Computers and Chemical Engineering 118 (2018) 1–13

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Model-based optimization of biopharmaceutical manufacturing in *Pichia pastoris* based on dynamic flux balance analysis

Victor N. Emenike[a,b,c], René Schenkendorf[a,b], Ulrike Krewer[a,b,*]

[a] Institute of Energy and Process Systems Engineering, Technische Universität Braunschweig, Franz-Liszt-Straße 35, Braunschweig 38106, Germany
[b] Center of Pharmaceutical Engineering (PVZ), Technische Universität Braunschweig, Franz-Liszt-Straße 35a, Braunschweig 38106, Germany
[c] International Max Planck Research School for Advanced Methods in Process and Systems Engineering, Sandtorstraße 1, Magdeburg 39106, Germany

ARTICLE INFO

Article history:
Received 30 October 2017
Revised 4 June 2018
Accepted 16 July 2018
Available online 19 July 2018

Keywords:
Biopharmaceutical manufacturing
Pichia pastoris
Dynamic flux balance analysis
Elementary process functions
Bilevel optimization
Complementarity constraints

ABSTRACT

Biologic drugs are promising therapeutics, and their efficient production is essential for a competitive pharma industry. Dynamic flux balance analysis (dFBA) enables the dynamic simulation of the extracellular bioreactor environment and intracellular fluxes in microorganisms, but it is rarely used for model-based optimization of biopharmaceutical manufacturing in *Pichia pastoris*. To bridge this gap, we present a model-based optimization approach based on dFBA to produce biologics in *P. pastoris* that combines ideas from bilevel optimization, penalization schemes, and direct dynamic optimization. As a case study, we consider the production of recombinant erythropoietin in *P. pastoris* growing on glucose, and predict a 66% improvement in the productivity of erythropoietin. We show that this improvement could be obtained by implementing an almost constant optimal feeding strategy which is different from typical exponential feeding strategies and that a high activity of most pathways in the central carbon metabolism is crucial for a high productivity.

1. Introduction

The biopharmaceutical industry is the fastest growing sector of the pharmaceutical industry with a steadily increasing market value which attained a total sales value of $140 billion in 2013 and will continue to increase in the near future (Aggarwal, 2014; Walsh, 2014). This immense growth of the biopharmaceutical industry can be attributed to the potency of biopharmaceuticals, their high specificity, fewer off-target effects, and their effectiveness in treating deadly diseases such as cancer and diabetes (Wells and Robinson, 2017). However, the cost of biologic drugs is extremely high and as such makes it difficult for developing countries to gain access to these drugs (Love et al., 2013). A possible solution to this challenge as suggested by Love et al. (2013) is to reduce the cost of manufactured goods (COGs) by increasing product yields while ensuring improved quality and potency per drug amount; this ensures reduced number of doses. Therefore, technological advances in biopharmaceutical manufacturing are required to drive down the COGs. In order to enable such technological advances particularly in fermentation, high-quality host cells and optimal bioreactor design are essential (Love et al., 2013; Wells and Robinson, 2017).

Typically, an ideal host cell line is one which ensures high cellular growth under economic media requirements, human-like glycosylation patterns, and the ability to efficiently excrete the recombinant protein of interest into the extracellular media (Love et al., 2013; Sreekrishna et al., 1997). The methylotrophic yeast *Pichia pastoris* fulfills the aforementioned qualities and as such is a popular and intensively studied host cell line since its development in the 1970s (Cereghino and Cregg, 2000; Potvin et al., 2012). Other features that makes *P. pastoris* a favorable host cell include its tightly regulated alcohol oxidase 1 promoter (pAOX1) and its preference for respiratory over fermentative-based growth; thus, it mitigates the formation of fermentative by-products such as ethanol which could lead to high toxic levels and negatively impact protein expression (Cereghino et al., 2002; Potvin et al., 2012).

Moreover, various strategies have been proposed to improve the productivity of recombinant proteins in *P. pastoris*. These include: intelligent design of expression vectors (Sreekrishna et al., 1997), use of different carbon sources (Xie et al., 2005), metabolic engineering (Saitua et al., 2017), efficient fermentation protocols (Potvin et al., 2012), and innovative bioreactor designs (Mozdzierz et al., 2015). Typically, these strategies are inves-

* Corresponding author at: Institute of Energy and Process Systems Engineering, Technische Universität Braunschweig, Franz-Liszt-Straße 35, Braunschweig 38106, Germany.
E-mail addresses: v.emenike@tu-braunschweig.de (V.N. Emenike), r.schenkendorf@tu-braunschweig.de (R. Schenkendorf), u.krewer@tu-braunschweig.de (U. Krewer).

https://doi.org/10.1016/j.compchemeng.2018.07.013
0098-1354/

tigated by empirical means, but various studies have shown that model-based approaches are key to uncovering strategies that lead to improved production of biopharmaceuticals in *P. pastoris* (Cos et al., 2006; d'Anjou and Daugulis, 2001; Jahic et al., 2002).

Amongst various modeling strategies for *P. pastoris*, first principles dynamic models are important because they predict temporal changes in relevant variables of the bioprocess (Çalık et al., 2011; Niu et al., 2013; Saitua et al., 2017). Dynamic models for *P. pastoris* fermentation can be classified into unstructured and structured models (Lu et al., 2015; Tziampazis and Sambanis, 1994). Unstructured models are phenomenological models that only consider the extracellular concentrations in a bioreactor and do not take into account the intracellular dynamics of the yeast cells (Lu et al., 2015; Tziampazis and Sambanis, 1994). Nevertheless, they are popular because they can be easily constructed (Tziampazis and Sambanis, 1994). However, unstructured models are limited because they cannot be extrapolated to operating conditions where there are significant cellular changes (Tziampazis and Sambanis, 1994; Zhang et al., 2000). Therefore, they might limit the possibility of identifying novel process windows (Hessel, 2009).

Structured models, on the other hand, are more detailed than unstructured models and consider both intracellular information and extracellular conditions (Höffner et al., 2013; Tziampazis and Sambanis, 1994). However, most structured models that are available for the recombinant expression of proteins in *P. pastoris* are based on compartmentalized models (Çelik et al., 2009b; Muñoz et al., 2008; Niu et al., 2013) that do not consider detailed intracellular fluxes.

In contrast to compartmentalized structured models, dynamic flux balance analysis (dFBA) models are structured models that enable the prediction of changes in the reaction pathways of a microorganism's metabolism due to changes in the external environment in a bioreactor (Höffner et al., 2013; Mahadevan et al., 2002). Several dFBA models have been used to simulate, control, and optimize the expression of proteins in other important microbial systems such as *Escherichia coli* (Meadows et al., 2010) and *Saccharomyces Cerevisiae* (Hjersted and Henson, 2006), but there is a scarcity of dFBA models for *P. pastoris* expression systems.

In an attempt to bridge this gap, Saitua et al. (2017) developed a dynamic genome-scale metabolic model for the production of recombinant human serum albumin (rHSA) in *P. pastoris*. Their model consists of the dynamic evolution of seven state variables, namely: glucose, biomass, ethanol, arabitol, citrate, pyruvate and the culture volume. They applied the dFBA framework by Sánchez et al. (2014) which involves an iteration between a dynamic block of the aforementioned states, a kinetic block that determines the substrate uptake kinetics, and a metabolic block which determines the flux distributions. However, the work by Saitua et al. (2017) involves a lot of dFBA simulations to optimize the protein productivity – this could be cumbersome given their iterative approach. Instead, algorithmic optimization could be used to optimize protein productivity by using the dFBA model.

Studies involving model-based algorithmic optimization for recombinant protein production in *P. pastoris* have been reported in the literature. Kobayashi et al. (2000) applied dynamic programming to determine the optimal methanol feeding profile for the maximization of rHSA. However, their approach could not predict the product concentrations accurately due to discontinuities in the methanol feed rate. Hence, they used trial-and-error simulations to get an optimal profile that predicts the product concentration – this is cumbersome and could lead to suboptimal results. Moreover, since their approach is based on dynamic programming, it suffers from the "curse of dimensionality" (Li et al., 2008; Sager, 2009) and is unsuitable for large scale problems. Even though the work of Kobayashi et al. (2000) represent progress in the use of algorithmic optimization for biopharmaceutical production in *P. pastoris*, they used an unstructured model that does not consider intracellular changes in the yeast cells. Furthermore, Llaneras et al. (2012) combined optimization algorithms, possibility theory, and stoichiometric models to estimate dynamic intracellular fluxes in *P. pastoris*. However, their work was geared towards state estimation and monitoring, and not the maximization of protein production. Therefore, there remains a need for more efficient algorithmic optimization approaches that utilize dFBA models to maximize biopharmaceutical production in *P. pastoris*.

In this paper, we present a model-based optimization approach for the recombinant production of the *P. pastoris* that is based on dFBA. As a case study, we consider the production of the glycoprotein erythropoietin (Jacobs et al., 1985) by *P. pastoris* growing on glucose. It has been reported that variations in the glycosylation patterns of recombinant erythropoietin can influence its potency (Schiestl et al., 2011). These variations are usually caused by changes in the cell line or process conditions (Schiestl et al., 2011). Therefore, it is important to utilize model-based approaches to predict the effects of process changes before these changes are implemented in the real bioprocess or even before the process is built.

Our dFBA model consists of an upper-level problem that is cast within the framework of elementary process functions (Freund and Sundmacher, 2008), and a lower-level problem posed as a flux balance analysis (FBA) model (Orth et al., 2010) – leading to a bilevel optimization problem. Here, the bi-level optimization problem is transformed into a single optimization problem by using the Karush–Kuhn–Tucker (KKT) conditions of the lower-level problem (Raghunathan et al., 2003). This transformation is done in order to solve the problem in one-shot without the need for interacting between multiple solvers, i.e., improving computational efficiency (Hjersted and Henson, 2006). The single optimization problem is a dynamic optimization problem that falls into a class of optimization problems called mathematical programs with equilibrium constraints (MPECs). MPECs present difficulties for state-of-art nonlinear programming (NLP) solvers because they violate constraint qualifications required by these solvers (Baumrucker et al., 2008). We address this issue by using the exact ℓ_1 penalization scheme (Baumrucker et al., 2008) instead of using mixed integer programming algorithms which are combinatorial in nature (Waldherr, 2016) or regularization schemes which involve solving relaxed MPECs iteratively (Joy and Kremling, 2010). Penalization schemes have been shown to be very efficient for solving MPECs resulting from FBA (Yang et al., 2008), but to our knowledge, this approach has not been used for MPECs arising from dFBA. Therefore, another contribution of our work is the extension of the penalization technique to handle the complementarity constraints stemming from dFBA bilevel optimization problems. The reformulated optimization problem is then solved at once by using the direct optimization approach (Biegler, 2007; Mahadevan et al., 2002) and avoids the iterative approach that was previously mentioned (Kobayashi et al., 2000; Saitua et al., 2017). We show that our solution approach is fast and efficient, and is able to maximize the productivity of erythropoietin in *P. pastoris*. Fig. 1 summarizes the tenets of the methodology we employed in this paper.

The rest of this paper is structured as follows: in Section 2, we describe the main features of the model-based approach and present the solution strategy used in Section 3. Following this, we introduce the optimization formulation for the case study considered in Section 4 and the accompanying results in Section 5. Lastly, we conclude and highlight possible future directions in Section 6.

