Computational and Structural Biotechnology Journal xxx (2018) xxx–xxx

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mini Review

In Silico Prediction of Large-Scale Microbial Production Performance: Constraints for Getting Proper Data-Driven Models

Julia Zieringer, Ralf Takors *

Institute of Biochemical Engineering, University of Stuttgart, Germany

ARTICLE INFO

Article history:
Received 26 March 2018
Received in revised form 11 June 2018
Accepted 12 June 2018
Available online xxxx

Keywords:
Gene regulatory networks
Scale-down devices
CFD
Compartment models
CFD-based compartment models

ABSTRACT

Industrial bioreactors range from 10.000 to 700.000 L and characteristically show different zones of substrate availabilities, dissolved gas concentrations and pH values reflecting physical, technical and economic constraints of scale-up. Microbial producers are fluctuating inside the bioreactors thereby experiencing frequently changing micro-environmental conditions. The external stimuli induce responses on microbial metabolism and on transcriptional regulation programs. Both may deteriorate the expected microbial production performance in large scale compared to expectations deduced from ideal, well-mixed lab-scale conditions. Accordingly, predictive tools are needed to quantify large-scale impacts considering bioreactor heterogeneities. The review shows that the time is right to combine simulations of microbial kinetics with calculations of large-scale environmental conditions to predict the bioreactor performance. Accordingly, basic experimental procedures and computational tools are presented to derive proper microbial models and hydrodynamic conditions, and to link both for bioreactor modeling. Particular emphasis is laid on the identification of gene regulatory networks as the implementation of such models will surely gain momentum in future studies.

Contents

1. Introduction

With the advent of metabolic engineering in the 1990s Bailey [1], the engineers' view on microbes changed. Process optimization no longer considered the extracellular environment (i.e. cultivation conditions)

* Corresponding author.
E-mail address: takors@ibvt.uni-stuttgart.de (R. Takors).

https://doi.org/10.1016/j.csbj.2018.06.002

Please cite this article as: Zieringer J, Takors R, In Silico Prediction of Large-Scale Microbial Production Performance: Constraints for Getting Proper Data-Driven Models, Comput Struct Biotechnol J (2018), https://doi.org/10.1016/j.csbj.2018.06.002

alone, but started to investigate intracellular mechanisms in addition Bailey [1]; Vallino and Stephanopoulos [2]. Since then, intracellular reaction rates have been quantified and models of regulatory processes finally aiming at identifying ta rgets for further strain and process improvement have been derived. To some extent driven by the observations that cellular engineering always results in multiple and complex systemic responses Bailey [1], furthermore catalyzed by the avalanche of omics data that were accessible, systems biology and systems metabolic engineering emerged in 2000. In essence, holistic models have been developed that aim to provide as sound and comprehensive a cellular view as possible.

The development clearly reflects the general engineering mindset of investigating the whole system by modularization, quantitative analysis, reassembling and studying the interaction of the networked modules. The earliest, simple examples may be given by the Monod growth model Jacob and Monod [3], followed by more sophisticated approaches like the lactose operon considering feedback regulation in *Escherichia coli*, finally leading to complex models comprising multiple levels of cellular regulation Kitano [4]. While such movements led to the birth of systems biology Westerhoff and Palsson [5] and systems metabolic engineering Lee et al. [6]; Park and Lee [7]; Becker et al. [8]; Wittmann and Lee [9] core engineering activities such as scale-up were a matter of steady development, too.

Scale-up is the procedure to transfer lab-bioprocesses in production (large) conditions, often covering 7 to 8 orders of magnitude of volume. Unfortunately, loss or even failure of large-scale performance may occur. Detailed knowhow is necessary to prevent unwanted production losses. Accordingly, Oosterhuis and Kossen were the first who presented a scale-up simulator (1983) for investigating the impact of oxygen gradients on *Gluconobacter oxydans* Oosterhuis et al. [10]. They further introduced bioreactor compartment models to achieve the coarse spatial resolution of local oxygen transfer rates to identify micro- and anaerobic zones Oosterhuis and Kossen [11]. This line of thinking was followed by a series of similar studies Neubauer et al. [12]; Buchholz et al. [13]; Löffler et al. [14,15]; von Wulffen et al. [16] and reached a new level of complexity by linking simulations of hydrodynamics and mass transports with simple metabolic models of *Saccaromyces cerevisiae* and *E. coli* Bylund et al. [17]; Lapin et al. [18,19]; Wang et al. [20]; Haringa et al. [21]. Notably, cellular dynamics were modeled by focusing on metabolism dynamics only. This is remarkable as systems biology has already shown that holistic models are able to cover a far broader range of complexity. Scale-up engineers have already pointed out Delvigne et al. [22] that profound knowhow is necessary to enable the best knowledge-based scale-up using in silico predictions.

