Journal Pre-proof

Biokinetics Modelling of Lycopene-Producing *E. coli* Fermentation using PAT Methodology

Vasudevan Vivek, Fitriani Nur Eka, Dennis Leung, Wee Chew

PII: S1369-703X(24)00162-1

DOI: https://doi.org/10.1016/j.bej.2024.109375

Reference: BEJ109375

To appear in: *Biochemical Engineering Journal*

Received date: 24 January 2024
Revised date: 21 April 2024
Accepted date: 27 May 2024

Please cite this article as: Vasudevan Vivek, Fitriani Nur Eka, Dennis Leung and Wee Chew, Biokinetics Modelling of Lycopene-Producing *E. coli* Fermentation using PAT Methodology, *Biochemical Engineering Journal,* (2024) doi:https://doi.org/10.1016/j.bej.2024.109375

Biokinetics Modelling of Lycopene-Producing *E. coli* Fermentation using PAT Methodology

Vasudevan Vivek, Fitriani Nur Eka, Dennis Leung, Wee Chew*

Singapore Institute of Food and Biotechnology Innovation, Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, Singapore 138669, Singapore.

Author to whom correspondence should be addressed. E-mail: Chew_Wee@sifbi.a-star.edu.sg

ABSTRACT

Process analytical technology (PAT), which espouses the use of in-line analytics to monitor critical process parameters (CPPs), critical quality attributes (CQAs) and mathematical analyses, was employed to elucidate underlying carbon-source metabolism of a lycopene-producing *E. coli* strain during non-induced and induced batch fermentations. In-line dielectric and Raman spectroscopies were deployed to monitor real-time culture biomass viability (X), primary substrate glucose (S), and intracellular lycopene production (P), together with off-line HPLC quantification of broth acetate (A). In-line PAT monitoring enabled identification of two distinct metabolic regimes during induced batch run with lycopene accumulation. Biokinetics mathematical models using differential-algebraic equations and Monod-type conversions were formulated to explain experimentally measured dynamic trends of the state variables X, S, A and P during the batch runs. Their optimized model parameters and seven dimensionless ratios reflected changes in concentrations or rates relative to glucose, acetate, or lycopene. The calculated rates of culture growth, glucose utilization, acetate accumulation, and lycopene production revealed salient underlying carbon metabolic shifts and transient trends that explained phenomenological observations: such as growth patterns, carbon source usage, overflow mechanism, cellular affinity or inhibitory effects, and lycopene cytotoxicity. These parameters and dimensionless ratios could be used to compare novel strains, develop *in silico* simulations for scale-up or troubleshoot batch-to-batch variations.

Keywords:

Biokinetics modelling; Differential algebraic equations; *Escherichia* coli metabolism; Process Analytical Technology (PAT); In-line monitoring; Lycopene

INTRODUCTION

Escherichia coli is a workhorse microbe in molecular biology. The relative ease of genetic manipulation has led to a large body of knowledge in its biochemistry, physiology, and genetics, resulting in rapid progress of *E. coli* metabolic engineering and synthetic biology [1-3]. Engineered *E. coli* is widely used for the biosynthesis of secondary metabolites and high-value chemicals [4]. The carotenoid, lycopene, which has anti-oxidative, anti-cancer and anti-inflammatory properties [5], is used as a pharmaceutical compound, nutraceutical, functional food, and cosmetics additive [6]. Lycopene has been expressed using metabolic engineering approaches in fungal, yeast and bacterial strains, and overproduced in *E. coli* and *S. cerevisiae* [6–9]. One of the efficient metabolic pathways in *E. coli* that produces lycopene is the mevalonate (MVA) pathway [10]. Part of acetyl-CoA, a central metabolite generated from pyruvate in the glycolytic process, is converted into a secondary metabolite, MVA, which is a precursor to lycopene formation through a series of metabolic reactions along this over-expressed pathway [7, 9]. Pyruvate is also directed via the respiratory or aerobic fermentative pathway to produce energy or acetate, respectively. This metabolic shift is attributed to an "overflow metabolism," caused by the saturation of the limited respiratory capacity of the cell [11]. Overflow mechanism refers to the seemingly wasteful strategy of incomplete oxidation of primary substrate (most often glucose) to produce metabolites such as lactate, acetate, or even ethanol (the 'overflow'), instead of using following through the respiratory pathway despite presence of oxygen [12]. Overflow mechanism occurs ubiquitously among fast-growing cells, including bacteria, fungi and mammalian cells. Some examples include the Crabtree effect observed in *S. cerevisiae* [13, 14]; bacterial Crabtree effect in *E. coli* [15] and *B. subtilis* [16]; and Warburg effect observed in tumor cells and cell lines [17, 18].

Metabolic pathways modelling in microorganism typically consists of complex interconnected networks of tens to hundreds of dynamic biochemical and enzymatic reactions. To understand such dynamic behaviour of an organism under different genetic and environmental conditions requires the dynamic quantification of intracellular and extracellular metabolites that participate in the metabolic pathways coupled to computational metabolic network models [19, 20]. These models typically estimate reaction rates and associated kinetic parameters using enzyme, metabolite and substrate concentrations measured from quantitative metabolomics experiments [20]. However, such parameter estimates depend on selecting the "correct set" of reactions that represent the cellular metabolism, linking culture nutrient compounds to key intracellular intermediates, final product and by-products via model parameters. Often,

metabolic pathways and enzymes involved in the expression of particular cellular metabolites could be i) unknown, ii) have competing hypotheses on adjacent metabolic pathways causing undecided flux dominance, or iii) contain experimentally unmeasurable compounds [20, 21]. Numerically, as the number of biochemical transformations, fluxes and involved cellular metabolites increase in metabolic network modelling, the aforesaid epistemological and experimental uncertainties worsen the reliability, interpretability, and applicability of estimated model parameters for future predictions. As such, it is not unusual to find competing metabolic network models attempting to mathematically describe a set of underdetermined system of biochemical reaction equations associated to similar experimental conditions, which is partially caused by data overfitting [20].

Alternatively, biokinetic models are phenomenological models derived from well-known cellular mechanisms that underlie the behaviour of the microbial system or metabolism studied [22]. Biokinetics, as the term suggests, describe the dynamic behaviour of a microorganism in an environment and relate rates of biomass growth and intra- /extra-cellular product formation to extracellular concentrations of substrate and other metabolites, adequately supported by time series measurements. In contrast to metabolic network modelling approaches, such phenomenological methodology reduces the number of biochemical reactions, chemical species, and related parameters by at least an order of magnitude. Moreover, estimated parameters in such phenomenological modelling can be interpreted with a physical meaning within the context of the biokinetic model equations, and relating to the fermentation operating conditions with the microorganism metabolism [23].

One of the major challenges encountered during parameter estimation in a numerical model is limited and noisy experimental data [24]. Off-line data from broth sampling is often tedious to collect should the fermentation extends for days, if not weeks [25]. Quantification often takes hours or days to measure such off-line samples through methods like chromatography, spectrometry, flow cytometry, etc., and therefore, are typically unusable in real-time bioprocess control. Moreover, off-line experimental data is sparsely sampled, especially if several state variables are involved in the biokinetics modelling. This leads to aforesaid underdetermined numerical conundrum and causes some intractability in estimating parameters in more complex biokinetic models, though the extent of mathematical under-determination is much less than those typically associated with metabolic flux equations.

The Process Analytical Technology (PAT) framework promoted by the U.S. Food and Drug Administration (FDA) aims to design, analyse and control manufacturing processes through timely on- or in-line measurements of critical process parameters (CPPs) with the goal of ensuring final product quality [26–30]. Through the PAT approach, in-line process analytics are often deployed, and together with multivariate data analysis allows monitoring of CPPs [31–33]. The availability of such abundance of real-time CPPs (or state variables) data aids towards overcoming the numerical conundrum of overfitting biokinetics model parameters.

To the best of our knowledge, there are no reported biokinetics model for *E. coli* in literature that describe its intracellular lycopene production with biomass growth and glucose consumption alongside the acetate overflow mechanism. A recent PAT implementation of dual in-line Raman and dielectric spectroscopies for tracking real-time CPPs of cell viability and intracellular lycopene successfully enabled dynamic decision-making for in-process optimizing of high culture viability and product yield. This enabled the yield of an in-house *E. coli* 322 strain to increase 12.3-fold via fed-batch strategy [25]. The biokinetics modelling herein utilizes aforesaid in-line spectroscopic data, together with off-line ones, to develop a mathematical description of substrate-limited batch cultures in a 10L bioreactor. The resultant biokinetics model parameter estimates, together with calculated fluxes and dimensionless ratios, shed light on the changes in the cellular metabolism of lycopene-producing engineered *E. coli* 322 strain during non-induced and induced fermentation. Such model development and phenomenological understanding is often a critical first step in scale-up experimentation after successful shake flask optimization of novel engineered strains.

MATERIALS AND METHODS

10L Bioreactor Fermentation Setup

The biokinetics studies herein was intended for elucidating a better understanding of *E. coli* 322 strain that had successfully undergone metabolic engineering to overproduce intracellular lycopene [7]. The shake flask protocol developed in-house was adapted for batch cultures in a 10L bioreactor with unique unbaffled geometry [25, 34], comprising of both non-induced and induced fermentations. Details on the unbaffled flat bottomed 10L bioreactor setup, its geometry, mixing patterns, steam-in-place sterilization, and process automation of impeller stirring, liquid dosing, pH, sterile airflow, culture temperature, and in-line PAT sensors deployed have been reported recently [25, 34].

The pH during fermentation was monitored by InPro® pH 3100 probe from Mettler Toledo, and maintained at 7.0 ± 0.1 using PID controller with sulfuric acid (106.6 $g\ L^{-1}$) and base mixture (ammonium hydroxide 14 $g\ L^{-1}$ and sodium hydroxide 20 $g\ L^{-1}$). Aeration with sterile compressed air was maintained at 1 vvm using a Brooks mass flow controller along with a stirring rate at 300 rpm throughout each batch. The impeller speed was demonstrated to achieve well-mixed fluid homogeneity in the 10L bioreactor using computational fluid dynamics (CFD) simulations and corroborated with tracer experiments for viscosities ranging 1 – 6 *cP*, equivalent to high *E. coli* cell densities *ca.* 48 $g\ L^{-1}$ [34].

The composition of the chemically defined RMedia has been described earlier [35], with stock mixtures prepared from chemical sources with at least analytical grade. Typical final biomass grown in shake flask on glucose-limited growth conditions at *ca.* 37 ˚C yielded OD_{600} less than 8.0 (measured at 600 nm with Synergy HT, BioTek (Agilent) microplate reader). Culture and media conditions were extended for two 8L batch cultures in the 10L bioreactor. An inoculum of *ca.* 0.3 OD_{600} of seed culture (prepared in shake flask) was added at the start of each 8L batch run with an initial glucose concentration of *ca.* 15 $g\ L^{-1}$ [35]. For the non-induced batch fermentation (**Run 1**), biomass was cultured in a total volume of *ca.* 8L without induction, with broth temperature maintained at 37.0 ± 0.5 °C using a Huber Unichiller (Ottenburg, Germany), pH at 7.0, controlled aeration and mixing as described above. A comparative induced 8L batch culture (**Run 2**) had the same bioreactor operation controls, with injection of 0.1% v/v (total volume basis) isopropyl β-D-1-thiogalactopyranoside (IPTG) when biomass reached *ca.* 1–2 OD_{600} to trigger lycopene production. Fermentation broth temperature was also immediately turned down after IPTG injection to *ca.* 30 ˚C and maintained till end of Run 2.

PAT In-line and Off-line Measurements

Process related data were measured with both in-line analytical probes and off-line instrumentation. The utility of dual in-line Raman and dielectric spectroscopies (DS) for reliably monitoring CPPs of *E. coli* intracellular lycopene estimates ($mg_P\ g_{VCC}^{-1}$) and culture viable cell concentration ($g_{VCC}\ L^{-1}$) was demonstrated and reported recently [25]. The same were deployed for the non-induced and induced 8L batch cultures herein for elucidating phenomenological biokinetics modelling. The Kaiser Optical System RXN2 Raman spectrometer was equipped with 500mW Invictus™ diode excitation laser at 785 nm wavelength, holographic notch filter and CCD detector. It has an effective spectral range of 100–3425 cm^{-1} with 4 cm^{-1} resolution from the Raman backscatter collected via a 220 mm

sapphire-tipped Bio-Optic immersion probe inserted through the bioreactor top plate. Raman spectral data was acquired at 5 min intervals, with 3 scans at 15 s laser exposure each. In-line dielectric signals of batch fermentation was continuously recorded with DS system from Aber *Futura Annular* probe (⌀12 mm, 450 mm) inserted through the top plate of 10L bioreactor. DS probe dual frequency scanning mode at 1120 and 1565 kHz with 90 scans filter to minimise signal aberration, along cleaning mechanism activated every 30 min.

Regular broth sampling from the 10L bioreactor vessel was made through aseptic rapid sampling tube [25] for off-line HPLC quantification and biomass readings. Centrifugation of aliquots were performed immediately after sampling to obtain cell pellet and clear supernatant (in Eppendorf Centrifuge 5810R at 3000 rpm for 10 min). Broth glucose and acetic acid chromatographic peaks were separated by 5 µL auto-injections into a Bio-rad Aminex HPX-87H 300 × 7.8 mm column on HPLC (Agilent Infinity II 1260) with VWD at 210 nm (G1314F) and II (G7162A) RID, and analysed by the OpenLab CDS Chemstation edition (Rev.C.01 07 SR4) software. Cell pellets centrifuged from pipetted 2.0 mL culture drawn immediately after sampling were dried in 80 ℃ oven overnight, before measuring the dry cell weight (g_{DCW}) in triplicate using the Mettler Toledo XSE105 precision weighing balance (± 0.01 mg).