2. Preliminaries

The dFBA modeling framework proposed in this work is a bilevel optimization problem that consists of an upper-level prob-

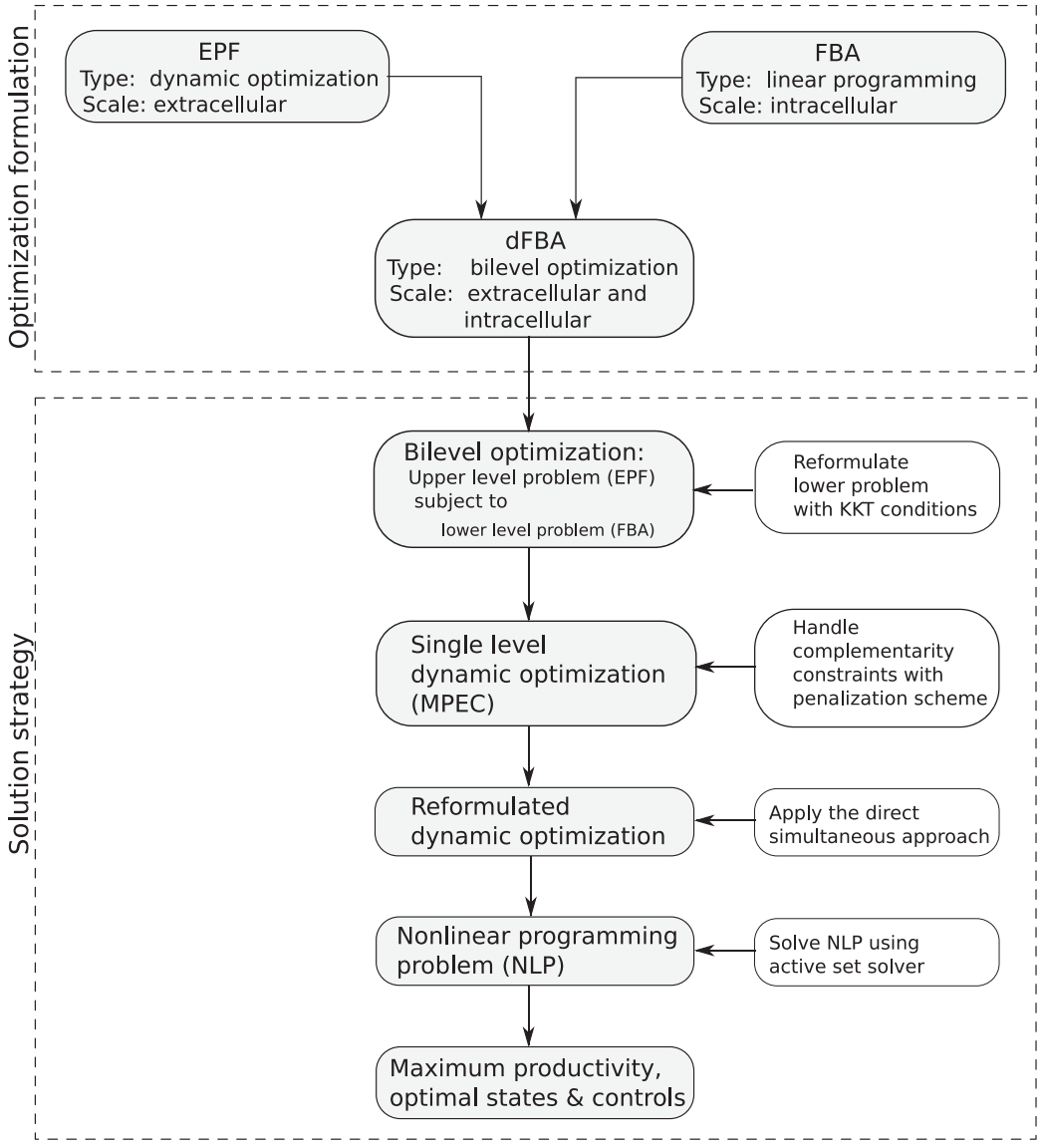

Fig. 1. Work flow of methodology from the model formulation to the solution strategy. EPF stands for elementary process function, FBA for flux balance analysis, dFBA for dynamic flux balance analysis, and MPEC for mathematical program with equilibrium constraints.

lem that is cast within the EPF paradigm and a lower-level problem that is modelled as FBA. Therefore, we describe FBA and then the EPF framework in more detail in the following sections.

2.1. Flux balance analysis

FBA is a stoichiometric modeling approach that is used for studying metabolic networks that range from small-scale to genome-scale metabolic reconstructions (Boghigian et al., 2010; Orth et al., 2010). FBA computes an optimal distribution of metabolic fluxes within the metabolic network of a microorganism. Thus, it leads to flux distributions that optimize phenotypes, e.g., cellular growth or the production of key metabolites (Orth et al., 2010).

Before performing FBA, a microorganism's metabolic network has to be represented mathematically by performing what is known as a metabolic reconstruction (Thiele and Palsson, 2010). This mathematical representation takes the form of a matrix with metabolic reactions and metabolites represented as columns and rows, respectively. The elements of this matrix are the stoichiometric coefficients of each reaction, and as such this matrix is called a stoichiometric matrix (Orth et al., 2010). The stoichiometric matrix in combination with a vector representing the fluxes leads to equality constraints that impose bounds on the system. Typically, the microorganism is assumed to be in a pseudo-steady state with respect to the external environment, and as such, there is no accumulation (Höffner et al., 2013).

Next, a phenotype in the form of a biological objective is defined. Typically, the biomass growth is selected as the objective function, but other objectives can also be considered (Schuetz et al., 2007; Zhao et al., 2017). Subsequently, the objective function is then maximized (or minimized) subject to the aforementioned equality constraints. Unfortunately, the stoichiometric matrix is underdetermined and therefore leads to different solutions that fulfill the same objective (Segre et al., 2002). To further constrain the solution space, bounds in the form of inequality constraints are imposed on the fluxes (Orth et al., 2010). Therefore, the solution set of these constraints is a convex polyhedron, i.e., an intersection of a finite number of half planes and half spaces (Boyd and Vandenberghe, 2004).

In sum, FBA is formulated as a linear programming (LP) problem with the biomass flux as the objective function, a steady-state

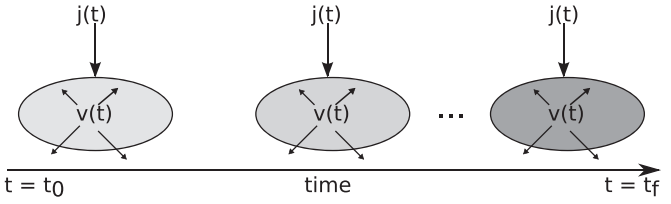

Fig. 2. Conceptual representation of biofluid element (including cells) in thermodynamic state space acted upon by dynamic fluxes such as the intracellular $v(t)$ and extracellular fluxes $j(t)$.

balance of n (intracellular) fluxes $\mathbf{v}$ through m metabolic reactions, the corresponding stoichiometric matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$, and bounds $\mathbf{v}^{\mathrm{L}}$ and $\mathbf{v}^{\mathrm{U}}$ on the fluxes (Orth et al., 2010). It is mathematically expressed as

$$\begin{aligned} \underset{\mathbf{v}}{\text{maximize}} \quad & \mathbf{c}^{\top}\mathbf{v} \\ \text{subject to} \quad & \mathbf{S}\mathbf{v} = \mathbf{0}, \\ & \mathbf{v}^{\mathrm{L}} \le \mathbf{v} \le \mathbf{v}^{\mathrm{U}}, \end{aligned} \tag{1}$$

where $\mathbf{c} \in \mathbb{R}^n$ is a weighting vector for the fluxes considered in the objective function, whose elements, as in our case, take the value of one for the element corresponding to the biomass flux and zero otherwise.

2.2. Elementary process functions

The EPF methodology proposed by Freund and Sundmacher (2008) differentiates itself from the conventional unit operation approach in process design that is based on and therefore limited to "off-the-shelf" processing units. Within the EPF framework, these units are replaced by functional modules in which the states of a passing fluid element are changed by fluxes such as heat, mass, component dosing, and diffusion fluxes. Mathematically, the fluid element is represented as (Freund and Sundmacher, 2008):

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \sum_{k=1}^{J} j_k^{\Phi}(\mathbf{x})\mathbf{e}_k, \tag{2}$$

where $\mathbf{x}$ is a state vector (e.g. mass, energy, concentration, etc.), j_k^{Φ} is the (extracellular) flux k of the functional module Φ, $\mathbf{e}_k$ is the elementary process function (EPF) of flux k, i.e., its basis vector in thermodynamic state space, and J is the total number of fluxes of the functional module Φ. The EPF $\mathbf{e}_k$ represents a certain direction of the flux k in thermodynamic state space, and the combined effect of the EPFs determines the region in thermodynamic state space that is attainable by the fluid element (see Freund and Sundmacher, 2008 for details). Alternatively, Eq. (2) could be re-written in state-space representation notation as:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{E}(\mathbf{x})\mathbf{j}^{\Phi}(\mathbf{x}), \tag{3}$$

where $\mathbf{j}^{\Phi}$ is the generalized flux vector of the functional module Φ, and $\mathbf{E}$ is the elementary process functions matrix which is the product of the inverse capacity matrix and a flux weighting matrix (see Peschel, 2012 for details).

The fluid element traveling through the functional modules is then tracked in order to find the optimal route in state space (see Fig. 2). Information of the optimal route is subsequently used for a technical realization with existing process units or gives rise to the development of new apparatuses. Recently, we have shown how to adapt the EPF approach for the design of optimal reactors for the synthesis of small molecule drugs (Emenike and Krewer, 2016; Emenike et al., 2018). In this work, we take a first step in extending EPF to the design of optimal bioreactors for the synthesis of biologic drugs (Emenike et al., 2017).

Within the EPF framework, an optimal bioreactor design problem can be formulated in such a way that an extracellular bioreaction functional module is used instead of a bioreactor unit. This results in the following dynamic optimization problem:

$$\begin{aligned} \underset{\mathbf{j}(t),\mathbf{z}(t)}{\text{minimize}} \quad & \mathcal{J} \\ \text{subject to} \quad & \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{E}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t)\mathbf{j}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t), \\ & \boldsymbol{g}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t) = \mathbf{0}, \\ & \boldsymbol{h}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t) \le \mathbf{0}, \\ & \mathbf{x}(t_0) = \mathbf{x}_0, \end{aligned} \tag{4}$$

on the time interval $\mathcal{T} \in [t_0, t_{\mathrm{f}}] \subset \mathbb{R}$ of the biochemical reaction with time $t \in [t_0,\ t_{\mathrm{f}}]$, where $\mathcal{J}$ is an objective function of biological relevance, e.g., yield or productivity, $\mathbf{x}(t) \in \mathbb{R}^{n_{\mathrm{x}}}$ is a vector of state variables such as extracellular metabolite concentrations or masses, $\mathbf{z}(t) \in \mathbb{R}^{n_{\mathrm{z}}}$ is a vector of algebraic variables such as substrate uptake or growth rates represented by Monod-type kinetics, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is a vector of time independent parameters, $\mathbf{j}(t) \in \mathbb{R}^{n_u}$ is the flux vector containing feeding rates and other key control variables, $\mathbf{E} \in \mathbb{R}^{n_{\mathrm{epf}}\times n_u}$ is the EPF matrix which contains the basis vectors in thermodynamic state space (Freund and Sundmacher, 2008), $\boldsymbol{g} : \mathcal{T}\times\mathbb{R}^{n_{\mathrm{x}}}\times\mathbb{R}^{n_\theta}\times\mathbb{R}^{n_{\mathrm{z}}} \to \mathbb{R}^{n_{\mathrm{g}}}$ is a function vector that defines the equality constraints, $\boldsymbol{h} : \mathcal{T}\times\mathbb{R}^{n_{\mathrm{x}}}\times\mathbb{R}^{n_\theta}\times\mathbb{R}^{n_{\mathrm{z}}} \to \mathbb{R}^{n_{\mathrm{h}}}$ is the inequality constraint function vector, and $\mathbf{x}_0$ is a vector of initial conditions of the states variables at initial time t_0.

