

25 within 10% of the measured data, and compared to a standard dFBA model, we found the
26 framework showed a 90% and 72% improvement in cell density and antibody titer prediction,
27 respectively. Thus, we demonstrate our hybrid modeling framework effectively captures cellular
28 metabolism and expands the applicability of dFBA to model the dynamic conditions in a
29 bioreactor.

30 Keywords: Bioprocess modeling; Machine learning; Metabolic models; Dynamic flux balance
31 analysis

32

33 **1. Introduction**

34 Maximizing recombinant protein titer in a pharmaceutical bioprocess can be facilitated by
35 optimizing nutrient feeding. Optimal conditions are commonly identified using time and
36 resource-intensive design of experiments (DOE) strategies (Kasemiire et al., 2021). Models built
37 on process data can help predict the trajectory of cellular states and control the process
38 environment (Sidoli et al., 2004). Predictive models have previously leveraged empirical Monod-
39 based equations to compute growth rates based on the extracellular concentrations of limiting
40 nutrients (Ben Yahia et al., 2021; Galleguillos et al., 2017). The uptake and secretion rates for
41 non-limiting nutrients are described by their relative uptake/secretion rates and/or kinetic rate
42 laws defined by concentrations (López-Meza et al., 2016). However, nutrient depletion and toxic
43 metabolite accumulation leads to metabolic shifts that cause uptake and secretion rates relative to
44 the limiting nutrient to change during the bioprocess (Sunley et al., 2008; Templeton et al.,
45 2013). This limitation motivates the inclusion of descriptive and mechanistic models of cellular
46 metabolism in dynamic bioreactor models.

47 Genome-scale metabolic models are comprehensive collections of all metabolic pathways for an
48 organism and are valuable for predicting product yields when nutrient uptake rates are specified.
49 Metabolic flux through the entire network can be predicted using constraint-based modeling,
50 such as flux balance analysis (Orth et al., 2010), which assumes that resource allocation in a cell
51 aims to fulfill specific cellular objectives. This capability is leveraged for dynamic flux balance
52 analysis (dFBA) (Mahadevan et al., 2002) and uses bioreactor substrate concentrations to
53 determine nutrient uptake by the metabolic model. Fluxes are then predicted with the metabolic
54 model to update metabolite concentrations in the bioreactor. Overall, this framework embeds the
55 FBA problem within a system of ODEs to predict metabolic and cellular dynamics in the reactor.

56 While dFBA is structurally simple, it has three disadvantages that limit its application to
57 mammalian bioprocessing. First, cellular metabolism is dynamic and therefore, the metabolic
58 model must be tailored to be consistent with the extracellular environment. Otherwise, the full
59 genome-scale model over-predicts intracellular fluxes as it affords the use of conditionally
60 inactivated pathways (Jerby et al., 2010). Second, changes in extracellular environments cause
61 cells to change the abundance of transporter proteins, which further changes kinetic parameters
62 governing nutrient uptake rates (Laakso et al., 2011). Third, cells exhibit metabolic shifts arising
63 from metabolite accumulation, such as lactate, wherein lactate production switches to lactate
64 consumption during the bioprocess (Torres et al., 2018). This is frequently seen in fed-batch
65 cultures with CHO cells and must be conditionally integrated into existing bioprocess models
66 (Nolan and Lee, 2011).

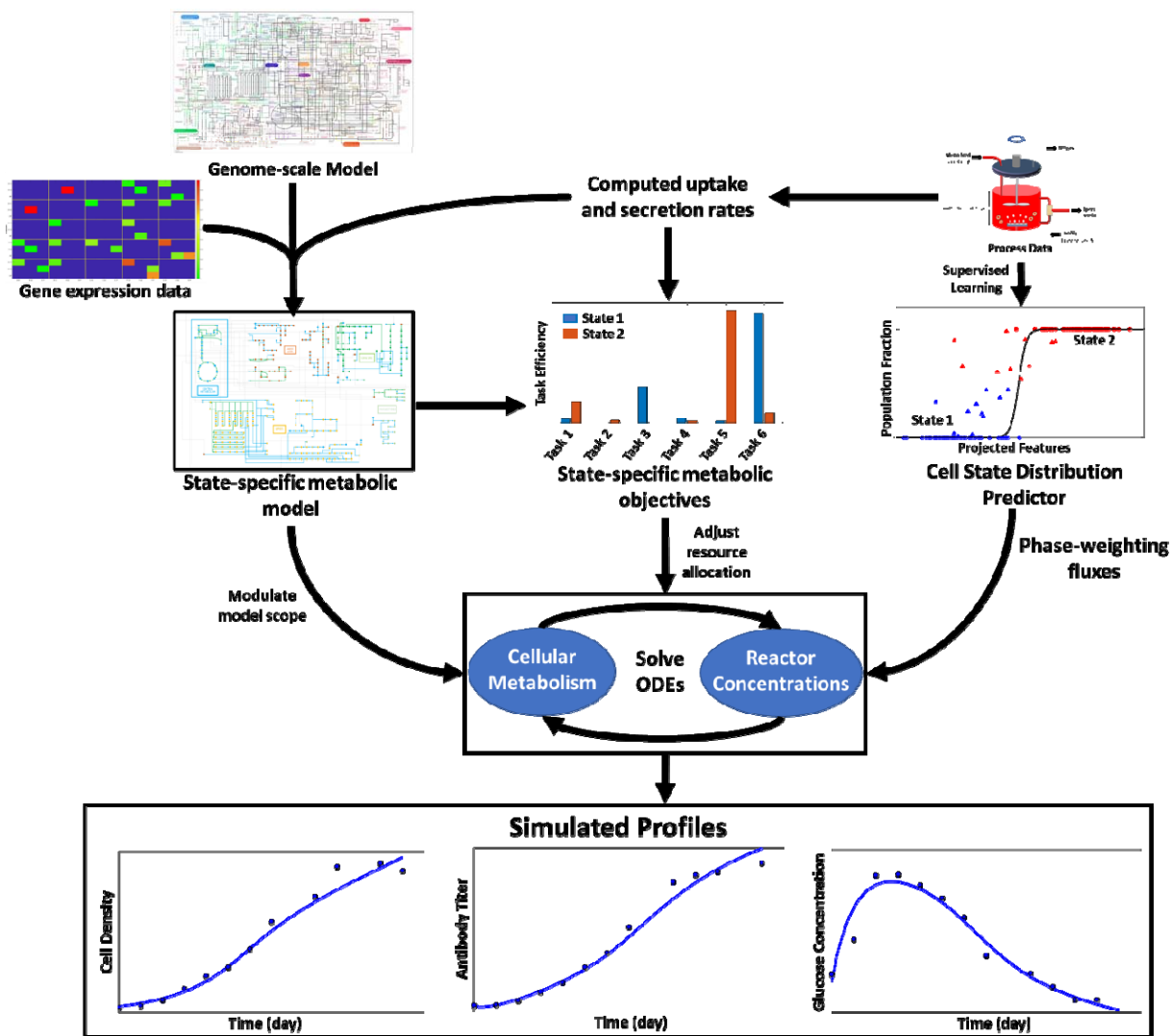
67 Capturing metabolic shifts requires us to first characterize them. Some algorithms rely on visual
68 inspection (Dean and Reddy, 2013) or piecewise linear regression (Ben Yahia et al., 2017) to
69 identify different process phases. However, these methods suffer from the drawback that the
70 model may reflect a single dataset or growth condition. Thus, they may not generalize to other
71 conditions prevalent in the bioreactor or states of a bioprocess. Finally, predicting product fluxes
72 requires us to know a cell's objectives for a given cellular state. Objective functions, such as
73 growth rate maximization, can be reliably applied to quantify metabolism in prokaryotes;
74 however, these objectives have limited relevance to mammalian cells, since they only partially
75 characterize the growth phase (Savinell and Palsson, 1992). To model the non-growing states,
76 alternative objective functions must be explored (Garcia Sanchez and Torres Saez, 2014). More
77 recently, parsimonious nutrient uptake was proposed as an objective (Chen et al., 2019), but it
78 does not capture the variation in amino acid allocation towards different recombinant proteins.