DS capacitance and conductivity data in CSV format were recorded, with subsequent analyses with MATLAB 9.8 (R2022a). Whereas post-run multivariate data analyses were performed on Raman spectral with MATLAB 9.8, following the same approach previously reported [25, 31–33]. For the batch studies herein, spectral estimates of glucose and lycopene were elucidated via the band-target entropy minimization (BTEM) curve resolution prior to performing multi-linear regression [25, 33], and eventually calibrated with offline data. The calibration plots are shown in *Supplementary Material* Figs. S1 and S2.

BIOKINETICS MODEL AND PARAMETERS OPTIMIZATION

A mechanistic model of E. coli metabolism based on previous works [36-38] was recently developed [39] with the inclusion of a continuous process of production and reassimilation of intracellular acetate, observed even under non-overflow conditions [40-42]. These have improved the understanding of acetate production and co-assimilation of both acetate and glucose in sugar-limited conditions [40, 41, 43]. In the current work, the biokinetics modeling of lycopene-producing E. coli 322 strain extends the macro-phenomenological approach developed by Anane et al. [39, 44] to mathematically express the carbon-source metabolism towards lycopene formation. A set of ordinary differential equations (ODEs) described three

time-dependent *state variables*, biomass $\boldsymbol{X(t)}$, extracellular concentrations of primary substrate, glucose $\boldsymbol{S(t)}$, and that of acetate $\boldsymbol{A(t)}$ (**Eqs. 1–3**), for both non-induced and induced batches. Post IPTG induction, additional equation to express specific rate of lycopene product formation, $\boldsymbol{P(t)}$, was augmented to the earlier system of ODEs. This quartet of differential equations mathematically described time evolution of the four state variables, *X*, *S*, *A* and *P*, that were measured experimentally using in-line PAT or off-line analytics described above. As there was an observed lag time (t_{lag}) before lycopene exponentially accumulated in the cells (**Figs. 3** and **4**), the lycopene production differential equation is modified into two distinct time regimes of **Eqs. 4a** and **4b**. The observed lag time indicated a metabolic switch to activate the lycopene producing MVA pathway. Details for the carbon metabolic switch are mathematically described by Eqs. 6a, 6b, 7 and 8 in section ***Glucose Uptake, Partitioning, Diversion, and Conversion*** below.

$$\frac{dX}{dt} = \mu(t) \cdot X \tag{1}$$

$$\frac{dS}{dt} = -r_S(t) \cdot X \tag{2}$$

$$\frac{dA}{dt} = r_A^{ex}(t) \cdot X \tag{3}$$

$$\frac{dP}{dt} = 0, \forall t < t_{lag} \tag{4a}$$

$$\frac{dP}{dt} = r_P(t), \forall t \geq t_{lag} \tag{4b}$$

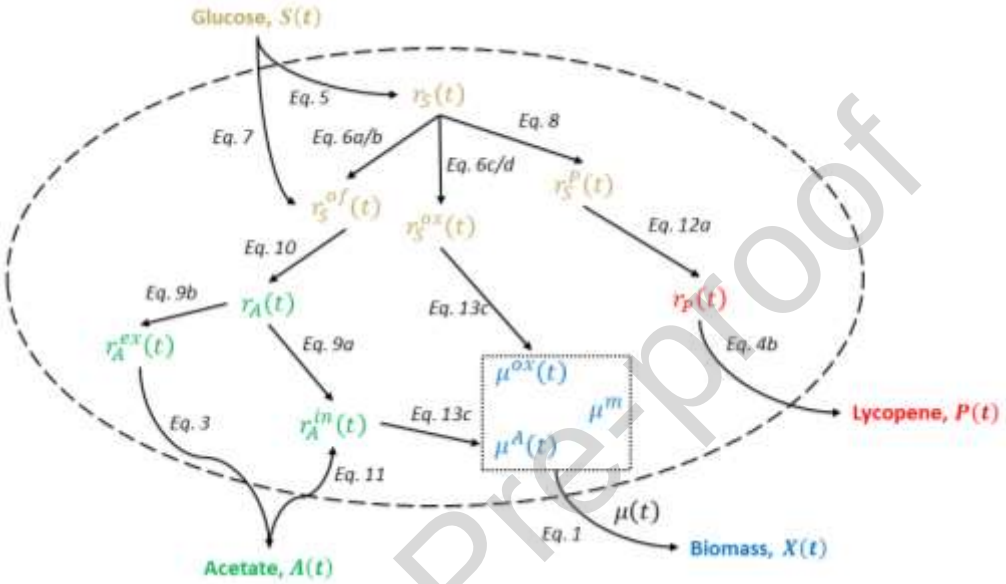

Figure 1: Biochemical reaction pathways in carbon metabolism of *E. coli* 322 strain

Furthermore, a set of auxiliary algebraic equations, including Monod-type kinetics, complemented the four ODEs to macroscopically describe glucose uptake, and partitioning, acetate formation through the overflow mechanism, oxidative pathways, and cell maintenance in the overall metabolism. The schematic diagram **Fig. 1** illustrates the network of biochemical reaction pathways from glucose to acetate, lycopene production and biomass formation, with their associated model equations. As described below, they were associated with the specific rates of glucose uptake and diversion, acetate formation, consumption or excretion, as well as lycopene production alongside biomass growth. These algebraic equations (**Eqs. 5–13**) along with earlier ODEs (**Eqs. 1–4**), its time-dependent biomass growth $\mu(t)$, glucose uptake $r_S(t)$, acetate excretion $r_A^{ex}(t)$, and lycopene production $r_P(t)$, formed sets of equations relating to *three distinct biokinetic models* that described non-induced (Run 1) and induced (Runs 2A and 2B) batch cultures. Full exposition on these biokinetics sets, including their model parameters and dimensionless ratios in relation to metabolic pathway affinities, inhibitions and limitations, is provided in section *Detailed Biokinetics Model* of the *Supplementary Material*.

Glucose Uptake, Partitioning, Diversion, and Conversion

Glucose was the primary carbon source of the *E. coli* 322 fermentation and its broth concentrations appear as the state variable $S(t)$ in the model equations. **Eq. 5** described the specific glucose uptake rate[§], $\boldsymbol{r_S(t)}$, in $g_S\ g_{VCC}^{-1}\ h^{-1}$ that accounted for acetate inhibition using a Monod-type kinetics with non-competitive inhibition [45]. This uptake rate further relates to glucose partitioning for non-induced and induced cases, via **Eqs. 6a** or **6b** respectively, towards separate rates for oxidative pathways, $\boldsymbol{r_S^{ox}(t)}$, overflow mechanism, $\boldsymbol{r_S^{of}(t)}$, and lycopene production, $\boldsymbol{r_S^{P}(t)}$. The conversion rates of glucose diverted for overflow and lycopene formation were modelled in **Eqs. 7** and **8** respectively. The rationale for using $r_S(t)$ in Eq. 8 is provided below under *Phenomenological and Metabolic Understanding Inferred from Biokinetics Models – Induced Run 2B.* Subsequently, glucose attributed to central metabolism involving aerobic respiration, $r_S^{ox}(t)$, was calculated with **Eqs. 6c** and **6d**.

$$r_S(t) = \frac{r_S^{max}}{1 + \frac{A(t)}{K_i^A}} \cdot \frac{S(t)}{S(t) + K_S} \tag{5}$$

Non-induced:

$$r_S(t) = r_S^{ox}(t) + r_S^{of}(t) \tag{6a}$$

Induced:

$$r_S(t) = r_S^{ox}(t) + r_S^{of}(t) + r_S^{P}(t) \tag{6b}$$

[§] The gravimetric dry cell weight of biomass $X(t)$ (in g_{DCW}) was estimated in real-time in its equivalent viable cell concentration (*i.e.* g_{VCC}) through the correlation of in-line dielectric capacitance *pF/cm* readings with off-line measured g_{DCW} (see *Supplementary Material* **Fig. S1**).

$$r_S^{of}(t) = r_{AS}^{max} \frac{S(t)}{S(t) + K_{AS}}$$

(7)

$$r_S^{P}(t) = r_{PS}^{max} \frac{r_S(t)}{r_S(t) + K_{PS}}$$

(8)

Non-induced:

$$r_S^{ox}(t) = r_S(t) - r_S^{of}(t)$$

(6c)

Induced:

$$r_S^{ox}(t) = r_S(t) - r_S^{of}(t) - r_S^{P}(t)$$

(6d)

Acetate Reassimilation, Excretion, and Accumulation

The specific rate of production of intracellular acetate via the overflow route, $\boldsymbol{r_A(t)}$, directly arose from glucose diverted via overflow mechanism, $r_S^{of}(t)$. This accounted for an intracellular reassimilation term, $\boldsymbol{r_A^{in}(t)}$, which explained broth acetate $A(t)$ consumed, and another term denoting excess acetate excreted into the extracellular medium, $\boldsymbol{r_A^{ex}(t)}$. The latter contributed directly to experimentally measured acetate $A(t)$ via **Eq. 3**. These acetate rates were expressed in $g_A\ g_{VCC}^{-1}\ h^{-1}$, and described by **Eqs. 9–11**.

$$r_A(t) = r_A^{in}(t) + r_A^{ex}(t)$$

(9a)

$$r_A^{ex}(t) = r_A(t) - r_A^{in}(t)$$

(9b)

$$r_A(t) = Y_{AS} \cdot r_S^{of}(t)$$

(10)

$$r_A^{in}(t) = \frac{r_A^{max}}{1 + \frac{S(t)}{K_i^S}} \cdot \frac{A(t)}{A(t) + K_A}$$

(11)

Lycopene Production

It was assumed that reassimilation of acetate does not contribute directly to lycopene production but contributed only towards biomass growth (see **Eq. 13** below). Thus, the intracellular lycopene productivity for *E. coli* 322 strain, or its specific rate of lycopene production, $\boldsymbol{r_P(t)}$, in $mg_P\ g_{VCC}^{-1}\ h^{-1}$, was modeled as **Eq. 12a**, and when combined with **Eq. 8** yielded **Eq. 12b**.

$$r_P(t) = Y_{PS} \cdot r_S^P(t)$$

(12a)

$$r_P(t) = (Y_{PS} \cdot r_{PS}^{max}) \frac{r_S(t)}{r_S(t) + K_{PS}} = \propto_{PS} \cdot \frac{r_S(t)}{r_S(t) + K_{PS}}$$

(12b)

$\boldsymbol{\propto_{PS}}$, or equivalently ($Y_{PS} \cdot r_{PS}^{max}$), was the regime specific intracellular lycopene productivity (in $mg_P\ g_{VCC}^{-1}\ h^{-1}$) achieved via MVA pathway given the bioreactor operating conditions fixed in this study. As seen in the post-induction phase (**Fig. 3)**, two distinct regimes of biomass growth $X(t)$ along glucose consumption $S(t)$ were observed as lycopene accumulated in the cell culture over time: **Run 2A** corresponded to an exponential biomass growth, while **Run 2B** corresponded to an inhibitory effect of lycopene on biomass growth (see details in *Results and Discussion*). These two regimes were modeled by introducing a linear term that promoted or inhibited the lycopene production, respectively [46] (**Eqs. 12c** and **12d**). The parameter P_{crit} in Run 2B (**Eq. 12d**) reflects predicted the maximum lycopene concentration that was achievable due to lycopene cytotoxicity.

