier of the inequality constraint for the primal problem. By rewriting the KKT conditions and comparing with Equation (8):

$$\begin{aligned} h_j(\mathbf{x}(\mu)^*) &= 0, \quad j = 1, \cdots, m \\ f_i\left(\mathbf{x}(\mu)^*\right) &\leq 0, \quad i = 1, \cdots, k \\ \nu_i(\mu)^* &\geq 0, \quad i = 1, \cdots, k \\ \nu_i(\mu)^* f_i\left(\mathbf{x}(\mu)^*\right) &= \mu, \quad i = 1, \cdots, k \end{aligned} \tag{18}$$

$$\nabla f_0(\mathbf{x}^*(\mu)) + \sum_{i=1}^{k} \nu^*(\mu) \nabla f_i(\mathbf{x}^*(\mu)) + \nabla \mathbf{h}\left(\mathbf{x}^*(\mu)\right)^{\mathrm{T}} \lambda^*(\mu) = 0$$

For each value of μ, the optimal value of $\mathbf{x}^*(\mu)$ ensures that the Lagrangian function reaches minimum with dual feasible solutions $\lambda^*(\mu)$ and $\nu^*(\mu)$. The dual function $g\left(\lambda^*(\mu), \nu^*(\mu)\right)$ is bounded, and it has the following form:

$$\begin{aligned} g\left(\lambda^*(\mu), \nu^*(\mu)\right) &= f_0\left(x^*(\mu)\right) + \sum_{i=1}^{k} \nu^*(\mu) f_i\left(x^*(\mu)\right) + \mathbf{h}\left(\mathbf{x}^*(\mu)\right)^{\mathrm{T}} \lambda^*(\mu) \\ &= f_0\left(x^*(\mu)\right) - k\mu \end{aligned} \tag{19}$$

Therefore, this subsection demonstrated that as the logarithmic barrier function is solved for several values of the penalty parameter μ, as μ decreases, the result is a sequence of feasible points converging to a solution of the original problem. In other words, as $\mu \to 0$ the optimal solutions, $\mathbf{v}(\mu)$, $\lambda^*(\mu)$, and $\nu^*(\mu)$ converge to that of the primal and dual function.

4.3 Interior Point Method to solve DFBA

In late 2018, Scott et al. (2018) proposed to solve DFBA using IPM method. This is achieved by transforming the the DFBA model into a system of implicit ODEs. This new

approach can solve large-scale DFBA problem accurately and efficiently, and due to its differentiable ability with respect to optimisation parameters, it can be used to determine the optimal control strategies by solving the optimal control problem, as according to Vassiliadis et al. (1994*a*,*b*), where the system satisfied the constraints imposed by the DFBA model. In this subsection shows and explains the detail procedure of implementing the IPM to solve a DFBA problem. Note that the procedure presented in this section is mainly based on Section 4.1 and 4.2 the problem formulation proposed by Scott et al. (2018).

The lower level of the DFBA model shown in Equation (5), also called the primal problem is rewritten as below to remove the chained inequality constraint:

$$\begin{aligned} \min_{\mathbf{v}} \quad & -\mathbf{c}^{\mathrm{T}}\mathbf{v} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{v} = \mathbf{b} \\ & \mathbf{v}^{\mathrm{up}} - \mathbf{v} \geqslant \mathbf{0} \\ & \mathbf{v} - \mathbf{v}^{\mathrm{lo}} \geqslant \mathbf{0} \end{aligned} \tag{20}$$

As suggested in the section 4.1, a dual problem of Equation (20) is defined as

$$\begin{aligned} \max_{\lambda,\mathbf{y},\mathbf{z}} \quad & -\mathbf{b}^{\mathrm{T}}\lambda + \mathbf{y}^{\mathrm{T}}\mathbf{v}^{\mathrm{lo}} - \mathbf{z}^{\mathrm{T}}\mathbf{v}^{\mathrm{up}} \\ \text{s.t.} \quad & \mathbf{A}^{\mathrm{T}}\lambda - \mathbf{y} + \mathbf{z} = \mathbf{c} \\ & \mathbf{y} \geqslant \mathbf{0} \\ & \mathbf{z} \geqslant \mathbf{0} \end{aligned} \tag{21}$$

where where $\lambda \in \mathbb{R}^m$, $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ are the dual variables corresponding to the Lagrange multiplier for lower bound and upper bound of primal problem respectively. Then, the logarithmic barrier function similar to Equation (12) is constructed to handle the bounds on variables:

$$\begin{aligned} \min_{\mathbf{v}} \quad & -\mathbf{c}^{\mathrm{T}}\mathbf{v} - \mu \sum_{i=1}^{n} \left[\ln\left(v_i^{\mathrm{up}}(\mathbf{x}) - v_i\right) + \ln\left(v_i - v_i^{\mathrm{lo}}(\mathbf{x})\right)\right] \\ \text{s.t.} \quad & \mathbf{A}\mathbf{v} = \mathbf{b} \end{aligned} \tag{22}$$

One can see that Equation (22) will always yield a unique solution, regardless if the embedded LP problem in Equation (6) is unique or not. Therefore, IPM provided the solution to one of the shortcoming of DFBA as listed in Section 2.4. The Lagrangian of the Equation (22) is

$$L(\mathbf{v}, \lambda, \mu) = -\mathbf{c}^{\mathrm{T}}\mathbf{v} - \mu \sum_{i=1}^{n} \left[\ln\left(v_i^{\mathrm{up}} - v_i\right) + \ln\left(v_i - v_i^{\mathrm{lo}}\right)\right] + \lambda^{\mathrm{T}}\left(\mathbf{A}\mathbf{v} - \mathbf{b}\right) \tag{23}$$

The first-order necessary conditions (KKT) are calculated by taking differentiation with respect to the three variables of Lagrange function:

$$\mathbf{A}^T\lambda - \mathbf{y} + \mathbf{z} - \mathbf{c} = 0 \tag{24a}$$

$$\mathbf{A}\mathbf{v} = \mathbf{b} \tag{24b}$$

$$z_i\left(v_i^{\mathrm{up}} - v_i\right) = \mu \tag{24c}$$

$$y_i\left(v_i - v_i^{\mathrm{lo}}\right) = \mu, \quad i = 1, 2 \ldots, n \tag{24d}$$

As mentioned above in Section 4.2, as the barrier parameter $\mu \to 0$, the logarithmic barrier function, as shown in Equation (22), converges to the original primal function, Equation (20). Scott et al. (2018) also provided the same proposition.

Therefore, the DFBA model, as shown in Equation (6), can be written as a form penalised by the logarithmic barrier function instead of the system of ODEs with embedded LP problem:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{v}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{25a}$$
$$\mathbf{v}^{\text{up}} = \mathbf{g}(\mathbf{x}(t)) \tag{25b}$$
$$\mathbf{v}^{\text{lo}} = \mathbf{g}'(\mathbf{x}(t)) \tag{25c}$$
$$\mathbf{A}\mathbf{v}(t) = \mathbf{b}(t) \tag{25d}$$
$$z_i(t)\left(v_i^{\text{up}}(\mathbf{x}) - v_i(t)\right) = \mu \tag{25e}$$
$$y_i(t)\left(v_i(t) - v_i^{\text{lo}}(\mathbf{x})\right) = \mu, \quad i = 1, 2 \ldots, n \tag{25f}$$
$$-\mathbf{A}^T\lambda(t) + \mathbf{y}(t) - \mathbf{z}(t) = -\mathbf{c} \tag{25g}$$

This Equation (25) is a system of differential algebraic equations (DAEs). It can be further converted into a system of implicit ODEs by differentiation:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{v}), \quad x(0) = x_0 \tag{26a}$$
$$\frac{d\mathbf{v}^{\text{up}}}{dt} = \frac{d\mathbf{g}(\mathbf{x})}{d\mathbf{x}} f = \frac{d\mathbf{g}}{dt} \tag{26b}$$
$$\frac{d\mathbf{v}^{\text{lo}}}{dt} = \frac{d\mathbf{g}'(\mathbf{x})}{d\mathbf{x}} f = \frac{d\mathbf{g}'}{dt} \tag{26c}$$
$$\frac{dy_i}{dt} = -\frac{\mu}{\left(v_i - v_i^{\text{lo}}(\mathbf{x})\right)^2}\left(\frac{dv_i}{dt} - \frac{dg_i'(\mathbf{x})}{dt}\right), \quad i = 1, 2 \ldots, n \tag{26d}$$
$$\frac{dz_i}{dt} = -\frac{\mu}{\left(v_i^{\text{up}}(\mathbf{x}) - v_i\right)^2}\left(\frac{dg_i(\mathbf{x})}{dt} - \frac{dv_i}{dt}\right), \quad i = 1, 2 \ldots, n \tag{26e}$$
$$0 = \mathbf{A}\frac{d\mathbf{v}}{dt} \tag{26f}$$
$$0 = -\mathbf{A}^T\frac{d\lambda}{dt} + \frac{d\mathbf{y}}{dt} - \frac{d\mathbf{z}}{dt} \tag{26g}$$

The implicit ODEs can be further reduced mathematically so as to decrease the number of differential variables in numerical integration. It is done by combining the Equation (26d) and (26e), and grouping the differential terms together. The reduced form of the ODEs or R-iODE problem can be formulated as follow, as shown by Scott et al. (2018):

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{v}), \quad x(0) = x_0 \tag{27a}$$
$$\widetilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_{m\times m} \\ \mathbf{D}(\mu, \mathbf{v}, \mathbf{x}) & \mathbf{A}^{\text{T}} \end{bmatrix} \begin{bmatrix} \frac{d\mathbf{v}}{dt} \\ \frac{d\lambda}{dt} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{m\times 1} \\ \mathbf{r}(\mu, \mathbf{v}, \mathbf{x}) \end{bmatrix} \tag{27b}$$
$$\lambda(0) = \lambda_0, \quad \mathbf{v}(0) = \mathbf{v}_0 \tag{27c}$$

where $\mathbf{D}(\mu, \mathbf{v}, \mathbf{x})$ is a positive definite diagonal matrix with its entries defined as:

$$D_{i,i}(\mu, \mathbf{v}, \mathbf{x}) = \frac{\mu}{\left(v_i^{\text{up}}(\mathbf{x}) - v_i\right)^2} + \frac{\mu}{\left(v_i - v_i^{\text{lo}}(\mathbf{x})\right)^2} \quad i = 1, 2 \ldots, n \tag{28}$$

and $\mathbf{r}(\mu, \mathbf{v}, \mathbf{x}) \in \mathbb{R}^n$ is a vector with entries:

$$r_i = \frac{\mu}{\left(v_i^{\text{up}}(\mathbf{x}) - v_i\right)^2} \frac{dg_i}{dt} + \frac{\mu}{\left(v_i - v_i^{\text{lo}}(\mathbf{x})\right)^2} \frac{dg_i'}{dt}, \quad i = 1, 2 \ldots, n \tag{29}$$

As recognised by Scott et al. (2018), consistency needs to be maintained that the initial value $x(0)$, as the measured extracellular environment, must correspond to the solution of the primal-dual barrier formulation of the FBA problem at the initial time $(\lambda(0) = \lambda_0, \mathbf{v}(0) = \mathbf{v}_0)$, derived from the DFBA model with the same barrier parameter (μ).