2.3. Dynamic flux balance analysis

The dFBA is formulated as a bilevel optimization problem where the EPF dynamic optimization problem (4) is the upper-level problem, and the FBA (1) is the lower-level problem:

$$\begin{aligned} \underset{\mathbf{j}(t),\mathbf{z}(t),\tilde{\mathbf{v}}(t)}{\text{minimize}} \quad & \mathcal{J} \\ \text{subject to} \quad & \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{E}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t)\mathbf{j}(\mathbf{x},\mathbf{z},\tilde{\mathbf{v}},\boldsymbol{\theta},t), \\ & \mathbf{g}(\mathbf{x},\mathbf{z},\tilde{\mathbf{v}},\boldsymbol{\theta},t) = \mathbf{0}, \\ & \mathbf{h}(\mathbf{x},\mathbf{z},\tilde{\mathbf{v}},\boldsymbol{\theta},t) \le \mathbf{0}, \\ & \tilde{\mathbf{v}}(t) \in \underset{\mathbf{v}(t)}{\arg\min}\ \{-\mathbf{c}^{\top}\mathbf{v}(t) \mid \mathbf{S}\mathbf{v}(t) = \mathbf{0},\ \mathbf{v}^{\mathrm{L}}(t) \\ & \qquad\qquad \le \mathbf{v}(t) \le \mathbf{v}^{\mathrm{U}}(t)\}, \\ & \mathbf{x}(t_0) = \mathbf{x}_0\,. \end{aligned} \tag{5}$$

Here, the rates of extracellular metabolites $\tilde{\mathbf{v}}$ such as the biomass growth rate and substrate uptake rate are computed by the inner FBA model at each time point. Note that $\tilde{\mathbf{v}}$ represent the rates (i.e., extracellular fluxes) that are computed by the lower-level FBA and used by the upper-level dynamic optimization problem (Höffner et al., 2013), while $\mathbf{z}(t)$ is the vector of algebraic variables (mostly rates) that are still computed by Monod-type kinetic equations and not by the lower-level FBA. In the next section, we present the solution strategy (as shown in Fig. 1) that we employed to solve the dFBA optimization problem efficiently.

3. Solution strategy

3.1. Bilevel optimization

In this section, we formulate the bilevel optimization problem (Eq. (5)) into a form that is convenient for most dynamic optimization solution strategies by transforming it into a single objective dynamic optimization problem. We also show how we handle the resulting complementarity constraints.

3.1.1. Karush–Kuhn–Tucker reformulation

Here, we transform the bilevel problem into a single objective optimization problem by replacing the lower-level FBA problem with its Karush–Kuhn–Tucker (KKT) conditions and complementarity constraints (Raghunathan et al., 2003):

$$\underset{\substack{\mathbf{j}(t),\mathbf{z}(t),\mathbf{v}(t),\\ \boldsymbol{\lambda}(t),\boldsymbol{\alpha}^{\mathrm{L}}(t),\boldsymbol{\alpha}^{\mathrm{U}}(t)}}{\text{minimize}} \ \mathcal{J} \tag{6a}$$

$$\text{subject to} \frac{d\mathbf{x}}{dt} = \mathbf{E}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t)\mathbf{j}(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t) \tag{6b}$$

$$\mathbf{g}(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t) = \mathbf{0} \tag{6c}$$

$$\mathbf{h}\big(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t\big) \le \mathbf{0} \tag{6d}$$

$$\nabla_{\mathbf{v}}\mathcal{L} = \boldsymbol{c} + \mathbf{S}^{\top}\boldsymbol{\lambda}(t) + \boldsymbol{\alpha}^{\mathrm{L}}(t) - \boldsymbol{\alpha}^{\mathrm{U}}(t) = \mathbf{0} \tag{6e}$$

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L} = \mathbf{S}\mathbf{v}(t) = \mathbf{0} \tag{6f}$$

$$\mathrm{diag}(\mathbf{v}(t) - \mathbf{v}^{\mathrm{L}}(t))\boldsymbol{\alpha}^{\mathrm{L}}(t) = \mathbf{0} \tag{6g}$$

$$\mathrm{diag}(\mathbf{v}(t) - \mathbf{v}^{\mathrm{U}}(t))\boldsymbol{\alpha}^{\mathrm{U}}(t) = \mathbf{0} \tag{6h}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{6i}$$

$$\boldsymbol{\alpha}^{\mathrm{L}}(t) \in \mathbb{R}^n, \boldsymbol{\alpha}^{\mathrm{U}}(t) \in \mathbb{R}^n, \boldsymbol{\lambda}(t) \in \mathbb{R}^m \ge \mathbf{0}. \tag{6j}$$

where $\mathcal{L}$ is the Lagrangian function given as:

$$\mathcal{L}(\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\alpha}^{\mathrm{L}},\boldsymbol{\alpha}^{\mathrm{U}},t) = -\mathbf{c}^{\top}\mathbf{v} - (\mathbf{S}\mathbf{v})^{\top}\boldsymbol{\lambda} - (\mathbf{v}^{\mathrm{U}} - \mathbf{v})^{\top}\boldsymbol{\alpha}^{\mathrm{U}} - (\mathbf{v} - \mathbf{v}^{\mathrm{L}})^{\top}\boldsymbol{\alpha}^{\mathrm{L}}, \tag{7}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers corresponding to $\mathbf{S}\mathbf{v} = \mathbf{0}$, $\boldsymbol{\alpha}^{\mathrm{L}}$ and $\boldsymbol{\alpha}^{\mathrm{U}}$ are vectors of Lagrange multipliers associated with the lower and upper bounds of the fluxes $\mathbf{v}$, and $\nabla_{\mathbf{v}}\mathcal{L}$ and $\nabla_{\boldsymbol{\lambda}}\mathcal{L}$ are first derivatives of the Lagrangian with respect to $\mathbf{v}$ and $\boldsymbol{\lambda}$, respectively.

The reformulation (6) is valid because the lower-level problem is convex and regular and as such the KKT conditions are sufficient and necessary conditions for optimality (Colson et al., 2007). With the occurrence of complementarity constraints (6g) and (6h), the problem represented by (6) falls into the class of problems called mathematical programs with equilibrium constraints (Baumrucker et al., 2008; Ralph and Wright, 2004).

Another approach to transforming the bilevel optimization problem into a single level optimization problem has been proposed by Nikdel et al. (2018). In their work, a systematic approach was used to identify a suitable objective function and limiting constraints by data fitting. Next, the identified objective function and limiting constraints are used to develop a predictive dFBA model. In our approach, no identification or experimental data fitting step is required. Instead, we directly use the KKT conditions of the inner FBA problem and then use the exact ℓ_1 penalization technique to handle issues with complementarity constraints as presented in Section 3.1.2.

3.1.2. Exact ℓ_1 penalization

For mathematical programs with equilibrium constraints (MPECs), even if the objective functions and the solution set of the fluxes $\mathbf{v} \in \mathcal{V}$ are both convex, they are still difficult to solve because of nonconvexities introduced by complementarity and Lagrangian constraints (Baumrucker et al., 2008; Colson et al., 2007). Moreover, in order to ensure that the replacement of the lower-level problem by its KKT conditions provides a necessary optimality condition, certain constraint qualifications (CQs) such as Guignard's CQ, Mangasarian-Fromovitz CQ and the linear independence CQ have to be fulfilled (Kyparisis, 1985; Wachsmuth, 2013).

Unfortunately, MPECs are known to violate these CQs and as such reformulations are often used to make them tractable for state-of-the-art NLP solvers to handle (Baumrucker et al., 2008; Ralph and Wright, 2004). Such reformulations can be classified into two main groups namely, regularization and penalization schemes (Baumrucker et al., 2008; Ralph and Wright, 2004). Regularization involves replacing the complementarity constraints with a constraint that has a positive parameter and then repeatedly solving the NLP as this value is successively decreased to a tolerance that is close to zero (Ralph and Wright, 2004). In the case of penalization, the complementarity constraints are transferred to the objective to form a penalty term (Ralph and Wright, 2004).

In this work, we performed preliminary studies in which we compared the regularization and penalization formulations, but found the penalization formulation to be the most stable for our purposes. We are also in favor of penalization schemes because provided that a sufficiently large penalty term is chosen, the reformulated MPEC can be solved efficiently in one-shot. This is in contrast with regularization schemes for dFBA where repeated iterations of relaxed MPECs is required to converge to a local optimal solution (Joy and Kremling, 2010). Furthermore, Baumrucker et al. (2008) conducted benchmark studies and showed that the usage of well-posed complementarity constraints coupled with the penalty formulation and an active set NLP solver is the most efficient strategy for solving MPECs arising from chemical engineering. This was also the case in our work, and as such, we used the exact ℓ_1 penalization technique (Baumrucker et al., 2008) in combination with the active set NLP solver CONOPT (Drud, 1994). The ℓ_1 penalization formulation of (6) is given as:

$$\underset{\substack{\mathbf{j}(t),\mathbf{z}(t),\mathbf{v}(t),\\ \boldsymbol{\lambda}(t),\boldsymbol{\alpha}^{\mathrm{L}}(t),\boldsymbol{\alpha}^{\mathrm{U}}(t)}}{\text{minimize}} \ \mathcal{J} + \rho\left\|\hat{\mathbf{v}}^{\top}\boldsymbol{\alpha}\right\|_1 \tag{8a}$$

$$\text{subject to} \ \frac{d\mathbf{x}}{dt} = \mathbf{E}(\mathbf{x},\mathbf{z},\boldsymbol{\theta},t)\mathbf{j}(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t) \tag{8b}$$

$$\mathbf{g}(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t) = \mathbf{0} \tag{8c}$$

$$\mathbf{h}\big(\mathbf{x},\mathbf{z},\mathbf{v},\boldsymbol{\theta},t\big) \le \mathbf{0} \tag{8d}$$

$$\nabla_{\mathbf{v}}\mathcal{L} = \boldsymbol{c} + \mathbf{S}^{\top}\boldsymbol{\lambda}(t) + \boldsymbol{\alpha}^{\mathrm{L}}(t) - \boldsymbol{\alpha}^{\mathrm{U}}(t) = \mathbf{0} \tag{8e}$$

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L} = \mathbf{S}\mathbf{v}(t) = \mathbf{0} \tag{8f}$$

$$\hat{\mathbf{v}} = \begin{bmatrix} \mathbf{v}(t) - \mathbf{v}^{\mathrm{L}}(t) \\ \mathbf{v}(t) - \mathbf{v}^{\mathrm{U}}(t) \end{bmatrix} \in \mathbb{R}^{2n} \tag{8g}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}^{\mathrm{L}}(t) \\ \boldsymbol{\alpha}^{\mathrm{U}}(t) \end{bmatrix} \tag{8h}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{8i}$$

$$\boldsymbol{\alpha} \in \mathbb{R}^{2n}, \boldsymbol{\lambda}(t) \in \mathbb{R}^{m} \geq \mathbf{0}. \tag{8j}$$

Here, the variables associated with the complementarity constraints Eqs. (6g) and (6h) were rearranged into the variables $\hat{\mathbf{v}}$ (8g) and $\boldsymbol{\alpha}$ (8h). Moreover, the complementarity constraints are transferred from the constraints to the objective function (8a) and multiplied by a sufficiently large penalty parameter ρ. The ℓ_1 penalization technique ensures that the complementarity constraints are satisfied provided that $\rho \geq \rho_c$, where ρ_c is a critical penalty parameter (Baumrucker et al., 2008). In this work, various values of ρ were compared and $\rho = 10^3$ was found to be sufficient.

3.2. Dynamic optimization

Eq. (8) is a dynamic optimization problem that can be solved by three main approaches: dynamic programming, indirect, and direct approaches (Sager, 2009). Dynamic programming (Bellman, 1952) typical is not suitable for large scale problems because it suffers from the "curse of dimensionality" (Li et al., 2008; Sager, 2009). Indirect ("optimize-then-discretize") methods such as Pontryagin's minimum principle (Pontryagin et al., 1962) are also known to be unapplicable to large scale problems because of difficulties in handling adjoints associated with path constraints (Sager, 2009). Direct ("discretize-then-optimize") methods are usually the method of choice for complex, highly nonlinear, large scale problems (Sager, 2009) such as the one considered in this paper. Direct methods include control vector parameterization (Vassiliadis et al., 1994a; 1994b), multiple shooting (Bock and Plitt, 1984), and the simultaneous approach (Biegler, 1984).

