Received: 17 July 2020 | Revised: 11 September 2020 | Accepted: 11 September 2020

DOI: 10.1002/bit.27568

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Data-driven in silico prediction of regulation heterogeneity and ATP demands of *Escherichia coli* in large-scale bioreactors

Julia Zieringer | **Moritz Wild** | **Ralf Takors**

Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany

Correspondence
Ralf Takors, Institute of Biochemical Engineering, University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany.
Email: takors@ibvt.uni-stuttgart.de

Abstract

Escherichia coli exposed to industrial-scale heterogeneous mixing conditions respond to external stress by initiating short-term metabolic and long-term strategic transcriptional programs. In native habitats, long-term strategies allow survival in severe stress but are of limited use in large bioreactors, where microenvironmental conditions may change right after said programs are started. Related on/off switching of genes causes additional ATP burden that may reduce the cellular capacity for producing the desired product. Here, we present an agent-based data-driven model linked to computational fluid dynamics, finally allowing to predict additional ATP needs of *Escherichia coli* K12 W3110 exposed to realistic large-scale bioreactor conditions. The complex model describes transcriptional up- and downregulation dynamics of about 600 genes starting from subminute range covering 28 h. The data-based approach was extracted from comprehensive scale-down experiments. Simulating mixing and mass transfer conditions in a 54 m^3 stirred bioreactor, 120,000 *E. coli* cells were tracked while fluctuating between different zones of glucose availability. It was found that cellular ATP demands rise between 30% and 45% of growth decoupled maintenance needs, which may limit the production of ATP-intensive product formation accordingly. Furthermore, spatial analysis of individual cell transcriptional patterns reveal very heterogeneous gene amplifications with hot spots of 50%–80% messenger RNA upregulation in the upper region of the bioreactor. The phenomenon reflects the time-delayed regulatory response of the

NOMENCLATURE: *a*, *b*, *c*, *d*, regression paramete; *C*, cluster; c_S, glucose concentration (mol L^{-1}); *D*, dilution rate (h^{-1}); *Deg*, degradation; *f*, fading genes; *gene*, gene; *i*, iteration index; k_{deg}, protein degradation rate (h^{-1}); K_S, substrate-specific uptake constant (g L^{-1}); $L_{mRNA,i}$, length of mRNA strand per gene *i* (nt); max_{TL}, maximum number of translations per mRNA; *Min size*, minimum number of observations per clusters; *MW*, molecular weight (g mol^{-1}); *Ntry*, number of trials per iteration; $n_{nucleotide}$, number of nucleotide monophosphates; *p*, persisting genes; *protein*, bulk protein; $q_{s,\ max}$, maximum substrate uptake rate $(g(gh)^{-1})$; *rmax*, maximum radius as a proxy for correlation; *S*, number of starvation passages; *Sc*, Schmidt number; S_{crit}, critical number of starvation passages; S_{min}, minimum number of starvation passages; *t*, time (general) (s); $t_{deg;\ med}$, median lifetime of mRNA (s); $TL_{i;\ j}$, initiation trigger for a translation j on mRNA *i*; v_{RIB}, velocity RIB (nts^{-1}); v_{RNAP}, velocity RNAP (nts^{-1}); x_{RIB}, location of RIB on mRNA (nt); x_{RNAP}, location of RNAP on DNA (nt); Δx_{RIB}, distance between ribosomes (nt); Δx_{RNAP}, distance between RNAPs (nt); Y_{XS}, substrate-biomass Yield $(g_S g^{-1}_{CDW})$; $\sigma_{RNAP}(t)$, number of active RNAPs; τ, residence time (s); μ, growth rate (h^{-1}).

cells that propagate through the stirred tank. After 4.2 h, cells adapt to environmental changes but still have to bear an additional 6% ATP demand.

KEYWORDS

agent-based modeling, ATP maintenance, computational fluid dynamics (CFD), *Escherichia coli*, glucose gradient, heterogeneities

1 | INTRODUCTION

To reduce the human CO_2 footprint in the atmosphere, sustainable bioprocesses replacing fossil resources by sugar may play a crucial role. Microbial production offers the potential to provide products for agricultural, biopharmaceutical, and chemical markets (Delvigne et al., 2017). As a prerequisite, such approaches need to be transferred successfully from laboratory to large-scale without loss of economic attraction, that is, without reduction of the sensitive TRY values (titer, rates, and yields) that served as constraints for economic evaluation. However, performance losses may occur, comprising increased by-product formation, reduced substrate-to-product conversion, reduced productivities, and so forth (Lara et al., 2006). They mirror cellular responses to large-scale heterogeneities that are induced by limited mass transfer and by insufficient mixing capacities (Noorman & Heijnen, 2017). Accordingly, research activities aimed to mimic large-scale conditions already in early-stage lab tests. One of the first examples is given by Oosterhuis and Kossen (1983), who repeatedly exposed cells to oxygen saturated and limiting conditions in a setting of two linked, stirred bioreactors. Multiple tests with alternate experimental scale-up simulators followed (overviews provided in Garcia-Ochoa & Gomez, 2009; Neubauer & Junne, 2010, 2016; Noorman, 2011; Takors, 2012; Zieringer & Takors, 2018) mimicking not only fluctuations of dissolved oxygen (DO) levels, but also nutrient availability and pH variations. Today, such approaches received key consideration to design robust microbial processes (Noorman & Heijnen, 2017). Still, the valid a priori prediction of large-scale heterogeneities' impact on cellular performance is of crucial importance for developing novel bioprocesses. Even further, findings of large-scale stress exposure may guide strain engineering to create particularly robust hosts. To reach this goal, Löffler et al. (2016) applied the so-called STR/PFR setup comprising a stirred tank reactor (STR) linked with a plug-flow reactor (PFR). Steady-state nutrient-limited, continuous cultivations were performed in STR before PFR was connected, frequently exposing cells to glucose-limiting conditions. Accordingly, cells repeatedly experienced temporal feast/famine conditions that were characterized by the residence time in the PFR. Comprehensive sampling in STR and PFR created a highly valuable data set of short- and long-term metabolic and transcriptional responses on repeated starvation stimuli (Löffler et al., 2016). The data set revealed that *Escherichia coli* not only react on extracellular stress by instantaneous metabolic shifts. Observations also revealed massive transcription of genes organized in operons (Nieß et al., 2017) and in fundamental regulons of strategic importance. For instance, the stringent response was repeatedly initiated by fast-rising intracellular (p)ppGpp levels in PFR, which were downregulated in STR. Löffler et al. (2016) reported an additional rise of growth-decoupled maintenance of up to 50%. So far, these findings were not yet used to predict the response of *E. coli* exposed to large-scale heterogeneities. First, a data-driven model is needed that describes the complex transcriptional response of *E. coli* to said stress conditions. Next, such a transcriptional model should be coupled with computational fluid dynamics (CFD) of a large-scale bioreactor to identify zones of different nutrient availability and to predict the cellular response of cells passing through those zones (Zieringer & Takors, 2018). Our study exactly tackles this two-step problem: Mixing heterogeneities and zones of different substrate availability of a 54 m^3 stirred bioreactor are predicted using CFD and assuming common operating conditions. The tracking of 120,000 *E. coli* cells finally yielded the prediction of additional ATP demands. Furthermore, spatially resolved transcriptional patterns of individual E. coli cells were predicted, unraveling the population heterogeneity in the industrial-scale bioreactor.

2 | MATERIALS AND METHODS

2.1 | Experimental setup

A glucose gradient was simulated in a stirred tank reactor (STR) coupled to a plug flow reactor (PFR), as depicted in Figure 1.

The experimental setup consists of an STR operated in continuous mode (dilution rate, $D = 0.2\ h^{-1}$) and connected with a plug flow reactor. Cells were grown under glucose-limited conditions in the STR (mean residence time of the cells in STR: 6.2 min) and experience starvation in the PFR (mean residence time of the cells in PFR: 125 s). The cells are circulating through the reactor system for 28 h process time, which equals, on average, around 200 passages of the starvation zone for each cell. In this way, the setup permits the analysis of transcriptional response for ongoing starvation passages through the PFR. Thereby, the tactical response is monitored via the PFR sample ports (P1–P5), while the strategic changes were tracked via the STR sample port (S). The cultivation was performed as biological triplicates under identical experimental conditions. For transcriptomic analysis, the samples were grouped by replicates, and sample time and location (STR or PFR) was chosen as a combined experimental design. Significantly expressed genes were determined using the described design and a generalized linear model within the