Run 2A:

$$r_P(t) = \propto_{PS} \cdot \frac{r_S(t)}{r_S(t) + K_{PS}} \left(1 + \frac{P(t)}{P_{crit}}\right)$$

(12c)

Run 2B:

$$r_P(t) = \propto_{PS} \cdot \frac{r_S(t)}{r_S(t) + K_{PS}} \left(1 - \frac{P(t)}{P_{crit}}\right)$$

(12d)

Biomass Growth

The overall specific growth rate of biomass (μ) could be modeled as a combination of several contributions, namely, (i) growth due to oxidation of glucose, $\mu^{ox}(t)$, (ii) retardation in growth due to maintenance of biomass, $\mu^{m}(t)$, (iii) growth due to reassimilation of acetate via the overflow route, $\mu^{of}(t)$, and (iv) growth due to consumption of extracellular acetate as secondary substrate, $\mu^{A}(t)$. The overall specific growth rate, $\mu(t)$, in h^{-1}, was initially modeled with various yield coefficients Y_{XS}^{ox}, Y_{XS}^{of}, and Y_{XA}, converting corresponding glucose and acetate diversion rates into biomass growth:

$$\mu(t) = (\mu^{ox}(t) - \mu^{m}) + \mu^{of}(t) + \mu^{A}(t)$$

(13a)

$$\mu(t) = Y_{XS}^{ox} \cdot (r_S^{ox}(t) - r_m) + Y_{XS}^{of} \cdot r_S^{of}(t) + Y_{XA} \cdot r_A^{in}(t)$$

(13b)

Preliminary simulations indicated that growth due to acetate reassimilation via the overflow route, $\mu^{of}(t)$, or $Y_{XS}^{of} \cdot r_S^{of}(t)$, was less than *ca.* 5 % when compared to the overall specific growth rate, $\mu(t)$. Phenomenologically, the diverted glucose rate, $r_S^{of}(t)$, should primarily account for acetate formation via overflow mechanism rather than biomass growth. Moreover,

the growth accounted by acetate reassimilation was represented by $r_A^{in}(t)$ leading to $\mu^A(t)$. Hence, the overall specific growth rate, $\mu(t)$, was simplified as:

$$\mu(t) = (\mu^{ox}(t) - \mu^m) + \mu^A(t) \tag{13c}$$

$$\mu(t) = Y_{XS}^{ox} \cdot (r_S^{ox}(t) - r_m) + Y_{XA} \cdot r_A^{in}(t) \tag{13d}$$

Table 1 Summary of experimentally measured state variables, parameters to be optimized and specific rates calculated from the biokinetic models

State variable	Model Parameters	Estimated Specific Rates**
Biomass (*X*)	r_m, Y_{XS}^{ox}, Y_{XA}	$\hat{\mu}$, $\widehat{\mu^m}$, $\widehat{\mu^{ox}}$, $\widehat{\mu^A}$
Glucose (*S*)	r_S^{max}, K_S, K_i^A, r_{AS}^{max}, K_{AS}, r_{PS}^{max}, K_{PS}	$\widehat{r_S}$, $\widehat{r_S^{of}}$, $\widehat{r_S^{ox}}$, $\widehat{r_S^P}$
Acetate (*A*)	Y_{AS}, r_A^{max}, K_i^S, K_A	$\widehat{r_A}$, $\widehat{r_A^{ex}}$, $\widehat{r_A^{in}}$
Lycopene (*P*)	Y_{PS}, P_{crit}	$\widehat{r_P}$

Table 1 compiles all aforesaid model parameters and specific rates associated with the three sets of model equations that mathematically describe experimentally observed state variables *X*, *S*, *A*, and *P*, for non-induced and induced cultures. The phenomenological influence of their optimised model parameters, *i.e.*, their relation to inhibitory or affinity effects of glucose and acetate, towards the specific rates, is further explained in the section *Phenomenological and Metabolic Understanding Inferred from Biokinetics Models*.

Numerical Optimization for Model Parameters Estimation

The macro-kinetic models developed above contained 4 ODEs and 13 auxiliary algebraic equations. To estimate the optimal parameter values in the set of nonlinear equations (**Eqs. 1–13**, and **Table 1**), corresponding initial values from literature studies of similar microbial strains under similar fermentation conditions [39, 44] were chosen as initial guesses (**Table S1** in Supplementary Material). The parameter search space was limited to all positive values

** These estimated specific rates are reported throughout the manuscript without the carat (^) symbol that typically signifies estimated values calculated from optimised model parameters after non-linear optimisation converges to a solution.

above 10^{-5} during optimization. A cost function that calculates the difference between estimated transient state variables *X*, *S*, *A*, *P*, and their batch fermentation data was used (**Eq. 15**). The model equations were integrated with the variable time-step ode15s solver and the cost function was minimized using the *lsqnonlin* subroutine with trust region reflective algorithm in MATLAB R2020b®.

With the use of literature chosen initial guesses for biokinetics model parameter optimization, the resulting solution would likely represent a local minimum solution. To find the global minimum, the interim local optimal parameter solution sets were used iteratively as guesses for subsequent parameters optimization. Initial trial optimizations indicated that 30-50 iterations were sufficient to reach a global minimum, wherein there was no longer any significant change in the overall RMSE. Therefore, an exhaustive number of iterations (*i*) equals 50 was chosen to arrive at the global optimum solution. This global convergence is demonstrated by plotting z-score parameter values against the number of iterations (**Eqn. 18**). The algorithm to estimate global optimal solution from given source of initial guess is summarized in a flowchart (**Fig. 2**). It is also noted that only parameter values wherein the change in overall RMSE (**Eqn. 16**) from the preceding iteration was less than 1% were used to calculate the corresponding mean, standard deviation and z-score of respective parameters.

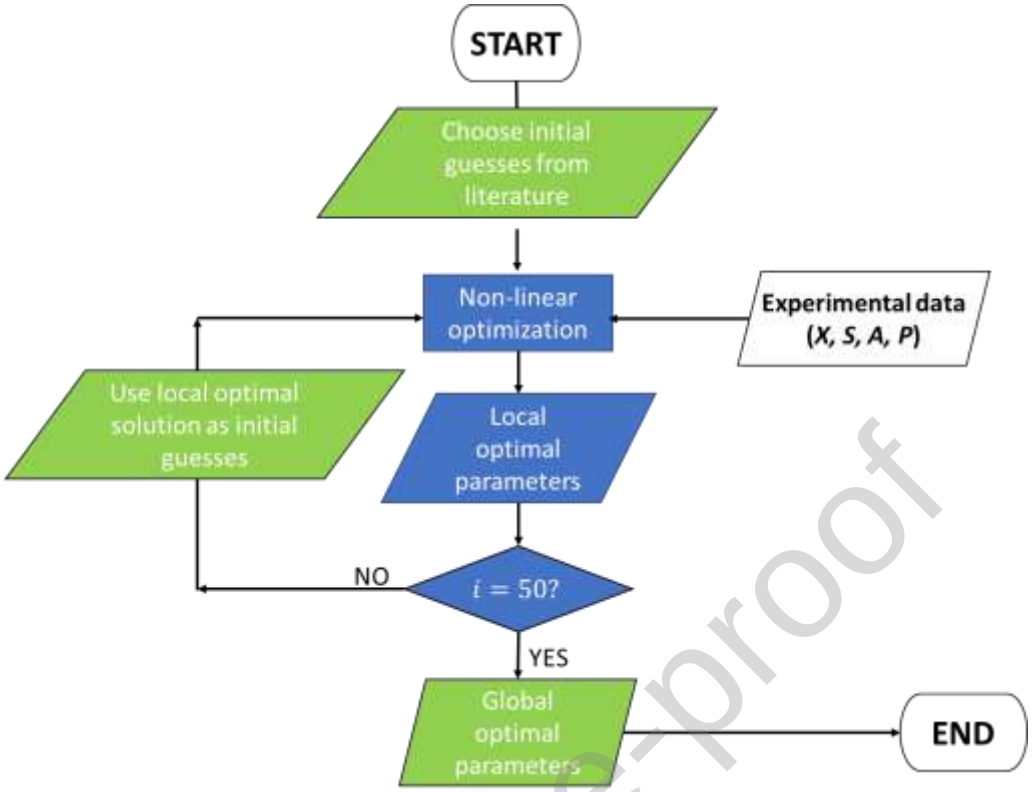

Figure 2: Schematic flowchart of algorithm for parameter estimation by global optimization.

Optimization Cost Function

The model parameters were estimated by solving the optimization problem:

$$\hat{\vec{\theta}} := \min_{\theta} \Phi(\vec{\theta})$$

(14)

$\hat{\vec{\theta}}$ was the vector of optimized parameters and $\Phi(\vec{\theta})$ was the cost function to be minimized.

$$\Phi(\vec{\theta}) = \sum_{k=1}^{n_p} \Phi_k(\vec{\theta}) = \sum_{k=1}^{n_p} \sum_{m=1}^{t_{expt}} \left\{ \frac{\hat{d}_k(\theta_1, \dots, \dots, \theta_p) - d_{k,m}^{expt}}{\max(d_{k,m}^{expt})} \right\}^2$$

(15)

$\Phi_k(\vec{\theta})$ was the cost function for individual state variables, n_p was the number of state variables, t_{expt} was the experimental sampling time-point and $d_{k,m}^{expt}$ were the experimental data points of the model estimated corresponding state variables data $\hat{d}_k$. A root-mean squared

error (RMSE) was computed to estimate the goodness-of-fit of the model solution to experimental data, and n_t^{expt} was the number of experimental datapoints, according to **Eq. 16**.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_p \times n_t^{expt}} \sum_{k=1}^{n_p} \sum_{m=1}^{t_{expt}} \left\{ \frac{\hat{d}_k(\theta_1, \dots, \dots, \theta_p) - d_{k,m}^{expt}}{\max(d_{k,m}^{expt})} \right\}^2}$$

(16)

The coefficient of determination, R^2, was also calculated to estimate the goodness of fit to experimental data (**Eq. 17**). It was calculated from the ratio of the residual sum of squares (RSS) and total sum of squares (TSS), with RSS being a measure of deviation of fitted model to experimental data, and TSS being a measure of the variance in the experimental data.

$$\mathrm{R}^2 = 1 - \frac{RSS}{TSS}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{k=1}^{n_p} \sum_{m=1}^{t_{expt}} \{ d_{k,m}^{expt} - \hat{d}_k(\theta_1, \dots, \dots, \theta_p) \}^2}{\sum_{k=1}^{n_p} \sum_{m=1}^{t_{expt}} \{ d_{k,m}^{expt} - \bar{d}_k^{expt} \}^2}$$

(17)

where $\bar{d}_k^{expt}$ is the mean of experimental data for each state variable, *X*, *S*, *A* and *P*.

During global optimization, a moving-average parameter value ($\overline{\theta_i}$) was calculated for each iteration whose percentage change in RMSE between successive iterations was less than 1%. After 50 iterations, a parameter z-score (**Eqn. 18**) was calculated for every iteration using the moving-average and moving-standard deviation of the parameter value at the 50th iteration.

$$z_{\theta_\mathrm{i}} = \frac{(\theta_\mathrm{i} - \overline{\theta_{50}})}{\overline{\sigma_{50}}}$$

(18)

where, θ_i is the parameter value at the *i*th iteration, $\overline{\theta_{50}}$ is the moving-average parameter value after 50 iterations, and $\overline{\sigma_{50}}$ is the moving-average standard deviation after 50 iterations.

There were 12 model parameters in the non-induced case (Run 1), and 16 such parameters in both the induced cases, Runs 2A and 2B (**Table 1**). Notably, there were 16 off-line measurements in both Runs 1 and 2A, and 12 off-line measurements in Run 2B. Due to insufficient off-line measurements, the model for Run 2B was an underdetermined system; as such, using off-line data alone no unique parameter solution set is achievable through optimization for Run 2B. Fortuitously, the abundance of data from in-line PAT analytics assisted to overcome this mathematical drawback and provided sufficient experimental measurements to uniquely determine the model parameters for Run 2B.

RESULTS AND DISCUSSION

Observation of Post-induction Metabolic Regime Transition via In-line PAT

One of the key aspects of this study highlighted the advantages of using PAT in-line analytics achieving continuous monitoring and high time-resolution of dynamic state variables. The combination of both in-line and off-line analytics data enabled numerical optimisation of the model parameters underpinning phenomenological biokinetics described in section *Biokinetics Model and Parameters Optimization* (**Eqs. 1–13**). In-line PAT monitoring of dielectric and Raman spectroscopies, with the latter analysed with multivariate chemometrics, elucidated high resolution dynamic data of critical process parameters (CPPs), *i.e.*, viable cell concentration (VCC), broth glucose and intracellular lycopene that corresponded to state variables *X*, *S* and *P*, respectively.

This analytical superiority was evident in the case of induced batch fermentation Run 2 (**Fig. 3**). Induction with IPTG was performed at batch time *ca.* 2.38 *h* when OD_{600} was *ca.* 1.4 (not shown in **Fig. 3**). An exponential growth phase was observed in both in-line dielectric correlated VCC and off-line dry cell weight (DCW) till batch time *ca.* 7 *h*, which was accompanied by rapid glucose consumption and lycopene production. However, after *ca.* 7 *h*, biomass growth was clearly slower along with lower glucose consumption and lycopene production. This pointed to a regime transition (*ca.* 6-8 *h*). To determine the approximate time at which the cell population switched over its metabolic regime, further data analysis was performed to in-line glucose and lycopene concentrations. Several significant changes in rates were observed: i) a kink in the decreasing glucose concentration profile, ii) a maximum in the calculated lycopene productivity ($mg_P\ g_{VCC}^{-1}\ h^{-1}$) and iii) a shift in the biomass growth rate away from a strong exponential trend to a slower one. Such decrease in rates indicate an

intracellular metabolic regime shift, possibly caused by the increased lycopene production that was detrimental to biomass growth [47]. Consequently, for computational purposes, the induced fermentation biokinetic model was split into two regimes: **Run 2A**, characterized by an exponential biomass growth, increasing lycopene productivity and rapid glucose consumption; and **Run 2B**, characterized by minimal biomass growth, decreasing lycopene productivity and slow glucose consumption. The switch between regimes was chosen at batch time *ca.* 6.5 *h*, when lycopene concentration was *ca.* 14 $mg_P\ g_{VCC}^{-1}$, as it corresponded to when both a kink in glucose profile and reduction in biomass exponential growth profile was simultaneously observed, alongside close proximity to the maximum lycopene productivity (**Fig. 3**).