79 Therefore, there is a need for a comprehensive framework that correctly and models the
80 biological characteristics of the cells in the bioreactor with high fidelity by addressing the
81 changes in cell states arising from constantly changing conditions in an industrial bioprocess.

82 Here we present Cellular Objectives and State Modulation In bioreaCtors (COSMIC)-dFBA, a
83 multi-scale modeling framework for predicting concentration profiles of glucose, metabolic
84 byproducts, antibody, amino acids, and cell density in a perfusion bioprocess (Figure 1). As with
85 standard dFBA, COSMIC-dFBA predicts concentration profiles of metabolites by solving a
86 system of ODE equations in which the uptake rates of metabolites are determined by kinetic rate
87 laws and product secretion rates are predicted by the metabolic model. To compute fluxes using
88 the metabolic model, COSMIC-dFBA first determines the number of metabolic states by
89 inspecting uptake and production fluxes between various sampling intervals. Using these data,
90 we then compute the fraction of cells in each phase, which provides a measure of state shift. We
91 then identify the metabolites that show a significant difference in concentration between the
92 identified states and train the cell state distribution predictor, a statistical model to predict state
93 shift based on the prevailing bioreactor conditions. Using uptake and secretion rates inferred
94 from spent media analysis, we then generate a priority list for metabolic tasks to determine the
95 order of resource allocation of various cellular objectives for each identified state. A
96 parameterized kinetic rate law is used to constrain nutrient uptake in each identified state. This
97 information is then used to solve the metabolic model and predict the net uptake and secretion
98 rates of all tracked metabolites. This framework accurately predicts concentration profiles and
99 antibody titers in a diverse range of bioreactor conditions including glucose, amino acid, and
100 oxygen depleted media. Therefore, this framework is a valuable resource for bioprocess
101 characterization and optimization.

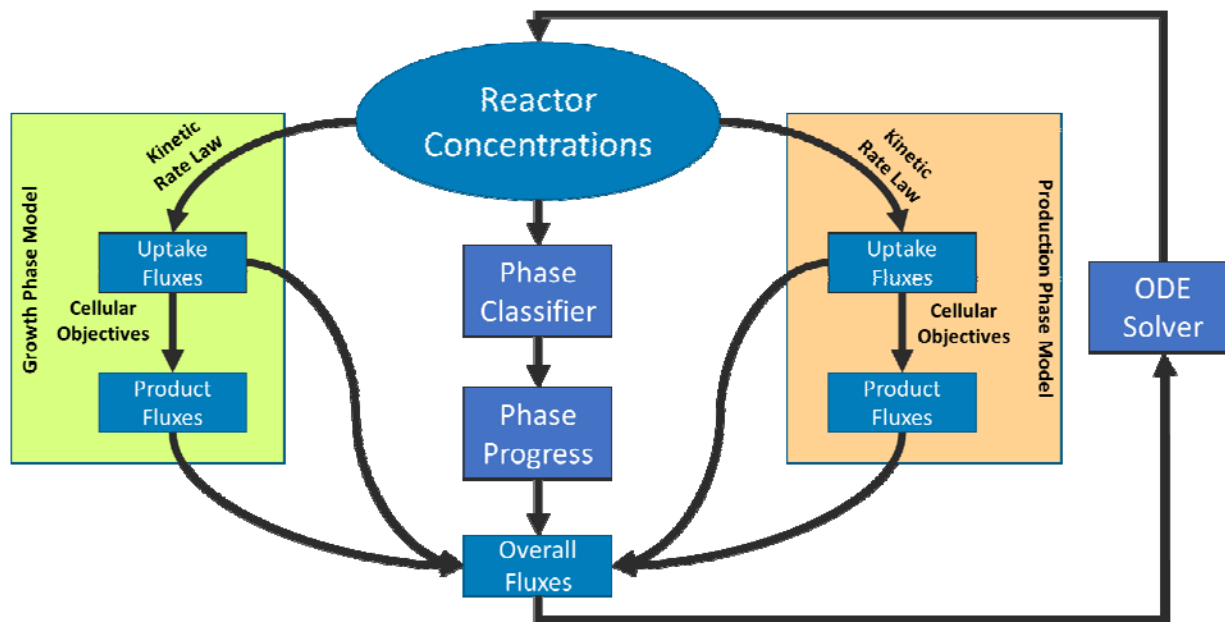
102 2. Results

103 2.1. The COSMIC-dFBA framework



104

105 **Figure 1A:** Overall workflow showing the pre-requisites and simulation approach used by
106 COSMIC-dFBA. COSMIC-dFBA predicts metabolite concentration, cell density, and antibody
107 titer profiles by solving a system of ordinary differential equations in which the rate of
108 metabolite uptake/secretion is determined using a metabolic model. In order to accomplish this,
109 three inputs must be specified. The first input is the state-specific metabolic model, which is
110 derived from a genome-scale metabolic model by overlaying different types of -omics data
111 (metabolomic, transcriptomic, or fluxomic data). The second requirement is the knowledge of
112 state-specific cellular objectives encoding the allocation of nutrients into various products, which
113 is inferred from metabolite uptake and secretion rates computed using spent media analysis. The
114 third requirement is a cell state distribution predictor, a machine learning model that predicts the
115 cell state based on prevailing conditions to adjust nutrient uptake by the metabolic model.