In this work, the simultaneous approach was selected because it allows us to transcribe Eq. (8) directly into NLPs without the need for successively calling a DAE solver as is the case for both control vector parameterization and multiple shooting (Biegler, 2007). Furthermore, the simultaneous approach has been shown to handle instabilities and path constraints efficiently (Biegler, 2007; 2010). Specifically, the DAEs were transcribed by using the method of orthogonal collocation on finite elements (Cuthrell and Biegler, 1987). The extracellular states were discretized on both collocation points and finite elements by using 20 finite elements with 3 Radau collocation points in each element, while the extracellular controls and intracellular fluxes were discretized on the 20 finite elements only by using a piecewise constant parameterization. The resulting NLP was implemented in the algebraic mathematical language AMPL (Fourer et al., 2003) and the CONOPT solver was used (Drud, 1994). All computations were performed on a Linux machine with an Intel (R) Core (TM) i7-4789 processor at 3.60 GHz, and 16 GB RAM.

4. Case study: recombinant protein production in *Pichia pastoris*

As already mentioned, the case study we consider is the recombinant production of the biologic drug erythropoeitin in *Pichia pastoris* with glucose as the sole carbon source (substrate). Here, we aim to maximize the productivity of erythropoeitin by using the computational approach presented previously. At the same time, we aim to obtain optimal dynamic controls at both the extracellular and intracellular levels that maximize productivity during the fermentation process. We chose the productivity of erythropoeitin as the objective function of the upper-level optimization problem as it is a widely used metric for accessing the economic viability of a bioprocess (Anesiadis et al., 2008; Kumar and Budman, 2017; St. John et al., 2017). Typically, high productivity implies lower operating and capital costs (Anesiadis et al., 2008).

As a benchmark to compare our optimization results, we chose the experimental work of Çelik et al. (2009a) in which erythropoeitin was produced in the presence of sorbitol and methanol. The maximum concentration of erythropoeitin reported by Çelik et al. (2009a) was 130 mg L^{-1} in 24 h and the maximum working volume was 2 L. Based on these values, the maximum possible productivity obtained by Çelik et al. (2009a) was estimated to be 10.83 mg h^{-1}. This productivity was obtained by utilizing the traditional exponential feeding strategy. In our study, we do not use any pre-defined feeding strategies, but obtain an optimal feeding strategy (control) by solving the optimization problem.

To model the dynamic intracellular flux distributions, a metabolic flux network adapted from the validated FBA model by Morales et al. (2014) was used. It consists of 37 intracellular metabolites and 47 intracellular reactions. The network consists of the TCA cycle, the pentose phosphate pathway (PPP), Embden-Meyerhoff-Parnas (EMP) pathway (i.e., glycolysis), pathways describing the metabolism of methanol, glycerol and glucose substrates, and transport reactions (see Fig. 3). For our purposes, we set the glycerol (flux 43) and methanol (flux 46) uptake fluxes to zero by enforcing the constraints $v_{\text{gly}} = 0$ and $v_{\text{meoh}} = 0$, respectively (Morales et al., 2014; Vercammen et al., 2014). Note that v_{gly} represents the glycerol uptake flux, while v_{meoh} represents methanol uptake flux. We also assume that sufficient oxygen is present for the cells to grow aerobically and as such do not include an oxygen balance in the dFBA model (Hjersted and Henson, 2006).

In contrast to available structured models for *P. pastoris* (Çelik et al., 2009b; Muñoz et al., 2008; Niu et al., 2013), we do not compartmentalize any of the aforementioned pathways, but consider all relevant intracellular metabolites and reactions. Moreover, we consider only the first level that was proposed in the reactor design framework proposed by Peschel et al. (2010), i.e., we do not constrain fluxes such as the substrate feeding rates or dissolved oxygen rate to mass or heat transfer limitations. Here, we consider an intensification strategy that involves the intermittent feeding of glucose substrate as the biofluid element progresses in time. This can be physically translated into a fed-batch bioreactor; however, we consider this as a functional module in order to be consistent with the EPF concept (Freund and Sundmacher, 2008). Nevertheless, since the approach presented in this work is based on the EPF framework, it can be extended to other intensification strategies.

Therefore, we obtain the following dFBA model:

$$\begin{aligned}
\underset{\phi(t)_{\text{gluc}}, \tilde{\mathbf{v}}(t), t_{\text{f}}}{\text{minimize}} \quad & \mathcal{J} \triangleq -m_{\text{epo}}(t_{\text{f}})/t_{\text{f}}, \\
\text{subject to} \quad & \frac{dm_{\text{biom}}(t)}{dt} = m_{\text{biom}} v_{\text{biom}}, \\
& \frac{dm_{\text{gluc}}(t)}{dt} = -m_{\text{biom}} v_{\text{gluc}} + C_{\text{gluc,in}} \phi_{\text{gluc}}, \\
& \frac{dm_{\text{epo}}(t)}{dt} = m_{\text{epo}} v_{\text{epo}}, \\
& \frac{dV(t)}{dt} = \phi_{\text{gluc}}, \\
& \tilde{\mathbf{v}}(t) \in \underset{\mathbf{v}(t)}{\arg\min} \ \{-\mathbf{c}^{\top}\mathbf{v}(t) \mid \mathbf{S}\mathbf{v}(t) = \mathbf{0}, \\
& \qquad\qquad -10^3 \leq \mathbf{v}(t) \leq 10^3\}, \\
& m_k = C_k V \ , \quad k \in \{\text{biom, gluc, epo}\}, \\
& C_{\text{gluc},0} = 50\,\text{mmol}\,\text{L}^{-1}, \\
& C_{\text{biom},0} = 1.0\,\text{g}\,\text{L}^{-1}, \\
& C_{\text{epo},0} = 0.0\,\text{g}\,\text{L}^{-1}, \\
& V_0 = 0.82\,\text{L},
\end{aligned}$$

$$
\begin{aligned}
&C_{\text{gluc,in}} = C_{\text{gluc},0},\\
&v_{\text{epo}} = a v_{\text{biom}} + b,\\
&m_k \geq 0, \quad k \in \{\text{biom, gluc, epo}\},\\
&m_{\text{biom}} \leq 400\text{g},\\
&0.0 \leq V \leq 3.0\,\text{L},\\
&0 \leq \phi_{\text{gluc}} \leq 1.0\,\text{L/h},\\
&v_{\text{gly}} = 0, v_{\text{meoh}} = 0,\\
&v_{\text{gluc}} \leq \frac{\mu_{\text{gluc,max}} C_{\text{gluc}}}{Y_{\text{biom/gluc}}(K_{\text{gluc}} + C_{\text{gluc}})} + \gamma_{\text{gluc}},\\
&20 \leq t_{\text{f}} \leq 60\,\text{h}\,.
\end{aligned} \tag{9}
$$

The presented dFBA optimization problem is then translated into the EPF formulation as follows:

$$
\begin{aligned}
\underset{\boldsymbol{\phi}(t),\tilde{\mathbf{v}}(t),t_{\text{f}}}{\text{minimize}} \quad & \mathcal{J} := -m_{\text{epo}}(t_{\text{f}})/t_{\text{f}}\\
\text{subject to} \quad & \frac{d\mathbf{x}}{dt} = \mathbf{E}(\mathbf{x}, \boldsymbol{\theta}, t)\mathbf{j}(\boldsymbol{\phi}, \tilde{\mathbf{v}}, t),\\
& \tilde{\mathbf{v}}(t) \in \underset{\mathbf{v}(t)}{\arg\min}\ \{-\mathbf{c}^{\top}\mathbf{v}(t) \mid \mathbf{S}\mathbf{v}(t) = \mathbf{0},\ \mathbf{v}^{\text{L}}(t)\\
& \qquad\qquad \leq \mathbf{v}(t) \leq \mathbf{v}^{\text{U}}(t)\},\\
& \mathbf{g}(\mathbf{x}, \boldsymbol{\phi}, \mathbf{v}, \boldsymbol{\theta}, t) = 0,\\
& \mathbf{h}(\mathbf{x}, \boldsymbol{\phi}, \mathbf{v}, \boldsymbol{\theta}, t) \leq 0,\\
& \mathbf{x}(t_0) = \mathbf{x}_0,
\end{aligned} \tag{10}
$$

where $\mathbf{x} = [m_{\text{biom}}, m_{\text{gluc}}, m_{\text{epo}}, V]^{\top}$,
$\mathbf{j} = [v_{\text{biom}}, v_{\text{gluc}}, v_{\text{epo}}, \phi_{\text{gluc}}]^{\top}$,

$$
\mathbf{E} = \begin{bmatrix} m_{\text{biom}} & 0 & 0 & 0\\ 0 & -m_{\text{biom}} & 0 & C_{\text{gluc,in}}\\ 0 & 0 & m_{\text{epo}} & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

where $\mathcal{J}$ is the productivity objective function in g h^{-1}, $\mathbf{S} \in \mathbb{R}^{37\times 47}$ is the stoichiometric matrix, $\mathbf{v}(t) \in \mathbb{R}^{47}$ is a vector representing the metabolic reaction fluxes in mmol g^{-1} h^{-1},$\mathbf{c} \in \mathbb{R}^{47}$ is the weighting vector as described in Section 2.1; m_{biom}, m_{gluc}, and m_{epo} are masses of the biomass, glucose, and erythropoietin, respectively. Similarly, C_k represents the concentration of component $k \in$ {biom, gluc, epo}, and $C_{k,0}$ is the initial concentration of component k. $C_{\text{gluc, in}}$ is the inlet concentration of the fed glucose, V is the volume of the biofluid element, V_0 is the initial condition of the volume at time, t_0, and ϕ_{gluc} is the feeding flux of glucose.

Furthermore, we bounded the glucose uptake flux with a Monod-type kinetic equation derived from Çelik et al. (2009b). This is to ensure that the glucose uptake flux at each time point is realistic (Höffner et al., 2013; Mahadevan et al., 2002). For the Monod-type kinetics bounding the glucose uptake flux v_{gluc}, $\mu_{\text{gluc, max}}$ represents the maximum specific growth rate on glucose, $Y_{\text{biom/gluc}}$ represents the biomass-to-glucose yield, K_{gluc} is the glucose saturation constant, and γ_{gluc} denotes the maintenance coefficient based on glucose. The specific erythropoietin production rate v_{epo} was assumed to follow the Luedeking-Piret model (Ren et al., 2003) and the coefficients a and b were determined by performing a least squares optimization on data from Çelik et al. (2009b) and Llaneras et al. (2012). For convenience, all kinetic parameters are summarized in Table 1. The bounds for final bioreaction time t_{f} and volume V were obtained from Çelik et al. (2009a,b) and were set between 20–60 h and 0.0–3.0 L, respectively, in order to ensure a fair comparison of our results to already published results in the literature.

Table 1
Model parameters.

Parameter	Value	Unit	Source
a	4.8×10^{-4}	–	This work
b	8×10^{-5}	–	This work
K_{gluc}	0.1	g L^{-1}	Jahic et al. (2002)
γ_{gluc}	0.025	g g^{-1} h^{-1}	Çelik et al. (2009b)
$\mu_{\text{gluc, max}}$	0.032	h^{-1}	Jungo et al. (2007)
$Y_{\text{biom/gluc}}$	0.62	g g^{-1}	Jungo et al. (2007)

5. Results and discussion

5.1. Extracellular states and metabolites

The maximum productivity of erythropoietin obtained after solving the problem in Section 4 was 17.97 mg h^{-1}, while the total cultivation time was at the lower bound of 20 h. This maximum productivity is approximately 66% higher than the benchmark experimental study described in Section 4 (Çelik et al., 2009a). Moreover, the optimization problem was solved within a total computational solution time of approximately 0.3 s. This shows that the model-based optimization approach presented in this paper is able to predict an improvement in the productivity of recombinant proteins in *P. pastoris*. Therefore, if experiments validate this improvement in productivity, the proposed model-based design framework will reduce the number of costly experiments and facilitate the design of efficient biopharmaceutical processes.