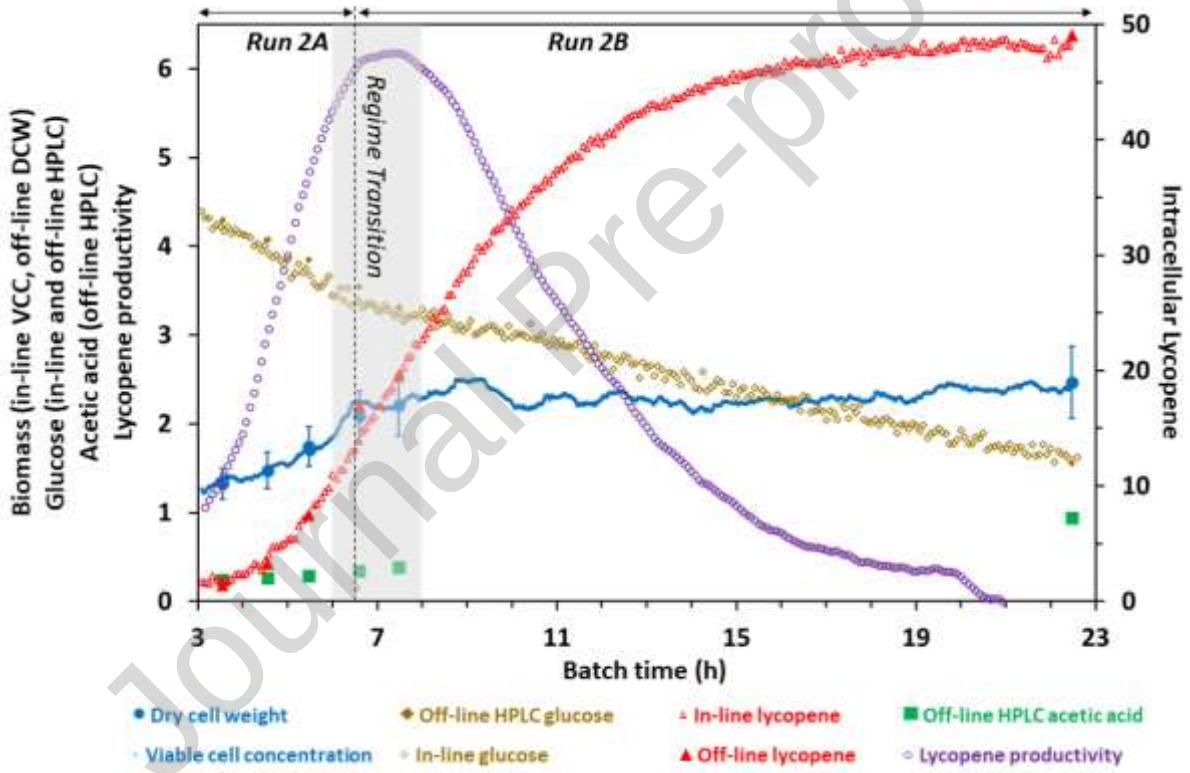

Figure 3: Profiles of in-line viable cell concentration ($g_{VCC}\ L^{-1}$), off-line dry cell weight ($g_{DCW}\ L^{-1}$), in-line and off-line HPLC glucose ($g_S\ L^{-1}$), off-line HPLC acetic acid ($g_A\ L^{-1}$), in-line and off-line intracellular lycopene ($mg_P\ g_{VCC}^{-1}$) and calculated lycopene productivity ($mg_P\ g_{VCC}^{-1}\ h^{-1}$**)** during induced batch fermentation of *E. coli* 322 strain in 10L bioreactor. Glucose concentrations are scaled down by a factor of 3. Open symbols indicate in-line measurements and closed symbols indicate off-line measurements.

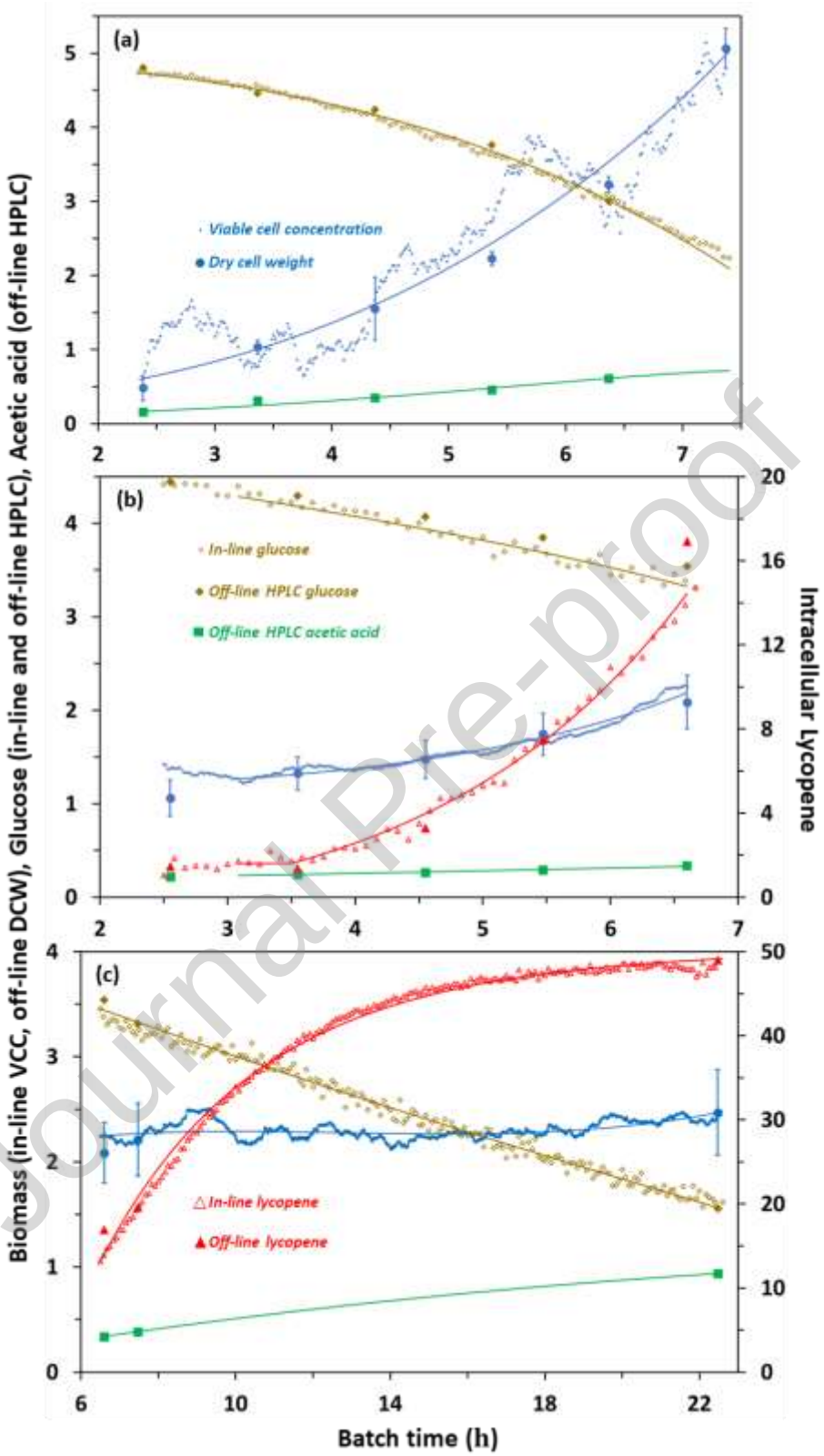

Figure 4: Profiles of in-line viable biomass or cell concentration ($g_{VCC}\ L^{-1}$), off-line dry cell weight ($g_{DCW}\ L^{-1}$), in-line and off-line HPLC glucose ($g_S\ L^{-1}$), off-line HPLC acetic acid ($g_A\ L^{-1}$) and in-line and off-line intracellular lycopene ($mg_P\ g_{VCC}^{-1}$) during **a)** non-induced Run 1, **b)** induced Run 2A and **c)** induced Run 2B batch fermentation of *E. coli* 322 strain in 10L bioreactor. Glucose concentrations are scaled down by a factor of 3. Open symbols indicate in-line measurements, closed symbols indicate off-line measurements and solid lines indicate biokinetics time-series predictions with optimized parameters.

Biokinetic Models Optimized Parameters

Collating the final optimization results for Runs 1, 2A and 2B, **Table S1** in *Supplementary Material* shows the initial guesses, the mean global optimum values and standard deviations of the parameters in the biokinetic models for non-induced (Run 1) and induced (Runs 2A and 2B) batch fermentation of *E. coli* in the 10L bioreactor. The initial guesstimates for the parameters common to Runs 1 and 2 were taken from literature values of *E. coli* culture [44]. The parameters r_{PS}^{max}, K_{PS} and Y_{PS} in Run 2 were uniquely associated with the lycopene-producing *E. coli* 322 strain, and thus have no literature values. Their initial guesses were assumed to be of similar magnitude to the literature values of maximum specific rates (r_S^{max}, r_A^{max}, r_{AS}^{max}), affinity constant (K_{AS}) and yields (Y_{XS}^{ox}, Y_{XA}, Y_{AS}) for glucose and acetate [44]. The initial guesses for critical lycopene concentration (P_{crit}) for Runs 2A and 2B were assumed equal to the initial and final in-line experimental concentrations, *ca.* 2 and 50 $mg_P\ g_{VCC}^{-1}$, respectively. The converged RMSE values for all the three optimization cases of Runs 1, 2A and 2B were below 5.2×10^{-3}, while the R^2 values were very close to 1.0 (rounded to 4 decimal places). The high R^2 values were due to the abundance of online experimental data compared to only a small number of model parameters (that is, 12–16) in the three biokinetic models leading to excellent goodness-of-fit.

The rate of convergence of the different parameters with iteration number was investigated using a plot of parameter z-score versus number of iterations (**Fig. 5**. For Run 1, **Figs. S3 and S4** for Runs 2A and 2B). As observed in our initial simulated trials, although the change in RMSE between successive iterations was less than 1% beyond 10 iterations (data not shown), the optimized parameters reached convergence only beyond 30 iterations. The spread in the optimised parameter values for all the iterations is shown as box plots in **Fig. 6** for Run 1.

Corresponding box plots for Runs 2A and 2B are shown in **Figs. S5 and S6** in the *Supplementary Material*, respectively.

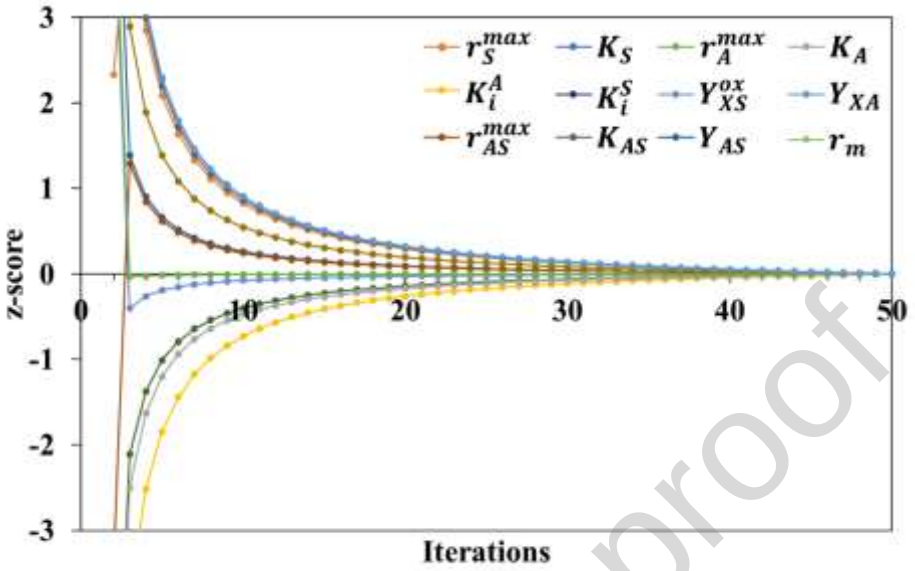

Figure 5: Rate of convergence for optimised parameters in Run 1 versus number of iterations.

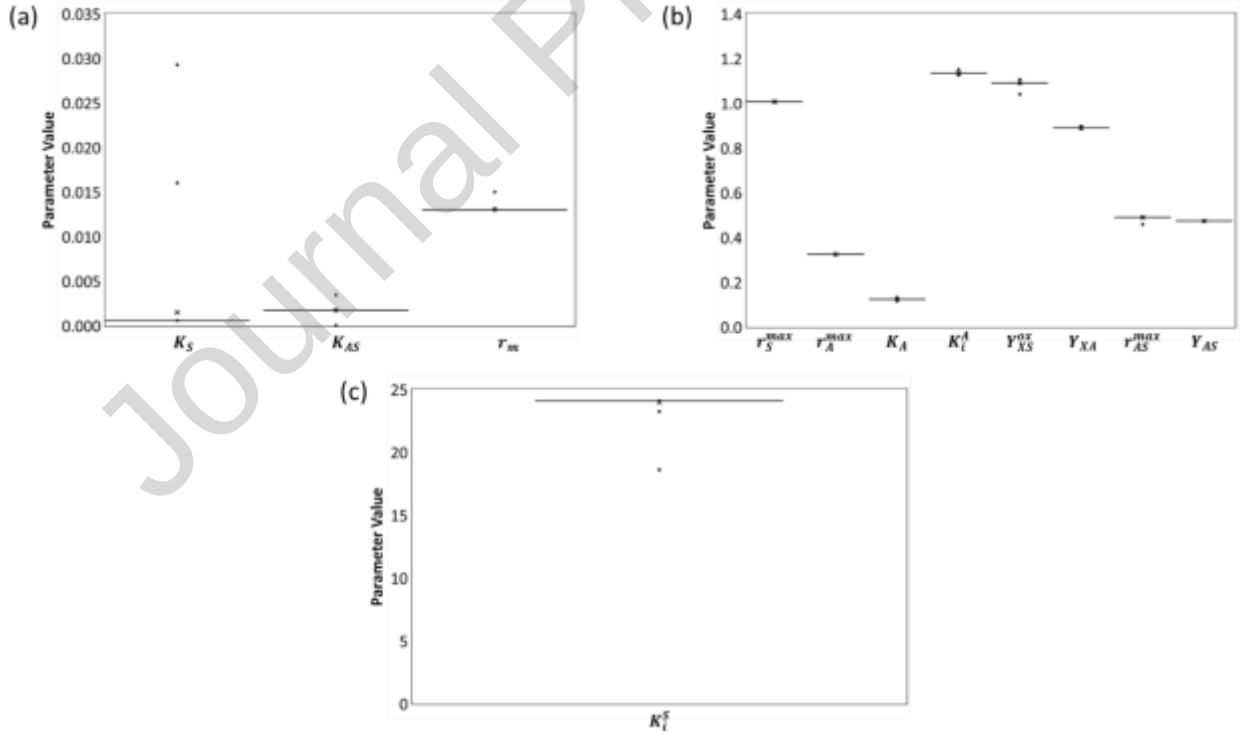

Figure 6: Spread in optimised parameter values for Run 1 over 50 iterations (a) Parameter values < 0.04 (b) Parameter values between 0.1 – 2 (c) Parameter values > 5.