116

117 **Figure 1B:** Computing instantaneous metabolic fluxes in COSMIC-dFBA. The system of ODEs
118 solved to update reactor metabolite concentrations requires uptake and secretion rates that are
119 computed as a weighted average of metabolism from all possible metabolic states (growth and
120 production states, in this case). The weights for the contributions are computed using the cell
121 state distribution predictor. The fluxes corresponding to each metabolic state are solved by
122 solving a multi-level flux balance analysis problem using the state-specific metabolic model,
123 provided state-specific uptake rates (determined by reactor metabolite concentrations using a
124 Monod-like equation), and specified cellular objectives. The net result is a set of flux
125 distributions corresponding to various metabolic states. These flux distributions are averaged
126 based on weights computed by the cell state distribution predictor to obtain the net uptake and
127 secretion rates.

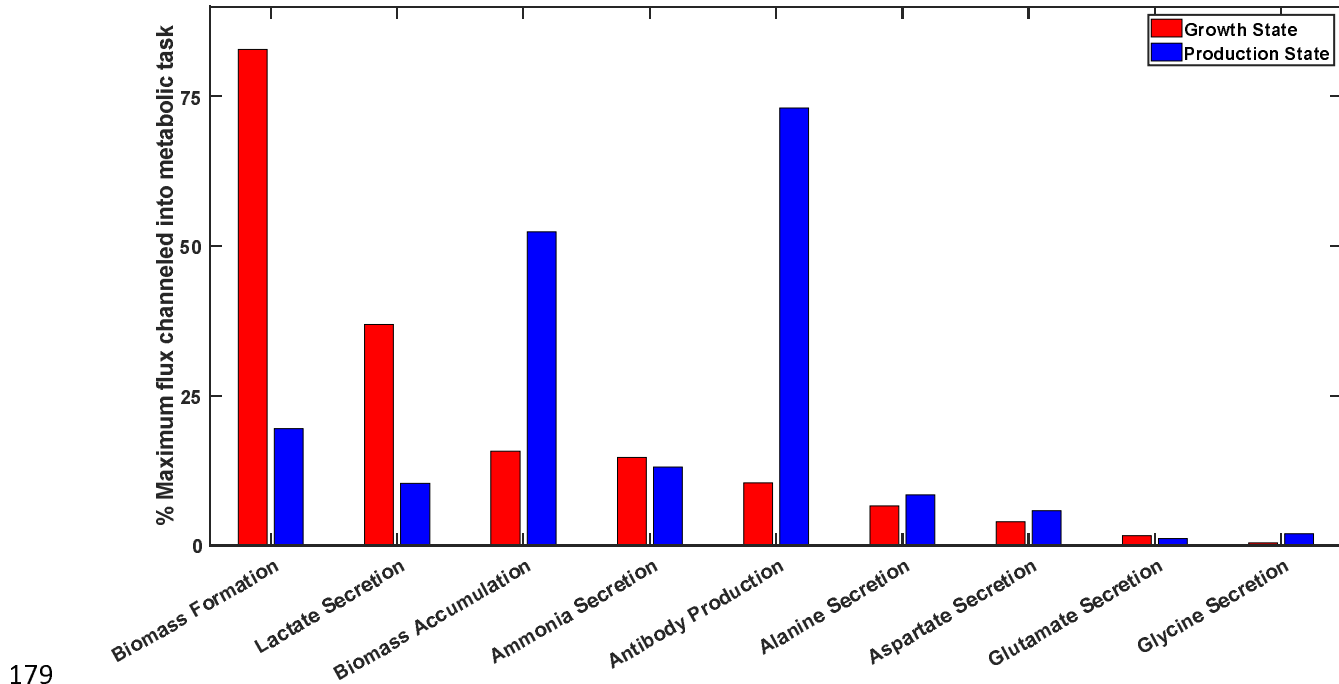
128 Cellular Objectives and State Modulation In bioreaCtors (COSMIC-dFBA) is a multi-scale
129 hybrid dynamic flux balance analysis framework that predicts total cell density, antibody titer,
130 and metabolite concentration profiles throughout a bioprocess. Figure 1 shows the schematic
131 representation of COSMIC-dFBA along with the pre-requisites and dynamic inputs required for
132 execution. We define a metabolic state (hereafter referred to as “state”) as the aggregate of
133 nutrient uptake, afforded pathways for metabolism, and flux distribution into various products.
134 The conceptual advancement by COSMIC-dFBA is the seamless transition between states in a
135 dynamic bioprocess without the need for condition-specific parametrization of state transition.
136 Because FBA is only applicable at metabolic steady-state, intracellular flux distributions are

137 constrained via nutrient uptake rates in traditional dFBA. COSMIC-dFBA overcomes this
138 limitation by assuming that overall metabolism in the reactor is a weighted average of
139 metabolism of cells in various states. The cell state at any time point is predicted by the Cell
140 State Distribution Predictor model based on instantaneous bioreactor conditions and feature
141 metabolite concentrations using a supervised machine-learning classifier (See Methods section
142 4.4. The four prerequisites for executing COSMIC-dFBA include (i) state-specific metabolic
143 models that contain limits on nutrient uptake and pathways available for metabolism, (ii) state-
144 specific uptake kinetics that reflect the effects of changing gene expression on nutrient uptake in
145 different cell states, (iii) state-specific metabolic tasks that encode the resource allocation in each
146 cell state, and (iv) a machine learning model to predict population distribution among cell states
147 based on the prevailing conditions in the bioreactor. The procedure for preparing these
148 prerequisites is described in the Supplementary Methods.