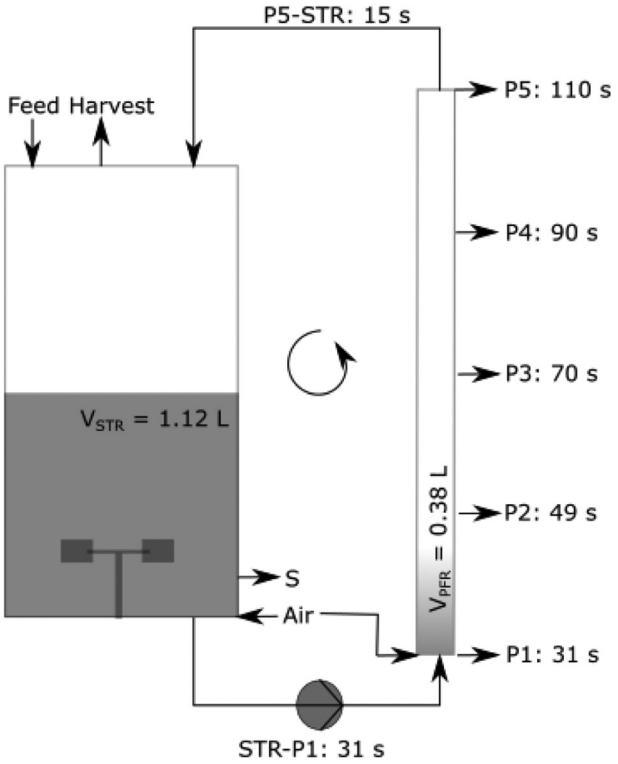

FIGURE 1 Scheme of a two-compartment system as used in the experimental setup (Löffler et al., 2016). The two-compartment device consists of a stirred tank reactor (STR) connected to a plug-flow reactor (PFR). Derived from non-ideally mixed large-scale industrial fermenters (Lapin et al., 2006), the setup mimics periodic substrate availability experienced by cells in large-scale bioreactors. The well-mixed STR is operated in glucose-limited continuous mode (dilution rate, $D = 0.2\ h^{-1}$). As soon as cells enter the PFR compartment, the residual substrate is consumed within seconds leading to starvation. The steady state before PFR onset at time zero was used as the reference state (S_0). Samples were taken at eleven distinct time points over 28 h. The system is equipped with five PFR sample ports (P1–P5) at defined residence times τ(s), as well as an STR sample port S. The total mean PFR residence time is $\tau_{PFR} = 125$ s

egdeR R-package (v.3.8.6) (Robinson et al., 2010). The detailed experimental implementation and RNA sequencing results used in this publication were published in the paper of Löffler et al. (2016).

2.2 | RNA-sequencing data cluster analysis

RNA-sequencing data contain time courses of messenger RNA (mRNA) abundance of 3908 genes. Thereof, three measurement sets sampled after 25 min, 2 h, and 28 h were chosen for further investigations. Significantly up- und downregulated genes of samples P1–P5 in PFR passing the threshold of *log*2 fold change (log2FC) > |0.58| and false discovery rate (FDR)-corrected *p* value < .05 were identified. Cluster analysis was performed using the *R*-package *flexclust* v. 1.4-0 (Leisch, 2006) applying *RStudio* v. 1.2.1335 (RStudio, Inc.) RNA-sequencing data contain time courses of mRNA abundance of 3908 genes. Thereof, three measurement sets sampled after 25 min, 2 h, and 28 h were chosen for further investigations. Significantly up- und downregulated genes of samples P1–P5 in PFR passing the threshold of *log*2 fold change (log2FC) > |0.58| and FDR-corrected *p* value < .05 were identified. Cluster analysis was performed using the *R*-package *flexclust* v. 1.4-0 (Leisch, 2006), applying *RStudio* v. 1.2.1335 to significantly reduce simulation efforts while including basic features of gene dynamics. The function *qtclust* (included in flexclust package) was used to perform stochastic quality-based clustering (SQBC) and *k*-means-clustering. Parameters of SQBC were set as follows (Table 1).

Ntry indicates the number of trials per iteration, while *rmax* is the maximum radius as a proxy for correlation. *Min size* defines the minimum number of observations per cluster. Data points not clustered by the algorithm are omitted. The setting of parameters ensured maximum comparability and five as the maximum number of clusters. Cluster properties are listed in the Supporting Information Appendix Tables D5–D8, and resulting clusters are displayed in Supporting Information C in the Appendix. The *k*-means algorithm was initialized with the centroids of the SQCB method.

2.3 | ATP calculation for single molecules

ATP requirements for the formation of amino acids and nucleotides were calculated using the results of Kaleta et al. (2013). The translational costs for protein formation and polymerization add up to 4 ATP per amino acid, including activation of the amino acid (1 ATP to 1 AMP) and peptide bond formation at the ribosome (2 GTP) (Stouthamer, 1973). Since there is a net production of 0.1 ATP per amino acid (for detailed calculation, see Löffler et al., 2016), the overall cost of amino acid synthesis and polymerization was estimated as four ATPs consumed per residue. The absolute numbers of synthesized and degraded nucleotides (*nts*) were estimated from experimental data. To recycle mono phosphorylated nucleotides (NMPs) to triphosphorylated nucleotides (NTPs), costs of 2 ATP were assumed. Assuming a P/O-ratio of 1.49 (ATP formation via NADH oxidation in respiration), the following ATP requirements were assumed for the bases (Table 2).

The growth-independent maintenance was used as 0.0027 mol $(g_{DW}h)^{-1}$, according to Taymaz-Nikerel et al. (2010).

2.3.1 | Calculation of mRNA abundance

Only additional ATP needs for transcription and translation were estimated considering the basic demands under nonperturbed conditions. Accordingly, total mRNA content was estimated following studies of Bremer and Dennis (2008) as 61.7 µg per 10^9 cells for a growth rate of $0.2\ h^{-1}$. 20% of the total dry weight was assumed to be RNA (Neidhardt et al., 1990), including 5% mRNA, a value which is

in line with conclusions from Stouthamer (1973). This results in total mRNA content of 6.17 g per 10^{16} mRNA per cell. The relative distribution of specific mRNAs is taken from the measured normalized counts as transcripts per million. The relative fraction multiplied with the total mRNA content gives the total mass of all mRNA encoded by a single gene. Dividing this number with the corresponding molecular weight yields the absolute number of molecules. Molecular weights of mRNAs were calculated with Equation 1 (Kibbe, 2007). Results are listed in Supporting Information Table A3. As the phosphate groups of two nucleotides are bound together, an OH-group is cleaved. This leads to reduced molecular weights of nucleotides in the polymer chain. Accordingly, an additional term was added to account for the 5'-triphosphate cleavage (159 g mol^{-1}).

$$MW_{mRNA} = n_{Guanine} \times 345.2 + n_{Cytosine} \times 305.2 + n_{Adenine} \times 329.2 + n_{Uracil} \times 306.2 + 159 \ (gmol^{-1}). \quad (1)$$

$n_{nucleotide}$ codes for the number of nucleotide monophosphates which is multiplied by their corresponding molecular weight.

2.4 | The biological model

The model was implemented using MATLAB v. 2019b, considering the four processes: transcription, translation, mRNA degradation, and protein degradation. The first three are implemented as agent-based approaches, while the last considers protein degradation as a decomposition in the continuum. The governing variable that controls the expression is the number of active RNA polymerases (RNAPs) per each cluster.

2.4.1 | Estimating the number of active RNAP

We assumed that gene expression levels follow sigmoidal courses. Hence, an equilibrium between synthesis and degradation may be achieved. Produced *nts* are given by

$$v_{RNAP} \cdot \int_0^t \sigma_{RNAP}(t)dt = Nt_{prod} \quad (2)$$

with $\sigma_{RNAP}(t)$ coding for the number of active RNAPs at time t. The shape of the sigmoidal function is defined as

$$\sigma(t) = \frac{a}{1 + e^{\{b \cdot (t+c)\}}} + d. \quad (3)$$

The parameters a, b, c, and d were fitted to the nucleotide synthesis, which was derived from experimental data by calculating the number of synthesized copies and by considering individual gene lengths. Consequently, steadily rising functions were obtained that allowed to estimate the number of active RNAPs per cluster. For the latter, a constant RNAP transcription velocity of 21 nucleotides per second was assumed (Chen et al., 2015).

2.4.2 | Transcription

After initiation, the continuous one-stranded movement of RNAP creates the mRNA transcripts measured. However, individual gene expression profiles were observed that could be grouped in clusters of similar transcription dynamics. Accordingly, only expression dynamics of representative, average genes per cluster are described in the model (Supporting Information Appendix Tables D5–D7). The minimum distance of 100 nts (Δx_{RNAP}) was considered between two subsequent RNAPs. Furthermore, all genes of one cluster were supposed to be randomly initiated with the functional

$$gene_{i,on|off}(t) = \begin{cases} 1 & \text{if free RNAP is avaliable and gene}_i \\ & \text{is randomly chosen,} \\ 0 & \text{else.} \end{cases} \quad (4)$$

Gene transcription is modeled as one-dimensional nucleotide extension with the relative movement of RNAP as

$$\frac{dx_{RNAP}}{dt} = \begin{cases} v_{RNAP} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \quad (5)$$

The constant transcriptional elongation rate, v_{RNAP}, is set to 21 nts^{-1}, which equals the average value found in *E. coli* during starvation (Chen et al., 2015). The variable x_{RNAP} indicates the relative position of RNAP on the DNA grid. The length of the resulting mRNA strand is equivalent as

$$x_{i,RNAP} = L_{i,mRNA}. \quad (6)$$

When the last nucleotide is reached, the mRNA is released. All fragments of mRNA are summed to get total mRNA amounts. While fractions of operons were found to be fully transcribed after initiation (Nieß et al., 2017), other scenarios coincided, too. For instance, only subsets of operons may be transcribed, or even opposing transcription reads in a single operon occurred (Mao et al., 2015). Accordingly, we assumed that only 10% of the initiated operons are transcribed completely. In other words, 10% of experimentally observed initiated operons were anticipated to finish full operon transcription even outside of PFR. The majority (90%) of other gene transcriptions was assumed to stop immediately after RNAPs have reached individual gene ends. For the sake of comparability, only relative mRNA enrichments are depicted in Figure 7c, referring to the mRNA level of individual cells after they have fluctuated through the bioreactor for 180 s. This time point was chosen to visualize the spatial distribution of already adapted cells and the ones which are still influenced by regime changes (Figure 7).