The box plots of some parameters show outliers that correspond to initial iterations (i < 30). However, the spread of the individual parameters lies within a narrow range indicating minimal deviations of parameter values post-convergence.

Corroboration of Biokinetics Model Predictions with Experimental Data

The dynamic growth, consumption, accumulation and production trajectories of state variables, *X*, *S*, *A* and *P*, predicted by the optimized biokinetics model, well represented the in-line PAT dielectric and Raman data, and off-line HPLC and UV-Vis measurements with minimal error (**Fig. 4**). Glucose concentrations in **Figs. 3 and 4** were reduced by a factor of 3 for visual presentation purposes.

The biokinetics model parameters for biomass (*X*), broth glucose concentration (*S*), and specific intracellular lycopene produced (*P*), were optimized based on in-line PAT data. The viable cell concentration in biomass ($g_{VCC}\ L^{-1}$) were estimated by calibrating off-line DCW ($g_{DCW}\ L^{-1}$) to in-line dielectric capacitance ($pF\ cm^{-1}$), **Fig. S1** in *Supplementary Material*. In-line Raman data yield both *S* and *P* time-series values through multivariate chemometrics (**Fig. S2** in *Supplementary Material*). As for acetate concentrations, its corresponding model parameters were optimized based on off-line HPLC acetate data since in-line measurements were not available. Moreover, the in-line PAT continuous monitoring advantage of state variables *X*, *S* and *P* in between sampling times, especially during overnight hours *ca.* 7.5–22.5 *h* (**Fig. 4c**), helped resolve the aforesaid underdetermined model parameter optimization issue in Run 2B (see section *Optimization Cost Function*).

During non-induced fermentation Run 1 (**Fig. 4a**), both in-line and off-line data, consumption and accumulation patterns of glucose (*S*) and acetate (*A*) concentrations followed an exponential viable biomass (*X*) growth pattern. The DCW off-line data at *ca.* 7.4 *h* is as shown in **Fig. 4a**, but corresponding off-line acetate and glucose concentrations were missing due to HPLC run problems. Fortunately, in-line Raman data estimates for glucose concentrations during the overnight period *ca.* 6.4–7.4 *h* were available for biokinetics model optimisation.

For Run 2A, the off-line DCW measurements monotonically increased, whereas the dielectric capacitance and corresponding viable biomass estimates indicated a slight decrease at batch time *ca.* 2.5–3.10 *h* (**Fig. 4b**). This corresponded to a lag time (t_{lag}) that was observed from the first sampling post IPTG induction (*ca.* batch time 2.55 *h*) to *ca.* 3.5 *h*. Although some lycopene was produced during this period, *ca.* 1.5 $mg_P\ g_{VCC}^{-1}$, minimal lycopene accumulation

was observed. Both t_{lag} and drop in viable biomass were indicative of the cellular metabolic switch to promote glucose diversion into the MVA pathway for lycopene production (**Eqs. 8** and **12c**). Several computational trials were performed for parameter optimization in Run 2A to account for both the in-line dielectric viable biomass and observed t_{lag}. The lowest RMSE and best fit was found by setting t_{lag} in **Eq. 4** to 3.5 *h* and optimizing parameters using in-line data starting from *ca.* 3.1 *h* that corresponds to minimum viable biomass (*ca.* 1.22 $g_{\mathrm{VCC}}\ L^{-1}$).

The batch times for Runs 1 and 2A were intentionally carried out to have an overlap between ca. 3–6.6 *h* for comparing the biokinetics without and with IPTG induction (**Figs. 4a** and **4b**). Run 2B continued beyond Run 2A for batch time *ca.* 6.5 – 22.5 *h* (**Fig. 4c**), with both sharing the same off-line data at *ca.* 6.6 *h*, since this sampling point fell within the aforesaid regime transition. The biomass grew to a higher concentration during Run 1 (*ca.* 5.1 $g_{DCW}\ L^{-1}$), as compared to the lower concentration during Run 2A (*ca.* 2.1 $g_{DCW}\ L^{-1}$), over *ca.* 4–5 *h* of fermentation time. This was due to a lower temperature (30 °*C*) during induced fermentation versus a higher temperature during non-induced fermentation (37 °*C*), as well as diversion of glucose from cellular growth to lycopene production post-induction [48]. The switch in cellular metabolic regime from Run 2A to Run 2B was accompanied by an increased cellular stress due to intracellular lycopene accumulation and cytotoxicity, thus exacerbating the decrease in growth rate [47]. This was reflected in a minimal biomass growth during Run 2B, reaching *ca.* 2.5 $g_{VCC}\ L^{-1}$, over *ca.* 16 *h* of fermentation time (**Fig. 4c**). More details associated with this metabolic switch and lycopene production stress from the elucidated biokinetics modelling is expounded in *Induced Run 2B* segment of the section *Phenomenological and Metabolic Understanding Inferred from Biokinetics Models*. Also, the in-line viable cell concentration (VCC) readings showed large oscillations during Run 1 (with periodicity *ca.* 3 *h*) and was contrasted by minimal oscillatory behavior during Run 2. The stronger viability oscillations were attributed to the synchronous and asynchronous growth cycles observed during the non-induced exponential growth [49]. The smaller periodic swing observed in the viable cell population during induced fermentation could be attributed to the simultaneous production of intracellular lycopene along with much slower cell growth and maintenance.

Glucose (*S*) and acetate (*A*) concentrations in the broth were compared near the end of Run 1 (at 6.4 *h*) and around the metabolic switch time in Run 2A (*ca.* 6.6 *h*), were observed at *ca.* 9.0 $g_S\ L^{-1}$ and *ca.* 10.6 $g_S\ L^{-1}$, respectively. This indicated a slightly faster glucose consumption during Run 1 to support a higher biomass growth. Relatively more acetate was produced in

Run 1 (*ca.* 0.61 $g_A\ L^{-1}$) than in Run 2A (*ca.* 0.34 $g_A\ L^{-1}$) during the overlapped batch time. Lower concentrations of biomass viability concentration (X) and acetate in Run 2A compared to Run 1 indicated the diversion of glucose and metabolic resources towards lycopene production at the expense of biomass growth and overflow route to acetate production. Furthermore, for Run 2B that extended beyond Run 2A for another *ca.* 16 *h*, glucose concentration fell to *ca.* 4.66 $g_S\ L^{-1}$, along a drastic reduction of biomass viability into a relatively "stationary" phase, *ca.* 2.25–2.48 $g_{VCC}\ L^{-1}$, and continued accumulation of both intracellular lycopene and broth acetate, reaching *ca.* 49.15 $mg_P\ g_{VCC}^{-1}$ and *ca.* 0.94 $g_A\ L^{-1}$ respectively.

In-line Raman monitoring of lycopene production during Runs 2A and 2B displayed a clear sigmoidal increase in specific lycopene concentration over time. The turning point of this sigmoid corresponded to the metabolic regime transition mentioned previously. In Run 2A, after an initial t_{lag}, there was a rapid production of lycopene along a concave ascent with slow exponential viable biomass growth. In contrast, during Run 2B, the lycopene production rise slackened along a convex trend accompanied by aforesaid "stationary" biomass phase measured, suggesting that the entire cell population had reached the maximum limit of lycopene production constrained by fermentation conditions of 1 vvm, 300 rpm and no addition of glucose or other essential nutrients.

In the next section, further understanding of changes in the state variables X, S, A and P with batch time is revealed from calculated specific rates of biomass growth, glucose consumption, acetate accumulation and lycopene production derived from respective biokinetics models.

Phenomenological and Metabolic Understanding Inferred from Biokinetics Models

The optimized parameters and calculated specific rates in the aforesaid three biokinetics models, under induced and non-induced conditions, provide phenomenological insights of intrinsic biokinetics underlying the engineered *E. coli* 322 strain during batch fermentation. **Fig. 7** plots the calculated specific rates of biomass growth, glucose utilization, acetate accumulation and lycopene production. The top, middle and bottom rows in **Fig. 7** correspond to Runs 1, 2A and 2B, respectively. The left column corresponds to various specific rates of biomass growth and lycopene production, while the right column shows specific rates of glucose uptake or diversion and acetate related ones in the respective runs.

The metabolic significance of seven optimized model parameters alongside temporal state variable concentrations glucose, $S(t)$, acetate, $A(t)$, and lycopene, $P(t)$, were expressed as *dimensionless relative ratios*, to either a measured state variable or one of their calculated specific rates, over the respective batch times for the different runs. Most of these ratios reflected the intrinsic metabolic influences, that is inhibitory or affinity effects, of glucose and acetate concentrations towards the specific rates, $r_S(t)$, $r_S^{of}(t)$, $r_S^P(t)$, $r_A(t)$, $r_A^{in}(t)$ and $r_P(t)$. Collectively, these seven ratios are listed with their metabolic significance:

1. **Eq. 5**, $A(t)/K_i^A$: inhibitory effect of extracellular acetate on glucose uptake;
2. **Eq. 5**, $K_S/S(t)$: microbial affinity for glucose uptake;
3. **Eq. 7**, $K_{AS}/S(t)$: microbial affinity for overflow mechanism, *i.e.* glucose diversion from external glucose uptake towards acetate production;
4. **Eq. 8**, $K_{PS}/r_S(t)$: tendency of microorganism to divert glucose towards product formation;
5. **Eq. 11**, $S(t)/K_i^S$: inhibitory effect of glucose on reassimilation of extracellular acetate;
6. **Eq. 11**, $K_A/A(t)$: microbial affinity for reassimilation of extracellular acetate;
7. **Eq. 12d**, $P(t)/P_{crit}$: lycopene production transient trajectory and maximum limit.

Non-induced Run 1

The overall specific rate of biomass growth (μ) decreased monotonically from *ca.* 0.52–0.31 h^{-1} with some tailing towards the end. Over the course of Run 1, the contribution of direct oxidation of glucose less maintenance ($\mu^{ox} - \mu^m$) to overall biomass growth (μ) decreased from *ca.* 79% to 36%, while that of reassimilation of acetate (μ^A) increased from *ca.* 21% to 64%. These trends indicate biomass primarily grew by direct oxidation of glucose rather than by reassimilation of acetate as a secondary substrate during most of Run 1, until *ca.* 6.0 *h* when μ^A exceeded ($\mu^{ox} - \mu^m$). The parallel transient monotonic decrease in the overall substrate uptake rate $r_S(t)$, *ca.* 30%, was reflected as a proportional monotonic decrease in overall specific growth rate $\mu(t)$, *ca.* 39%. Moreover, from **Eq. 5**, based on the optimized parameter values for K_i^A and K_S for Run 1, the decrease in overall glucose uptake, $r_S(t)$, was due to an overpowering acetate inhibition effect $A(t)/K_i^A$ (ranges *ca.* 0.14–0.63) with increasing extracellular or broth acetate concentration. This was coupled with excess broth glucose indicated by the strong microbial affinity for glucose, as observed by the dimensionless ratio $K_S/S(t)$ ranging *ca.* $2 \times 10^{-3} - 5 \times 10^{-3}$.