This review addresses the current need for knowledge-based process scale-up by elucidating the putative contributions of modeling. The existing plethora of modeling approaches will be structured with respect to granularity and usefulness to (i) identify and (ii) model key regulatory phenomena and (iii) to link cellular models with predictions of large-scale hydrodynamics. It will be shown that the time is right to approach the challenging goal of in silico predicted large-scale performance of microbial producers.

2. Data-driven Approach

Comprehensive data sets are necessary to develop gene regulatory models, generated to answer the biological question of interest. This also holds true for elucidating complex metabolic and regulatory responses of producer cells that are exposed to industrial production conditions. One approach to collect representative data is to mimic large-scale conditions and to capture time series of regulatory dynamics as a basis for unraveling dynamic regulatory models. Such approaches usually require rapid sampling experiments that 'freeze' metabolic states monitored in scale-down experiments. Examples of experimental procedures are given in the following.

2.1. Experimental Set-Ups Mimicking Large-Scale Heterogeneities

In large-scale production processes micro-environmental inhomogeneities often occur. Insufficient mixing leads to severe axial and horizontal concentration gradients. Producer cells frequently cross these poorly mixed zones which triggers metabolic and transcriptional responses accordingly Takors [23]. Because large-scale experimental data are rarely accessible, experimental scale-up simulators are typically applied, reflecting large-scale conditions Delafosse et al. [24]. Pioneering studies were performed by Oosterhuis and Kossen Oosterhuis et al. [10] using a two compartment system comprising two stirred tank reactors (STRs) to investigate the effect of different oxygen levels upon the gluconic acid fermentation of *Gluconobacter oxydans*. Since then, variations of the two compartment set up considered the combination of an STR and a plug flow reactor (PFR). Reviews have been given by Delvigne et al. and Neubauer et al. Delvigne et al. [22]; Neubauer and Junne [26]. Fig. 1 depicts selected examples for several STR-STR and STR-PFR applications.

Experimental scale-up simulators do not merely consist of two compartments. Three compartment approaches have been studied as well. Examples are the STR-STR-STR cascade of Buchholz et al. Buchholz et al. [13] and the PFR-STR-PFR set-up of Lemoine et al. [28]. Accordingly, more complex scale-up scenarios could be analyzed.

Notably, two and three compartment scale-up simulators mirror the cellular responses on repeated, frequent stimuli. In contrast, investigations of single perturbations may be a proper tool for deriving distinct stimulus/response correlations, see Fig. 1 for examples. On this basis, explicit metabolic and transcriptional dynamics can be deduced that, when properly superimposed, result in the complex cellular response observed. However, signal transduction is highly networked in the cells which may cause the cross-interference of multiple stimuli. The coincidence of multiple stimuli in large-scale fermentation is the rule rather than the exception Xu et al. [29]; Egli [30]. Accordingly, multiple stimulus/response studies are likely to gain importance in the future.

2.2. Experimental Access to Metabolic and Transcriptional Responses

Samples taken from the scale-up simulators need to be processed so that metabolic and transcriptional states are 'frozen' immediately. Metabolic inactivation and purification can be achieved via several approaches Oldiges et al. [31]; Teleki et al. [32]; Pfizenmaier et al. [33]; Matuszczyk et al. [34] and requires individual optimization for the given problem. Blocking intracellular transcription is achieved by sampling into RNA protect kits Löffler et al. [14]. Correctly prepared, samples can be treated further to identify metabolic compositions via metabolic profiling or fingerprinting techniques Fernie et al. [35]; Winder et al. [36]; Fiehn [37], protein contents via affinity tags Gygi et al. [38] or mass spectrometry Aebersold and Mann [39] and transcript levels, either applying microarrays or, more preferred, next generation sequencing technologies analyzing mRNAs Nagalakshmi et al. [40]; Nookaew et al. [41]; Wang et al. [42]. To reduce the overall sequencing expenses, library preparation usually is done via a rRNA depletion or poly-A enrichment step to remove non-coding rRNA.

Various methods for RNA Seq analysis are available and have been reviewed recently by Conesa et al. Conesa et al. [43]. Regarding modeling, time series of transcripts are particularly important which requires methods of differential gene expression analysis. Fig. 2 provides an overview of a typical workflow making use of public R packages.

Once time series of transcripts are available, modelers may be interested in unraveling gene clusters showing similar transcription dynamics and data integration in dynamic models. Applicants may be guided via evaluating reports of Rapaport et al. Rapaport et al. [49], Hecker et al. and Banf et al. Hecker et al. [50]; Banf and Rhee [51]. Currently, algorithms such as DeSeq2 Love et al. [52] and MaSigPro Conesa et al. [53] are often applied.

Please cite this article as: Zieringer J, Takors R, In Silico Prediction of Large-Scale Microbial Production Performance: Constraints for Getting Proper Data-Driven Models, Comput Struct Biotechnol J (2018), https://doi.org/10.1016/j.csbj.2018.06.002