5 Co-culture System

5.1 Motivation

It is known that microorganisms in nature tend to form a consortium of different species, the same can be observed in industrial bioprocesses, either naturally or artificially. One can expect interactions within the community influence the cellular kinetics of each other (Braga et al. 2016). Some research, for example Henson & Hanly (2014), suggested synergistic effect may occur within the consortia of microorganisms. Therefore, it would be preferable to understand the effect of these interactions during bioprocess design.

While it would be impractical to carry out *in vivo* experiments for different combinations of every strain of microbial species, computational simulations open doors to fast prescreening through the possible microbial combinations (Höffner et al. 2013, Henson & Hanly 2014). The integration of the computational and experimental approach lead to better product quality as well as optimal resource utilisation through potential synergistic effect (Hanly & Henson 2014).

There is only few research have conducted on co-culture system with DFBA model. A research group at the University of Massachusetts has investigated this topic before. Hanly & Henson (2011) studied two non-interacting species co-cultured in a dynamical environment and Hanly & Henson (2014) investigated the inhibitory effect of one species on the other in a batch. However, both studies used the SOA method as described in Section 3.1. Its modelling capability has been limited. Most of the other simulation work carried out in the past have been focus on developing metabolic networks on an isolated species. Hence more work needs to be done on simulating microbial community with novel simulation toolkits.

Section 2 concludes that DFBA model can be a good candidate in exploring biotechnological potentials of organisms; and in Section 4, a novel, efficient and accurate simulation method, the interior point method, has been formulated to solve the DFBA model. The author of this work considers improving the simulation of DFBA co-culture microbial system with a IPM reformulation. In this subsection, undergoing work has been carrying out by the author is detailed.

Note, at this stage, no experimental data has been provided to the author for comparing to. Therefore, only simulation work based on well established microbial DFBA models has been presented in order to validate and support the reliability of the IPM method. In the following subsections, microbial model selection has been partially based on Scott et al. (2018) and a master thesis (Srisuma 2018), though all the results are done by the author. Future study with experimental data can provide predictive insights of industrial-scale bioprocesses.

5.2 Single culture simulation

Firstly, the simulation has been carried out on single species. The original DFBA model was adapted from Equation (6):

$$\frac{dx(t)}{dt} = v_b(t)x(t), \qquad x(0) = x_0 \tag{30a}$$

$$\frac{dS(t)}{dt} = v_s(t)x(t), \qquad S(0) = S_0 \tag{30b}$$

$$\frac{dP(t)}{dt} = m_P v_p(t)x(t), \qquad P(0) = P_0 \tag{30c}$$

$$v_s^{\text{up}} = g(\mathbf{x}(t)) = -m_S \frac{q_{S,max} S(t)}{S(t) + K_S} \tag{30d}$$

$$v_s^{\text{lo}} = \mathbf{g}'(\mathbf{x}(t)) = m_S \frac{q_{S,max} S(t)}{S(t) + K_S} \tag{30e}$$

$$\mathbf{v}(t) \in \underset{\mathbf{w}}{\text{argmin}} \left\{ -\mathbf{c}^{\mathrm{T}}\mathbf{w} | \mathbf{A}\mathbf{w} = \mathbf{b}(t), \quad \mathbf{v}^{\text{lo}}(\mathbf{x}) \leq \mathbf{w} \leq \mathbf{v}^{\text{up}}(\mathbf{x}) \right\} \tag{30f}$$

where $\mathbf{x} = \{x, S, P\}$ represents the extracellular environment containing the concentration of biomass, x; the concentration of sole substrate, in the case study below, the glucose S; and the concentration of sole product, P. All three variables have units in g/L. v_b, v_s, v_p are elements of the vector $\mathbf{v}$, they represents the flux of the biomass, substrate and product respectively with unit h^{-1}. For simplicity, cell death is not considered in this report despite it can be included in future work. As mentioned in Section 2.5, the species uptake rates are expressed as the lower and/or upper bounds of their respective fluxes $v_s^{\text{lo}} < v_s(t) < v_s^{\text{up}}$. Equation (30d) shows the upper limit $v_s^{\text{up}} \in \mathbf{v}^{\text{up}}$ of the uptake rate of the substrate, to be assumed to follow the Michaelis-Menten equation as suggested by Scott et al. (2018), with $q_{S,max}$ the maximum specific uptake rate, K_S the affinity constant and m_S is the molecular mass of substrate. Similarly, the m_P in the above equation is the molecular mass of the product. Lastly, the lower bound in Equation (30e) of the uptake flux in this report is taken as the negative of the upper bound value, while a lower bound of zero might be better: it make physical sense because the consumption of the substrate is irreversible reaction. However, both are correct mathematically since the substrate uptake rate would always hit the upper bound in order to maximise the biological objective. A zero value for the lower bound would be presented in the future work.

A strain of *Komagataella Phaffii* with 1221 metabolites and 1474 reaction fluxes from Cankorur-Cetinkaya et al. (2017) is modelled with the extra data shown in the Table 1.

Table 1: Parameter used for the *K. phaffii* model

Variable	Value	Unit	Reference
m_S	0.18	g/mmol	-
x_0	10^{-3}	g/L	-
S_0	20	g/L	-
$q_{S,max}$	6	mmol/gDWh	(Heyland et al. 2011)
K_S	0.385	g/L	(Prielhofer et al. 2013)

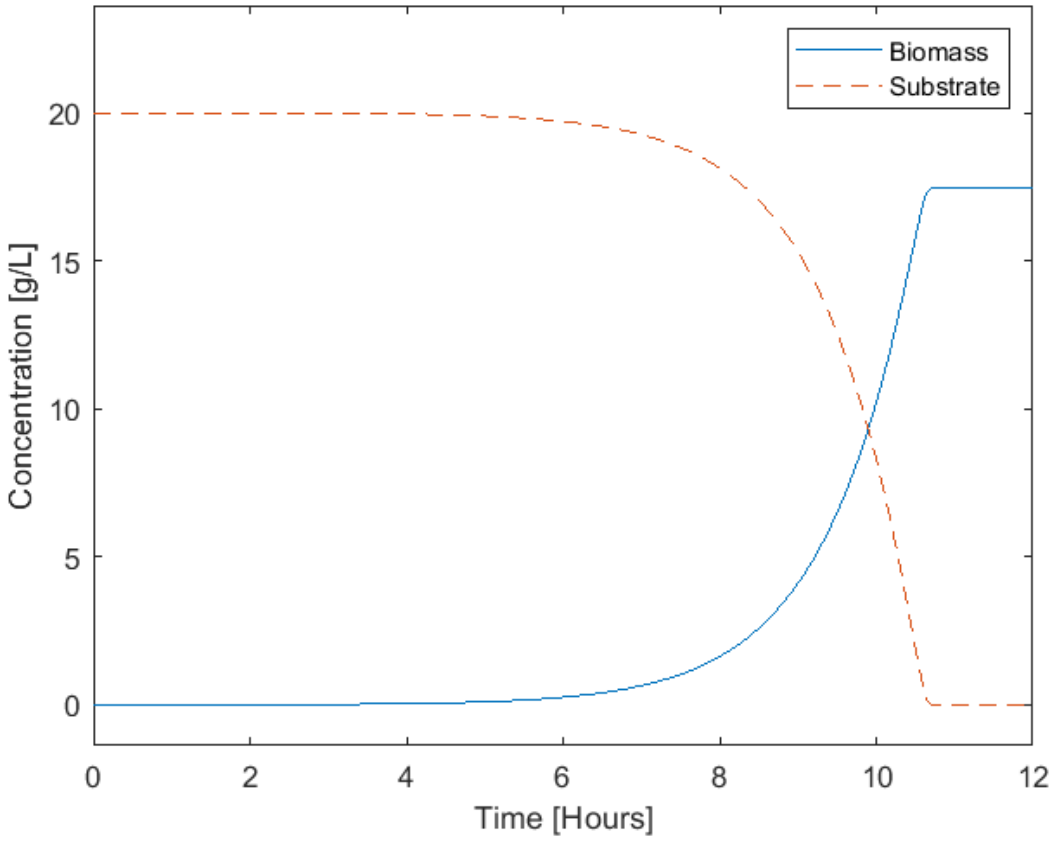

Figure 2: The concentrations time profile of isolated *K. phaffii* against glucose concentration

Another species model used is *Saccharomyces cerevisiae* with 2220 metabolites and 3493 reaction fluxes, as reported by Aung et al. (2013). Similarly, the extra data are collected and shown Table 2.