149 COSMIC-dFBA simulates bioreactor metabolite and product concentrations by solving a system
150 of ODEs describing the feeding, removal, and metabolism of nutrients and products in the
151 bioreactor (Figure 1B). At each time point, the uptake rates and secretion rates are computed in
152 three steps. First, uptake rates for all nutrients are calculated using computed kinetic rate laws for
153 each state. Next, the computed nutrient uptake rates are used to constrain the respective state-
154 specific metabolic models. The state-specific secretion rates are computed by solving the state-
155 specific metabolic model using a multi-objective FBA. Finally, the average uptake and secretion
156 rates are computed by weighting the computed state-specific uptake and secretion rates by the
157 fraction of cells in each state, predicted by the cell state distribution predictor model based on
158 prevalent bioreactor condition. These overall rates are then used to update the nutrient
159 concentrations in the bioreactor.

160 **2.2. Cellular objectives are cell state-specific**

161 A PCA of computed fluxes revealed two distinct cellular metabolic states (state 1 and state 2)
162 representing metabolism before day 3 and after day 10. We analyzed the computed state-specific
163 uptake and secretion rates (see Methods section 4.3 and Supplementary methods section 1) in the
164 context of the *iCHO1766* metabolic model to quantify the changes in resource allocation
165 associated with state shift. We first computed the task efficiencies (defined as the ratio of
166 measured flux to maximum flux predicted by the metabolic model) for each secreted product and
167 assigned priorities to each metabolic task (see Supplementary Methods). Figure 2 shows the task
168 efficiency averaged across all reactor conditions for all measured metabolic byproducts in both
169 states. We found that biomass formation and lactate secretion were the top two metabolic tasks in
170 state 1, accounting for 88% of the consumed carbon and 40% of the consumed nitrogen, as
171 quantified by FBA. Based on this, we call state 1 the “growth state”. The primary metabolic task
172 in the production phase was antibody production, accounting for 73% of the consumed nitrogen.
173 Based on this, we call state 2 the “production state”. Although the total cell density did not
174 change for cells in the production state, the cell size steadily increased, suggesting that biomass
175 precursors were being synthesized and accumulated. These findings demonstrate that metabolism
176 qualitatively changes upon state shift in the bioprocess and motivates the need to incorporate
177 approaches to account for cell state shifts and changing metabolic objectives in a dFBA
178 simulation.



179

180 **Figure 2:** Resource allocation towards various metabolic tasks in the growth and production
181 phases. Cells were predominantly in the growth state before day 3 and transitioned to the
182 production state between day 3 and day 10. Past day 10, cells were primarily in the production
183 state. Most of the cellular resources were channeled into biomass formation in the growth state
184 and towards antibody production in the production state. Lactate was produced from glucose via
185 glycolysis and from asparagine and glutamine via the anaplerotic pathways. Additional carbons
186 were channeled into synthesizing biomass precursors in the production state, which were
187 accumulated intracellularly. A similar fraction of consumed nitrogen was channeled into
188 ammonia generation (via glutaminolysis and asparagine degradation) and alanine production via
189 transamination in both states. Glycine production was significantly reduced in the production
190 state.

191

192

193

194

195

196

197

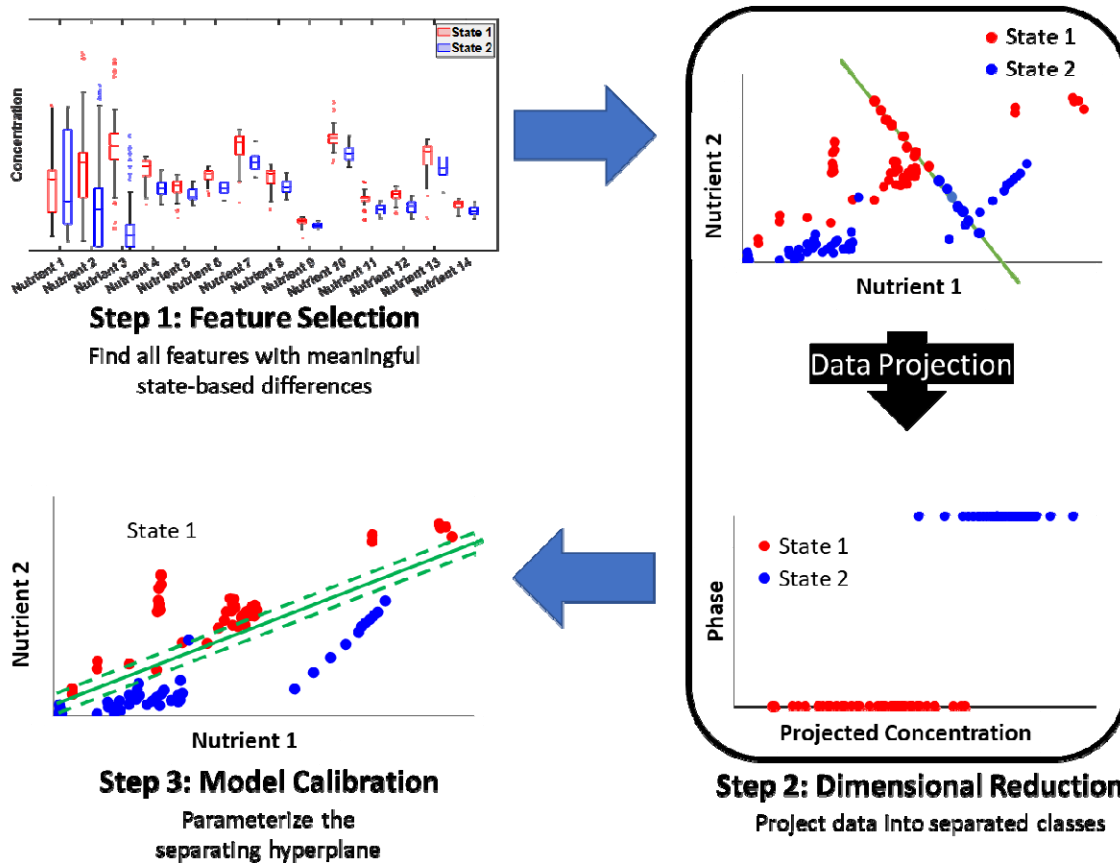
198

199

200

201 **2.3.The cell state distribution predictor captures phase-shifts driven by nutrient and**

202 **oxygen depletion**



203

204 **Figure 3:** Training the phase classifier model to predict cell state based on bioreactor conditions