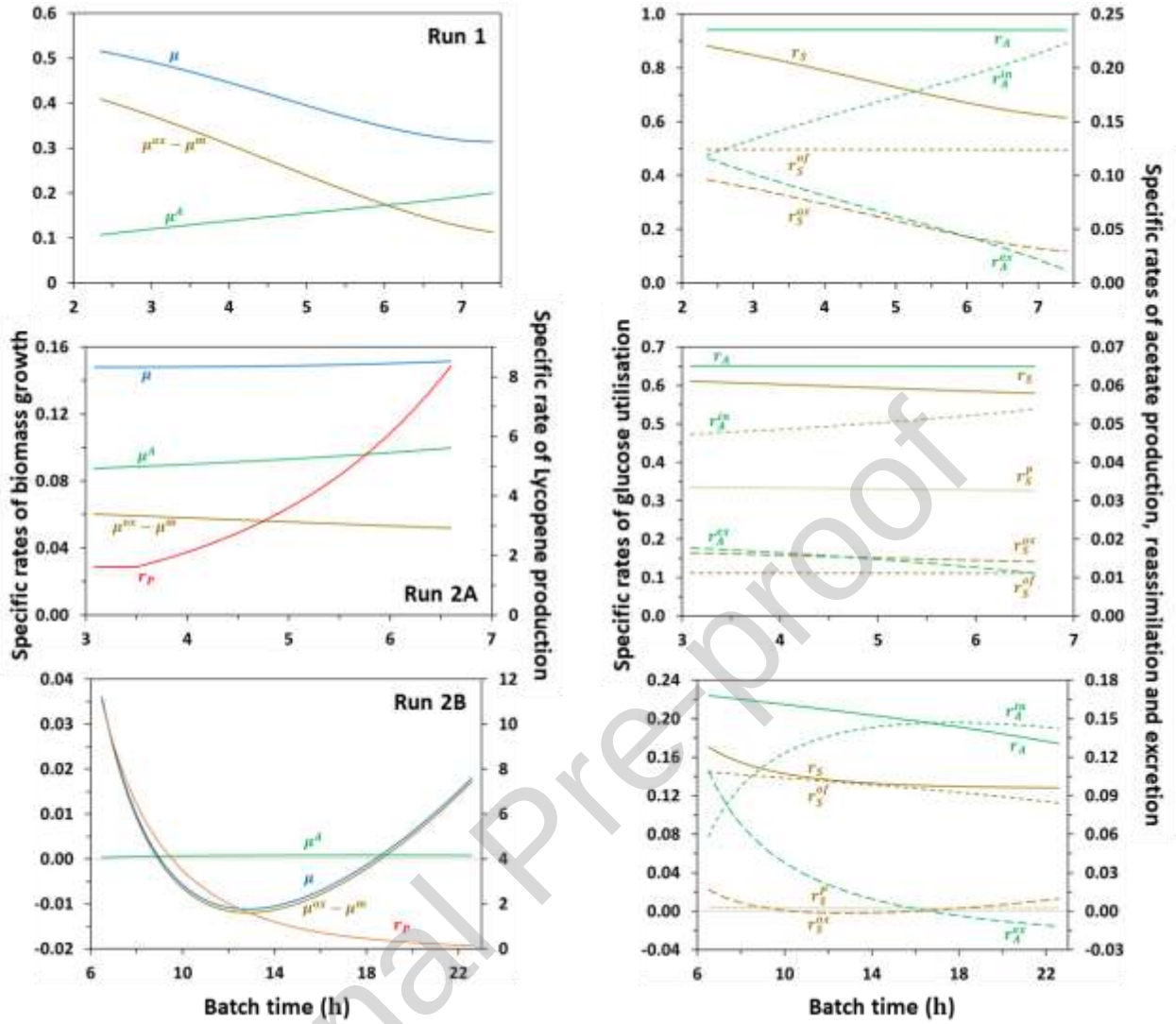

Figure 7 Calculated specific rates from biokinetics model. Left column: Simulated specific rates of biomass growth (h^{-1}), including biomass growth due to glucose consumption via direct oxidation, except cellular maintenance ($\mu^{ox} - \mu^{m}$); reassimilation of acetate (μ^{A}) and overall specific growth rate (μ); and lycopene production ($mg_P\ g_{VCC}^{-1}\ h^{-1}$). Right column: Simulated specific rates of glucose utilization ($g_S\ g_{VCC}^{-1}\ h^{-1}$), including overall glucose consumption (r_S), glucose consumption for acetate production via overflow route (r_S^{of}), direct oxidation of glucose into the TCA cycle (r_S^{ox}) and glucose redirected into lycopene production via MVA pathway (r_{SP}); and production (r_A), reassimilation (r_A^{in}) and excretion of acetate (r_A^{ex}) in $g_A\ g_{VCC}^{-1}\ h^{-1}$. Top Row: Non-induced Run 1. Middle Row: Induced Run 2A. Bottom Row: Induced Run 2B.

Details on how the aforesaid dimensionless relative ratios, specific rates and biokinetics models mathematically describe the phenomenological transient changes of state variables observed experimentally, and relating to intrinsic metabolic inferences, for both non-induced and induced cases, are expounded below.

As for the overflow mechanism in Run 1, its negligible K_{AS} value of 1.89 × 10^{-5} $g_S\,L^{-1}$ indicated excess broth glucose concentrations triggered the overflow metabolism, which is well-known in literature [12, 14]. Furthermore, as described by **Eq. 7**, a constant glucose rate into overflow mechanism, $r_S^{of}(t)$, *ca.* 0.50 $g_S\,g_{VCC}^{-1}\,h^{-1}$, was a consequence of a negligible acetate production affinity over time, as reflected by the values of $K_{AS}/S\,(t)$ below *ca.* 3.01 × 10^{-6}. This led to a constant rate of intracellular acetate production, $r_A(t)$, *ca.* 0.24 $g_A\,g_{VCC}^{-1}\,h^{-1}$ (**Eq. 10**, **Fig. 7** top row). This $r_A(t)$, though constant, comprised of transient rates, $r_A^{in}(t)$ and $r_A^{ex}(t)$, according to **Eq. 9a**. From **Eq. 11**, rate of acetate reassimilation, $r_A^{in}(t)$, increased due to a decrease in glucose inhibition effect $S(t)/K_i^S$ (ranges *ca.* 0.61–0.27), with decreasing glucose concentrations, coupled to an increase in microbial affinity for acetate reassimilation via the ratio $K_A/A(t)$ (ranges *ca.* 0.72–0.16). This rising $r_A^{in}(t)$ trend consequently decreases the rate of acetate excretion, $r_A^{ex}(t)$, as in **Eq. 9b** (**Fig. 5** top row). **Eq. 3** relates experimental HPLC values of broth acetate, $A(t)$, to $r_A^{ex}(t)$ and biomass, $X(t)$. Despite the decrease in $r_A^{ex}(t)$, broth acetate concentrations increased due to more significant biomass growth.

According to **Eqs. 13c** and **13d**, the specific growth rates, $\mu^{ox}(t)$and $\mu^A(t)$, were directly proportional to the specific glucose consumption rate via direct oxidation, $r_S^{ox}(t)$, and acetate reassimilation rate, $r_A^{in}(t)$, via their respective yield coefficients, Y_{XS}^{ox} and Y_{XA}. Consequently, in **Fig. 7** top row, the monotonic decrease of biomass growth in Run 1 resulting from direct oxidation of glucose, $(\mu^{ox}(t)-\mu^m)$ from *ca.* 0.41–0.11 h^{-1} mirrored the intracellular monotonic decrease of glucose consumption via direct oxidation, $r_S^{ox}(t)$, from *ca.* 0.38–0.12 $g_S\,g_{VCC}^{-1}\,h^{-1}$. Likewise, the monotonic increase in biomass growth due to acetate reassimilation, $\mu^A(t)$, from *ca.* 0.11–0.20 h^{-1}, paralleled the monotonic increase in extracellular acetate reassimilation rate, $r_A^{in}(t)$, from *ca.* 0.12–0.22 $g_A\,g_{VCC}^{-1}\,h^{-1}$.

Induced Run 2A

In contrast to Run 1, Run 2A was conducted at 30℃ post IPTG induction to produce intracellular lycopene. The overall specific rate of biomass growth, $\mu(t)$, in Run 2A was 2–3 times lower and remained relatively constant at *ca.* 0.15 h^{-1} (**Fig. 7** middle row). Moreover, partly due to the production of lycopene, $r_p(t)$ in Run 2A, contribution of direct oxidation of glucose less maintenance $(\mu^{ox}-\mu^m)$ to overall biomass growth (μ) was surprising lower than that from reassimilation of extracellular acetate (μ^A). Also, the percentage contribution of

($\mu^{ox} - \mu^{m}$) towards overall growth rate (μ) decreased gradually from *ca.* 41% to 34%, which was balanced out by μ^{A} rise from *ca.* 59% to 66%. These predictions showed that during the initial post-induction Run 2A, biomass growth was attributed by preferential reassimilation of acetate as a substrate, rather than directly from glucose. This was corroborated by $r_S^P(t)$ values in **Fig. 7** middle row, which accounted for *ca.* 55% of $r_S(t)$, indicating a large part of extracellular glucose uptake rate was diverted towards lycopene production (**Eq. 6b**).

Both the parameters K_i^A and K_S values were comparable in magnitude and value in Runs 1 and 2A (**Table S1** in *Supplementary Material*). However, due to IPTG influence, the acetate concentrations, $A(t)$, around similar batch times in these two runs were different (**Fig. 3a** and **3b**). In Run 1, $A(t)$ ranged from ca. 0.16–0.61 $g_A\ L^{-1}$, whereas in Run 2A, they ranged from *ca.* 0.24–0.34 $g_A\ L^{-1}$. In Run 2A, there was a muted acetate inhibition effect compared to Run 1, as the dimensionless ratio $A(t)/K_i^A$ in Run 2A was much lower at *ca.* 0.18–0.26. The values for the ratio $K_S/S(t)$ in Run 2A ranged *ca.* 0.01–0.02, indicating a 3-6 fold higher glucose affinity than in Run 1. In combination, these contributed towards Run 2A glucose uptake rate, $r_S(t)$, with a slight decline (*ca.* 0.60 $g_S\ g_{VCC}^{-1}\ h^{-1}$), which was notably lesser than the calculated $r_S(t)$ for Run 1 (ranges *ca.* 0.88–0.61 $g_S\ g_{VCC}^{-1}\ h^{-1}$). It is inferred for non-induced Run 1, its stronger biomass growth arose from the larger total of glucose uptake without IPTG influence.

The optimized value for parameter K_{AS} in Run 2A was *ca.* 0.06 $g_S\ L^{-1}$, which was *ca.* three orders of magnitude larger than that in Run 1 (**Table S1** in *Supplementary Material*). However, values for the ratio $K_{AS}/S(t)$ were less than *ca.* 5.8×10^{-3} during Run 2A. Similar to Run 1, this indicated a relatively high affinity for diverting glucose towards acetate production (**Eq. 7**), therefore triggering the experimentally observed overflow mechanism (**Fig. 3b**). Consequently, this led to a constant glucose consumption via the overflow route, $r_S^{of}(t)$, *ca.* 0.112 $g_S\ g_{VCC}^{-1}\ h^{-1}$, and subsequently, a constant rate of acetate production, $r_A(t)$, *ca.* 0.065 $g_A\ g_{VCC}^{-1}\ h^{-1}$ (**Eq. 10**). Compared to Run 1, $r_S^{of}(t)$ in Run 2A was *ca.* 4.4 times lower, resulting in *ca.* 3.6 times lower $r_A(t)$.

The optimized value of parameter K_i^S in Run 2A was *ca.* 40% lower than that in Run 1 (**Eq. 11**; **Fig. 7** middle row; **Table S1**), leading to a more significant glucose inhibition effect $S(t)/K_i^S$ (ranging *ca.* 0.90 – 0.70). On the other hand, parameter K_A in Run 2A was vanishingly small leading to remarkably low values of the ratio $K_A/A(t)$ (*ca.* 1.1×10^{-4} – 0.8×10^{-4}). This led to a marked increase in the microbial tendency to reassimilate acetate. These ratios

contributed to a more gradual increase in $r_A^{in}(t)$ in Run 2A at a lower magnitude (*ca.* 0.047–0.054 $g_A\ g_{VCC}^{-1}\ h^{-1}$), as compared to the corresponding sharper increase and higher values in Run 1 (*ca.* 0.12–0.23 $g_A\ g_{VCC}^{-1}\ h^{-1}$). Despite the much lower $r_A^{in}(t)$ in Run 2A, as previously mentioned, there was a major contribution of μ^A (*ca.* 0.09–0.10 h^{-1}) towards a smaller μ that was almost constant (*ca.* 0.15 h^{-1}). This was because firstly, the larger *ca.* 66% drop in Y_{XS}^{ox} in Run 1 (*ca.* 1.107 $g_{VCC}\ g_S^{-1}$) as compared to Run 2A (*ca.* 0.368 $g_{VCC}\ g_S^{-1}$), and the *ca.* 52% difference in Y_{XA} values of Run 1 (*ca.*0.897 $g_{VCC}\ g_A^{-1}$) to Run 2A (*ca.* 1.854 $g_{VCC}\ g_A^{-1}$); and secondly, a much lower and relatively constant r_S^{ox} value in Run 2A (*ca.* 0.15 $g_S\ g_{VCC}^{-1}\ h^{-1}$), due to lycopene production, in contrast to Run 1 (decreasing *ca.* 0.38–0.12 $g_S\ g_{VCC}^{-1}\ h^{-1}$). The low $r_A(t)$ in Run 2A produced less pronounced divergent trends of $r_A^{in}(t)$ and $r_A^{ex}(t)$ (**Eq. 9a**), which eventually led to the aforementioned lower $A(t)$ observed (**Eq. 3**).

As shown in **Fig. 4** middle row and according to **Eq. 6b**, $r_S^P(t)$ values accounted for *ca.* 55% of $r_S(t)$, which indicated that a large part of glucose was diverted towards lycopene production. Since the ratio of $K_{PS}/r_S(t)$ ranged *ca.* 0.92 – 0.97 in Run 2A, which was close to unity, this indicated some resistance of the engineered *E. coli* to divert glucose towards lycopene production (**Eq. 8**). Experimentally, post IPTG induction in Run 2A, there was a time lag (t_{lag}) before a slow initial sigmoidal rise in the lycopene concentration. This was most likely due to the switching on of *crt* genes within the *E. coli* 322 metabolism, which was mathematically captured by the resistance ratio $K_{PS}/r_S(t)$.