Table 2: Parameter used for the *S. cerevisiae* model

Variable	Value	Unit	Reference
m_S	0.18	g/mmol	-
x_0	10^{-3}	g/L	-
S_0	20	g/L	-
$q_{S,max}$	12.9	mmol/gDWh	(van Dijken et al. 1993)
K_S	6.95	g/L	(Maier et al. 2003)

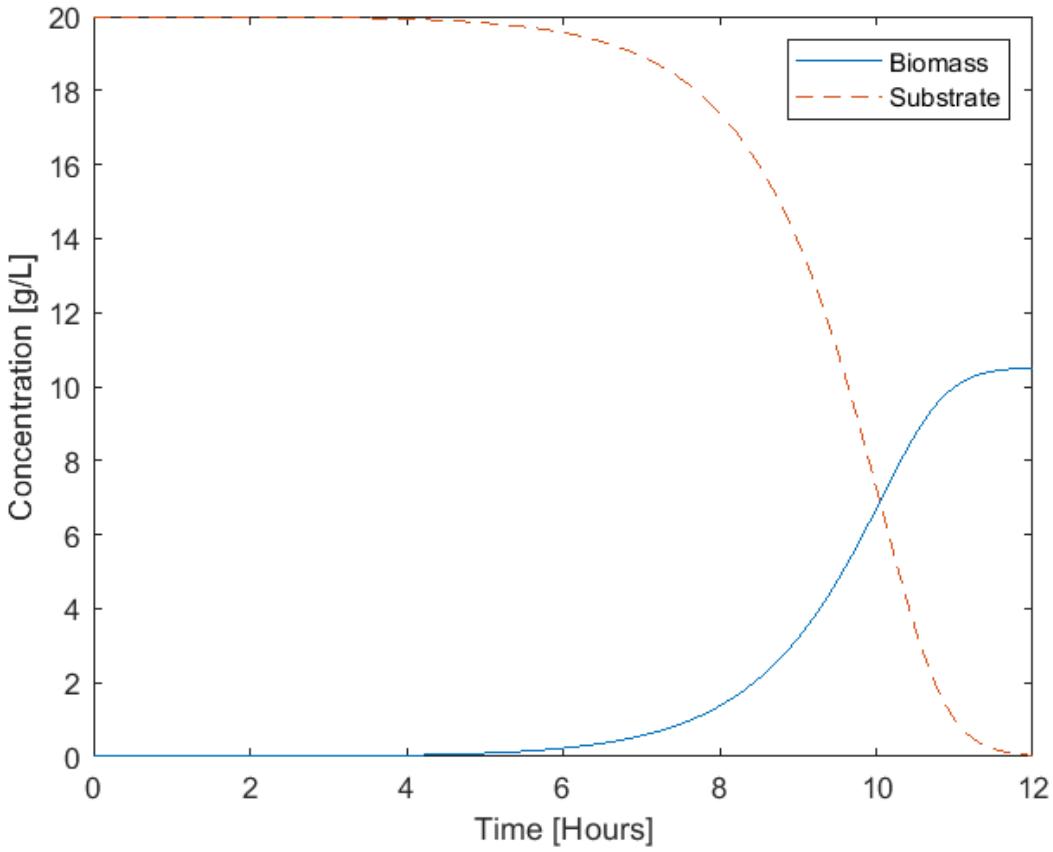

Figure 3: The time profile of isolated *S. cerevisiae* against glucose concentration

The simulation is run in MATLAB, the initial condition is calculated with CPLEXLP Optimizer developed by IBM.

The biomass concentration evolution plotted against the glucose concentration spanning the 12 hours operating time of isolated systems of *K. phaffii* and *S. cerevisiae* are shown in Figure 2 and 3 respectively.

5.3 Co-culture simulation

After the single-cell systems have been simulated. a co-culture model can be developed by combining both metabolic networks. In cases that the species are not interacting with each other but share the common medium, as modelled by Hanly & Henson (2011) with SOA method, the substrate uptake term in Equation (30b) need to expand to accommodate the sharing of common nutrient. While in other cases, where there is (chemical) interaction between two species, the effect of the interaction can be expressed as in the uptake rate term.

In this subsection, a batch process contains a co-culture of *K. phaffii* and *S. cerevisiae* under oxygen-limited conditions is simulated. In a typical batch process, the initial substrate is provided, while no extra substrate is provided. Oxygen is limited, although in this example it is not being modelled, it is sufficiently low that *S. cerevisiae* will produce ethanol (P) as a competitive strategy to inhibit the substrate uptake of other species, in this case *K. phaffii*, follows a non-competitive pattern (Brown et al. 1981).

In one study, it is shown that the ethanol production rate of *S. cerevisiae* deceases as the concentration of the glucose in the environment deceases (Krishnan et al. 1999). The

maximal ethanol production rate of *S. cerevisiae* is 10 mmol/(gDWh) (Heux et al. 2006), which is used to set the lower bound v_P^{lo} of the ethanol production flux. v_P^{lo} is gradually relaxed as the glucose concentration decreases.

The reason for setting the ethanol production rate as the lower bound may seem counter-intuitive. However, since the problem objective is to maximise the total biomass production, the simulation model will always prioritise using glucose to maximise biomass than to produce ethanol. Therefore, the ethanol production flux will always hits the lower bound.

Therefore the DFBA model, adapted from Equation (6) is:

$$\frac{dx_1(t)}{dt} = v_{b1}(t)x_1(t), \qquad x_1(0) = 10^{-3}\text{g/L} \tag{31a}$$

$$\frac{dx_2(t)}{dt} = v_{b2}(t)x_2(t), \qquad x_2(0) = 10^{-3}\text{g/L} \tag{31b}$$

$$\frac{dS(t)}{dt} = v_{s1}(t)x_1(t) + v_{s2}(t)x_2(t), \qquad S(0) = 20\text{g/L} \tag{31c}$$

$$\frac{dP(t)}{dt} = m_P v_p(t)x(t), \qquad P(0) = 0\text{g/L} \tag{31d}$$

$$v_{s1}^{\text{up}} = g_1(\mathbf{x}_1(t)) = -m_S \frac{q_{S1,max}S(t)}{(S(t)+K_{S1})(1+\frac{P(t)}{K_i})} \tag{31e}$$

$$v_{s1}^{\text{lo}} = \mathbf{g'}_1(\mathbf{x}_1(t)) = m_S \frac{q_{S1,max}S(t)}{(S(t)+K_{S1})(1+\frac{P(t)}{K_i})} \tag{31f}$$

$$v_{s2}^{\text{up}} = g_2(\mathbf{x}_2(t)) = -m_S \frac{q_{S2,max}S(t)}{S(t)+K_{S2}} \tag{31g}$$

$$v_{s2}^{\text{lo}} = \mathbf{g'}_2(\mathbf{x}_2(t)) = m_S \frac{q_{S2,max}S(t)}{S(t)+K_{S2}} \tag{31h}$$

$$\mathbf{v_1}(t) \in \underset{\mathbf{w_1}}{\text{argmin}} \left\{ -\mathbf{c_1}^{\text{T}}\mathbf{w_1} | \mathbf{A_1}\mathbf{w_1} = \mathbf{b_1}(t), \quad \mathbf{v_1}^{\text{lo}}(\mathbf{x_1}) \le \mathbf{w_1} \le \mathbf{v_1}^{\text{up}}(\mathbf{x_1}) \right\} \tag{31i}$$

$$\mathbf{v_2}(t) \in \underset{\mathbf{w_2}}{\text{argmin}} \left\{ -\mathbf{c_2}^{\text{T}}\mathbf{w_2} | \mathbf{A_2}\mathbf{w_2} = \mathbf{b_2}(t), \quad \mathbf{v_2}^{\text{lo}}(\mathbf{x_2}) \le \mathbf{w_2} \le \mathbf{v_2}^{\text{up}}(\mathbf{x_2}) \right\} \tag{31j}$$

$$m_P v_P^{\text{lo}} = \begin{cases} 10 & S > 7.5\text{g/L} \\ 2 & 2 < S \le 7.5\text{g/L} \\ 0 & S \le 2\text{g/L} \end{cases} \tag{31k}$$

where P is the concentration of ethanol produced, v_P is one element of the *S. cerevisiae* flux representing the ethanol production flux in h^{-1}. Note that subscript 1 represents *K. phaffii* and subscript 2 represents *S. cerevisiae*. To avoid confusion, subscript numbering denotes different species in this section rather than the elemental-wise numbering of a vector unless mention otherwise.

Equation (31e) defines the non-competitive inhibition of the ethanol produced on the substrate uptake of *K. phaffii* with K_i represents the ethanol inhibition constant of *K. phaffii*.

Equation (31k) is adapted from Srisuma (2018), where the molecular mass of ethanol $m_P = 0.046g/mmol$.

In the Figures shown below (Figure 4, Figure 5 and Figure 6) the interaction between *K. phaffii* and *S. cerevisiae* are presented with $K_i = 3$.

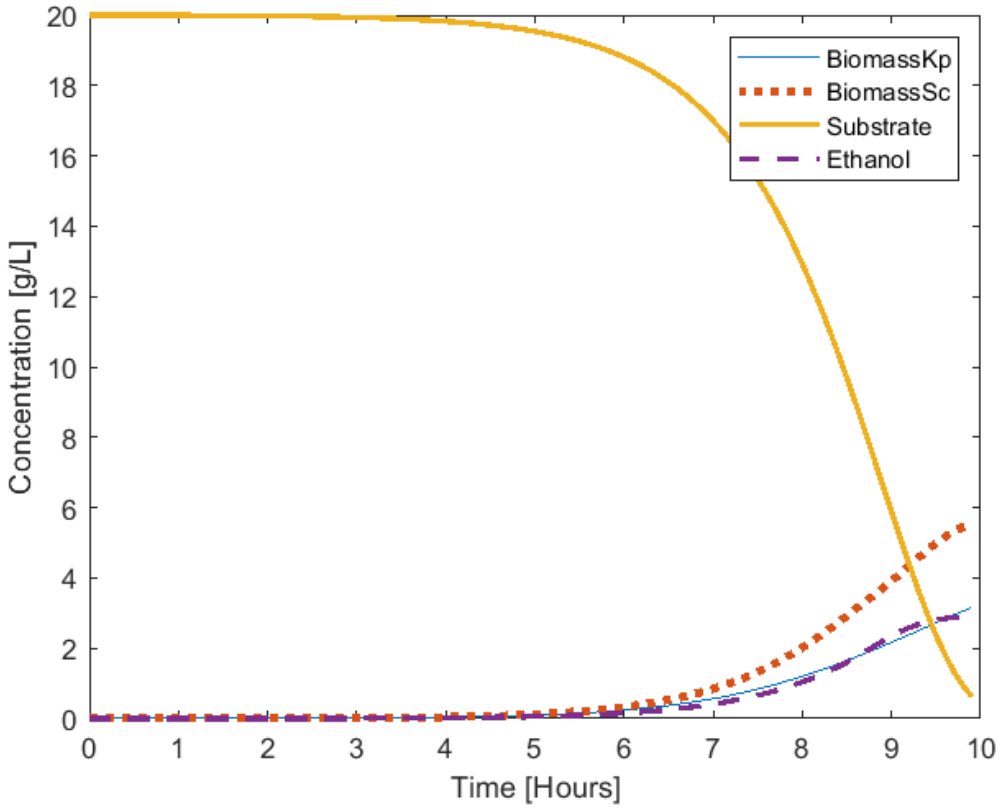

Figure 4: The concentrations time profile of *K. phaffii* and *S. cerevisiae*, substrate and ethanol