205

206 To ensure that cell state is properly predicted by changes in reactor conditions when simulating a

207 bioprocess, we developed a state classification model and trained it through a three-step

208 workflow (Figure 3). The first step is to identify metabolites whose depletion correlates with the

209 observed state shift. To accomplish this, we label each data measurement as either growth state,

210 production state, or mixed state based on the state progression parameter computed concurrently

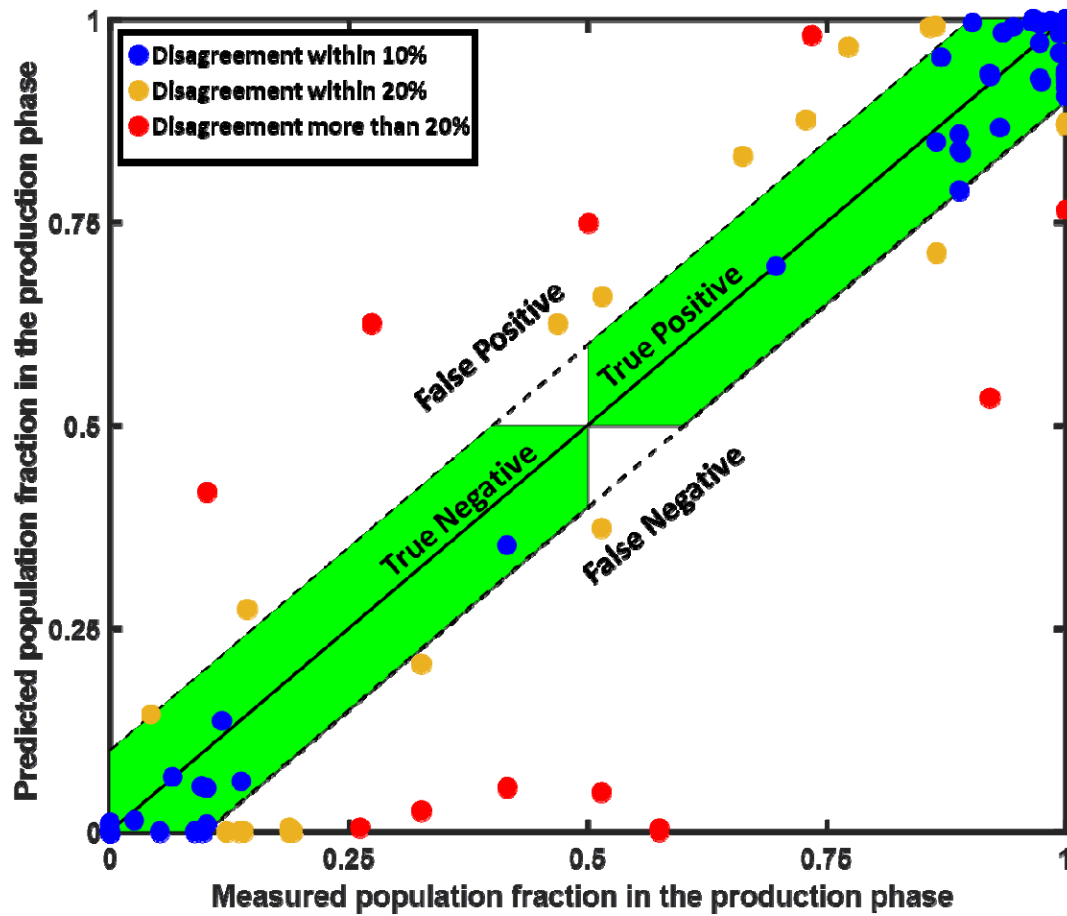
211 with uptake and secretion rates (see Supplementary Methods section 1). The state progression

212 parameter, p , represents the distribution of cell populations in each metabolic state with $p = 0$
213 indicating that all cells are in the growth state and $p = 1$ indicating that all cells are in the
214 production state. Metabolite concentrations at time points with $p < 0.2$ (less than 20% of the cells
215 in the production state) were considered to represent the growth state and at time points with $p >$
216 0.8 (more than 80% of the cells in the production state) were considered to represent the
217 production state. We excluded metabolites whose media concentration increased over time as
218 metabolic byproducts were constantly cleared from the bioreactor by perfusion and retain only
219 those metabolites whose media concentration decreases by at least 50%. For the cell line and
220 process considered here, the full list of features includes glucose, asparagine, and glutamine as
221 the potential metabolite candidates in addition to oxygen level and bioreactor temperature
222 (Supplementary Figure S1).

223 The second step in developing the phase classifier model is dimensional reduction with Linear
224 Discriminant Analysis (LDA) to project the features to a lower dimensional space, such that
225 projected features are correctly classified into the growth and production state. For this, we
226 consider features corresponding to growth state and production states ($p < 0.2$ or $p > 0.8$) and
227 ignore features corresponding to the mixed state. In the final step, we fit a logistic curve to model
228 the relationship between projected features from all three states and the predicted state
229 progression parameter. The resulting machine learning model intakes the prevailing bioreactor
230 features and predicts cell state to determine the net uptake and secretion rates in the reactor (See
231 Figure 1B).

232 The cell state distribution predictor model correctly predicted the state for 94 of 130 time points
233 across all bioreactor growth conditions with an accuracy of 0.1 (difference between predicted
234 and computed state is less than 0.1) and 118 of 130 time points with an accuracy of 0.2 (Figure

235 4). The model had a specificity of 0.78 and a sensitivity of 0.681. The F1-score was 0.731 and
236 Matthews' correlation coefficient was 0.454. In contrast, models based on a random classifier
237 (cell state distribution assumed to be a random number between 0 and 1) had a Matthews'
238 correlation coefficient of -0.72, indicating that the state prediction by the trained model
239 significantly outperformed random chance (permutation test, p-value $< 10^{-6}$). The model
240 correctly identified state shifts associated with the depletion of asparagine, glutamine, and
241 glucose, even in growth conditions with altered amino acid and glucose availability. In cases
242 with altered oxygen availability, the model correctly identified the cell state for all data except
243 those between days 6 to 8. This was because the cells had already transitioned into the
244 production state in the oxygen-depleted condition before the feature metabolites were
245 sufficiently depleted for the model to identify and predict a state shift, leading to a false negative
246 prediction. Other cells had not transitioned to the production state despite depletion of the feature
247 metabolites in the high-oxygen condition, leading to a false positive prediction by the phase
248 classifier. Except for a small number of extreme conditions, the model robustly predicts cell state
249 shifts arising from nutrient depletion within the bioreactor across a wide range of conditions.