2.4.3 | Translation

The translational process is modeled by describing the movement of ribosomes (RIB) on the mRNA strand. The process is assumed to take place in cotranscriptional manner (Proshkin et al., 2010): After

synthesized mRNA reached a minimum length of 80 nts (Δx_{RIB}) (Bremer & Dennis, 2008), the first ribosome attaches to the free 5'-cap end. Further ribosomes may bind too, provided that minimum distance between two subsequent ribosomes and the maximum number of translations per mRNA (max_{TL}) are fulfilled. Released ribosomes may be reused following the same scenario. The initiation trigger $TL_{i,j}$ for a translation j on mRNA i can be described as

$$TL_{i,j,on|off}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,j-1} \text{is } >80 \text{ and } j < max_{TL}, \\ 0 & \text{else.} \end{cases} \tag{7}$$

The movement of ribosomes is analogous to the movement of RNAP as

$$\frac{dx_{RIB}}{dt} = \begin{cases} v_{RIB} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \tag{8}$$

With the translational elongation rate $v_{RIB} = v_{RNAP}$ (Nieß et al., 2017; Proshkin et al., 2010) and the number of maximum translations $max_{TL} = 11$ (Bremer & Dennis, 2008). Because one cluster of genes revealed rapid degradation after 50 s in PFR at 28 h, only one translation per transcript was assumed for this group of genes. As soon as the ribosomes pass the last nucleotide, the protein is released and is assigned to the group of bulk proteins.

2.4.4 | Degradation

Degradation of transcripts is initiated as soon as ribosomal protection of the 5'-cap vanishes, which is a co-translational process. The velocity of the RNASE was adapted to the gene length $L_{mRNA,i}$, that is, individual $v_{RNASE,i}$ was estimated considering the experimentally observed mRNA median lifetime $t_{deg;med}$ of 2.8 min in nutrient-rich and of 4.6 min in starvation zones (Chen et al., 2015). Degradation was initiated for mRNA i as

$$Deg_{i,j,on|off}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,TL_{max}} \text{is } > 80, \\ 0 & \text{else.} \end{cases} \tag{9}$$

with the movement of RNASE.

$$\frac{dx_{RNASE}}{dt} = \begin{cases} v_{RNASE} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \tag{10}$$

$$v_{RNASE,i} = \frac{L_{mRNA,i}}{t_{deg,med} - \frac{TL_{max} * \Delta x_{RIB}}{v_{RIB}}} \tag{11}$$

For transcripts longer than ≈ 1000 nts, degradation is initiated already when transcription is not finished yet. Chen et al. (2015) found that 88 of 263 mRNAs showed lifetimes of the 5'-cap shorter than the synthesis time of the transcript. Accordingly, co-transcriptional degradation was considered for long transcripts.

Protein degradation is described using a constant rate degradation rate k_{deg} for the bulk proteins (Maurizi, 1992). First-order degradation kinetics were assumed depending on the nutrient condition, as

$$k_{deg} = \begin{cases} 0.01 h^{-1} \text{innutrientrichzones,} \\ 0.08 h^{-1} \text{instarvationzones.} \end{cases} \tag{12}$$

Consequently, bulk protein of a subsequent time step $t + 1$ equals

$$protein(t + 1) = protein(t) \times (1 - k_{deg}). \tag{13}$$

2.5 | Geometry and reactor setup

To consider a relevant industrial fed batch fermentation scenario, a 54-m^3 stirred tank bioreactor was chosen. The main geometry was derived from Haringa et al. (2016) with precise dimensions and information about the inner geometry from Kuschel et al. (2017). The reactor setup included four baffles and a stirrer with two Rushton stirrers equipped with eight blades at the bottom and six blades at the top. With a stirring rate of 100 rpm the tip speed of 6.75 m s^{-1} was reached. The impeller Reynolds number was 2.77×10^6 and the required power was 225.69 kW, equaling a power number of 13.64. The feeding rate was set to 3.68 kg m^{-3} s^{-1} for an average growth rate of 0.2 h^{-1}. Aeration, gas transfer, and oxygen uptake were neglected in the study. The simplifying focus on the mono-phase conditions mirrors the basic strategy to showcase the propagation of transcript dynamics and the occurrence of additional large-scale ATP demands. Noteworthy, experimental data of Löffler et al. (2016) were measured particularly excluding any impacts of oxygen limitation. Furthermore, said power inputs and cultivation conditions were chosen such that oxygen limitation is unlikely in the large-scale scenario. Cell concentration of 31.8 $kg_{CDW} \cdot m^{-3}$ was assumed and a simple Monod-like kinetic was used to simulate the substrate uptake q_s:

$$q_S = q_{S,max} \cdot \frac{c_S}{K_S + c_S}, \tag{14}$$

where $q_{s,max}$ is the maximum uptake rate, c_s is the glucose concentration, and the approximated substrate-specific uptake constant K_s with 4 mg L^{-1}. The maximum uptake rate was calculated with the biomass substrate yield $Y_{XS} = 0.25\ g_S \cdot g^{-1}_{CDW}$ and the maximum growth rate $\mu = 0.2\ h^{-1}$ (Villadsen et al., 2011). Based on the experimental observations in Löffler et al. (2016) we concluded that stringent response is the predominant regulatory scheme initiated by repeated starvation. As a key characteristic stringent response reduces ATP consuming procedures trying to keep carbon supply on the maximum level achievable under stress conditions. Accordingly, we consider glucose uptake as a Monod-type function not being affected by the stringent response observed.

2.6 | Simulation setup

For the numerical simulation, the commercial calculation tool ANSYS Fluent version 19.1 was used. With 872,232 hexahedral numerical

cells resulting in an aspect ratio of 12.6 and a minimal orthogonal quality of 0.34, high mesh quality was achieved. Schmidt (*Sc*) number tuning with *Sc* = 0.2 leads to the same circulation time as achieved by Haringa et al. (2016). The flow field was approximated by solving the Reynolds-averaged Navier-Stokes (RANS) equations in combination with the realizable k-ε model for turbulence. All surfaces were set as no-slip boundaries except for the no-shear top area, which equaled the reactor filling height. Baffles and impellers were modeled as 0-thickness walls. Both impeller units were set to sliding mesh motion to generate a more realistic flow field. For glucose feed, a separate volume at the top of the reactor was defined, and a constant mass flow was set. The flow field and uptake kinetics were calculated every 10 ms until the glucose concentration was constant and a pseudo stationary gradient was reached, showing constant metabolic activity. The conditions were "frozen" for 180 s to track bacterial movements. These *lifelines* were simulated as massless Lagrangian particles with a discrete random walk (DRW) model passing through the flow field. Every 30 ms, the position and glucose concentration of each bacterium were recorded. In total, 120,000 bacterial cells were tracked over around 180 s (residence time distribution: Supporting Information Appendix Figure E14). According to the ergodic theorem, the same average values are obtained by tracking 1,080,000 bacteria for 20 s (for more information, see Appendix: Supporting information F). However, due to the limitation of simulation time and capacity, the simulation results were extended by repeating the single lifelines (Figure 2b) every 180 s while preserving the lifeline cluster groups. RNAP activities, mRNA levels, ribosomal activities, and protein formation were calculated as described in Section 2.4 considering each lifeline cluster (RNAP and mRNA profile of one lifeline group: Figure 2c; mRNA content of individual cells after 180 s: Figure 7c). Finally, additional ATP demands were estimated (Figure 2d).