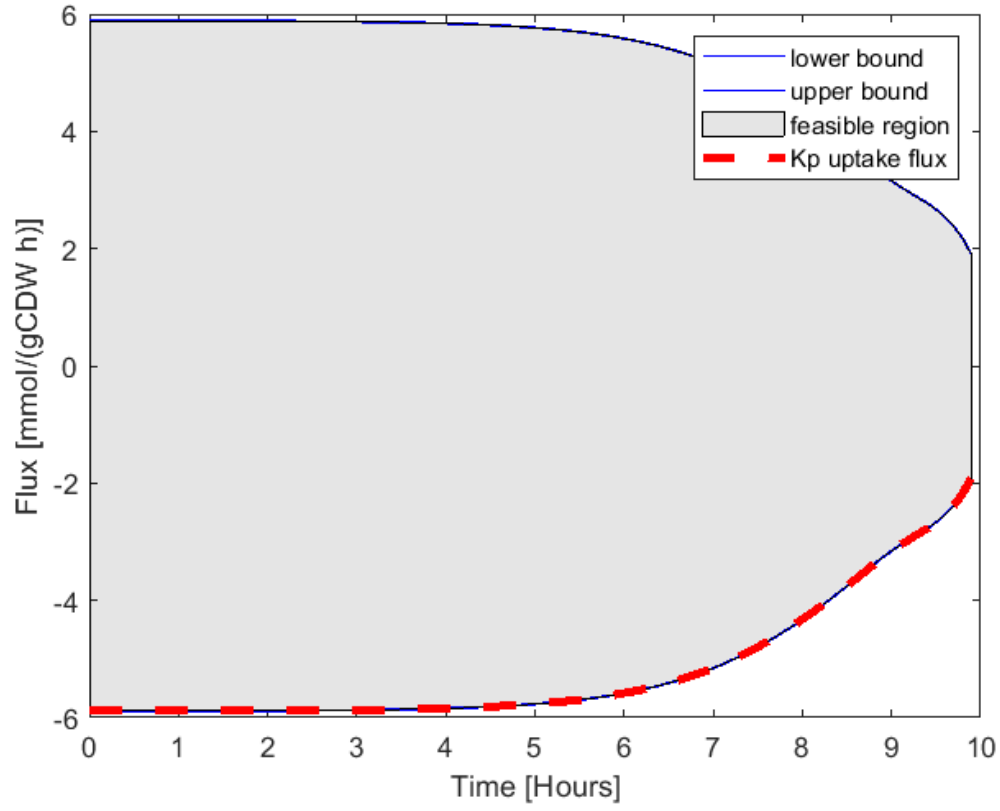

Figure 5: The time profile of *K. phaffii* substrate uptake rate and its bound

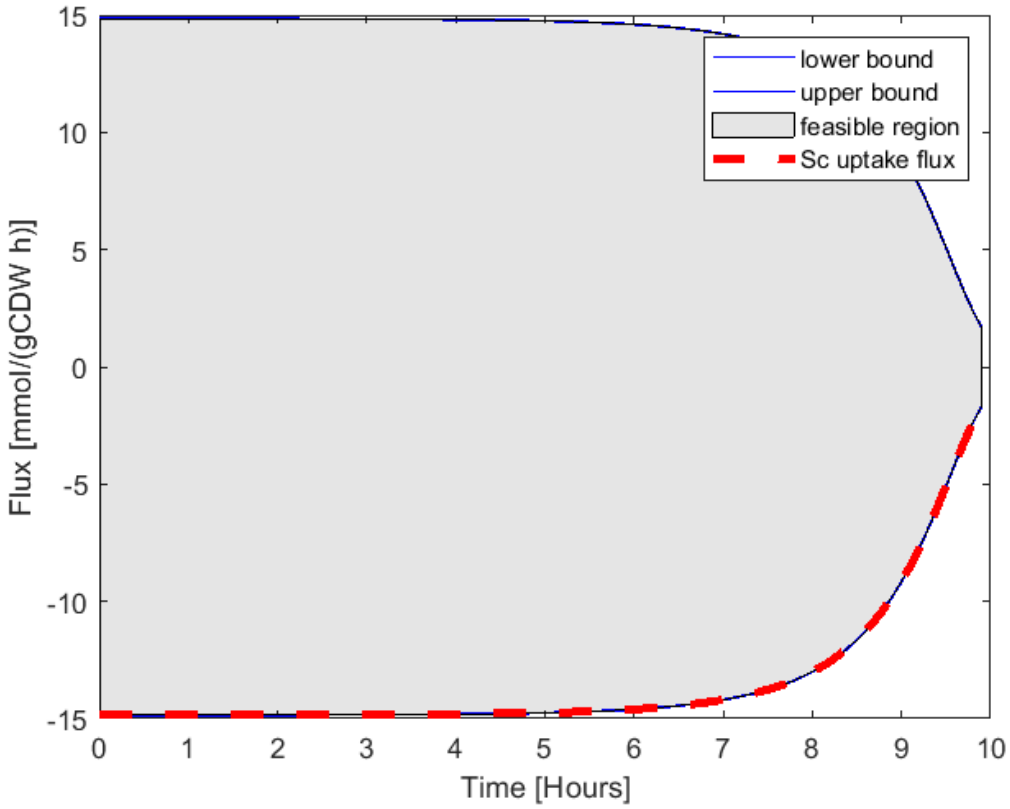

Figure 6: The time profile of *S. cerevisiae* substrate uptake rate and its bound

Noted that in DFBA model, the materials going into the cell are generally denoted as a negative value, therefore the upper bounds of the substrate uptake flux are negative.

5.4 Dynamic co-culture simulation

The co-culture system in Section 5.3 is a batch process. In this section, a fed-batch process is considered where the substrate is continuously fed into the reactor. The same co-culture system described in the above section is maintained, while more complex factors are considered:

$$\frac{dV}{dt} = F, \quad V(0) = 0.5\text{L} \tag{32a}$$

$$\frac{d(x_1(t)V)}{dt} = v_{b1}(t)x_1(t)V, \qquad x_1(0) = 10^{-3}\text{g/L} \tag{32b}$$

$$\frac{d(x_2(t)V)}{dt} = v_{b2}(t)x_2(t)V, \qquad x_2(0) = 10^{-3}\text{g/L} \tag{32c}$$

$$\frac{d(S(t)V)}{dt} = FS_f v_{s1}(t)x_1(t)V + v_{s2}(t)x_2(t)V, \qquad S(0) = 20\text{g/L} \tag{32d}$$

$$\frac{d(P(t)V)}{dt} = m_P v_p(t)x(t)V, \qquad P(0) = 0\text{g/L} \tag{32e}$$

$$v_{s1}^{\text{up}} = g_1(\mathbf{x}_1(t)) = -m_S \frac{q_{S1,max}S(t)}{(S(t) + K_{S1})(1 + \frac{P(t)}{K_i})} \tag{32f}$$

$$v_{s1}^{\text{lo}} = \mathbf{g'}_1(\mathbf{x}_1(t)) = m_S \frac{q_{S1,max}S(t)}{(S(t) + K_{S1})(1 + \frac{P(t)}{K_i})} \tag{32g}$$

$$v_{s2}^{\text{up}} = g_2(\mathbf{x}_2(t)) = -m_S \frac{q_{S2,max}S(t)}{S(t) + K_{S2}} \tag{32h}$$

$$v_{s2}^{\text{lo}} = \mathbf{g'}_2(\mathbf{x}_2(t)) = m_S \frac{q_{S2,max}S(t)}{S(t) + K_{S2}} \tag{32i}$$

$$\mathbf{v_1}(t) \in \underset{\mathbf{w_1}}{\text{argmin}} \left\{ -\mathbf{c_1}^{\text{T}}\mathbf{w_1} | \mathbf{A_1}\mathbf{w_1} = \mathbf{b_1}(t), \quad \mathbf{v_1}^{\text{lo}}(\mathbf{x_1}) \leq \mathbf{w_1} \leq \mathbf{v_1}^{\text{up}}(\mathbf{x_1}) \right\} \tag{32j}$$

$$\mathbf{v_2}(t) \in \underset{\mathbf{w_2}}{\text{argmin}} \left\{ -\mathbf{c_2}^{\text{T}}\mathbf{w_2} | \mathbf{A_2}\mathbf{w_2} = \mathbf{b_2}(t), \quad \mathbf{v_2}^{\text{lo}}(\mathbf{x_2}) \leq \mathbf{w_2} \leq \mathbf{v_2}^{\text{up}}(\mathbf{x_2}) \right\} \tag{32k}$$

$$m_P v_P^{\text{lo}} = \begin{cases} 10 & S > 7.5\text{g/L} \\ 2 & 2 < S \leq 7.5\text{g/L} \\ 0 & S \leq 2\text{g/L} \end{cases} \tag{32l}$$

where V is the (liquid) volume of the reactor, F is the feed volumetric flowrate, and S_f is the inlet stream glucose concentration in g/L. The differential equations can be simplified by product rule so that the resultant equation can be used in MATLAB.

In the following simulation run, the feed flowrate was chosen to be $0.044L/h$, inlet glucose concentration was $100g/L$, and the K_i value was set to 5. All other parameters can be found in Table 1 and Table 2.

Figure 7 shows the concentration profile of *K. phaffii*, *S. cerevisiae* substrate and ethanol in a fed-batch process. While Figure 8 shows the substrate uptake rate and its upper and lower bounds for *S. cerevisiae*. The substrate uptake rate for *K. phaffii* follows the same trend and hence omitted from this report.