250

251 **Figure 4:** Comparison of model-predicted and measured population fractions in the production
252 phase. The Blue dots represent the data points that were correctly identified to be in either the
253 growth or production phase with a 10% margin of error. The orange dots represent the correctly
254 predicted phases with a 20% margin of error. The red dots represent data that were incorrectly
255 predicted by the phase classifier model.

256

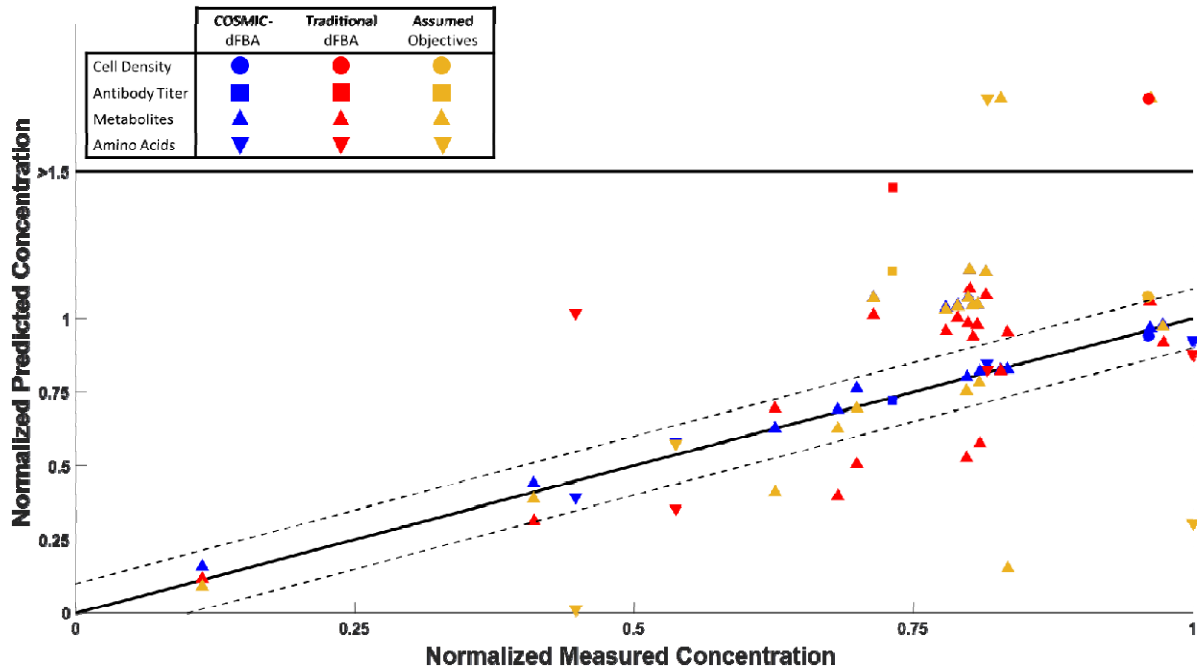
257

258

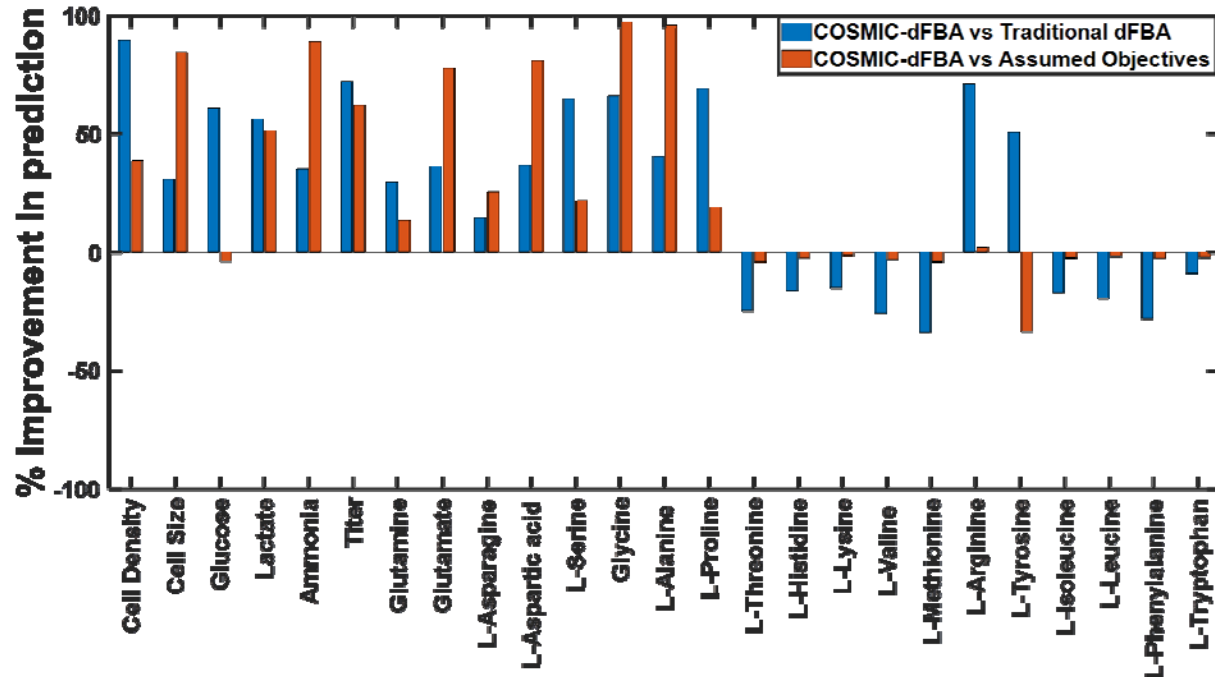
259

260

261 **2.4. The dFBA algorithm accurately predicts concentration profiles**



262 **Figure 5A:** Consistency of measured and predicted concentrations on day 13 for amino acids
263 (downward triangles), glycolytic metabolites (upward triangles), cell density (circles) and
264 antibody titer (square) using COSMIC-dFBA (blue markers), a standard dFBA algorithm with
265 specified cellular objectives and phase switch at a fixed time point (red markers), and a standard
266 dFBA algorithm with the phase classifier from COSMIC-dFBA but assuming maximize biomass
267 objective during the growth phase and maximize antibody production objective in the production
268 phase (orange markers).
269



270
271 **Figure 5B:** Improvement in predictions by COSMIC-dFBA compared to models with no
272 classifier or assumed objective functions.