3 | RESULTS

3.1 | Simulation results biological model

To identify data-driven parameters for the model of *E. coli* K12 W3110, clusters of mean mRNA levels were identified (Supporting information Appendix Section C). Figures 3–5 show the simulations (blue lines) and mean experimental values (red dots) for mRNA levels, active RNAPs, and the number of translated proteins per cell during PFR passage. The

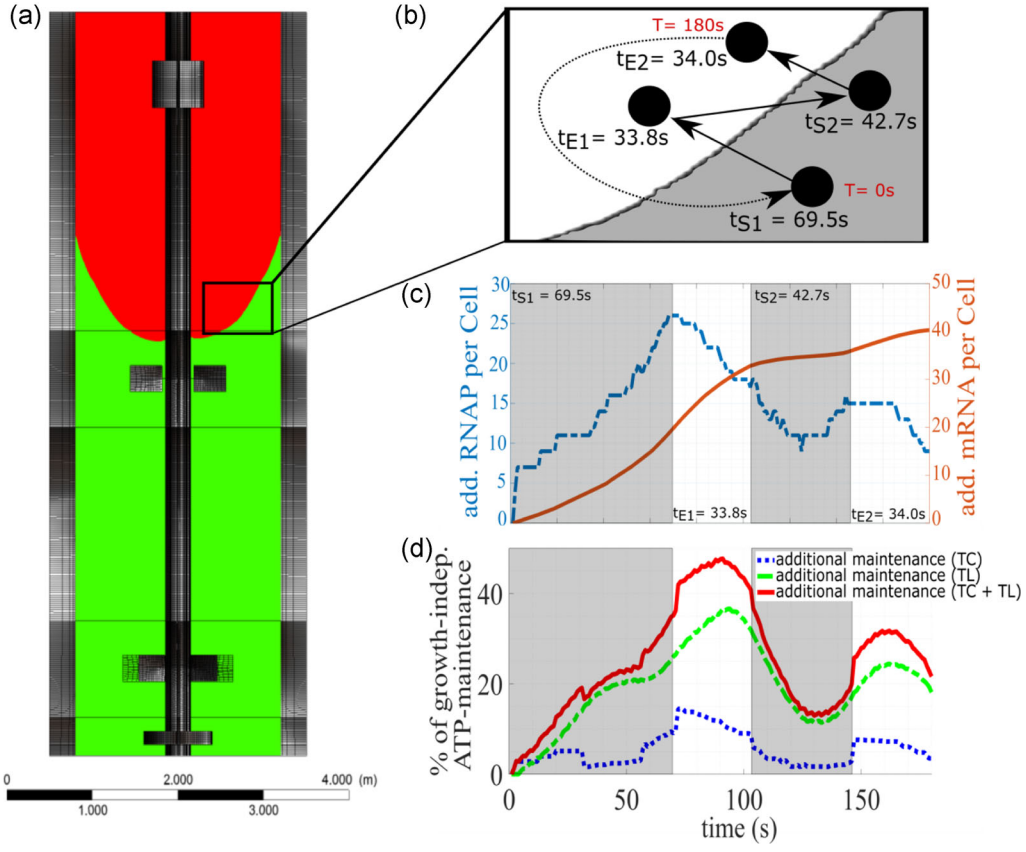

FIGURE 2 Impact of frequent exposure to feast and famine conditions in a large-scale bioreactor (a). White areas reflect nutrient excess, while gray areas indicate starvation. The size of the areas reflects the corresponding residence time indicated with t_S for starvation and t_E for excess residence time (c: Bar plot for one cluster of particle trajectories). The starvation-induced regulatory responses are propagated into the glucose excess zone, causing a maximum growth-independent ATP-maintenance in the glucose excess regime (d) based on additional active RNA polymerase(RNAP) for transcription (TC) (c) and ribosomes for translation (TL) [Color figure can be viewed at wileyonlinelibrary.com]

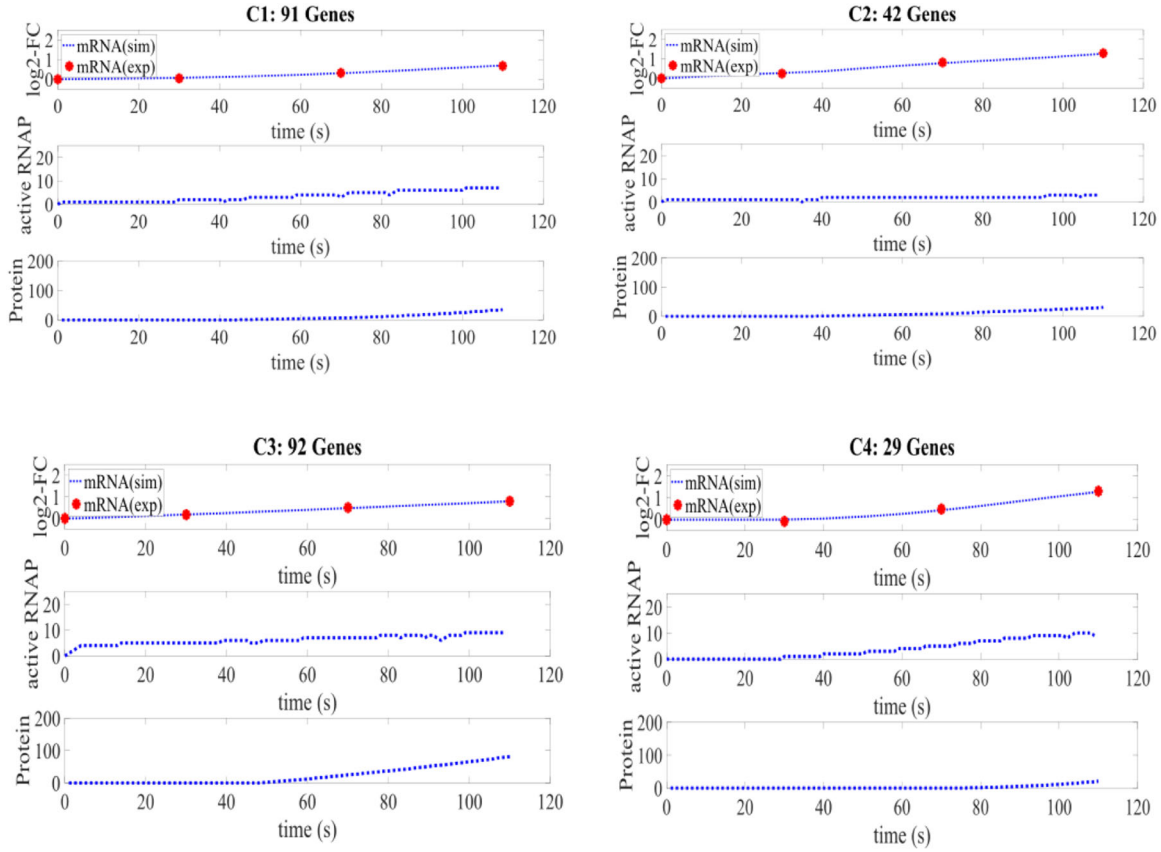

FIGURE 3 Simulated, cell-specific number of *additional* messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–4 (C1–C4) along starvation passage (t = 0–110 s, Figure 1) at 25 min. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

synthesis rate accelerates over time as more RNAP molecules are involved in the transcriptional response to the starvation stimulus. After a transcript is completely synthesized and the RNAP is released the number of active RNAP shortly drops before it rises again. Synthesized proteins appear with a delay that corresponds to the required translation time. Activities for transcription and translation result in additional ATP demands which are indicated as add-ons to nongrowth-dependent maintenance (NAM), shown in Supporting information Appendix Figure B9. For transcription, costs are derived from nucleotide balancing, including the release of nucleotides by mRNA degradation and the need for mRNA synthesis. Cost for translation mirrors the amino acid needs and integration according to ribosomal activity. As indicated, translation costs are more than 2.5-fold higher than those of transcription. At maximum, cells have to bear 36.8% additional NAM, 10.4% coding for transcription, 26.4% for translation. This happens during the early phase of frequent starvation exposure, that is, after three starvation passages (25 min process time). After 2 h process time, the ATP demand still increases. More than 45% NAM increase is observed, illustrating the remaining high number of active RNAP. Later, after 28 h, NAM add-ons reduce more than fivefold compared with maximum needs. Then, transcription accounts for about 1% NAM rise only. The total NAM increase only mirrors 9.5%.

3.2 | Linking cluster kinetics

Cellular adaptations to frequent environmental stimuli are mirrored in the cluster dynamics of differentially expressed genes (DEGs) that were measured after 25 min, 2 h, and 28 h. Only 81 of 521 DEGs are conserved over the entire process time. This reflects the replacement of the initial sigma factor 70 dominated response by σ_{38} mediated regulatory programs (Löffler et al., 2016). To simulate the transition, so-called "damping" and "amplification" factors were identified using mean gene expressions as a reference based on the simulated log2FC mRNA dynamics in Figures 3–5. Clusters were subdivided in "persisting," "subsiding," and "non-active" fractions. The first collected genes with continuing high expression levels, while the second comprised genes with declining expression levels. The last summed those genes that were either not yet or no more expressed between subsequent time points (Figure 6). The damping factor is the ratio of the mean $\log_2$FC of subsiding and persisting genes for each cluster between two time points. The amplification factor is the ratio of the mean $\log_2$FC of genes at 25 min, which are active at time point 2 h divided by the mean $\log_2$FC ratio of genes activated after 2 h (see exemplary calculation Figure 6). The factors are used to calculate the amount of active RNAP:

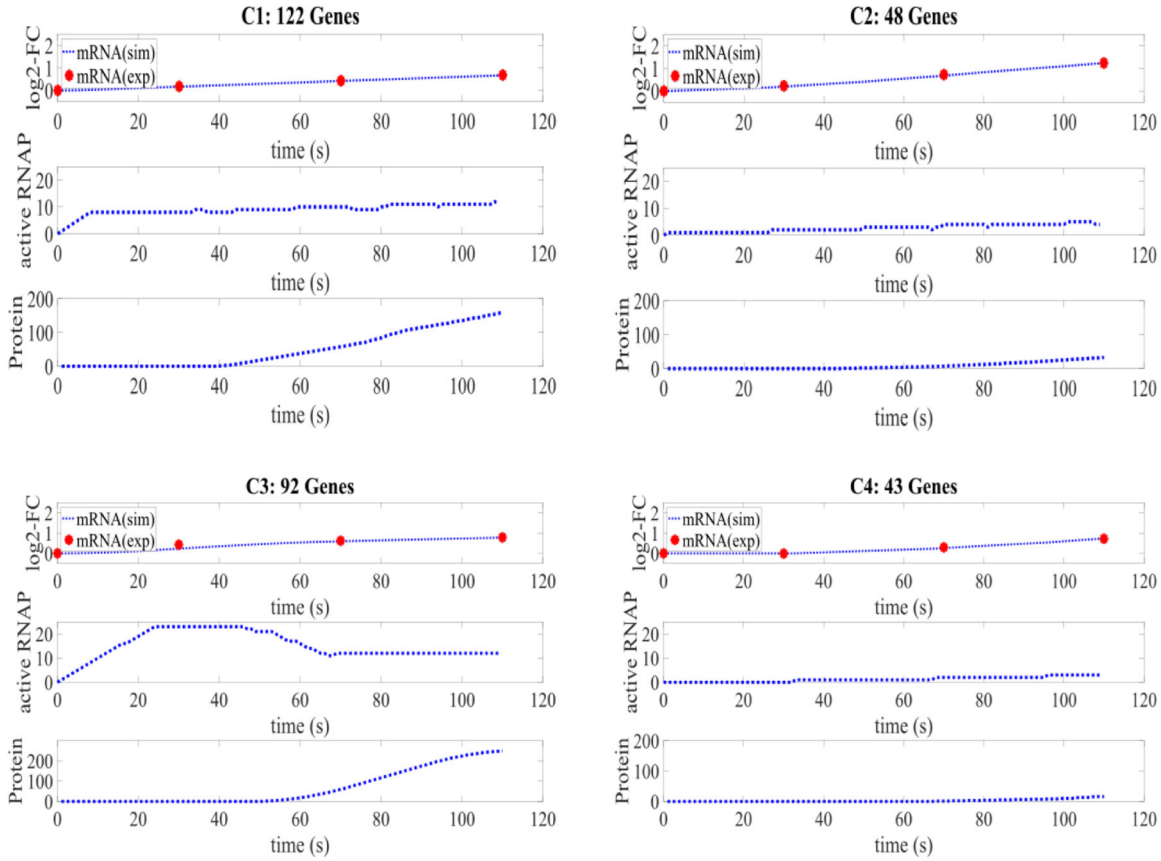

FIGURE 4 Simulated, cell-specific number of *additional* messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–4 (C1–C4) along starvation passage (t = 0–110 s, Figure 1) at 2 h. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

$$x_{RNAP}(s,t) = x_{RNAP}(t)*\left(1 - \frac{Damping\ factor}{(S_{crit} - S_{min})}*(S(t) - S_{min})\right) \quad (15)$$

$$x_{RNAP}(s,t) = x_{RNAP}(t)*\left(\frac{Amplification\ factor}{(S_{crit} - S_{min})}*(S(t) - S_{min})\right) \quad (16)$$

The entire transition process is guided by the number of starvation passages $S(t)$ per time. S_{min} encodes the minimum and S_{crit} the critical number of passages. Whereas the first is a regression parameter, the later reflects experimental observations of Löffler et al. (2016) as follows: 25 min equal 3 PFR (starvation) passages, 2 h equal 14, and 28 h equal 176. Noteworthy, the 28 h benchmark is chosen as a new steady state was observed already then (Löffler et al., 2016). The modeling approach allows transferring of the STR/PFR observations to other conditions using the frequency of feast/famine exposure $S(t)$ as a key criterion.

3.3 | Numerical simulation

3.3.1 | Glucose gradient

Applying the criterion of converged turbulent dissipation rate/power input, the pseudo-stationary glucose gradient of Figure 7 was obtained (Figure 7). Accordingly, no further changes in glucose concentrations simulated at five locations occurred. The average concentration in the bioreactor was 23.74 mg L^{-1}. For comparison, the average glucose level observed by the Lagrangian particles ("cells") was 22.79 mg L^{-1}. Consequently, only 4% deviation was found, which is qualified as a small difference indicating good homogeneous distribution and reflecting impacts of the turbulence model and of particle lifeline filtering. The volumetric distribution between starvation and excess zone is 73%–27%, respectively. Again, similar percentages were calculated by integrating mean residence times of all lifelines. The mean residence time of the cells in the starvation regime is 9.46 s (Supporting Information Appendix Figure E14), which is in the same range as published by Haringa et al. (2016).

Ideally, large-scale simulations should have been compared with real in situ measurements to challenge the predictions. However, such data are missing, which represents a common problem often faced by academic groups. Nevertheless, applying CFD simulations still offers the best chances for getting highly accurate large-scale predictions as complex hydrodynamics, even including overlapping flow fields between stirrers, are well predictable. Notably, the latter may hamper the application of simplifying compartment-based estimations, which basically assume separated flow fields between stirrers (see Supporting Information Appendix H).

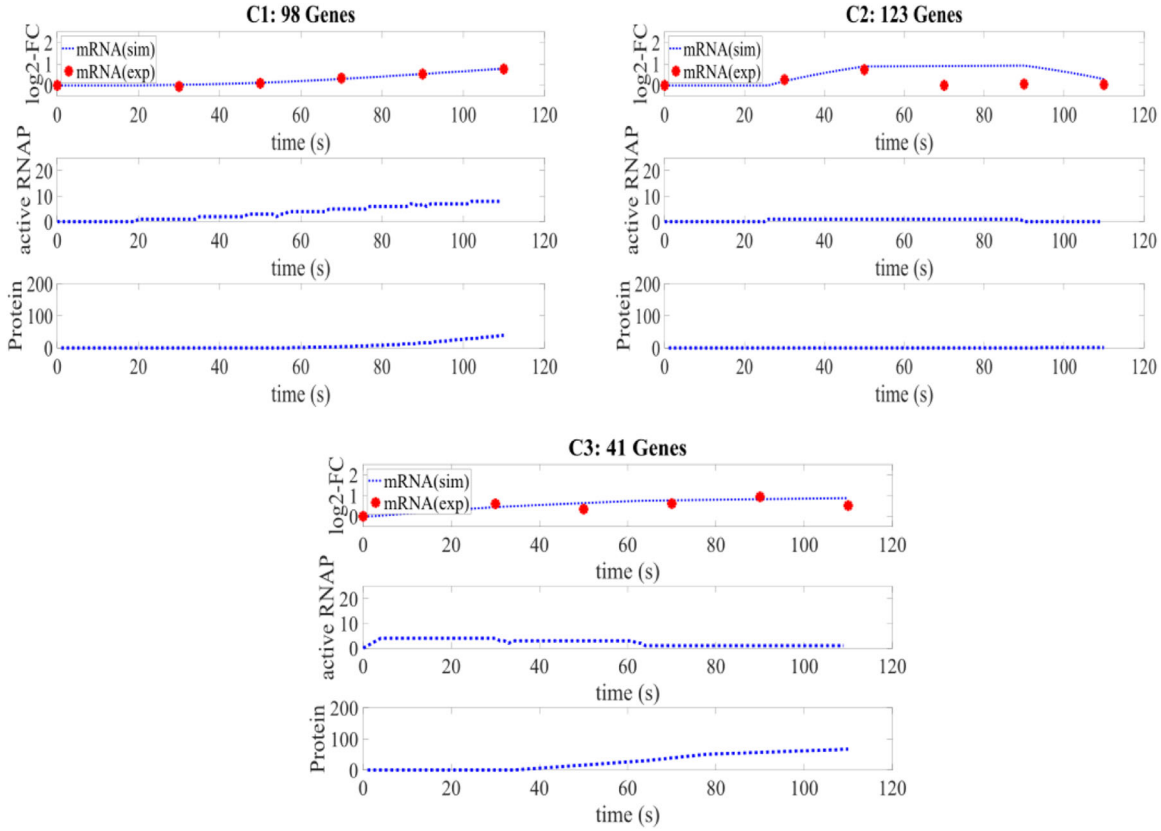

FIGURE 5 Simulated, cell-specific number of additional messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–3 (C1–C3) along starvation passage (t = 0–110 s, Figure 1) at 28 h. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

For simplicity, Eulerian simulations only considered the liquid phase, thereby assuming sufficient oxygen supply in the bioreactor without calculating DO concentrations explicitly. Furthermore, additional turbulence due to bubble interaction was neglected, too. Substrate consumption followed Monod-type kinetics taking place in each numerical cell. This implied that bacterial cells were homogeneously distributed at each time step.