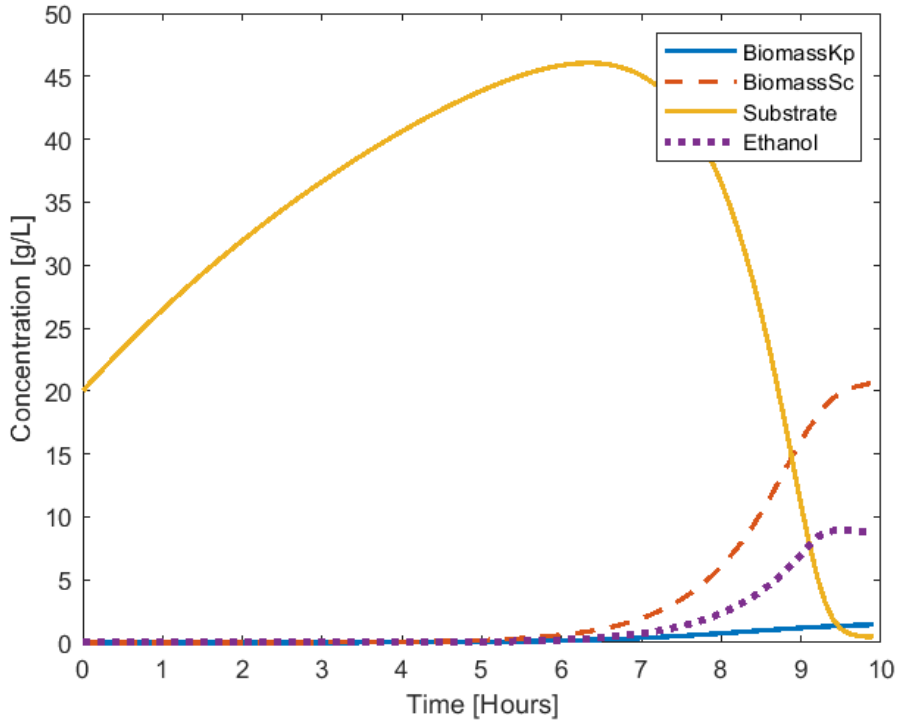

Figure 7: The concentrations time profile of *K. phaffii* and *S. cerevisiae*, substrate and ethanol for a fed-batch process

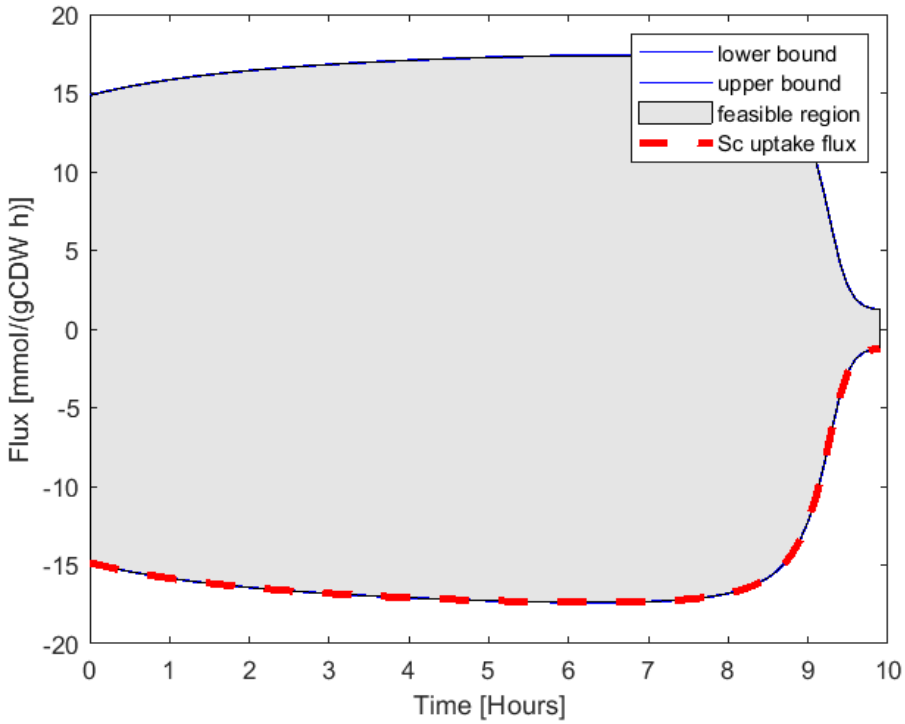

Figure 8: The time profile of *S. cerevisiae* substrate uptake rate and its bound for a fed-batch process

In this example, the volumetric volume is increased linearly with time, as steered by the

constant inlet flowrate. The constant influx of glucose sees an initial surge in the glucose (substrate) concentration. The deluge of glucose in the extracellular environment causes the cells to uptake glucose almost at maximal rate–mathematically, the upper bounds of both species relaxed, the uptake rates increase driven by the objective function. This can be seen as the surge in the biomass concentrations in Figure 7, especially that of *S. cerevisiae*; and as the widening of the feasible region in Figure 8.

At the same time, as the glucose concentration is high, the ethanol production rate keeps at the highest rate. Consequently its inhibitory effect suppresses the growth of *K. phaffii*. By comparing Figure 7 with Figure 4, with the continuous introduction of substrate, the dominance of *S. cerevisiae* over *K. phaffii* can be achieved.

5.5 Optimal control of the fed-batch co-culture

From the example simulation of Section 5.3 and 5.4, it can be seen that the final concentrations of the two species can be varied due to operating conditions, *i.e.* in this case due to the introduction of a fixed inlet flow of substrate.

In real-life industrial processes, there are some criteria need to meet during (path constraint) or at the end (end-point constraint)of the process. In bioprocess industry, an end-point constraint can be crucial for quality control. For example, the final composition of yeasts in the culture can affect the quality of the wine in wine-making industry (Antonelli et al. 1999). Therefore, to successfully yield the desired outcome by control variables is a subject of interest.

In this example, the fed-batch co-culture model as presented in Section 5.4 is optimised with a piece-wise constant feed flowrate control. An arbitrary control objective, the final concentrations of *S. cerevisiae* and *K. phaffii* are the same, is proposed. The least square of the final concentration difference is the problem objective to be minimised (Srisuma 2018). The problem objective is:

$$\min_{F(\cdot)} \left(1 - \frac{x_1}{x_2}\right)^2 \Bigg|_{t_f} \tag{33}$$

The constraints of this example is system of implicit ODEs from Equation (32). An extra bound constraints for the control variable was placed

$$0 \leq F(t) \leq 0.2 \quad L/h$$

What is more, the total simulation time was set as 14 hours, during which 7 control intervals were set. In other words, the inlet flowrate can be adjusted in every two hours. The optimal control problem presented in this section was solved using MATLAB embedded nonlinear solver "fmincon". The solution provided are local optimal solution obtained by "brutal force"– to test all the possible solutions in the solution space and to present the optimal solution within its neighbour. The simulation took several hours to complete, with more than 40 local optima found. Figure 9 and 10 only shows one of the local optima control and its concentration time profile.

In this control example, one can see that in order the reduce the gap between biomass concentrations, the extracellular substrate concentration should be kept at a lower value. One control strategy is to introduce almost no inlet flowrate of substrate in the first 8 hours and increased amount in the final 4 hours. It makes sense as in the absence of large

amount of substrate enables the growth of the *K. phaffii*. This can be seen as the increase in the *K. phaffii* concentration after 8 hours.

In future work, advance optimisation and control strategy, for example dynamic programming, can be coupled with the IPM methodology to improve the efficiency of searching the global solution for control, should that be required. However, the author may argue that in this example, some of the local optima are, within a certain tolerance, achieving the objective set by the problem. In reality, the marginal differences between the local optima are not as critical in making control strategies. Noted that this example is only a simplified problem of the real-life, complex process, where more constraints and a more realistic objective function are in place.

With the introduction of more constraints, either end-point constraints or path constraints, the feasible optimal solution will decrease. This would limit the set of control strategies to choose from. For example the concentration of ethanol should always below a certain value, or the total volume of the reactor should be lower than a certain value, *etc.* Moreover, the change of objective function may also make the simulation more practical. For example, the final total biomass concentration, is optimised, as in Section 5.4, while the control objective in this example is formulated into a constraint.

In conclusion, it is very important to note that the simulation work is acting as a prescreening procedure, but not a surrogate to *In situ* experiment. The model should be built based on the industrial data and realistic objectives.

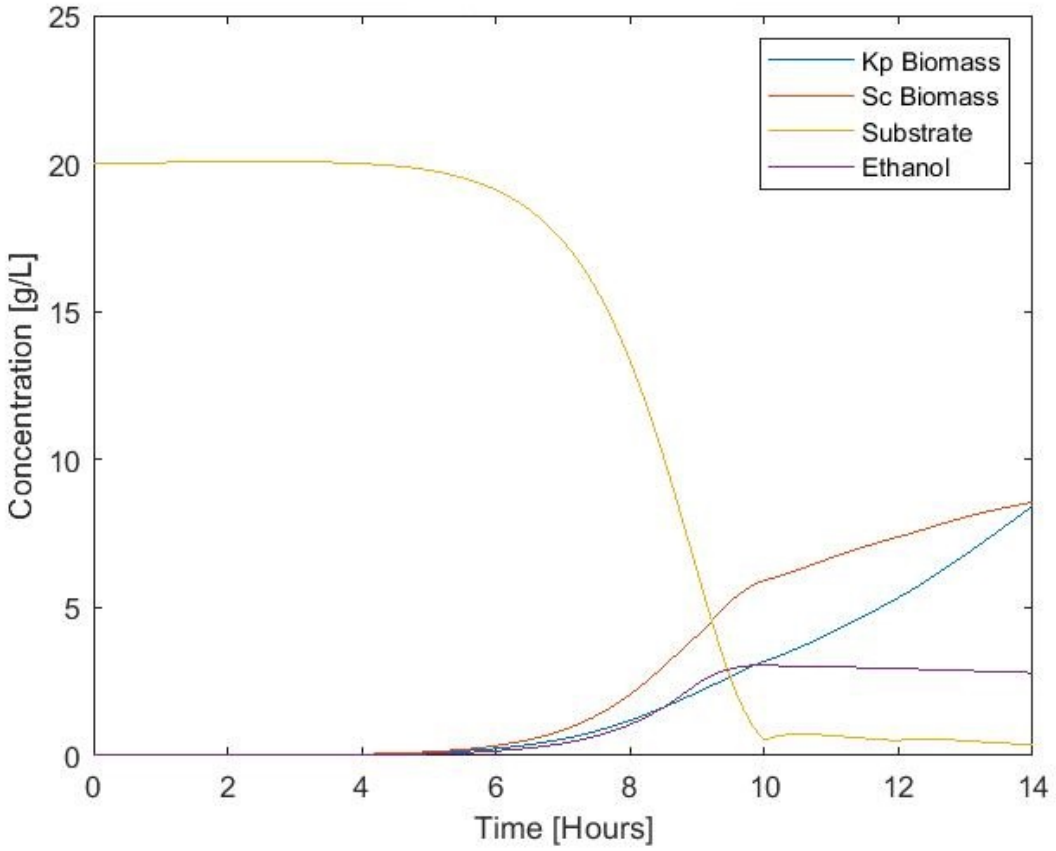

Figure 9: The concentrations time profile of *K. phaffii* and *S. cerevisiae*, substrate and ethanol with control