Induced Run 2B

The specific rate values of biomass growth, $\mu(t)$, its various contributors, $(\mu^{ox}(t) - \mu^m)$ and $\mu^A(t)$, and $r_S^{ox}(t)$ were significantly lower in magnitude in Run 2B than those in Run 2A and Run 1 (**Fig. 7**). Furthermore, the small negative values of these specific rates in Run 2B could be numerical artefacts from the optimization process. Since the in-line dielectric viable biomass ($g_{VCC}\ L^{-1}$) was relatively stationary in Run 2B (**Fig. 4c**), the transient trends of calculated μ, $(\mu^{ox} - \mu^m)$, μ^A and r_S^{ox} values did not hold much phenomenological inferences (**Eqs. 13c** and **13d**). However, it was noteworthy that the calculated specific rate of maintenance, μ^m, in Run 1 (*ca.* 165.93 × 10^{-4} h^{-1}) was significantly larger than its corresponding value for Run 2A (*ca.* 0.87 × 10^{-4} h^{-1}) and larger than that in Run 2B (*ca.* 81.62 × 10^{-4} h^{-1}). This could suggest higher maintenance requirement during non-induced fermentation. In Run 2B, the other specific rates of glucose utilization and acetate-related changes had phenomenological significance since they had minimal negative values, and were of comparable magnitude with those in Run 2A.

Referring to **Fig. 3**, glucose $S(t)$ was consumed at two distinct rates in Runs 2B (*ca.* 10–5 $g_S\ L^{-1}$) and 2A (*ca.* 13–10 $g_S\ L^{-1}$). Further, the overall glucose consumption rate, r_S, (*ca.* 0.17–0.13 $g_S\ g_{VCC}^{-1}\ h^{-1}$) in Run 2B was *ca.* 4.3 times smaller than that in Run 2A, *ca.* 0.61–0.58 $g_S\ g_{VCC}^{-1}\ h^{-1}$. Although the acetate inhibition effect, $A(t)/K_i^A$ (*ca.* 0.13–0.36), in Run 2B was comparable to that in Run 2A (*ca.* 0.18–0.26), a lower affinity for glucose, $K_S/S(t)$ (*ca.* 0.10–0.23), in Run 2B decreased the overall glucose uptake, r_S (**Eq. 5**). Despite this difference in $r_S(t)$, the rate of glucose diversion towards the overflow route, $r_S^{of}(t)$, was similar for both regimes (**Fig. 7**), viz. Run 2B (*ca.* 0.14–0.11 $g_S\ g_{VCC}^{-1}\ h^{-1}$) and Run 2A (constant at *ca.* 0.11 $g_S\ g_{VCC}^{-1}\ h^{-1}$). The mathematical insight to understanding the similarity in $r_S^{of}(t)$ in both Runs 2A and 2B regimes is given by **Eq. 7**. The two parameters K_{AS} and r_{AS}^{max} were key contributors in diverting a part of the glucose consumed, $r_S(t)$, towards the overflow route, $r_S^{of}(t)$. The affinity ratio $K_{AS}/S(t)$ ranged *ca.* 0.46–1.03 in Run 2B, whereas in 2A, the ratio was negligibly small and less than *ca.* 5.8×10^{-3}. This indicated a marked affinity towards glucose diversion for Run 2A, in contrast to a moderated one in Run 2B. Although the optimized parameter r_{AS}^{max} was *ca.* 1.88 times higher in Run 2B than in Run 2A (**Table S1** in *Supplementary Material*), the lower glucose affinity in Run 2B cancelled out the potential maximum diversion of glucose towards overflow mechanism, resulting in similar rates of glucose overflow, $r_S^{of}(t)$, in both the regimes.

However, the yield of acetate produced via overflow route, Y_{AS}, was different in Runs 2A and 2B, viz. *ca.* 0.58 and 1.16 $g_A\ g_S^{-1}$, respectively. The conversion of glucose into acetate from overflow mechanism was more efficient in Run 2B than in Run 2A, leading to larger $r_A(t)$ values in 2B (**Fig. 7**). Interestingly, although Y_{AS} for Run 1 was only 0.47 $g_A\ g_S^{-1}$, since there was no lycopene production, $r_S^{of}(t)$ was *ca.* four times higher, leading to a higher $r_A(t)$ in the non-induced case. Furthermore, $r_A(t)$ in Run 2B decreases linearly *ca.* 0.17 to 0.13 $g_A\ g_{VCC}^{-1}\ h^{-1}$, in contrast to relatively constant $r_A(t)$ in both Runs 1 and 2A. The observed curvature in $r_A^{in}(t)$ for Run 2B could be attributed to the lower glucose concentrations, $S(t)$, and a high K_S^i parameter that led to lower inhibition $S(t)/K_S^i$, *ca.* 0.14–0.06, as compared to Runs 1 and 2A. Also, $K_A/A(t)$ range, *ca.* 5.5–2.0, in Run 2B displays lowest affinity for acetate reassimilation amongst the three regimes studied. For $r_A^{ex}(t)$ in Run 2B, its decreasing profile mirrors the increase in $r_A^{in}(t)$ (**Eq. 9a**; **Fig. 7** bottom row).

The biokinetics models for lycopene production assume carbon flux solely from glucose diverted, $r_S^P(t)$, towards MVA pathway (**Eq. 8**). The ratio $K_{PS}/r_S(t)$ ranged *ca.* 20.5–27.3 in Run 2B, which is exceeding 20 times more resistance to glucose diversion compared to Run 2A, together with the dramatic drop of transient lycopene productivity curve in **Fig. 3** for Run 2B, corresponded to the cytotoxicity of intracellular lycopene accumulation that transpired. Moreover, from **Eq. 6b** and **Fig. 7** bottom row, in Run 2B, $r_S^P(t)$ values accounted for only *ca.* 2.4% of glucose uptake rate, $r_S(t)$. This explains the reason behind the success of using $r_S(t)$ instead of $S(t)$ during our initial model trials for **Eq. 8**, which captures the phenomenological minute metabolic flux through the secondary MVA pathway that is dependent on $r_S(t)$ rather than glucose concentration in the broth. Interestingly, during initial optimization trials for Run 2A, where **Eq. 8** was formulated with $S(t)$ instead of $r_S(t)$, the optimization converged with low RMSE, but was not successful for Run 2B. This is because $r_S^P(t)$ accounts for a significant portion (*ca.* 55%) of $r_S(t)$ in Run 2A.

The time profiles of $r_P(t)$ in Runs 2A and 2B corroborate with the transient intracellular lycopene productivity observed (**Fig. 3**). Two other important numbers relating to lycopene production are the critical lycopene concentration, P_{crit}, and regime specific intracellular lycopene productivity, $\propto_{\boldsymbol{PS}}$. For Run 2B, its P_{crit} value *ca.* 48.95 $mg_P\ g_{VCC}^{-1}$, indicated the theoretical maximum lycopene concentration achievable under the batch fermentation conditions of constant aeration, *ca.* 1 vvm, and agitation, *ca.* 300 rpm, without further addition of glucose and other nutrients. This is very close to the final experimental lycopene concentration of 49.0 $mg_P\ g_{VCC}^{-1}$ (**Fig. 3**). The estimated productivities, $\propto_{PS}$, in Runs 2A and 2B were 1.35 and 323 $mg_P\ g_{VCC}^{-1}\ h^{-1}$, respectively. The larger productivity in Run 2B arose from relatively minimal biomass growth, whereas there is a slight exponential growth in Run 2A (**Figs. 4b** and **4c**).

CONCLUSION

The advantages of applying the QbD / PAT framework, in-line PAT analytics and biokinetics modelling to the engineered *E. coli* 322 strain were demonstrated in this work. First, higher time resolution data from in-line process monitoring was achieved for state variables *X*, *S*, and *P*. Specifically, in-line dielectric spectroscopy revealed real-time cell population viability concentration or biomass (*X*), in-line Raman spectroscopy and subsequent multivariate chemometrics elucidated concentrations of broth glucose (*S*) and specific intracellular lycopene

(*P*). Additionally, in-line PAT data provided assured data back-up when off-line data were unavailable, due to sampling or analytical instrument issues (*e.g.* HPLC).

Second, the PAT data abundance corroborated well with off-line measurements (**Figs. 3** and **4**). Together with off-line HPLC broth acetate concentrations (*A*), the PAT data provided abundant experimental datapoints to overcome the underdetermined scenarios frequently encountered in nonlinear optimization of parameters in biokinetics models (**Eqs. 1**–**16**). Further, the three optimized biokinetics models of non-induced Run 1, and induced Runs 2A and 2B yielded excellent fit (low RMSE values) to the abundant experimental data, thus providing confidence towards the metabolic insights derived from them. Computed specific rates of cell growth, cellular metabolism of glucose uptake and diversion, acetate formation, reassimilation and excretion, intracellular lycopene production (**Fig. 7**) were derived from the optimized model parameter values (**Table S1** in *Supplementary Material*). All these fitted like pieces of an "abstract" jigsaw puzzle to mathematically explain macroscopic time-dependent changes in *E. coli* culture viability and growth alongside overflow mechanism, under the impact of intracellular product formation causing a switch in carbon flux metabolic regime.

Third, metabolic insights from model parameter estimates were harnessed through combining 1) phenomenological observations of experimentally measured state variables, 2) biokinetics modelling with appropriate model formulation, and 3) suitable non-linear optimization method. The time-dependent rates of intra- or extra-cellular glucose and acetate fluxes were distinctly related to the specific fermentation conditions, *i.e.*, for non-induced case or the two-regime induced scenario producing lycopene.

A highlight was the use of seven dimensionless ratios relating to transient changes in state variables, *S*, *A*, and *P*, *i.e.* $A(t)/K_i^A$, $K_S/S(t)$, $K_{AS}/S(t)$, $K_{PS}/r_S(t)$, $S(t)/K_i^S$, $K_A/A(t)$, and $P(t)/P_{crit}$. Their metabolic significance on affinity, inhibitory or limiting effects to key Monod-type kinetics equations (**Eqs. 5**–**13**), together with comparison of their calculated transient values, provided a deeper understanding into the complex intertwined metabolism of *E. coli* 322 strain. For example, low values of K_{AS} and $K_{AS}/S(t)$ below *ca.* 10^{-5} reflected a strong overflow mechanism, the dimensionless ratio $K_S/S(t)$ *ca.* 10^{-3} indicated strong microbial affinity for glucose, or changes in ratios $A(t)/K_i^A$ and $S(t)/K_i^S$ across batch-time regimes signified metabolic changes in relative effect of acetate and glucose inhibition on glucose uptake and acetate reassimilation respectively. Moreover, for lycopene production, the extent of cytotoxicity was reflected by the ratio $K_{PS}/r_S(t)$; wherein values close to unity (*ca.*

1.0) indicated rapid lycopene production with a slow exponential culture growth (Run 2A), in contrast to high ratios greater than *ca.* 20 leading to stagnant biomass and slowing down of lycopene productivity towards its limiting concentration predicted by P_{crit} *ca.* 48.95 $mg_P\ g_{VCC}^{-1}$ (Run 2B).

Another new finding was the dependency of the overall cell growth rate (μ), switching from direct glucose via oxidation path for non-induced (μ^{ox}) to the acetate reassimilation route (μ^{A}) during the initial post-induction regime Run 2A. The latter can be alluded to the diversion of intracellular metabolites from glucose oxidation metabolism into lycopene production $r_S^P(t)$, causing the need to utilize secondary carbon source acetate to sustain a slower biomass growth.

Evidently, this work has demonstrated the usefulness of the estimated values of model parameters, various calculated rates of change associated with state variables, and their seven dimensionless ratios alongside PAT methodology. With this approach, lab-bench bioreactor studies at varied conditions of bespoke engineered microbial strain can be compared for understanding metabolic aspects. Useful classification of strain phenotypical metabolic patterns and biokinetics could also be derived. In so doing, this aids toward *in silico* design for fermentation conditions testing, scaling-up, or troubleshooting batch-to-batch variations.

AUTHOR CONTRIBUTIONS

Vasudevan, V. – Biokinetics mathematical formulation, numerical optimisation, manuscript writing and revision.
Fitriani, N.E. – *E. coli* culture, in-line Raman and dielectric spectroscopies implementation, manuscript revision.
Leung, D. – 10L bioreactor experimental set-up, planning, and execution, manuscript revision.
Chew, W. – Lead investigator, biokinetics mathematical formulation, bioreactors experimental design, set-up, and execution, PAT analytics and multivariate chemometrics, manuscript writing and revision.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DECLARATION OF COMPETING INTEREST

The authors declare no financial or personal conflict of interests that negatively impact the results and outcome reported herein.

ACKNOWLEDGEMENTS

This work was supported by the A*STAR grant IAF-PP: BioTrans Phase 3 (H20H6a0028) secured by Dr. Nicholas David Lindley, who is very supportive of our research with PAT analytics and biokinetics modelling. The authors also thank Ms Sudha Shukal, Dr. Zhang Congqiang and Dr. Chen Xixian, for their growth media protocol and use of *E. coli* 322 strain.