3.3.2 | Statistical evaluation

Lifeline statistics reflect the imprint of changing micro-environmental conditions on cells fluctuating through the bioreactor. To be precise, cellular residence times in different concentration zones and shifts between proximal regimes were studied. At the start, cells were "inserted" into the bioreactor along a straight line reaching from top to bottom. After a few simulation steps, cells were distributed homogeneously before individual cell tracking started for 180 s. Lifeline records were cured by percolating only those with residence times longer than 0.13 s. The latter represent unrealistically turbulent fluctuations. The following threshold was defined for regime analysis: If cells experience lower or higher glucose levels than K_S for at least one second, the period is labeled as starvation or saturation time, respectively. Noteworthy, the minimum residence time of 1 s equals the average metabolite turnover time in *E. coli* (Shamir et al., 2016; Taymaz-Nikerel et al.,2011). The implementation of the harsh regime boundary K_S finally leads to rapid and somewhat artificial regime shifts. They were excluded from analysis by ignoring the upper and lower 1% of regime changes. Alternately, the consideration of alarmones such as (p)ppGpp serving as intracellular triggers to initiate transcriptional regulation may yield continuous models. Unfortunately, understanding of alarmone formulation, degradation, and alarmone induced regulation is still too fragmented to build dynamic transcriptional models. In total, measures for residence time percolation and shift filtering only reduced the data set by 3% (Supporting Information Appendix Figure E14).

At maximum, 41 regime shifts were observed during the 180 seconds observation period. Most frequently, 20 regime changes occurred and cells rested in a single zone no longer than 30 seconds. As a key characteristic, cells once exposed to glucose starvation reset their regulation signal. But RNAP and ribosomes remain active, propagating the starvation response into the glucose excess regime (Figure 2c,d). According to their starvation pattern the cell lifelines were assigned to 70 different clusters. Thereby each cluster represents a specific fluctuation pattern, reflecting the amount and duration of changes between starvation and excess zone (Figure 2b).

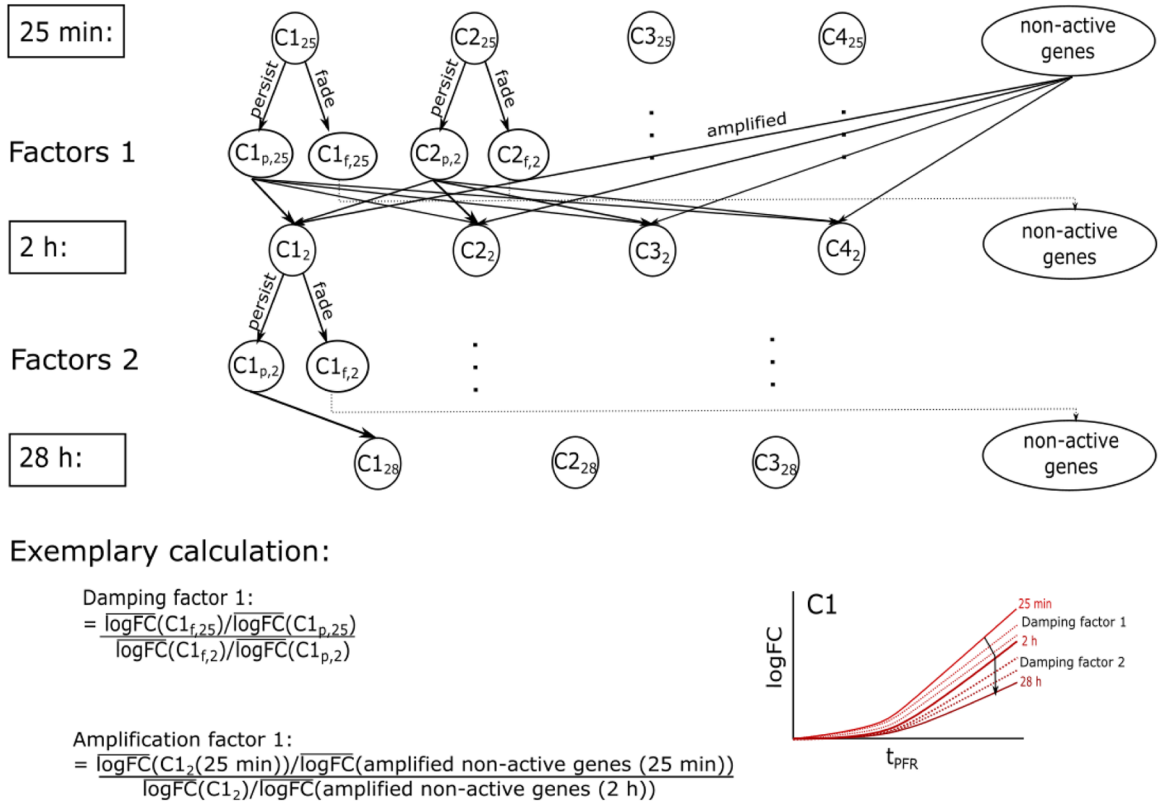

FIGURE 6 Scheme illustrating how "damping" and "amplification" factors are derived from experimental cluster data (C1–C4) measured at 25 min, 2 h, and 28 h after plug-flow reactor passage. In general, active genes of individual clusters may continue amplification (persisting, *p*), reduce expression (fading, *f*), or even be activated from the group of non-active genes. To bridge the gene expression dynamics from 25 min to 2 h and from 2 to 28 h, so-called damping and amplification factors are calculated as illustrated in the example. They use mean logFC values of the relevant time points. The damping factor correlates the ratios of fading-to-persisting genes of the two subsequent measurements. By analogy, the amplification factor calculates ratios of mean numbers of gene expression versus "first time" amplified genes for each time point and correlates the same for two subsequent events [Color figure can be viewed at wileyonlinelibrary.com]

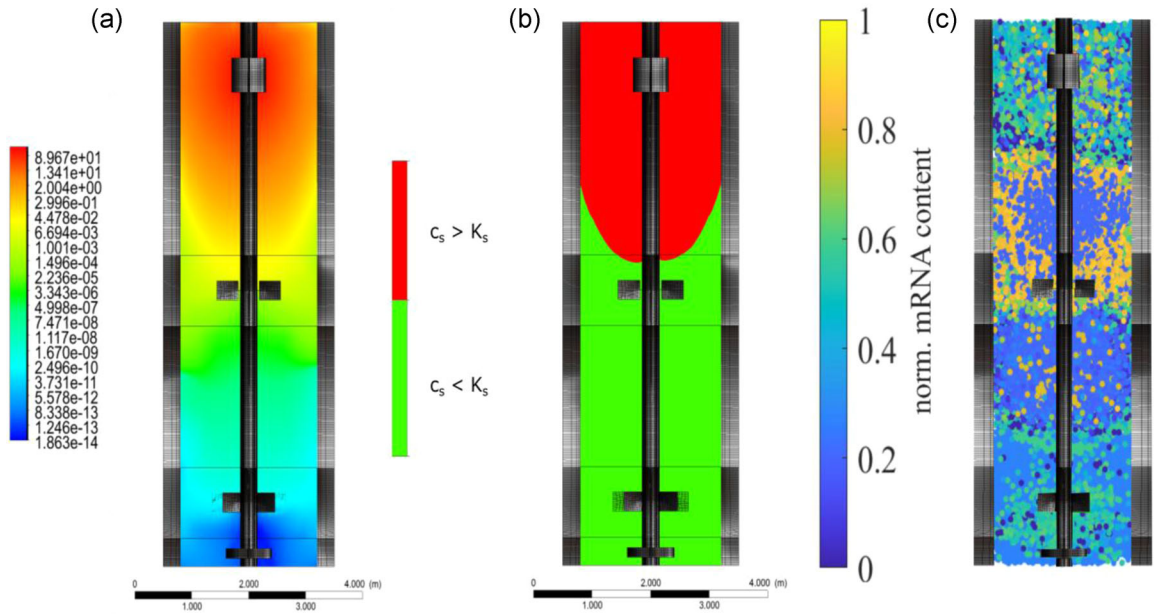

FIGURE 7 Simulation results of the 54 m^3 stirred tank reactor. (a, left) Log-contours of c_s/K_s gradients. (b, middle) Assignment to regimes (limitation, green: $c_s < K_s$; saturation, red: $c_s > K_s$). (c, right) Normalized messenger RNA (mRNA) content analyzed after cells fluctuated 180 s through the bioreactor [Color figure can be viewed at wileyonlinelibrary.com]

3.3.3 | Coupling the biological model with lifelines

To minimize the computational efforts, particle lifelines were exported from ANSYS Fluent. Starvation patterns of lifelines (Figure 2b) served as input for the biological model. A workflow scheme is provided in the Supporting Information Appendix Figure H16. It was assumed that each entry of the starvation zone activated the expression of distinct gene clusters as experimentally observed. Accordingly, individual expression patterns were estimated for each cell, mirroring their particular tracking history.