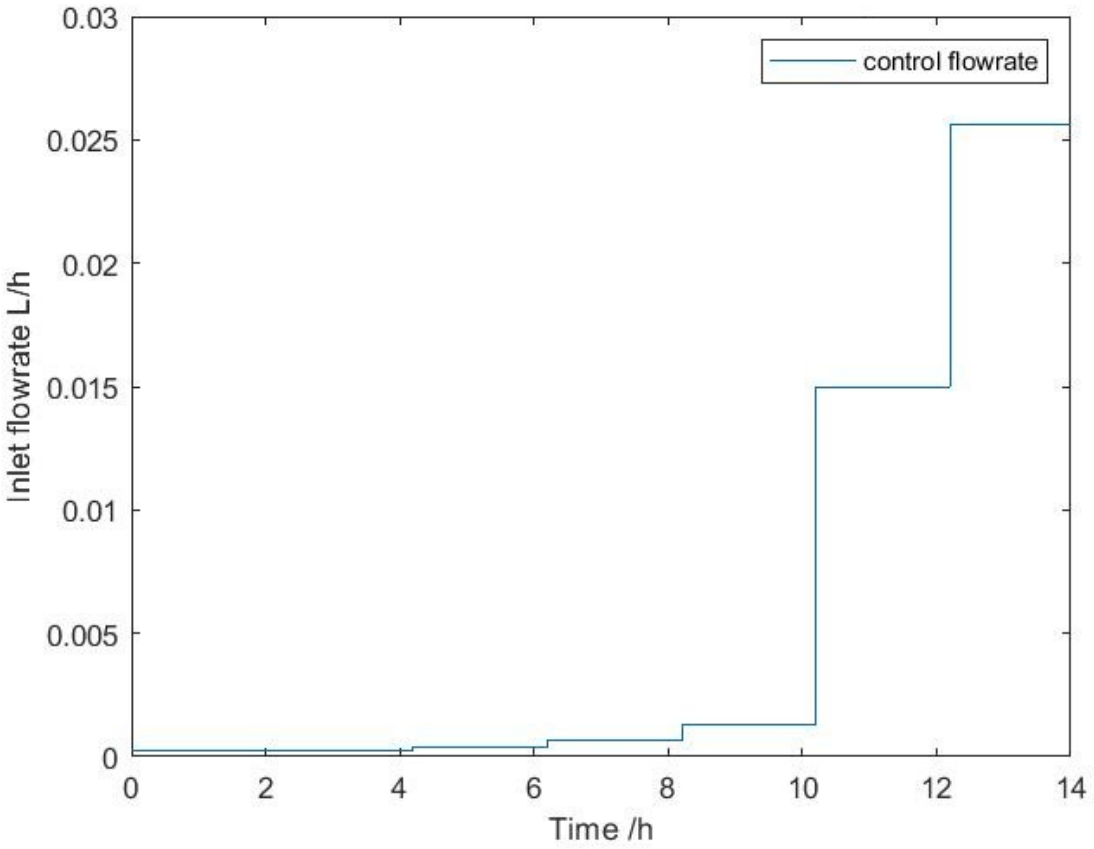

Figure 10: Control steps–the optimal feed flowrate

It can be shown that the using IPM reformulation to solve the DFBA model can be used in co-culture design, optimisation and control. This project can potentially be useful in several ways of a real-world complex microbial process. Noted that this methodology needs to be simulated and validated against industrial-scale data. Only in this case the reliability of this methodology can be evaluated and its applicability in this field can be established.

6 Conclusions and future direction

6.1 Conclusions

This report presents the construction of the dynamic flux balance analysis (DFBA) model to simulate the bioprocess system. With both advantages limitation given, the work argues for the industrial application of the DFBA model.

Afterwards, the current DFBA solution techniques are listed, highlighting a mathematical formulation of the interior point method (IPM), which has only been proposed to solve DFBA very recently. The solution of the IPM is smooth and differentiable, which is promising to be used as for optimal control application and possibly many more. The IPM reformulation is detailed in this work, with the equation provided.

With the methodology of using IPM to solve the DFBA model, its applicability has been test on co-culture microbial systems. The interactions between species in the culture are incorporated into the model as the substrate uptake rates, which in turns affect the cell's metabolism and cell growth. Examples have been provided for batch and fed-batch co-culture systems in Section 5.3 and 5.4 respectively. What is more, due to the differentiability of the IPM methodology, the optimal control over the fed-batch co-culture has been tested in Section 5.5. While the results seemed to make sense, an important note to take is that simulation work can only act as a prescreening procedure, where its decisions should based on the industrial or experimental information. Work can be done to refine the equations being incorporated into the model, improving the efficiency of the optimal control problem, or exploring other application of the IPM methodology.

The future work can be divided into two parts, one short-term work focus on co-culture systems and one long-term work concerns improving the methodology with an expansion of application.

6.2 Short-term future work

In the next two to three quarters, co-culture systems are the main focus. As mentioned in Section 5, there is still refining work needed in terms of problem formulation and numerical solver selection, especially in terms of optimal control simulation. More importantly, the simulation should be run on new sets of microbial consortia with industrial-scale data to compare with. Ultimately, the simulation results will be compiled into a paper submitted for publication.

There could be additional features to be included in this work. The potential inclusion of genetic mutants of a certain species in the co-culture(Hanly & Henson 2014). Note that in this short-term plan, links between genetic modification and the phenotype are not consider, while it may be a topic for long-term study. During short-term phase, genetic modifications are only to be expressed as the extracellular uptake equations.

The co-culture systems with the genetically manipulated species can offer information in maximisation of bioprocesses in the future.

6.3 Long-term future work

Long-term future work mainly concerns with the completion of the IPM methodology. As suggested by Gianchandani et al. (2010), the DFBA model facilitates the hypothesis-driven research in developing novel biotechnologies. Hjersted et al. (2007), Raman & Chandra (2009) investigated the effect of genetic perturbation, such as gene insertion and

deletion, and the effect on the productivity of the cell culture. The links between the gene and reactions in the metabolic network, and how manipulation of the stoichiometric matrix, as described in Section 2.3 are to be investigated. The information can be incorporated to simulation genetically manipulated bio-systems. It will be a interesting research topic as it can be used to design of more efficient metabolic pathways for the production of target products.

Another area of research is in Sensitivity analysis of each reaction pathway, as reported by Conejeros & Vassiliadis (2000). How sensitive the reaction pathway obtained by the IPM methodology is towards final process objectives can offer extra information towards designing bioprocess. Future work on the construction of sensitivity analysis over the DFBA model can be based on previous work done by Conejeros & Vassiliadis (1998*b*, *a*).

Apart from linear objective functions for the DFBA model, non-linear objective functions discussed in Section 2 can also be investigated. What is more, the model predictive control (MPC) scheme can be incorporated in the optimal control strategy in order to expand the applicability of the DFBA model. If time allows, cell signalling, can be incorporated into the study as part of the metabolic network.

All the results gathered during the development of each stage of the project will be compiled for the publication of peer-reviewed papers. The objectives of this project are the following:

- in-depth simulation of co-culture model
- the expansion of the capability of the IPM methodology by incorporating additional simlation scheme

The proposed work plan is presented in Figure 11 as a Gantt chart. Times intervals are divided into year quarters: Q1 corresponds to the period from January to March, Q2 from April to June, *etc.*

Furthermore, the author has participated in the writing of a review paper, just submitted for the publication in the journal "Computers and Chemical Engineering", titled "Optimal Control on Chemical Processes: Past, Present and Future". Besides, there is a list of the lectures attended during the nine-month period at Cambridge, which can be found in Section 7.