The concentration profiles of the glucose substrate and biomass are shown in Fig. 4a. It can be seen that the glucose concentration is gradually consumed by the cells to ensure growth. Concurrently, the glucose substrate feeding starts at a rate of 50 mL/h at the initial time point and decreases slightly to 45 mL/h at t = 18 h to indicate a slow transitioning from growth to production phase (cf. Fig. 5). This is to ensure the rapid growth of the *P. pastoris* cells in the shortest possible time, thereby ensuring maximum productivity of recombinant erythropoietin. Rapid cellular growth is crucial for the maximization of the productivity because growth rate is proportional to the erythropoietin production rate as mathematically represented by the Luedeking-Piret model in Section 4. From a practical point of view, maximum productivity is favored by rapid cellular growth because recombinant erythropoietin is heterologously expressed and secreted in the host cells (Park and Fred Ramirez, 1988). Furthermore, the glucose concentration in the extracellular environment is almost exhausted after 18 h to indicate the complete transition from the growth phase to the production phase (cf. Fig. 4a). Next, a sudden increase in the glucose substrate feeding from 45 to 195 mL/h is observed at the 18-h mark, maintained at this value for 1 h before decreasing to approximately zero at the final time $t_{\text{f}} = 20$ h (cf. Fig. 5). This sudden increase in glucose feed rate could be to counteract the exhaustion of glucose and cell death in fermentation media at time, $t = 18$ h (see Fig. 4a) while ensuring that erythropoietin productivity is maximized towards the end of the process (see Fig. 4b). This shows that our approach can handle trade-off issues between growth and productivity that typically arise in biopharmaceutical production. It can also be seen on Fig. 5 that the highest volume obtained is approximately 1.8 L and therefore within the working volume of 0.5 - 2.0 L stipulated by Çelik et al. (2009a).

We can also see from Fig. 4 that cells grow exponentially for the first 18 h and then the biomass concentration decreases slightly towards the end of the fermentation time of 20 h. This decrease indicates the commencement of proteolytic degradation as described in other studies (Çelik et al., 2009b; Muñoz et al., 2008). Thus, the model-based optimization approach presented in this work is able

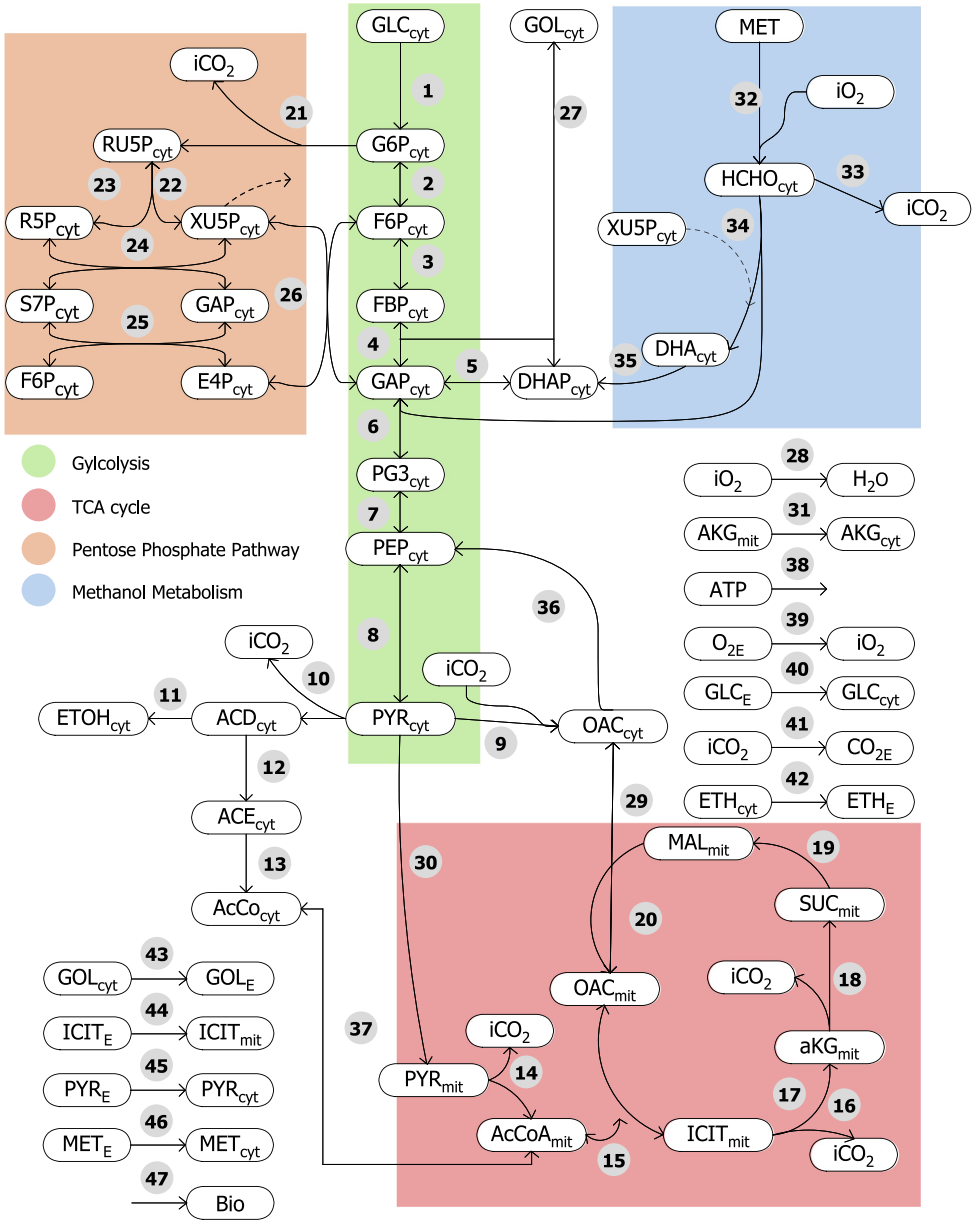

Fig. 3. Metabolic network of *Pichia pastoris*, adapted from Morales et al. (2014).

to predict both the exponential growth and proteolytic degradation that are akin to *P. pastoris* (Muñoz et al., 2008; Potvin et al., 2012).

Moreover, our approach simultaneously predicts all considered intracellular fluxes of *P. pastoris* along with the extracellular fluxes. Fig. 6 shows that the intracellular glucose uptake flux follows a similar trend as the extracellular glucose concentration profile (cf. Fig. 4).

Furthermore, the slight decrease in external biomass concentration after 18 h (Fig. 4) is caused by the intracellular growth rate flux which reaches zero at the same time (Fig. 6). On the other hand, the biomass growth rate gradually decreases as the extracellular biomass and substrate concentrations increase and decrease, respectively. However, we notice that the biomass concentration increases to 221.27 g/L in 18 h after which it slightly decreases to 220.20 g/L at the final time. This can be explained by the intracellular growth rate flux which reaches zero at this same time thus implying that cellular growth has stopped (cf. Fig. 6).

Table 2
Flux-gene-enzyme mapping (adapted from Nocon et al., 2014).

Flux	Gene	Enzyme	Functional category
11	ADH2	Alcohol dehydrogenase	Fermentative pathway
20	MDH1	Malate dehydrogenase	TCA
21	SOL3	6-phosphogluconolactonase	PPP
21	ZWF1	Glucose-6-phosphatedehydrogenase	PPP
22	RPE1	Ribulose-5-phosphate-3- epiremase	PPP
23	RPE1	Ribulose-5-phosphate-3- epiremase	PPP

5.2. Intracellular fluxes

The optimal dynamic activity of the intracellular fluxes in *P. pastoris* are shown in Fig. 7. Details on the metabolic reactions (fluxes) are shown in Fig. 3. These dynamic flux distributions correspond to the optimal metabolic physiological activity required for maximum productivity of erythropoeitin.

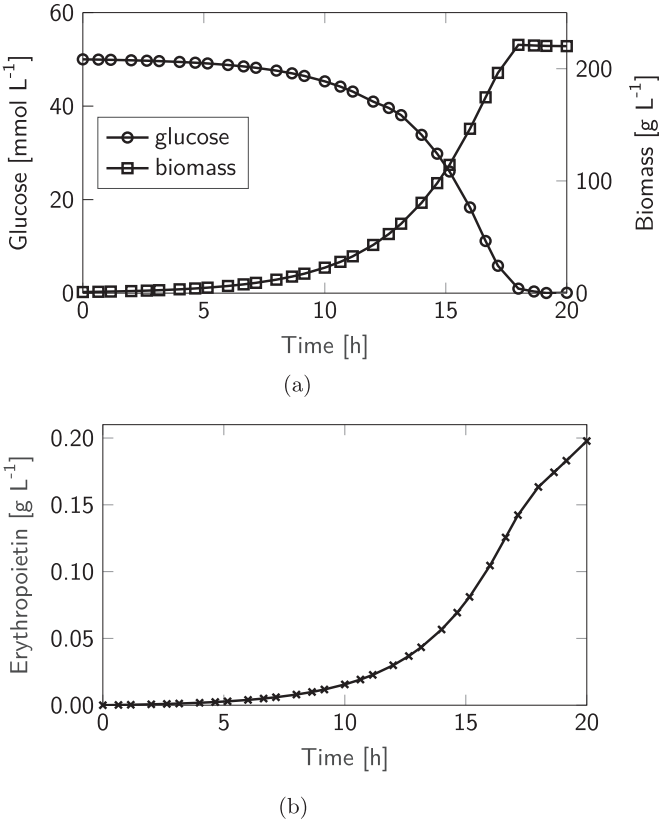

Fig. 4. Extracellular concentration profiles for the glucose substrate and biomass (a); and the recombinant erythropoietin product (b).

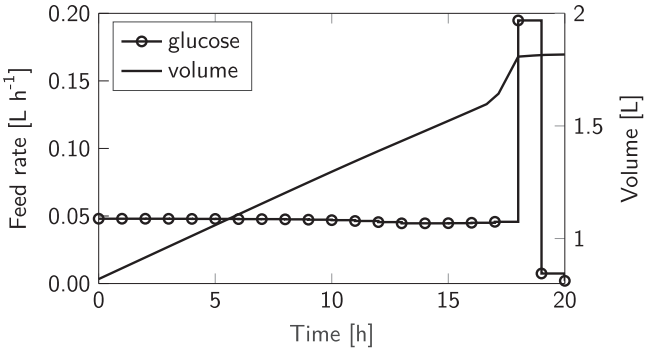

Fig. 5. Substrate feed rate control and volume profiles.

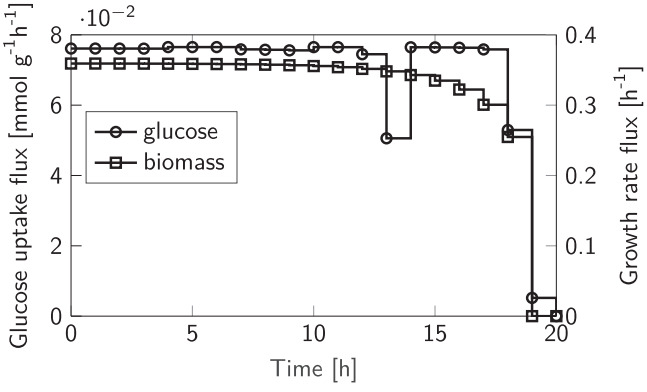

Fig. 6. Intracellular substrate uptake and growth rate fluxes.