As indicated in Figure 7c, the basic expression level of the cellular population is increased by 37.7% compared with the reference. This reflects the additional cellular needs to adapt to the heterogeneous mixing conditions in the reactor. Noteworthy, high mRNA levels, induced by preceding starvation, are propagated into glucose-rich zones (Figure 7c). Expression patterns reflecting starving conditions occur in saturating glucose zones and vice versa. The phenomenon mirrors the delayed transcriptional response that is slower than convective movements of cells in the bioreactor. Consequently, a high transcriptional heterogeneity occurred in the tank. A maybe surprising pattern is disclosed: In the lower part of the reactor, cells envisage low glucose concentrations but show reduced mRNA levels (norm. mean mRNA level: 0.33). On the contrary, cells facing high glucose levels in the upper part reveal high mRNA levels (norm. mean mRNA level: 0.42).

About 25% of the cells permanently stay in the starvation zone. This fraction even adapts to the limiting conditions, which reduce the transcriptional response gradually. Cells located close to zones of glucose excess highly fluctuate between starving and saturating conditions. Consequently, strong gene expression responses are observed.

The average ATP-demand of a newborn, not preconditioned population of 120,000 cells exposed to the 54 m^3 bioreactor is shown in Figure 8. Basically, plot 8 illustrates the cyclic passing of 180 s lasting lifelines. To filter related peaks, only average values are indicated using a moving median filter over 700 data points. Furthermore, a moving standard deviation with a window size of 400 data points is added as a shadow. Synchronization-like patterns reflect the clustering of particles in groups. 70 bins were used with passable computational effort. The maximum of 45% additional maintenance is predicted shortly (0.03 h) after cells were exposed to the bioreactor condition. After about 0.5 h most of the population has adapted to the heterogeneous environment reducing the additional ATP demand to 6.5%. After 4.22 h the last cell fraction is in the adaption state.

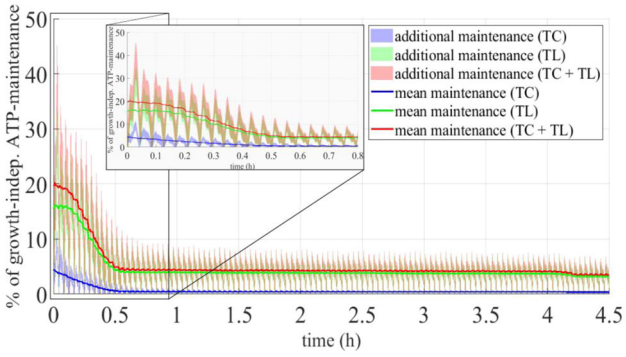

FIGURE 8 Additional ATP demands of a population of 120,000 "newborn," not preconditioned cells in a 54 m^3 reactor over 4.5 h process time. Courses for mean transcription (blue: TC), translation, (green: TL) and the sum of both are depicted (red: TC + TL). The shaded areas display the standard deviation [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Parameters and omitted genes of the SQBC algorithm

Time point	25 min	2 h	28 h
ntry	1000	1000	1000
rmax	0.5	0.5	0.9
Min size	3	3	5
Genes omitted	0	1	6

Abbreviations: *Ntry*, number of trials per iteration; *rmax*, maximum radius as a proxy for correlation; *Min size*, minimum number of observations per clusters.

4 | DISCUSSION

To disclose spatial heterogeneities of regulation patterns and additional ATP demands of *E. coli* K12 W3110 exposed to a 54 m^3 STR CFD based lifeline analysis was coupled with agent-based modeling for transcription and translation.

As a prerequisite, a proper model describing stress-induced dynamics of transcription and translation is needed. Applying a clustering approach, it was possible to properly describe the experimentally observed transcription dynamics (Löffler et al., 2016) of 821 genes using 16 parameters. Under steady-state conditions, newly synthesized and recycled *nts* equilibrate in cells before they enter PFR. However, ATP demands transcription rise inside PFR as mRNA synthesis exceeds the recycling rate. The introduction of 'amplification' and 'damping' factors managed to model the transition from fast response to long-term adaptation, as visible in the experimental data. Accordingly, modeling succeeded to mirror the cellular

TABLE 2 ATP costs for de novo nucleotide synthesis, based on Löffler et al. (2016)

Base	ATP	NADH	NADPH	Overall ATP
Guanin	11	−3	1	8.53
Cytosin	13	−3	0	9
Adenin	9	0	1	6.53
Uracil	7	0	1	7
Average	10	−1.5	0.75	7.7

efforts shifting control from σ_{70} to σ_{38}, more and more (Löffler et al., 2016). Given that *E. coli* cells with doubling times of 3.3 h contain about 120 active RNAPs (Bremer & Dennis, 2008), approximately one-fifth of the available RNAPs at 25 min is used in the transcriptional response (Figure 3).

The number of involved RNAPs slowed down after 28 h for two reasons. First, the absolute number of initiated transcripts is reduced, which mirrors cellular adaptation. Second, transcription even stopped after 50 s for a large group of genes (Figure 5). Accordingly, reduced synthesis costs occur. This is amplified by the prolonged lifetime of transcripts during starvation, which further reduces the amount of synthesis to obtain a certain level of mRNA abundance as described in Section 2.4.4 mRNA lifetime is proportional to their distance from the 5′ end of the transcript according to Chen et al. (2015), which is in line with the observation that the 5′ end contains important determinants that regulate RNA lifetime (Arnold et al., 1998).

Supporting Information Appendix Figure B9 shows that protein synthesis accounts for the major part of ATP consumption. Peak values reach about 45% NAM at 2 h process time. Later on, the demand steadily reduces as less mRNA transcripts are synthesized and less proteins are translated. At 25 min, the number of active ribosomes involved in the stress response rises steadily during starvation, reaching 10% of all available ribosomes (Reference: 10^3 ribosomes/cell; Bremer & Dennis, 2008). The fraction decreases to 5% after 28 h reflecting the adaptation process.

Linking the transcription and translation model to large-scale lifelines reveals the impact of delayed cellular response to fast external changes. In essence, cellular responses of transcription and translation are slower than convective zone shifts. Consequently, they are transported from one location to another, basically decoupled from external changes. Spatial analysis of all cells after 180 s (Figure 7c) reveals the highest transcript levels close to or even inside the glucose excess regime (50%–80% mRNA upregulation), while the lowest are found at the bottom. Once initiated, the starvation response propagated into the glucose excess zone. There, additional needs for transcription, translation, and ATP may limit the targeted formation of industrial products in microbial cells.

Noteworthy, it is exactly this feature that distinguishes the current model from previous lifeline studies (Haringa et al., 2016, 2017; Kuschel et al., 2017). There, metabolic and cell cycle responses were considered as an *instantaneous* cellular responses. Fast external changes are immediately translated into cellular replies. Later, Haringa et al. (2018) implemented metabolically buffered responses by considering variable enzyme pools (Tang et al., 2017). In this context, the current study proceeds by additionally integrating downstream transcriptional responses incorporating another level of cellular control. Our approach introduces the *non-instantaneous*, delayed response by considering intracellular programs of longer time scales than external changes. Accordingly, responses may propagate in different zones of the reactor, causing nonexpected transcriptional regulation programs there.

The approach was exploited further by estimating the entire add-on ATP demand for 120,000 newborn cells monitoring 4.5 h process time (Figure 8). As shown, the adaptation of the population is finished after 4.2 h disclosing a remaining ATP add-on of about 6% NAM compared to the 45% max NAM at the beginning. In terms of microbial productivity, these ATP needs simply reduce the amount of available ATP for product formation; that is, they limit biomass-specific productivities. The phenomenon has often be described in large-scale fermentation (Lara et al., 2006). Noteworthy, it is likely to be pronounced in hyper-producing cells with ATP intensive product formation. Often enough, such production processes run in fed-batch mode, installing reduced, limiting metabolic activity to stay within the technical limits of aeration and cooling. Consequently, those additional ATP needs hit cells with limited ATP forming capacities.

To evaluate the impact of particle simulation time, an additional simulation was conducted using a high-performance computation cluster studying 60,000 particles for about 460 s. Similar results were obtained for the key readouts, that is, time courses of ATP maintenance demands and residence time distributions remained. The simulated adaptation time reduced from 4.2 to 3.7 h, mirroring the lowered amount of particles staying in the starvation zone for the entire process time (around 5%) (see Supporting Information J). However, modelers need to consider that long simulation times are likely to violate the intrinsic constraint of one-way coupling, neglecting particle-environment interactions for the sake of simplicity. In this dilemma, we decided for the analysis of 180 s to capture key dynamics while still fulfilling the one-way coupling constraint.

As pointed out by Löffler et al. (2016), the majority of transcription dynamics are caused by the frequent on/off switching of stringent response, mediated by rising intracellular (p)ppGpp levels. Hence, creating stringent response deficient strains (Michalowski et al., 2017) opens the door to prevent non-wanted NAM increase. Besides, other cellular stress programs may be targeted as well (Supporting Information Appendix Table D4).

5 | CONCLUSION

The current modeling approach marries computational lifeline analysis with cellular regulation models, thereby introducing the *non-instantaneous* cellular response to changing extracellular conditions. Consequently, the spot of stress induction and the location of cellular phenotype do not need to be the same. Accordingly, heterogeneities in large-scale bioreactors comprise the physical levels linking local conditions tightly with metabolic responses and the cellular regulation level encompassing delayed responses such as transcriptional or translational effects.