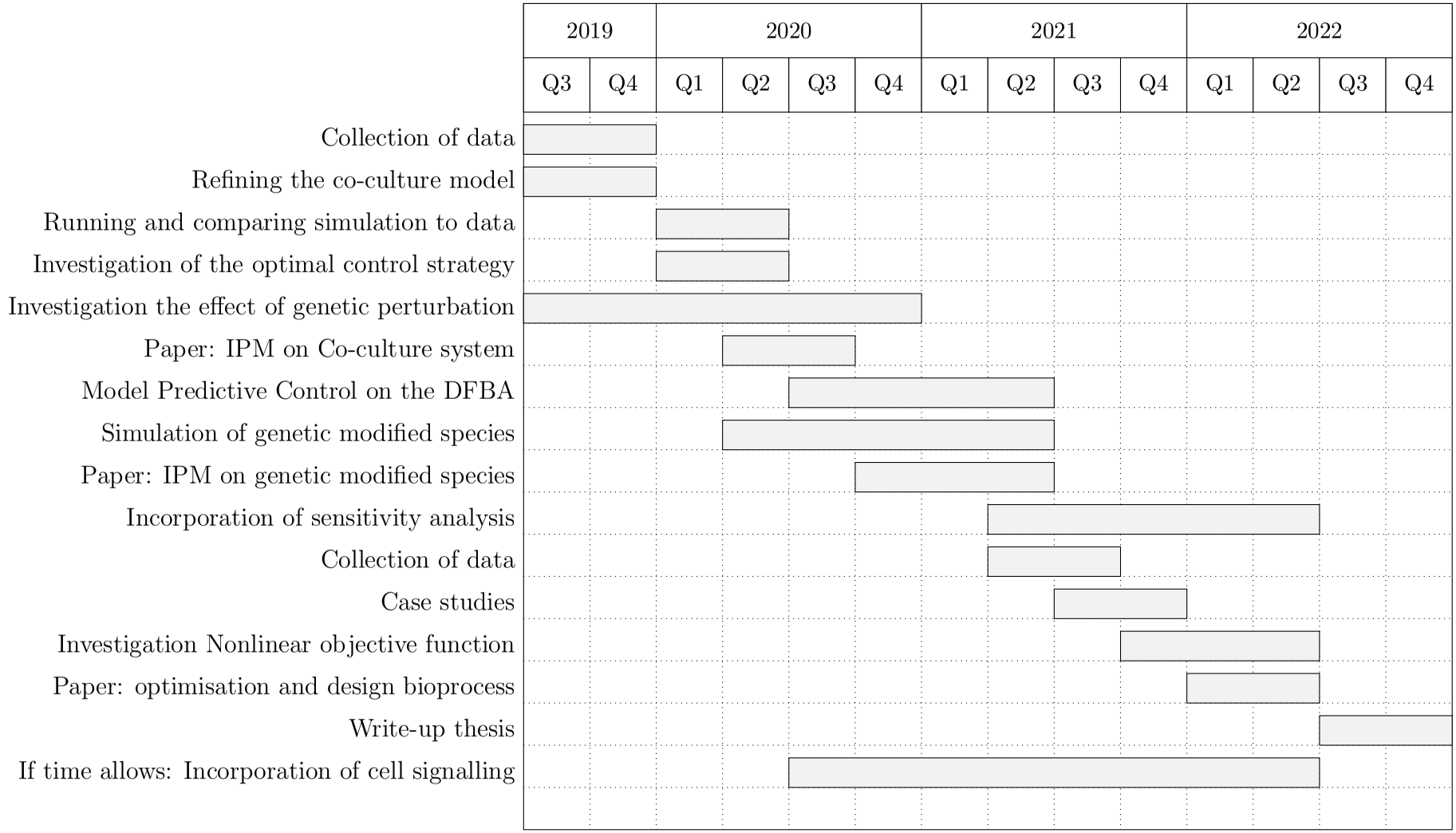

Figure 11: Gantt chart of future work.

7 Lecture list

Date	Title	Venue
31 Oct	Healthy eating: Emulsion microstructure engineering for salt or sugar reduction - Prof Bettina Wolf	CEB
6 Dec	CPSE annual consortium meeting	Imperial College London
1 Mar	Biotechnology meets breeding in baker's yeast - Prof Ed Louis	CEB
19 Mar	Bio-inspired photonics: from nature to applications - Dr Silvia Vignolini	CEB
Apr	AI - Dr Sean Holden	Cambridge
4 Apr	Roger Sargent Memorial	Imperial College London
17 Apr	The Ribosome - from far past to near future - Prof Ada E. Yonath	Cambridge
24 Apr	Atomistic Machine Learning between Physics and Data - Prof Michele Ceriotti	Cambridge
1 May	An Editor's Insight into Publishing in Nature Journals - Dr Giulia Pacchioni	Cambridge
30 May	Modes of Collective Cell Migration on 2- and 3-D Substrata - Prof Chwee Teck (C.T.) Lim	Cavendish Laboratory
12 Jun	Process Systems Engineering (PSE) tools and big data for bio-resource recovery - Prof Brent Young	CEB
Incoming	Engineering Language Unit training course	Department of Engineering
Incoming	CEB Graduate conference	CEB

Publication submitted		
Date	Title name	Journal name
15 Jun	Optimal Control on Chemical Processes: Past, Present and Future	Computers and Chemical Engineering

Nomenclature

Abbreviations

DW	Dry weight of biomass
DA	Direct Approach
DAE	Differential Algebraic Equation
DFBA	Dynamic Flux Balance Analysis
DOA	Dynamic Optimisation Approach
FBA	Flux Balance Analysis
IPM	Interior Point Method
KKT	Karush-Kuhn-Tucker (condition) , page 10
LP	Linear Programming
ODE	Ordinary Differential Equation
QSS	Quasi-Steady State
SOA	Static Optimisation Approach

Greek characters

λ, λ^*	Lagrange multiplier for equality constraints, and its optimal value
μ	Barrier (penalty) parameter
ν, ν^*	Lagrange multiplier for inequality constraints, and its optimal value

Roman characters

$\widetilde{\mathbf{A}}$	Stoichiometric matrix of the R-iODE problem, see equation (27)
$\mathbf{A}$	Stoichiometric matrix of the DFBA model
$\mathbf{b}$	Modified vector containing equality constraints
$\mathbf{c}$	Weighing factor of the biological objective
$\mathbf{D}$	Matrix of the R-iODE problem , see equation (27)
$\mathbf{E}$	Stoichiometric matrix of fluxes with equal lower and upper bounds
$\mathbf{g}'$	Uptake rate lower bound functions
$\mathbf{g}$	Uptake rate upper bound functions
$\mathbf{H}$	Hessian matrix
$\mathbf{h}$	Generic form of equality constraints

$\mathbf{J}$	Jacobian matrix	
$\mathbf{N}$	Stoichiometric matrix of metabolic network	
$\mathbf{r}$	Vector of the R-iODE problem , see equation (27)	
$\mathbf{v}, \mathbf{w}$	Vector of the flux distribution	
$\mathbf{v}^{\text{lo}}$	Lower bound vector of the flux distribution	
$\mathbf{v}^{\text{up}}$	Upper bound vector of the flux distribution	
$\mathbf{v}_{\text{e}}$	Fluxes with equal lower and upper bounds	
$\mathbf{x}$	Time dependent concentration of extracellular species	
$\mathbf{x}^*$	Optimal objective solution , page 10	
$\mathbf{x}_0$	Initial concentration of extracellular species	
$\mathbf{y}$	Dual variable	
$\mathbf{z}$	Dual variable	
$\tilde{m}$	Number of metabolites in the metabolic network	
e	Number of fluxes with equal lower and upper bounds	
F	Feed volumetric flowrate	L/h
$f(.)$	Generic form of continuous function	
$I(u)$	Indicator function of u	
$I_\mu(u)$	Logarithmic barrier function of u	
k	Number of total inequality constraints	
K_i	Inhibition constant	g/L
K_S	Affinity constant	g/L
L	Lagrangian function	
m	Number of total equality constraints	
m_S, m_P	Molecular weight of substrate and product	$g/mmol$
n	Number of fluxes (reactions) in the metabolic network	
P, P_0	Product (also ethanol in this work), and initial product concentration	g/L
$Q_{S,max}$	Maximum specific uptake rate	$mmol/gDWh$
S, S_0	Substrate (also glucose in this work), and initial substrate concentration	g/L
S_f	Inlet stream substrate concentration	g/L
V	Volume of the reactor	L
v_b, v_S, v_P	Element of the flux distribution	h^{-1}

References

Acevedo, A., Aroca, G. & Conejeros, R. (2014), 'Genome-Scale NAD(H/+) Availability Patterns as a Differentiating Feature between Saccharomyces cerevisiae and Scheffersomyces stipitis in Relation to Fermentative Metabolism', *PLOS ONE* **9**(1), 1–12.
URL: *https://doi.org/10.1371/journal.pone.0087494*

Antonelli, A., Castellari, L., Zambonelli, C. & Carnacini, A. (1999), 'Yeast Influence on Volatile Composition of Wines', *Journal of Agricultural and Food Chemistry* **47**(3), 1139–1144.
URL: *https://doi.org/10.1021/jf9807317*

Aung, H. W., Henry, S. A. & Walker, L. P. (2013), 'Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism', *Industrial Biotechnology* **9**(4), 215–228.
URL: *https://doi.org/10.1089/ind.2013.0013*

Baart, G. J. E. & Martens, D. E. (2012), Genome-Scale Metabolic Models: Reconstruction and Analysis, *in* M. Christodoulides, ed., 'Neisseria meningitidis: Advanced Methods and Protocols', Humana Press, Totowa, NJ, pp. 107–126.

Biddy, M. J. G., Scarlata, C. & Kinchin, C. (2016), Chemicals from Biomass: A Market Assessment of Bioproducts with Near-Term Potential, Technical report, United States.
URL: *https://www.osti.gov/servlets/purl/1244312*

Biegler, L. T. (2007), 'An overview of simultaneous strategies for dynamic optimization', *Chemical Engineering and Processing: Process Intensification* **46**(11), 1043 – 1053.
URL: *http://www.sciencedirect.com/science/article/pii/S0255270107001122*

Biegler, L. T. (2010), *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Braga, R. M., Dourado, M. N. & Araújo, W. L. (2016), 'Microbial interactions: ecology in a molecular perspective.', *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]* **47 Suppl 1**(Suppl 1), 86–98.

Brown, S. W., Oliver, S. G., Harrison, D. E. F. & Righelato, R. C. (1981), 'Ethanol inhibition of yeast growth and fermentation: Differences in the magnitude and complexity of the effect', *European journal of applied microbiology and biotechnology* **11**(3), 151–155.
URL: *https://doi.org/10.1007/BF00511253*

Cankorur-Cetinkaya, A., Dikicioglu, D. & Oliver, S. G. (2017), 'Metabolic modeling to identify engineering targets for Komagataella phaffii: The effect of biomass composition on gene target identification', *Biotechnology and Bioengineering* **114**(11), 2605–2615.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.26380*

Conejeros, R. & Vassiliadis, V. S. (1998*a*), 'Analysis and Optimization of Biochemical Process Reaction Pathways. 1. Pathway Sensitivities and Identification of Limiting

Steps', *Industrial & Engineering Chemistry Research* **37**(12), 4699–4708.
URL: *https://doi.org/10.1021/ie980410k*

Conejeros, R. & Vassiliadis, V. S. (1998*b*), 'Analysis and Optimization of Biochemical Process Reaction Pathways. 2. Optimal Selection of Reaction Steps for Modification', *Industrial & Engineering Chemistry Research* **37**(12), 4709–4714.
URL: *https://doi.org/10.1021/ie980411c*

Conejeros, R. & Vassiliadis, V. S. (2000), 'Dynamic biochemical reaction process analysis and pathway modification predictions.', *Biotechnology and bioengineering* **68**(3), 285–297.