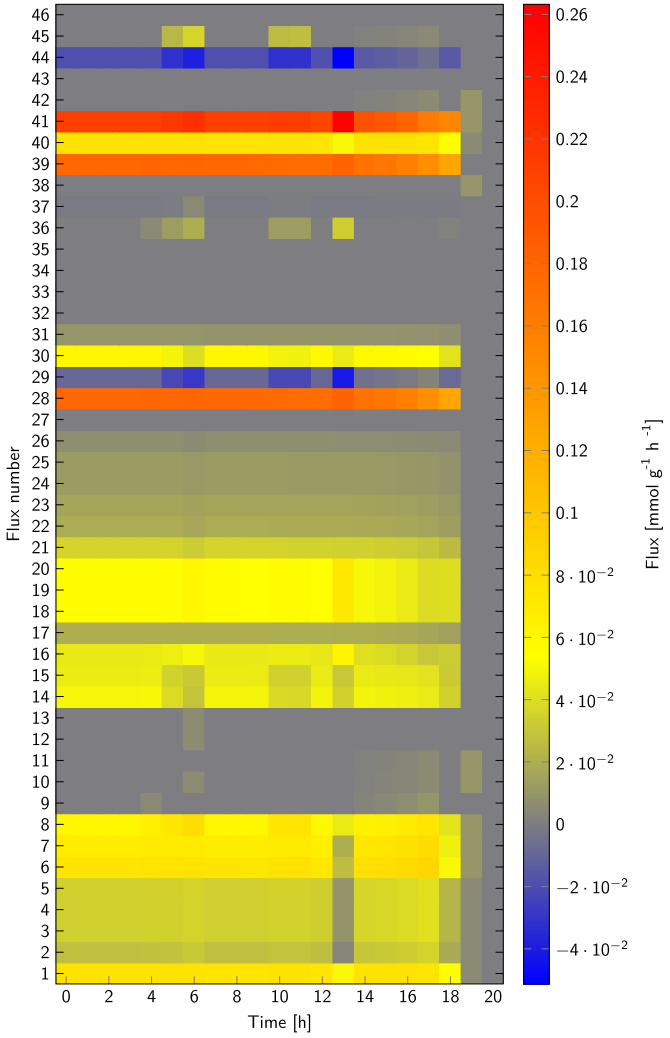

Fig. 7. Optimal dynamic intracellular fluxes in *P. pastoris*. Flux numbers correspond to the intracellular reactions in Fig. 3.

Note that since methanol and glycerol uptake fluxes are set to zero as described in Section 4, fluxes 32–35 and 46 corresponding to methanol metabolism and flux 27 corresponding to glycerol formation are inactive throughout the process. In the subsequent sections, we discuss in detail optimal flux evolution of the glycolysis, tricarboxylic acid (TCA) cycle, fermentative, pentose phosphate pathways, and the transport fluxes.

5.2.1. Embden–Meyerhoff–Parnas (glycolysis) pathway

First, we consider the flux distributions at the initial time point (cf. Fig. 8a). We observe that fluxes 1–8 corresponding to the Embden Meyerhoff Parnas (EMP) pathway (i.e., glycolysis) are higher at the initial time point $t = 0$ h than at time, $t = 19$ h. This implies that for optimal productivity to be achieved, a higher activity of the glycolysis pathway at the onset of the fermentation is required during the exponential growth phase when high cell growth is crucial. This high activity of the EMP pathway plays an important role in the production of energy in the form of ATPs required for the synthesis of organic intermediates and amino acids (Vercammen et al., 2017).

Apart from the onset of the fermentation process, the EMP fluxes (fluxes 1–8) are active for the majority of the fermentation process until 18 h where the growth rate decreases in order to facilitate higher erythropoietin production in the shortest possible time (see Fig. 7). On the contrary, towards the end of the process,

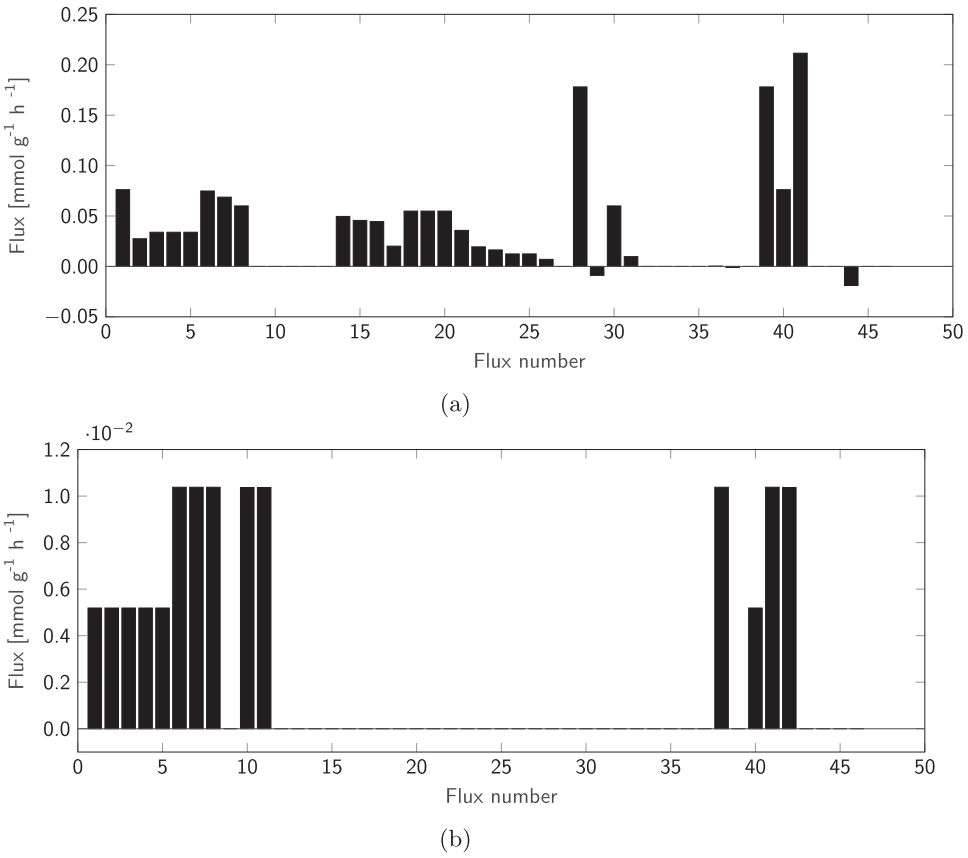

Fig. 8. Flux distributions at the initial time, $t_0 = 0$ h (a); and at time, $t = 19$ h close to the final time (b). Flux numbers correspond to the intracellular reactions in Fig. 3.

relatively lower glycolysis fluxes are required for cellular maintenance in order to counteract the cellular burden posed by recombinant protein production (Heyland et al., 2011).

We also note that the fluxes 9 and 10 are inactive for the first 14 h of the process, but increases slightly towards the end of the process (see Figs. 8b and 7). Intracellular reactions 9 and 10 correspond to the pathway which drives the formation and consumption of pyruvate, and therefore implies that pyruvate is produced in trace amounts as a by-product at the end of the bioreaction. This validates the claim by other authors that trace amounts of pyruvate are sometimes present in the media when using *P. pastoris* (Heyland et al., 2010; Isidro et al., 2016). This implies that for optimal productivity to be ensured, *P. pastoris* strain should be metabolically engineered in such a way that fluxes corresponding to pyruvate formation are inactive. However, more experimental work is required to validate this finding.

5.2.2. Tricarboxylic acid (TCA) cycle

The TCA cycle is an important part of the central carbon metabolism of *P. pastoris* because it produces a majority of energy required for the synthesis of amino acids used for cell growth (Vercammen et al., 2017). The dynamic intracelluar fluxes in TCA cycle are shown in Fig. 7. It can be concluded that the activity of the TCA cycle (fluxes 14–20) implies that enough energy is produced during the exponential growth phase, and this energy is simultaneously used to produce amino acids required for cellular growth and recombinant proteins during the production phase (cf. Fig. 7).

By looking at Figs. 6 and 7, we see that the TCA cycle and biomass flux are positively correlated with the exception of time $t = 13$ h. Here, we see that the biomass flux decreases while fluxes 18–20 increase slightly. This slight increase suggests an increase in energy generation through the TCA cycle for productivity maximization while compensating for low growth rate (Heyland et al., 2011).

We also notice that the malate dehydrogenase (MDH1) gene (see Table 2) which drives flux 20 is active for the most part of the process. Nocon et al. (2014) state that overexpression and activity of MDH1 might be beneficial for protein production, while Driouch et al. (2012) suggest that the downregulation or knockout of MDH1 could improve protein synthesis in the fungus *Aspergillus niger*. Due to this discrepancy, Nocon et al. (2014) have reported that there is no lucid correlation between MDH1 activity and protein production. Our results suggest that the activity or inactivity of the MDH1 could depend on the exact time point in the fermentation in which a measurement is taken. Specifically, we notice a dynamic switching between activity and inactivity of the MDH1 gene (flux 20) at the 19-h mark. It might be that the dynamic switching between activity and inactivity of MDH1 could favor optimal erythropoietin production. Furthermore, this could be implemented experimentally by dynamically overexpressing and knocking out the MDH1 gene in the growth and production phases, respectively (Brockman and Prather, 2015a; 2015b).

5.3. Fermentative pathway

The fermentative pathway is represented by fluxes 11–13. Fluxes 12 and 13 of the fermentative pathway are inactive throughout time horizon (cf. Fig. 7). Flux 11 corresponds to the activity of the alcohol dehydrogenase (ADH2) gene (cf. Table 2) which drives the intracellular reaction that results in the production of ethanol in the cytoplasm. Similar to the fluxes 9 and 10 of the pyruvate branch point, flux 11 slightly increases towards the end of the process due to the activity of ADH2; implying that ethanol is formed

towards the end of process where cellular burden is high. Since ethanol production could lead to lower erythropoeitin productivity, our results show that *P. pastoris* can optimally adjust its physiology properly.

This is done in order to ensure that erythropoeitin productivity is still maximized towards the end of the process. Our results support earlier studies which claim that *P. pastoris* host cells favor the respiratory pathway and as such produce little or no by-products and are less prone to follow the fermentative pathways as in *E. coli* or *S. cerevisiae* (Fazenda et al., 2013).

5.3.1. Pentose phosphate pathway (PPP)

The pentose phosphate pathway (PPP) is a part of the central carbon metabolism of *P. pastoris* that utilizes the energy produced from the EMP pathway and TCA cycle to synthesize the amino acids and organic precursors required for biomass synthesis and recombinant protein production. The results for the optimal dynamic intracellular fluxes in PPP (fluxes 21–26) are shown in Fig. 7.

Here, fluxes 21–26 are active throughout the first 18 h when growth rate is highest (see Figs. 6 and 7). This is logical as the PPP is a major producer of the amino acids and organic precursors required for biomass synthesis (Vercammen et al., 2017). A closer look at Fig. 7 reveals that flux 21 which corresponds to the 6-phosphogluconolactonase (SOL3) and glucose-6-phosphatedehydrogenase (ZWF1) genes has the strongest impact on the PPP, and consequently, the biomass and the recombinant erythropoietin protein production. This implies that for maximum productivity, more emphasis should be placed on engineering *P. pastoris* strains such that the activity of the ZWF1 and SOL3 genes are higher than other PPP-associated genes such as ribulose-5-phosphate-3-epiremase (see Table 2). Lastly, we observe that flux 28 is very high for the first 17 h but decreases slightly at the 18-h mark. This is logical because flux 28 represents oxidative phosphorylation which is a reaction that produces energy in the form of ATP that is required for both biomass synthesis and maximum erythropoietin productivity.

5.3.2. Transport fluxes

Transport fluxes (fluxes 29–46) facilitate the exchange of metabolites between the *P. pastoris* cell membrane and the extracellular environment, and between the mitochondrion and the cytoplasm within the cell (cf. Fig. 7). Flux 31 shows the transport of α-ketoglutarate which serves as an organic precursor for the synthesis of biomass. Even though, flux 31 is very low throughout the process, its presence throughout the growth phase implies that it is still important for maximum biomass synthesis.