REFERENCES

[1] S. Öztürk, P. Çalik, T.H. Özdamar, Fed-Batch Biomolecule Production by Bacillus subtilis: A State of the Art Review, Trends Biotechnol. 34 (2016) 329–345. https://doi.org/10.1016/j.tibtech.2015.12.008.

[2] L. Sanchez-Garcia, L. Martín, R. Mangues, N. Ferrer-Miralles, E. Vázquez, A. Villaverde, Recombinant pharmaceuticals from microbial cells: A 2015 update, Microb. Cell Factories 15 (2016) 1–7. https://doi.org/10.1186/s12934-016-0437-3.

[3] S. Pontrelli, T.Y. Chiu, E.I. Lan, F.Y.H. Chen, P. Chang, J.C. Liao, Escherichia coli as a host for metabolic engineering, Metab. Eng. 50 (2018) 16–46. https://doi.org/10.1016/j.ymben.2018.04.008.

[4] G. Stephanopoulos, H. Alper, J. Moxley, Exploiting biological complexity for strain improvement through systems biology, Nat. Biotechnol. 22 (2004) 1261–1267. https://doi.org/10.1038/nbt1016.

[5] A. Herzog, U. Siler, V. Spitzer, N. Seifert, A. Denelavas, P.B. Hunziker, W. Hunziker, R. Goralczyk, K. Wertz, Lycopene reduced gene expression of steroid targets and inflammatory markers in normal rat prostate, FASEB J. 19 (2005) 1–24. https://doi.org/10.1096/fj.04-1905fje.

[6] F. Zhu, L. Lu, S. Fu, X. Zhong, M. Hu, Z. Deng, T. Liu, Targeted engineering and scale up of lycopene overproduction in Escherichia coli, Process Biochem. 50 (2015) 341–346. https://doi.org/10.1016/j.procbio.2014.12.008.

[7] C. Zhang, V.Y. Seow, X. Chen, H.P. Too, Multidimensional heuristic process for high-yield production of astaxanthin and fragrance molecules in Escherichia coli, Nat. Commun. 9 (2018). https://doi.org/10.1038/s41467-018-04211-x.

[8] B. Shi, T. Ma, Z. Ye, X. Li, Y. Huang, Z. Zhou, Y. Ding, Z. Deng, T. Liu, Systematic Metabolic Engineering of Saccharomyces cerevisiae for Lycopene Overproduction, J. Agric. Food Chem. 67 (2019) 11148–11157. https://doi.org/10.1021/acs.jafc.9b04519.

[9] Z. Wang, J. Sun, Q. Yang, J. Yang, Metabolic engineering escherichia coli for the production of lycopene, Molecules. 25 (2020) 3136. https://doi.org/10.3390/molecules25143136.

[10] J. Yang, G. Zhao, Y. Sun, Y. Zheng, X. Jiang, W. Liu, M. Xian, Bio-isoprene production using exogenous MVA pathway and isoprene synthase in Escherichia coli, Bioresour. Technol. 104 (2012) 642–647. https://doi.org/10.1016/j.biortech.2011.10.042.

[11] B. Sonnleitner, O. Käppeli, Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis, Biotechnol. Bioeng. 28 (1986) 927–937. https://doi.org/10.1002/bit.260280620.

[12] 44. D. Molenaar, R. van Berlo, D. de Ridder, B.Teusink, Shifts in growth strategies reflect tradeoffs in cellular economics, Mol. Syst. Biol. 5 (2009) 323. https://doi.org/10.1038/msb.2009.82

[13] J.P. van Dijken, R.A. Weusthuis, J.T. Pronk, Kinetics of growth and sugar consumption in yeasts. Antonie van Leeuwenhoek 63 (1993) 343. https://doi.org/10.1007/BF00871229

[14] 43. H.T. Pham, G. Larsson, S.O. Enfors, Growth and energy metabolism in aerobic fed-batch cultures of Saccharomyces cerevisiae: simulation and model verification, Biotechnol. Bioeng. 60 (1998) 474–482. https://doi.org/10.1002/(sici)1097-0290(19981120)60:4<474::aid-bit9>3.0.co;2-j

[15] G.N. Vemuri, E. Altman, D.P. Sangurdekar, A.B. Khodursky, M.A. Eiteman, Overflow metabolism in Escherichia coli during steady-state growth: transcriptional regulation and effect of the redox ratio. Appl. Environ. Microbiol. 72 (2006) 3653. https://doi.org/10.1128/AEM.72.5.3653-3661.2006

[16] A.L. Sonenshein, Control of key metabolic intersections in Bacillus subtilis. Nat. Rev. Microbiol. 5 (2007) 917. https://doi.org/10.1038/nrmicro1772

[17] W.L. McKeehan, Glycolysis, glutaminolysis and cell proliferation. Cell Biol. Int. Rep. 6 (1982) 635. https://doi.org/10.1016/0309-1651(82)90125-4

[18] P.P. Hsu, D.M. Sabatini, Cancer cell metabolism: Warburg and beyond. Cell. 134 (2008) 703. https://doi.org/10.1016/j.cell.2008.08.021

[19] D. Machado, R.S. Costa, E.C. Ferreira, I. Rocha, B. Tidor, Exploring the gap between dynamic and constraint-based models of metabolism, Metab. Eng. 14 (2012) 112–119. https://doi.org/10.1016/j.ymben.2012.01.003.

[20] J.K. Khanijou, H. Kulyk, C. Bergès, L.W. Khoo, P. Ng, H.C. Yeo, M. Helmy, F. Bellvert, W. Chew, K. Selvarajoo, Metabolomics and modelling approaches for systems metabolic engineering, Metab. Eng. Commun. 15 (2022) e00209. https://doi.org/10.1016/j.mec.2022.e00209.

[21] W.D. Jang, G.B. Kim, Y. Kim, S.Y. Lee, Applications of artificial intelligence to enzyme and pathway design for metabolic engineering, Curr. Opin. Biotechnol. 73 (2022) 101–107. https://doi.org/10.1016/j.copbio.2021.07.024.

[22] L. Lema-Perez, R. Muñoz-Tamayo, J. Garcia-Tirado, H. Alvarez, On parameter interpretability of phenomenological-based semiphysical models in biology, Inform. Med. Unlocked 15 (2019) 100158. https://doi.org/10.1016/j.imu.2019.02.002.

[23] E. Aguirre-Zapata, H. Alvarez, C.V. Dagatti, F. di Sciascio, A.N. Amicarelli, Parametric interpretability of growth kinetics equations in a process model for the life cycle of Lobesia botrana, Ecol. Modell. 482 (2023) 110407. https://doi.org/10.1016/j.ecolmodel.2023.110407.

[24] D.C. López, T. Barz, M. Peñuela, A. Villegas, S. Ochoa, G. Wozny, Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production, Biotechnol. Prog. 29 (2013) 1064–1082. https://doi.org/10.1002/btpr.1753.

[25] N.E. Fitriani, V. Vasudevan, D. Leung, Y.X. Wong, W. Wu, W. Chew, In-line dielectric and Raman monitoring of intracellular lycopene production in E. coli fermentation, Submitted to Biotechnol. Bioeng.

[26] W. Chew, P. Sharratt, Trends in process analytical technology, Anal. Methods. 2 (2010) 1412. https://doi.org/10.1039/c0ay00257g.

[27] Food and Drug Administration, Pharmaceutical cGMPs for the 21st century - a risk-based approach final report, 2004.

[28] M. Molony, C. Undey, Pat Tools for Biologics: Considerations and Challenges, Quality by Design for Biopharmaceuticals: Principles and Case Studies. (2008) 211–253. https://doi.org/10.1002/9780470466315.ch12.

[29] S.T. Tan, D. Wang, I. Halim, S.K. Teoh, P. Sharratt, G. Loh, R.L. Wong, S.M.C. Yee, C.Y. Loke, W. Chew, Integrated Platform at ICES Kilo-Lab for Process Quality by Design, Comput. Aided Chem. Eng. 31 (2012) 595–599. https://doi.org/10.1016/B978-0-444-59507-2.50111-6.

[30] A.S. Rathore, QbD/PAT for bioprocessing: moving from theory to implementation, Curr. Opin. Chem. Eng. 6 (2014) 1–8. https://doi.org/10.1016/j.coche.2014.05.006.

[31] M. Assirelli, W. Xu, W. Chew, Reactor Kinetics Studies via Process Raman Spectroscopy, Multivariate Chemometrics, and Kinetics Modeling, Org. Proc. Res. Dev. 15, (2011) 610–621. https://doi.org/10.1021/op100337v.

[32] W. Chew, Information-theoretic chemometric analyses of Raman data for chemical reaction studies, J. Raman Spectrosc., 42, (2011) 36–47. https://doi.org/10.1002/jrs.2660.

[33] W. Rusli, P.K. Naraharisetti, W. Chew, Calibration methods to circumvent unknown component spectra for quantitative in situ Raman monitoring of co-polymerization reactions, React. Chem. Eng., 6, (2021) 538–547. https://doi.org/10.1039/d0re00424c.

[34] V. Vivek, F.N. Eka, W. Chew, Mixing studies in an unbaffled bioreactor using a computational model corroborated with in-situ Raman and imaging analyses, Chem. Eng. J. Adv. 9 (2022) 100232. https://doi.org/10.1016/j.ceja.2021.100232.

[35] P. Ng, L.W. Khoo, A. Thong, W. Chew, Optimization of extraction conditions for LC-ToF-MS analysis of mevalonate pathway metabolites in engineered E. coli strain via statistical experimental designs, Talanta 254 (2023) 124182. https://doi.org/10.1016/j.talanta.2022.124182.

[36] P. Neubauer, H.Y. Lin, B. Mathiszik, B, Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in Escherichia coli. Biotechnol. Bioeng. 83 (2003) 53–64. https://doi.org/10.1002/bit.10645

[37] H.Y Lin, B. Mathiszik, B. Xu, S.O. Enfors, P. Neubauer, Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli. Biotechnol. Bioeng. 73 (2001) 347–357. https://doi.org/10.1002/bit.1068

[38] B. Xu, M. Jahic, S.O. Enfors, Modeling of overflow metabolism in batch and fed-batch cultures of Escherichia coli. Biotechnol. Prog. 15 (1999) 81–90. https://doi.org/10.1021/bp9801087

[39] E. Anane, C. Diana, C. López, P. Neubauer, M.N. Cruz-Bournazou, Modelling overflow metabolism in Escherichia coli by acetate cycling, Biochem. Eng. J. 125 (2017) 23-30. https://doi.org/10.1016/j.bej.2017.05.013.

[40] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J.R. Williamson, T. Hwa, Overflow metabolism in Escherichia coli results from efficient proteome allocation. Nature 528 (2015) 99–104. https://doi.org/10.1038/nature15765

[41] K. Peebo, K. Valgepea, A. Maser, R. Nahku, K. Adamberg, R. Vilu, Proteome reallocation in Escherichia coli with increasing specific growth rate. Mol. Biosyst. 11 (2015) 1184–1193. https://doi.org/10.1039/c4mb00721b

[42] K. Valgepea, K. Adamberg, R. Nahku, P.J. Lahtvee, L., Arike, R. Vilu, Systems Biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of Acetyl-CoA synthetase. BMC Syst. Biol. 4 (2010) 166. https://doi.org/10.1186/1752-0509-4-166

[43] V. Bernal, S. Castaño-Cerezo, M. Cánovas, Acetate Metabolism regulation in Escherichia coli: carbon overflow, pathogenicity, and beyond. Appl. Microbiol. Biotechnol. 100 (2016) 8985–9001. https://doi.org/10.1007/s00253-016-7832-x

[44] E. Anane, A. Sawatzki, P. Neubauer, M.N. Cruz-Bournazou, Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments, J. Chem. Technol. Biotechnol. 94 (2019) 516-526. https://doi.org/10.1002/jctb.5798.

[45] J. Shiloach, R. Fass, Growing E. coli to high cell density - a historical perspective on method development, Biotechnol. Adv. 23 (2005) 345–357. https://doi.org/10.1016/j.biotechadv.2005.04.004

[46] A.J.J. Straathof, Modelling of end-product inhibition in fermentation, Biochem. Eng. J. 191 (2023) 108796. https://doi.org/10.1016/j.bej.2022.108796.

[47] A.M. Miguez, M.P. McNerney M.P. Styczynski, Metabolomics Analysis of the Toxic Effects of the Production of Lycopene and Its Precursors, Front. Microbiol. 9 (2018) 760. https://doi.org/10.3389/fmicb.2018.00760.

[48] C. Zhang, X. Chen, N.D. Lindley, H.-P. Too, A “plug-n-play” modular metabolic system for the production of apocarotenoids, Biotechnol. Bioeng. 115 (2018) 174−183. https://doi.org/10.1002/bit.26462.

[49] K. Asami, Characterization of biological cells by dielectric spectroscopy, J. Non Cryst. Solids 305 (2002) 268–277. https://doi.org/10.1016/S0022-3093(02)01110-9

Highlights

- Apply PAT methodology with in-line monitoring and mathematical modelling to E. coli strain culture overproducing lycopene.
- Biokinetics modelling using differential-algebraic equations (DAE) with Monod-type conversions described state variables.
- Optimised model parameters and 7 dimensionless ratios reflected time-dependent changes in rates or concentrations.
- Calculated change in rates of fluxes and concentrations revealed salient underlying metabolic shifts and transient trends.
- Models explain cell growth, carbon source use, overflow mechanism, affinity or inhibitory effects, lycopene cytotoxicity.