Feng, X., Xu, Y., Chen, Y. & Tang, Y. J. (2012), 'Integrating Flux Balance Analysis into Kinetic Models to Decipher the Dynamic Metabolism of Shewanella oneidensis MR-1', *PLOS Computational Biology* **8**(2), 1–11.
URL: *https://doi.org/10.1371/journal.pcbi.1002376*

Fernandes, S., Robitaille, J., Bastin, G., Jolicoeur, M. & Wouwer, A. V. (2016), 'Application of Dynamic Metabolic Flux Convex Analysis to CHO-DXB11 Cell Fed-batch Cultures', *IFAC-PapersOnLine* **49**(7), 466 – 471.
URL: *http://www.sciencedirect.com/science/article/pii/S2405896316305936*

García Sánchez, C. E. & Torres Sáez, R. G. (2014), 'Comparison and analysis of objective functions in flux balance analysis', *Biotechnology Progress* **30**(5), 985–991.
URL: *https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/btpr.1949*

Gianchandani, E. P., Chavali, A. K. & Papin, J. A. (2010), 'The application of flux balance analysis in systems biology', *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**(3), 372–382.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.60*

Gomez, J. A., Höffner, K. & Barton, P. I. (2014), 'DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis', *BMC Bioinformatics* **15**(1), 409.
URL: *https://doi.org/10.1186/s12859-014-0409-8*

Hanly, T. J. & Henson, M. A. (2011), 'Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures', *Biotechnology and Bioengineering* **108**(2), 376–385.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.22954*

Hanly, T. J. & Henson, M. A. (2014), 'Dynamic model-based analysis of furfural and HMF detoxification by pure and mixed batch cultures of S. cerevisiae and S. stipitis', *Biotechnology and Bioengineering* **111**(2), 272–284.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.25101*

Hatzimanikatis, V., Emmerling, M., Sauer, U. & Bailey, J. E. (1998), 'Application of mathematical tools for metabolic design of microbial ethanol production.', *Biotechnology and bioengineering* **58**(2-3), 154–161.

Henson, M. A. & Hanly, T. J. (2014), 'Dynamic flux balance analysis for synthetic microbial communities', *IET Systems Biology* **8**(5), 214–229.
URL: *https://digital-library.theiet.org/content/journals/10.1049/iet-syb.2013.0021*

Heux, S., Sablayrolles, J.-M., Cachon, R. & Dequin, S. (2006), 'Engineering a Saccharomyces cerevisiae Wine Yeast That Exhibits Reduced Ethanol Production during Fermentation under Controlled Microoxygenation Conditions', *Applied and Environmental Microbiology* **72**(9), 5822–5828.
URL: *https://aem.asm.org/content/72/9/5822*

Heyland, J., Fu, J., Blank, L. M. & Schmid, A. (2011), 'Carbon metabolism limits recombinant protein production in Pichia pastoris', *Biotechnology and Bioengineering* **108**(8), 1942–1953.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.23114*

Hiesinger, P. R. & Hassan, B. A. (2005), 'Genetics in the Age of Systems Biology', *Cell* **123**(7), 1173–1174.
URL: *http://www.sciencedirect.com/science/article/pii/S0092867405013814*

Hjersted, J. L., Henson, M. A. & Mahadevan, R. (2007), 'Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture', *Biotechnology and Bioengineering* **97**(5), 1190–1204.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.21332*

Hollywood, K., Brison, D. R. & Goodacre, R. (2006), 'Metabolomics: Current technologies and future trends', *PROTEOMICS* **6**(17), 4716–4723.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200600106*

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

Jones, K. D. & Kompala, D. S. (1999), 'Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures', *Journal of Biotechnology* **71**(1), 105 – 131.
URL: *http://www.sciencedirect.com/science/article/pii/S0168165699000176*

Karlsen, E. (2017), *A method for dynamic flux balance analysis with global constraints.*

Knorr, A. L., Jain, R. & Srivastava, R. (2006), 'Bayesian-based selection of metabolic objective functions', *Bioinformatics* **23**(3), 351–357.
URL: *https://doi.org/10.1093/bioinformatics/btl619*

Kresnowati, M. T. A. P., Suarez-Mendez, C. M., Winden, W. A. v., Gulik, W. M. v. & Heijnen, J. J. (2008), 'Quantitative physiological study of the fast dynamics in the intracellular pH of Saccharomyces cerevisiae in response to glucose and ethanol pulses', *Metabolic Engineering* **10**(1), 39 – 54.
URL: *http://www.sciencedirect.com/science/article/pii/S1096717607000651*

Krishnan, M. S., Ho, N. W. Y. & Tsao, G. T. (1999), 'Fermentation kinetics of ethanol production from glucose and xylose by recombinant Saccharomyces 1400(pLNH33)', *Applied Biochemistry and Biotechnology* **78**(1), 373–388.
URL: *https://doi.org/10.1385/ABAB:78:1-3:373*

Leighty, R. W. & Antoniewicz, M. R. (2011), 'Dynamic metabolic flux analysis (DMFA): A framework for determining fluxes at metabolic non-steady state', *Metabolic Engineering* **13**(6), 745 – 755.
URL: *http://www.sciencedirect.com/science/article/pii/S1096717611001005*

Mahadevan, R., Edwards, J. S. & Doyle, F. J. r. (2002), 'Dynamic flux balance analysis of diauxic growth in Escherichia coli.', *Biophysical journal* **83**(3), 1331–1340.

Maier, A., Völker, B., Boles, E. & Fred Fuhrmann, G. (2003), 'Characterisation of glucose transport in Saccharomyces cerevisiae with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters', *FEMS yeast research* **2**, 539–50.

Mashego, M. R., Gulik, W. M. v., Vinke, J. L., Visser, D. & Heijnen, J. J. (2006), 'In vivo kinetics with rapid perturbation experiments in Saccharomyces cerevisiae using a second-generation BioScope', *Metabolic Engineering* **8**(4), 370 – 383.
URL: *http://www.sciencedirect.com/science/article/pii/S1096717606000206*

Meadows, A. L., Karnik, R., Lam, H., Forestell, S. & Snedecor, B. (2010), 'Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation', *Metabolic Engineering* **12**(2), 150 – 160.
URL: *http://www.sciencedirect.com/science/article/pii/S1096717609000627*

Monteiro, R. D. C. & Adler, I. (1989), 'Interior path following primal-dual algorithms. part I: Linear programming', *Mathematical Programming* **44**(1), 27–41.
URL: *https://doi.org/10.1007/BF01587075*

Orth J, Fleming R, P. B. (2010), 'Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide', *EcoSal Plus* .
URL:
http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.10.2.1

Price, N. D., Reed, J. L. & Palsson, B. O. (2004), 'Genome-scale models of microbial cells: evaluating the consequences of constraints.', *Nature reviews. Microbiology* **2**(11), 886–897.

Prielhofer, R., Maurer, M., Klein, J., Wenger, J., Kiziak, C., Gasser, B. & Mattanovich, D. (2013), 'Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris', *Microbial Cell Factories* **12**(1), 5.
URL: *https://doi.org/10.1186/1475-2859-12-5*

R Antoniewicz, M. (2013), 'Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks', *Current opinion in biotechnology* **24**.

Raman, K. & Chandra, N. (2009), 'Flux balance analysis of biological systems: applications and challenges', *Briefings in Bioinformatics* **10**(4), 435–449.
URL: *https://doi.org/10.1093/bib/bbp011*

Schuetz, R., Kuepfer, L. & Sauer, U. (2007), 'Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli.', *Molecular systems biology* **3**, 119.

Scott, F., Wilson, P., Conejeros, R. & Vassiliadis, V. S. (2018), 'Simulation and optimization of dynamic flux balance analysis models using an interior point method reformulation', *Computers & Chemical Engineering* **119**, 152 – 170.
URL: *http://www.sciencedirect.com/science/article/pii/S0098135418309190*

Sonntag, D., Scandurra, F. M., Friedrich, T., Urban, M. & Weinberger, K. M. (2011), 'Targeted metabolomics for bioprocessing.', *BMC proceedings* **5 Suppl 8**(Suppl 8), P27.

Srisuma, P. (2018), Dynamic Simulation and Optimisation of Biochemical Engineering Systems Using the Interior Point Approach, PhD thesis, University of Cambridge, Cambridge.

Steinmeyer, D. E. & Shuler, M. L. (1989), 'Structured model for Saccharomyces cerevisiae', *Chemical Engineering Science* **44**(9), 2017 – 2030.
URL: *http://www.sciencedirect.com/science/article/pii/0009250989851383*

Toya, Y., Kono, N., Arakawa, K. & Tomita, M. (2011), 'Metabolic Flux Analysis and Visualization', *Journal of Proteome Research* **10**(8), 3313–3323.
URL: *https://doi.org/10.1021/pr2002885*

Tziampazis, E. & Sambanis, A. (1994), 'Modeling of cell culture processes', *Cytotechnology* **14**(3), 191–204.
URL: *https://doi.org/10.1007/BF00749616*

van Dijken, J. P., Weusthuis, R. A. & Pronk, J. T. (1993), 'Kinetics of growth and sugar consumption in yeasts', *Antonie van Leeuwenhoek* **63**(3), 343–352.
URL: *https://doi.org/10.1007/BF00871229*

van Gulik, W. M. & Heijnen, J. J. (1995), 'A metabolic network stoichiometry analysis of microbial growth and product formation', *Biotechnology and Bioengineering* **48**(6), 681–698.
URL: *https://doi.org/10.1002/bit.260480617*

Vassiliadis, V. S., Sargent, R. W. H. & Pantelides, C. C. (1994*a*), 'Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems without Path Constraints', *Industrial & Engineering Chemistry Research* **33**(9), 2111–2122.
URL: *https://doi.org/10.1021/ie00033a014*

Vassiliadis, V. S., Sargent, R. W. H. & Pantelides, C. C. (1994*b*), 'Solution of a Class of Multistage Dynamic Optimization Problems. 2. Problems with Path Constraints', *Industrial & Engineering Chemistry Research* **33**(9), 2123–2133.
URL: *https://doi.org/10.1021/ie00033a015*

Vatcheva, I., Jong, H. d., Bernard, O. & Mars, N. J. I. (2006), 'Experiment selection for the discrimination of semi-quantitative models of dynamical systems', *Artificial Intelligence* **170**(4), 472 – 506.
URL: *http://www.sciencedirect.com/science/article/pii/S0004370205002092*

Zhao, X., Noack, S., Wiechert, W. & Lieres, E. v. (2017), 'Dynamic flux balance analysis with nonlinear objective function', *Journal of Mathematical Biology* **75**(6), 1487–1515.
URL: *https://doi.org/10.1007/s00285-017-1127-4*