Moreover, the cell maintenance is represented by flux 38. It can be seen that there is no need for cellular maintenance during the growth phase. However, as soon as the cell transitions to the production phase at $t = 18$ h, we observe the activity of the cellular maintenance flux. This sudden need for cellular maintenance can be attributed to the need for the cell to counteract proteolytic degradation. This is logical and attests to the key insights that can be obtained by using the approach presented in this paper. We also observe that the glucose uptake flux (flux 40) is relatively high until $t = 18$ h (see also Fig. 6). Flux 39 which represents the oxygen uptake flux is active and high for the first 18 h and as a result justifies our assumption that the process is well aerated (see Section 4). This implies that for maximum recombinant production, high oxygen rates are required.

Another important finding is the presence of fermentative by-products. Most studies suggest that fermentative by-products such as ethanol, citrate acid or pyruvate are either not produced or produced in little quantities (Heyland et al., 2010; Isidro et al., 2016). This is because *P. pastoris* has been known to follow oxidative respiration instead of the fermentative pathway (Fazenda et al., 2013; Isidro et al., 2016). However, our results show that ethanol (flux 42) is excreted to the extracellular environment during the production phase, but not produced during the growth phase (0–13 h). This could be the reason why some authors do not detect ethanol in their studies (Çelik et al., 2010; Sola et al., 2007). Furthermore, our results suggest that citric acid (flux 44) and pyruvate (flux 45) are present during the process. Nonetheless, the presence of fermentative by-products is supported by other studies (Heyland et al., 2010; 2011) and NMR experiments (Isidro et al., 2016).

Even though fermentative by-products such as ethanol are considered to be disadvantageous during heterologous protein production, it might be that a little ethanol in the culture media serves as a carbon source when the main carbon sources such as glucose are exhausted, thus improving productivity (Wegerhoff and Engell, 2016). This shows that our dFBA approach can provide insights into the underlying biological occurrences in the *P. pastoris* cells; thus, serving as a detailed modeling approach for the design of optimal bioreactors for biopharmaceutical manufacturing.

6. Conclusions

In this paper, we have presented a model-based strategy to optimize the productivity of recombinant protein production by *P. pastoris* that is based on dFBA. In our work, the external environment is modeled within the EPF framework, while the intracellular metabolic network is modeled with FBA. Our approach enables us to gain insights into what dynamic strategies can be implemented at both bioreactor and microorganism scales in order to maximize productivity.

Besides the maximization of productivity, we believe that the approach presented herein can be extended to other objectives that are of biological relevance. Another key contribution in this work is the efficient solution strategy for the dFBA. Here, we have shown that dFBA problems can be solved efficiently and directly by replacing the lower-level with its KKT conditions (Raghunathan et al., 2003), handling the belligerent complementarity constraints with the ℓ_1 penalization technique (Baumrucker et al., 2008), and solving the dynamic optimization by the simultaneous approach (Biegler, 2007).

Nevertheless, there are some open questions and research opportunities to be tackled. For instance, the EPF formulation of the bioreactor design problem presented needs to be explored in greater detail in future studies. A possible direction will be to combine the EPF bioreactor formulation with the three-level reactor design approach proposed by Peschel et al. (2010). This could lead to the design of innovative bioreactors that consider the dynamic manipulations of extracellular and intracellular controls for improving the production of biologic drugs in *P. pastoris*.

Another interesting aspect to consider is the oxygen consumption rate. In this work, we have assumed that sufficient oxygen is available for the cells to grow. Nevertheless, it might be interesting in the future to include an oxygen balance and dissolved oxygen as an additional control variable (Güneş and Çalık, 2016). In addition, some of the trajectories of the intracellular fluxes are not smooth. Hence, B-splines can be used to parameterize the intracellular fluxes profiles so as to obtain smoothened curves (Vercammen et al., 2014). All in all, we believe that the model-based optimization strategy presented in this paper is a valuable contribution to the growing literature on strategies for improving the heterologous expression of proteins in *P. pastoris* and could facilitate the design of next generation biopharmaceutical processes.

Acknowledgment

We gratefully acknowledge the funding from Ministry of Science and Culture (MWK), Grant no. ZN 2887, Lower Saxony, Germany under the SynFoBiA project “Novel synthesis and formulation method for poorly soluble drugs and sensitive pharmaceuticals”.

References

Aggarwal, S.R., 2014. What's fueling the biotech engine–2012 to 2013. Nature 32 (1), 32–39. doi:10.1038/nbt.2794.

Anesiadis, N., Cluett, W.R., Mahadevan, R., 2008. Dynamic metabolic engineering for increasing bioprocess productivity. Metab. Eng. 10 (5), 255–266. doi:10.1016/j.ymben.2008.06.004.

Baumrucker, B., Renfro, J., Biegler, L.T., 2008. MPEC problem formulations and solution strategies with chemical engineering applications. Comput. Chem. Eng. 32 (12), 2903–2913. doi:10.1016/j.compchemeng.2008.02.010.

Bellman, R., 1952. On the theory of dynamic programming. Proc. Natl. Acad. Sci. 38 (8), 716–719.

Biegler, L.T., 1984. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. Comput. Chem. Eng. 8 (3–4), 243–247. doi:10.1016/0098-1354(84)87012-X.

Biegler, L.T., 2007. An overview of simultaneous strategies for dynamic optimization. Chem. Eng. Process. 46 (11), 1043–1053. doi:10.1016/j.cep.2006.06.021.

Biegler, L.T., 2010. Nonlinear Programming. Society for Industrial and Applied Mathematics, Philadelphia doi:10.1137/1.9780898719383.

Bock, H.G., Plitt, K.-J., 1984. A multiple shooting algorithm for direct solution of optimal control problems. IFAC Proc. Vol. 17 (2), 1603–1608. doi:10.1016/S1474-6670(17)61205-9.

Boghigian, B.A., Seth, G., Kiss, R., Pfeifer, B.A., 2010. Metabolic flux analysis and pharmaceutical production. Metab. Eng. 12 (2), 81–95. doi:10.1016/j.ymben.2009.10.004.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York.

Brockman, I.M., Prather, K.L., 2015a. Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. Metab. Eng. 28, 104–113. doi:10.1016/j.ymben.2014.12.005.

Brockman, I.M., Prather, K.L., 2015b. Dynamic metabolic engineering: new strategies for developing responsive cell factories. Biotechnol. J. 10 (9), 1360–1369. doi:10.1016/10.1002/biot.201400422.

Çalık, P., Şahin, M., Taşpınar, H., Soyaslan, E.Ş., İnankur, B., 2011. Dynamic flux balance analysis for pharmaceutical protein production by *Pichia pastoris*: human growth hormone. Enzyme Microb. Technol. 48 (3), 209–216. doi:10.1016/j.enzmictec.2010.09.016.

Çelik, E., Çalık, P., Oliver, S.G., 2009a. Fed-batch methanol feeding strategy for recombinant protein production by *Pichia pastoris* in the presence of co-substrate sorbitol. Yeast 26 (9), 473–484. doi:10.1002/yea.1679.

Çelik, E., Çalık, P., Oliver, S.G., 2009b. A structured kinetic model for recombinant protein production by Mut+ strain of *Pichia pastoris*. Chem. Eng. Sci. 64 (23), 5028–5035.

Çelik, E., Çalık, P., Oliver, S.G., 2010. Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: effects of methanol feeding rate. Biotechnol. Bioeng. 105 (2), 317–329. doi:10.1002/bit.22543.

Cereghino, G.P.L., Cereghino, J.L., Ilgen, C., Cregg, J.M., 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr. Opin. Biotechnol. 13 (4), 329–332. doi:10.1016/S0958-1669(02)00330-0.

Cereghino, J.L., Cregg, J.M., 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. 24 (1), 45–66. doi:10.1111/j.1574-6976.2000.tb00532.x.

Colson, B., Marcotte, P., Savard, G., 2007. An overview of bilevel optimization. Ann. Oper. Res. 153 (1), 235–256. doi:10.1007/s10479-007-0176-2.

Cos, O., Ramon, R., Montesinos, J.L., Valero, F., 2006. A simple model-based control for *Pichia pastoris* allows a more efficient heterologous protein production bioprocess. Biotechnol. Bioeng. 95 (1), 145–154. doi:10.1002/bit.21005.

Cuthrell, J.E., Biegler, L.T., 1987. On the optimization of differential-algebraic process systems. AIChE J. 33 (8), 1257–1270.

d'Anjou, M.C., Daugulis, A.J., 2001. A rational approach to improving productivity in recombinant *Pichia pastoris* fermentation. Biotechnol. Bioeng. 72 (1), 1–11. doi:10.1002/1097-0290(20010105)72:1<1::AID-BIT1>3.0.CO;2-T.

Driouch, H., Melzer, G., Wittmann, C., 2012. Integration of *in vivo* and *in silico* metabolic fluxes for improvement of recombinant protein production. Metab. Eng. 14 (1), 47–58. doi:10.1016/j.ymben.2011.11.002.

Drud, A.S., 1994. CONOPT–a large-scale GRG code. ORSA J. Comput. 6 (2), 207–216. doi:10.1287/ijoc.6.2.207.

Emenike, V.N., Krewer, U., 2016. Model-based optimal design of continuous-flow reactors for the synthesis of active pharmaceutical ingredients. Chem. Ing. Tech. 88 (9), 1215–1216. doi:10.1002/cite.201650267.

Emenike, V.N., Schenkendorf, R., Krewer, U., 2017. Model-based optimization of the recombinant protein production in *Pichia pastoris* based on dynamic flux balance analysis and elementary process functions. In: Antonio Espuña, M.G., Puigjaner, L. (Eds.), 27th European Symposium on Computer Aided Process Engineering. In: Computer Aided Chemical Engineering. Elsevier, pp. 2815–2820. doi:10.1016/B978-0-444-63965-3.50471-2.

Emenike, V.N., Schenkendorf, R., Krewer, U., 2018. A systematic reactor design approach for the synthesis of active pharmaceutical ingredients. Eur. J. Pharm. Biopharm. 126, 75–88. doi:10.1016/j.ejpb.2017.05.007.

Fazenda, M.L., Dias, J.M., Harvey, L.M., Nordon, A., Edrada-Ebel, R., LittleJohn, D., McNeil, B., 2013. Towards better understanding of an industrial cell factory: investigating the feasibility of real-time metabolic flux analysis in *Pichia pastoris*. Microb. Cell Fact. 12, 51. doi:10.1186/1475-2859-12-51.

Fourer, R., Gay, D., Kernighan, B., 2003. AMPL. A Modelling Language for Mathematical Programming, 2nd Ed. Brooks/Cole Publishing Company.

Freund, H., Sundmacher, K., 2008. Towards a methodology for the systematic analysis and design of efficient chemical processes: Part 1. From unit operations to elementary process functions. Chem. Eng. Process. 47 (12), 2051–2060. doi:10.1016/j.cep.2008.07.011.

Güneş, H., Çalık, P., 2016. Oxygen transfer as a tool for fine-tuning recombinant protein production by *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter. Bioprocess Biosyst. Eng. 39 (7), 1061–1072. doi:10.1007/s00449-016-1584-y.

Hessel, V., 2009. Novel process windows–gate to maximizing process intensification via flow chemistry. Chem. Eng. Technol. 32 (11), 1655–1681. doi:10.1002/ceat.200900474.

Heyland, J., Fu, J., Blank, L.M., Schmid, A., 2010. Quantitative physiology of *Pichia pastoris*during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol. Bioeng. 107 (2), 357–368. doi:10.1002/bit.22836.

Heyland, J., Fu, J., Blank, L.M., Schmid, A., 2011. Carbon metabolism limits recombinant protein production in *Pichia pastoris*. Biotechnol. Bioeng. 108 (8), 1942–1953. doi:10.1002/bit.23114.