To detect the latter and to describe them properly in data-driven models, experimental scale-up simulators are necessary that mirror transcriptional and translational cellular replies as performed by Kuschel and Takors (2020). The setting of such devices may differ from “conventional” scale-up simulators that typically mimic circulation times. Because the entire transcriptional responses should be clearly detectable, rather long stress exposure periods should be installed and read.

ACKNOWLEDGMENTS

The authors would like to thank the group of Computational Biology at the Institute of Biochemical Engineering for the use of the Galaxy-Server and Maike Kuschel and Flora Siebler for their support and inspiring discussions during the simulation process. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Prof. Dr.-Ing. Ralf Takors advised the study, while Moritz Wild implemented major parts of the agent-based model. Simulation, evaluation, and writing of the manuscript were conducted by Julia Zieringer.

ORCID

Ralf Takors http://orcid.org/0000-0001-5837-6906

REFERENCES

Arnold, T. E., Yu, J., & Belasco, J. G. (1998). mRNA stabilization by the ompa 5' untranslated region: Two protective elements hinder distinct pathways for mRNA degradation. *RNA*, *4*(3), 319–330.

Bremer, H., & Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, *3*(1), 1–49. https://doi.org/10.1128/ecosal.5.2.3

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Molecular Systems Biology*, *11*(1), 781. https://doi.org/10.15252/msb.20145794

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., & Noorman, H. (2017). Bioprocess scale-up/down as integrative enabling technology: From fluid mechanics to systems biology and beyond. *Microbial Biotechnology*, *10*(5), 1267–1274. https://doi.org/10.1111/1751-7915.12803

Garcia-Ochoa F., & Gomez E. 2009. Bioreactor-scale-up and oxygen transfer in microbial processes: An overview. *Biotechnology Advances*, *27*, 153–176. https://doi.org/10.1016/j.biotechadv.2008.10.006

Haringa, C., Deshmukh, A. T., Mudde, R. F., & Noorman, H. J. (2017). Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chemical Engineering Science*, *16*, 652–663. https://doi.org/10.1016/j.ces.2017.01.014

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., Mudde, R. F., & Noorman, H. J. (2016). Euler-Lagrange computational fluid dynamics for (bio) reactor scale down: An analysis of organism lifelines. *Engineering in Life Sciences*, *16*(7), 652–663. https://doi.org/10.1002/elsc.201600061

Haringa, C., Tang, W., Wang, G., Deshmukh, A. T., van Winden, W. A., Chu, J., van Gulik, W. M., Heijnen, J. J., Mudde, R. F., & Noorman, H. J. (2018). Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chemical Engineering Science*, *175*, 12–24. https://doi.org/10.1016/j.ces.2017.09.020

Kaleta, C., Schäuble, S., Rinas, U., & Schuster, S. (2013). Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnology Journal*, *8*(9), 1105–1114. https://doi.org/10.1002/biot.201200267

Kibbe, W. A. (2007). Oligocalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(Web Server issue), W43–W46. https://doi.org/10.1093/nar/gkm234

Kuschel, M., Siebler, F., & Takors, R. (2017). Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. *Bioengineering*, *4*(2), 27. https://doi.org/10.3390/bioengineering402002

Kuschel, M., & Takors, R. (2020). Simulated oxygen and glucose gradients as a prerequisite for predicting industrial scale performance a priori. *Biotechnology and Bioengineering*, *117*, 2760–2770. https://doi.org/10.1002/bit.27457

Lapin, A., Schmid, J., & Reuss, M. (2006). Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor: A structured–segregated approach. *Chemical Engineering Science*, *61*(14), 4783–4797. https://doi.org/10.1016/j.ces.2006.03.003

Lara, A. R., Galindo, E., Ramírez, O. T., & Palomares, L. A. (2006). Living with heterogeneities in bioreactors: Understanding the effects of environmental gradients on cells. *Molecular Biotechnology*, *34*(3), 355–382. https://doi.org/10.1385/MB:34:3:355

Leisch, F. (2006). A toolbox for k-centroids cluster analysis. *Computational Statistics and Data Analysis*, *51*(2), 526–544. https://doi.org/10.1016/j.csda.2005.10.006

Löffler, M., Simen, J. D., Jäger, G., Schäferhoff, K., Freund, A., & Takors, R. (2016). Engineering *E. coli* for large-scale production–strategies considering atp expenses and transcriptional responses. *Metabolic Engineering*, *38*, 73–85. https://doi.org/10.1016/j.ymben.2016.06.008

Mao, X., Ma, Q., Liu, B., Chen, X., Zhang, H., & Xu, Y. (2015). Revisiting operons: an analysis of the landscape of transcriptional units in *E. coli*. *BMC Bioinformatics*, *16*(1), 356. https://doi.org/10.1186/s12859-015-0805-8

Maurizi, M. R. (1992). Proteases and protein degradation in *Escherichia coli*. *Experientia*, *48*(2), 178–201. https://doi.org/10.1007/BF01923511

Michalowski, A., Siemann-Herzberg, M., & Takors, R. (2017). *Escherichia coli* HGT: Engineered for high glucose throughput even under slowly growing or resting conditions. *Metabolic Engineering*, *40*, 93–103. https://doi.org/10.1016/j.ymben.2017.01.005

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: A molecular approach*. Sinauer.

Neubauer, P., & Junne, S. (2010). Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Current Opinion in Biotechnology*, *21*(1), 114–121. https://doi.org/10.1016/j.copbio.2010.02.001

Neubauer, P., & Junne, S. (2016). Scale-up and scale-down methodologies for bioreactors. In *Bioreactors: Design, operation and novel applications*. (pp. 323–354) Wiley. https://doi.org/10.1002/9783527683369.ch11

Nieß, A., Löffler, M., Simen, J. D., & Takors, R. (2017). Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: Experimental evidence and mathematical model. *Frontiers in Microbiology*, *8*, 1195. https://doi.org/10.3389/fmicb.2017.01195

Noorman, H. J. (2011). An industrial perspective on bioreactor scale-down: What we can learn from combined large-scale bioprocess and model fluid studies. *Biotechnology Journal*, *6*(8), 934–943. https://doi.org/10.1002/biot.201000406

Noorman, H. J., & Heijnen, J. J. (2017). Biochemical engineering's grand adventure. *Chemical Engineering Science*, *170*, 677–693. https://doi.org/10.1016/j.ces.2016.12.065

Oosterhuis, N. M. G., & Kossen, N. W. (1983). Dissolved oxygen concentration profiles in a production-scale bioreactor. *Biotechnology and Bioengineering*, *26*, 546–550. https://doi.org/10.1002/bit.260260522

Proshkin, S., Rahmouni, A. R., Mironov, A., & Nudler, E. (2010). Cooperation between translating ribosomes and rna polymerase in transcription elongation. *Science*, *328*(5977), 504–508. https://doi.org/10.1126/science.1184939

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edger: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Shamir, M., Bar-On, Y., Phillips, R., & Milo, R. (2016). Snapshot: Timescales in cell biology. *Cell*, *164*(6), 1302. https://doi.org/10.1016/j.cell.2016.02.05

Stouthamer, A. H. (1973). A theoretical study on the amount of atp required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek*, *39*(1), 545–565. https://doi.org/10.1007/BF02578899

Takors, R. (2012). Scale-up of microbial processes: Impacts, tools and open questions. *Journal of Biotechnology*, *160*(1), 3–9. https://doi.org/10.1016/j.jbiotec.2011.12.010

Tang, W., Deshmukh, A. T., Haringa, C., Wang, G., van Gulik, W., van Winden, W., Reuss, M., Heijnen, J. J., Xia, J., Chu, J., Noormann, H. J. (2017). A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by *Penicillium chrysogenum*. *Biotechnology Bioengineering*, *114*(8), 1733–1743. https://doi.org/10.1002/bit.26294

Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J. T., Heijnen, J. J., & van Gulik, W. M. (2010). Genome-derived minimal metabolic models for *Escherichia coli* mg1655 with estimated in vivo respiratory atp stoichiometry. *Biotechnology and Bioengineering*, *107*(2), 369–381. https://doi.org/10.1002/bit.22802

Taymaz-Nikerel, H., Van Gulik, W. M., & Heijnen, J. J. (2011). *Escherichia coli* responds with a rapid and large change in growth rate upon a shift from glucose-limited to glucose-excess conditions. *Metabolic Engineering*, *13*(3), 307–318. https://doi.org/10.1016/j.ymben.2011.03.003

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction engineering principles*. Springer Science & Business Media.

Zieringer, J., & Takors, R. (2018). In silico prediction of large-scale microbial production performance: Constraints for getting proper data-driven models. *Computational and Structural Biotechnology Journal*, *16*, 246–256. https://doi.org/10.1016/j.csbj.2018.06.002

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zieringer J, Wild M, Takors R. Data-driven in silico prediction of regulation heterogeneity and ATP demands of *Escherichia coli* in large-scale bioreactors. *Biotechnology and Bioengineering*. 2020;1–14. https://doi.org/10.1002/bit.27568