273 We simulated the cell density, glucose, lactate, antibody, and 17 amino acid concentration
274 profiles over the 13-day perfusion bioprocess run across 8 different media conditions
275 (Supplementary Figures S2 – S11). To evaluate its performance, we compared the concentration
276 profiles, predicted using COSMIC-dFBA with two implementations of traditional dFBA. In the
277 first case (referred to as the “traditional dFBA case”), we retained the cellular objectives, but
278 assumed that phase transition coincided with the hypothermic shift. In the second case (referred
279 to as the “assumed objective case”), we retained the phase classifier from COSMIC-dFBA, but
280 assumed that the cells only maximize biomass during the growth phase and maximize antibody
281 production in the production phase. Figure 4A compares the day 13 concentration predictions by
282 COSMIC-dFBA and the two test cases. We found that COSMIC-dFBA significantly
283 outperformed both test cases based on traditional dFBA, thus highlighting the need to account
284 for changing bioreactor conditions and metabolic tasks. We also evaluated the improvement in
285 prediction (defined as the mean fractional reduction in disagreement between predicted and

286 measured concentrations over the course of the bioprocess) as a measure of how well the
287 concentration profiles predicted by each algorithm agree with the experimental data. From this
288 we found that the concentration profiles predicted by COSMIC-dFBA for cell density, antibody
289 titer, glucose, lactate, glutamine, and glutamate were in better agreement with the measured data
290 than the traditional dFBA cases (Figure 4B). However, the standard dFBA test cases better
291 predicted the consumption of several essential amino acids.

292 The traditional dFBA case greatly overestimated the final cell density in all eight growth
293 conditions. This was because the traditional dFBA case assumed that the entire cell population
294 transitioned from the growth phase to the production phase when the hypothermic shift was
295 applied, regardless of the bioreactor conditions. Thus, this case failed to account for the
296 redistribution of metabolic fluxes and a shift from cell growth to antibody production when key
297 metabolites were depleted early, particularly in the low glucose and low amino acid cases. This
298 led to an extended growth phase in all eight conditions, and a higher cell density at the end of the
299 growth phase. Consequently, this approach predicted a higher antibody titer in all growth
300 conditions. On the other hand, the final cell density predicted using the “assumed objectives”
301 case was only 14.8% higher than those predicted using COSMIC-dFBA. This agreement
302 between COSMIC-dFBA and the “assumed objectives” case arises from the fact that the
303 assumed maximization of biomass formation is close to the actual metabolism of the cells, which
304 channels, on average, 82% of the resources towards biomass production in the growth phase
305 (Figure 2). This implementation also assumed that all available resources were channeled into
306 antibody production in the production phase, whereas the experimental data suggests that 25% of
307 the resources were channeled into other cellular processes. This led to a dramatic overprediction
308 of antibody titer in the bioreactor. Overall, these comparisons demonstrate the importance of the

309 two integral components of COSMIC-dFBA (the phase classifier and comprehensive accounting
310 of metabolic tasks), which contribute to the algorithm's superior predictive capabilities compared
311 to existing dFBA-based bioprocess modeling frameworks.

312

313 **3. Discussion**

314 This study presents COSMIC-dFBA, a multi-scale dynamic flux balance analysis framework that
315 combines machine learning and mechanistic modeling techniques to simulate cell behavior in a
316 perfusion bioprocess and predict metabolic shifts in response to changing bioreactor conditions.
317 This framework operates at two scales: the bioreactor scale and the cellular scale. The cellular
318 scale interfaces with the bioreactor scale using the cell state distribution predictor that determines
319 the distributions of cell populations in various states based on prevailing bioreactor conditions.
320 Based on the determined cell state, nutrients are consumed according to previously
321 parameterized kinetic rate laws, and consumed nutrients are channeled into appropriate state-
322 specific metabolic tasks (e.g., cell growth, antibody production, etc.). This yields the net
323 instantaneous production and consumption rates of all metabolites in the bioreactor, which are
324 then used to update the bioreactor concentrations by solving a system of ODEs. Leveraging the
325 metabolic model provides a mechanistic relationship between nutrient uptake and product
326 secretion as well as additional pathways through which metabolic flux is diverted to generate
327 byproducts. By dynamically adjusting product yields, this framework always ensures that
328 nutrient consumption and product formation in the bioreactor satisfy conservation of mass and
329 are thermodynamically feasible, which is not always the case when modeling a bioprocess using
330 empirical models. Unlike previous dFBA approaches (Nolan and Lee, 2011), COSMIC-dFBA
331 does not need to solve any quadratic programming problems, which considerably decreases the

332 computational cost. This permits the use of genome-scale metabolic models for dFBA, which
333 increases generalizability. Incorporating the means to modulate cellular resource allocation using
334 a hybrid modeling paradigm improves fidelity without the need for developing detailed
335 mechanistic models such as whole-cell models or ME-models. Furthermore, by using an adaptive
336 time step, a desired integration accuracy can be ensured without resorting to collocation (St John
337 et al., 2017), which significantly reduces the number of time-steps and by extension, the number
338 of times the FBA problem must be solved (de Oliveira et al., 2023; Zhuang et al., 2011).

339 COSMIC-dFBA is particularly versatile in that it only requires the usual data typically collected
340 during a bioprocess to train the model. Uptake and secretion rates were computed from
341 metabolite concentration profiles and analyzed to determine phase-specific resource allocation to
342 identify the major metabolic tasks prioritized by the cell in various states, whereas phase shifts
343 were predicted based on reactor metabolite concentrations and temperature shifts. Other types of
344 omics data can be readily incorporated to minimize manual interventions. For example,
345 integrating gene expression data enables extraction of context-specific metabolic models
346 (Gustafsson et al., 2023; Opdam et al., 2017), which have been previously shown to vary
347 between process phases. Transcriptomics data can also suggest metabolic tasks not captured by
348 exo-metabolomics data (Helen et al., 2022; Masson et al., 2023; Richelle et al., 2021).
349 Proteomics data can be incorporated to correlate changes in transporter abundance with phase
350 shifts (Colijn et al., 2009; Sanchez et al., 2017; Tian and Reed, 2018; Yeo et al., 2020), which
351 modulates the maximum uptake rate of nutrients in each process phase.