Hjersted, J.L., Henson, M.A., 2006. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. Biotechnol. Prog. 22 (5), 1239–1248. doi:10.1021/bp060059v.

Höffner, K., Harwood, S., Barton, P., 2013. A reliable simulator for dynamic flux balance analysis. Biotechnol. Bioeng. 110 (3), 792–802. doi:10.1002/bit.24748.

Isidro, I.A., Portela, R.M., Clemente, J.J., Cunha, A.E., Oliveira, R., 2016. Hybrid metabolic flux analysis and recombinant protein prediction in *Pichia pastoris* X-33 cultures expressing a single-chain antibody fragment. Bioprocess Biosyst. Eng. 39 (9), 1351–1363. doi:10.1007/s00449-016-1611-z.

Jacobs, K., Shoemaker, C., Rudersdorf, R., Neill, S.D., Kaufman, R.J., Mufson, A., Seehra, J., Jones, S.S., Hewick, R., Fritsch, E.F., et al., 1985. Isolation and characterization of genomic and cDNA clones of human erythropoietin. Nature 313 (1), 806–810. doi:10.1038/313806a0.

Jahic, M., Rotticci-Mulder, J., Martinelle, M., Hult, K., Enfors, S.-O., 2002. Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess Biosyst. Eng. 24 (6), 385–393. doi:10.1007/s00449-001-0274-5.

Joy, J., Kremling, A., 2010. Study of the growth of *Escherichia coli* on mixed substrates using dynamic flux balance analysis. IFAC Proc. Vol. 43 (6), 401–406. doi:10.3182/20100707-3-BE-2012.0059.

Jungo, C., Schenk, J., Pasquier, M., Marison, I.W., von Stockar, U., 2007. A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with *Pichia pastoris*. J. Biotechnol. 131 (1), 57–66. doi:10.1016/j.jbiotec.2007.05.019.

Kobayashi, K., Kuwae, S., Ohya, T., Ohda, T., Ohyama, M., Tomomitsu, K., 2000. High level secretion of recombinant human serum albumin by fed-batch fermentation of the methylotrophic yeast, *Pichia pastoris*, based on optimal methanol feeding strategy. J. Biosci. Bioeng. 90 (3), 280–288. doi:10.1016/S1389-1723(00)80082-1.

Kumar, D., Budman, H., 2017. Applications of polynomial chaos expansions in optimization and control of bioreactors based on dynamic metabolic flux balance models. Chem. Eng. Sci. 167, 18–28. doi:10.1016/j.ces.2017.03.035.

Kyparisis, J., 1985. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. Math. Program. 32 (2), 242–246. doi:10.1007/BF01586095.

Li, D., Wang, Q., Wang, J., Yao, Y., 2008. Mitigation of curse of dimensionality in dynamic programming. IFAC Proc. Vol. 41 (2), 7778–7783. doi:10.3182/20080706-5-KR-1001.01315.

Llaneras, F., Tortajada, M., Ramón, D., Picó, J., 2012. Dynamic metabolic flux analysis for online estimation of recombinant protein productivity in *Pichia pastoris* cultures. IFAC Proc. Vol. 45 (2), 629–634. doi:10.3182/20120215-3-AT-3016.00112.

Love, J.C., Love, K.R., Barone, P.W., 2013. Enabling global access to high-quality biopharmaceuticals. Curr. Opin. Chem. Eng. 2 (4), 383–390. doi:10.1016/j.coche.2013.09.002.

Lu, A.E., Paulson, J.A., Mozdzierz, N.J., Stockdale, A., Versypt, A.N.F., Love, K.R., Love, J.C., Braatz, R.D., 2015. Control systems technology in the advanced manufacturing of biologic drugs. In: Control Applications (CCA), 2015 IEEE Conference on. IEEE, pp. 1505–1515. doi:10.1109/CCA.2015.7320824.

Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys. J. 83 (3), 1331–1340. doi:10.1016/S0006-3495(02)73903-9.

Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab. Eng. 12 (2), 150–160. doi:10.1016/j.ymben.2009.07.006.

Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F., 2014. Validation of an FBA model for *Pichia pastoris* in chemostat cultures. BMC Syst. Biol. 8, 142. doi:10.1186/s12918-014-0142-y.

Mozdzierz, N.J., Love, K.R., Lee, K.S., Lee, H.L., Shah, K.A., Ram, R.J., Love, J.C., 2015. A perfusion-capable microfluidic bioreactor for assessing microbial heterologous protein production. Lab Chip 15 (14), 2918–2922. doi:10.1039/C5LC00443H.

Muñoz, D.F.M., Enciso, N.A.A., Ruiz, H.C., Avellaneda, L.A.B., 2008. A simple structured model for recombinant IDShr protein production in *Pichia pastoris*. Biotechnol. Lett. 30 (10), 1727–1734. doi:10.1007/s10529-008-9750-1.

Nikdel, A., Braatz, R.D., Budman, H.M., 2018. A systematic approach for finding the objective function and active constraints for dynamic flux balance analysis. Bioprocess Biosyst. Eng. 41 (5), 641–655. doi:10.1007/s00449-018-1899-y.

Niu, H., Daukandt, M., Rodriguez, C., Fickers, P., Bogaerts, P., 2013. Dynamic modeling of methylotrophic *Pichia pastoris* culture with exhaust gas analysis: from cellular metabolism to process simulation. Chem. Eng. Sci. 87, 381–392. doi:10.1016/j.ces.2012.11.006.

Nocon, J., Steiger, M.G., Pfeffer, M., Sohn, S.B., Kim, T.Y., Maurer, M., Rußmayer, H., Pflügl, S., Ask, M., Haberhauer-Troyer, C., et al., 2014. Model based engineering of *Pichia pastoris*central metabolism enhances recombinant protein production. Metab. Eng. 24, 129–138. doi:10.1016/j.ymben.2014.05.011.

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248.

Park, S., Fred Ramirez, W., 1988. Optimal production of secreted protein in fed-batch reactors. AIChE J. 34 (9), 1550–1558. doi:10.1002/aic.690340917.

Peschel, A., 2012. Model-based Design of Optimal Chemical Reactors. Shaker Verlag, Aachen.

Peschel, A., Freund, H., Sundmacher, K., 2010. Methodology for the design of optimal chemical reactors based on the concept of elementary process functions. Ind. Eng. Chem. Res. 49 (21), 10535–10548. doi:10.1021/ie100476q.

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.F., Mishechenko, E.F., 1962. Mathematical Theory of Optimal Processes. Wiley, New York.

Potvin, G., Ahmad, A., Zhang, Z., 2012. Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: a review. Biochem. Eng. J. 64, 91–105. doi:10.1016/j.bej.2010.07.017.

Raghunathan, A.U., Pérez-Correa, J.R., Bieger, L.T., 2003. Data reconciliation and parameter estimation in flux-balance analysis. Biotechnol. Bioeng. 84 (6), 700–709. doi:10.1002/bit.10823.

Ralph, D., Wright, S.J., 2004. Some properties of regularization and penalization schemes for MPECs. Optim. Methods Softw. 19 (5), 527–556. doi:10.1080/10556780410001709439.

Ren, H., Yuan, J., Bellgardt, K.-H., 2003. Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance. J. Biotechnol. 106 (1), 53–68. doi:10.1016/j.jbiotec.2003.08.003.

Sager, S., 2009. Reformulations and algorithms for the optimization of switching decisions in nonlinear optimal control. J. Process Control 19 (8), 1238–1247. doi:10.1016/j.jprocont.2009.03.008.

Saitua, F., Torres, P., Pérez-Correa, J.R., Agosin, E., 2017. Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. BMC Syst. Biol. 11, 27. doi:10.1186/s12918-017-0408-2.

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. Metab. Eng. 25, 159–173. doi:10.1016/j.ymben.2014.07.004.

Schiestl, M., Stangler, T., Torella, C., Čepeljnik, T., Toll, H., Grau, R., 2011. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat. Biotechnol. 29 (4), 310–312. doi:10.1038/nbt.1839.

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol. Syst. Biol. 3, 119. doi:10.1038/msb4100162.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99 (23), 15112–15117. doi:10.1073/pnas.232349399.

Sola, A., Jouhten, P., Maaheimo, H., Sanchez-Ferrando, F., Szyperski, T., Ferrer, P., 2007. Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. Microbiology 153 (1), 281–290. doi:10.1099/mic.0.29263-0.

Sreekrishna, K., Brankamp, R.G., Kropp, K.E., Blankenship, D.T., Tsay, J.-T., Smith, P.L., Wierschke, J.D., Subramaniam, A., Birkenberger, L.A., 1997. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. Gene 190 (1), 55–62. doi:10.1016/S0378-1119(96)00672-5.

St. John, P.C., Crowley, M.F., Bomble, Y.J., 2017. Efficient estimation of the maximum metabolic productivity of batch systems. Biotechnol. Biofuels 10, 28. doi:10.1186/s13068-017-0709-0.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5 (1), 93–121. doi:10.1038/nprot.2009.203.

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

Vassiliadis, V., Sargent, R., Pantelides, C., et al., 1994. Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints. Ind. Eng. Chem. Res. 33 (9), 2111–2122. doi:10.1021/ie00033a014.

Vassiliadis, V., Sargent, R., Pantelides, C., et al., 1994. Solution of a class of multistage dynamic optimization problems. 2. Problems with path constraints. Ind. Eng. Chem. Res. 33. doi:10.1021/ie00033a015. 2123–2123

Vercammen, D., Logist, F., Van Impe, J., 2014. Dynamic estimation of specific fluxes in metabolic networks using non-linear dynamic optimization. BMC Syst. Biol. 8, 132. doi:10.1186/s12918-014-0132-0.

Vercammen, D., Telen, D., Nimmegeers, P., Janssens, A., Akkermans, S., Fernandez, E.N., Logist, F., Van Impe, J., 2017. Application of a dynamic metabolic flux algorithm during a temperature-induced lag phase. Food Bioprod. Process. 102, 1–19. doi:10.1016/j.fbp.2016.10.003.

Wachsmuth, G., 2013. On LICQ and the uniqueness of Lagrange multipliers. Oper. Res. Lett. 41 (1), 78–80. doi:10.1016/j.orl.2012.11.009.

Waldherr, S., 2016. State estimation in constraint based models of metabolic-genetic networks. In: American Control Conference (ACC), 2016. IEEE, pp. 6683–6688. doi:10.1109/ACC.2016.7526723.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32 (10), 992–1000. doi:10.1038/nbt.3040.

Wegerhoff, S., Engell, S., 2016. Control of the production of *Saccharomyces cerevisiae* on the basis of a reduced metabolic model. IFAC-PapersOnLine 49 (26), 201–206. doi:10.1016/j.ifacol.2016.12.126.

Wells, E., Robinson, A.S., 2017. Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement. Biotechnol. J. 12 (1), 1860–7314. doi:10.1002/biot.201600105.

Xie, J., Zhou, Q., Du, P., Gan, R., Ye, Q., 2005. Use of different carbon sources in cultivation of recombinant *Pichia pastoris* for angiostatin production. Enzyme Microb. Technol. 36 (2), 210–216. doi:10.1016/j.enzmictec.2004.06.010.

Yang, L., Mahadevan, R., Cluett, W.R., 2008. A bilevel optimization algorithm to identify enzymatic capacity constraints in metabolic networks. Comput. Chem. Eng. 32 (9), 2072–2085. doi:10.1016/j.compchemeng.2007.10.015.

Zhang, W., Bevins, M.A., Plantz, B.A., Smith, L.A., Meagher, M.M., 2000. Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. Biotechnol. Bioeng. 70 (1), 1–8. doi:10.1002/1097-0290(20001005)70:1<1::AID-BIT1)3.0.CO;2-Y.

Zhao, X., Noack, S., Wiechert, W., von Lieres, E., 2017. Dynamic flux balance analysis with nonlinear objective function. J. Math. Biol. 75 (6–7), 1487–1515. doi:10.1007/s00285-017-1127-4.