352 The key strength of COSMIC-dFBA is the ability to learn from additional experimental data that
353 allows it to predict newer states. Our analyses indicate that the cell state distribution predictor is
354 a vital component of this framework that smoothly modulates state shifts using a single layer

355 perceptron (linear combination of inputs combined with a logistic activation function). The
356 choice of activation function was based on previous efforts to model cellular signal transduction
357 (Samaga and Klamt, 2013; Wynn et al., 2012) and gene activation (Ay and Arnosti, 2011) in
358 response to changing environmental conditions within the bioreactor. The main drawback of this
359 approach is that the framework cannot automatically determine the cause of the state shift
360 (arising from nutrient depletion, temperature shift, oxygen limitation, etc.) and assumes that all
361 phase shifts are of the same nature. In the current implementation of COSMIC-dFBA, we
362 circumvent this by defining the cellular objectives for each type of phase shift in advance.
363 However, automated prediction of changes in metabolic task priorities in response to phase shifts
364 will require an overlay of the signaling (Lin et al., 2022; Sompairac et al., 2019) and gene
365 expression networks (Pio et al., 2022) on to existing models of metabolism in the absence of
366 fully descriptive whole-cell models (Ahn-Horst et al., 2022; Karr et al., 2012). Such models will
367 expand the predictive capabilities of COSMIC-dFBA to predict heterogeneity in cell populations
368 in large-scale bioreactors arising from non-homogeneous mixing and poor local oxygen transfer.
369 That will allow the framework to predict and correct the potential detriments to process yield and
370 productivity upon scale-up to manufacturing scales. Despite these limitations, COSMIC-dFBA
371 significantly outperforms traditional dFBA in its current form. The ability to model dynamic
372 metabolism uniquely positions this framework for applications in bioprocesses with metabolic
373 shifts.

374 **4. Methods**

375 **4.1. Cell culture and process data acquisition**

376 A stable, clonally derived Chinese hamster ovary (CHO) cell line expressing a non-glycosylated
377 recombinant protein was thawed and scaled up in proprietary growth media to generate sufficient

378 cell mass to inoculate a production perfusion bioreactor. The production bioreactors were
379 operated in 3 L stirred tank bioreactors with a 1.5 L working volume for 13 days using
380 proprietary chemically defined media. Bioreactors were inoculated in the same basal production
381 media. Perfusion was performed using alternating tangential flow filtration starting at Day 0 at a
382 perfusion rate of 1 bioreactor volume per day for a duration of 13 days. On Day 8, the
383 temperature setpoint was decreased for the remaining duration of the experiment. The
384 experimental conditions were set up following a Box Behnken DOE varying dissolved oxygen,
385 perfusion media amino acid levels, and perfusion media glucose concentration as shown in
386 Supplementary Table **ST1**.

387 Bioreactor parameters, such as agitation, dissolved oxygen concentration, pH, and temperature
388 were monitored and controlled through a DeltaV controller (Emerson, St. Louis, MO, USA). The
389 pH was controlled through CO₂ or 1 M Na₂CO₃ addition. Dissolved oxygen was maintained by
390 sparging oxygen through a drilled pipe and a sintered sparger. Additionally, inline off-gas O₂ and
391 CO₂ were monitored using the BlueSens BlueVary gas sensor (BlueSens, Wood Dale, IL, USA).
392 The daily sampling procedure consisted of cell density and viability using a Cedex HiRes
393 analyzer (Roche Diagnostics, Indianapolis, IN, USA), metabolites (lactate, glucose, glutamine,
394 glutamate, and ammonium) from a Cedex Bio HT analyzer (Roche Diagnostics, Indianapolis, IN,
395 USA), osmolality using the Advanced Instruments OsmoPRO (Advanced Instruments, Norwood,
396 MA, USA), and external pH, pCO₂, and pO₂ using a Siemens RAPIDLab 1260 (Siemens
397 Healthineers, Erlangen, Germany). Daily clarified samples for each reactor were analyzed for
398 titer via HPLC. Amino acid concentrations were determined as follows: cell culture supernatant
399 samples were filtered through a 0.2µm filter then diluted properly with 18 mM HCl and mixed
400 with the internal standard mixture containing heavy isotope labeled amino acids. An UHPLC

401 system Agilent 1290 (Agilent Technologies, Santa Clara, CA, USA) equipped with a reversed
402 phase C18 column (Agilent Poroshell 120 SB-C18, 1.9 μm , 2.1 mm \times 100 mm) was used for
403 components separation. The mobile phases used were water (A) and acetonitrile (B) in 0.2%
404 heptafluorobutyric acid (HFBA). Targeted quantitation data were acquired using the dynamic
405 Multiple Reaction Monitoring (MRM) mode on an Agilent 6490 Triple Quadrupole mass
406 spectrometer. Agilent MassHunter B.08.00 was used for data acquisition and data analysis.

407 **4.2. Metabolic model and data processing**

408 iCHO1766 was used as the base metabolic model (Hefzi et al., 2016). The protein secretory
409 pathway (Gutierrez et al., 2020) was appended to iCHO1766 to accurately model the precursor
410 and energy demands for antibody synthesis and secretion. Two phases were identified using the
411 concentration data. The growth rate, antibody specific productivity, uptake and secretion rates of
412 all measured metabolites, and the fraction of cell population in each phase were computed from
413 the concentration profiles using nonlinear regression as described in the supplementary methods.
414 The computed fluxes in each growth condition are reported in Supplementary Table ST3.

415 **4.3. Inferring state-specific metabolic task objectives and priorities**

416 State-specific metabolic flux distributions were modulated in terms of metabolic tasks and task
417 efficiencies. Each state-specific model was calibrated as described in the supplementary material.
418 Briefly, all measured quantities were classified into either nutrients (consumed by cells) or
419 byproducts (generated by cells) in each phase. All secreted byproducts were considered
420 “metabolic tasks” and their priority order was determined in an iterative manner. First, the uptake
421 rates of nutrients were fixed in the metabolic model. Following this, the flux through each
422 metabolic task was individually maximized using Flux Balance Analysis (FBA) (Varma and

423 Pálsson, 1994). Task efficiency for each metabolic task was computed as the ratio of measured
424 flux through the metabolic task to the maximum flux predicted using FBA. The task with the
425 highest efficiency was considered the highest priority task as it reflects the maximal nutrient
426 utilization towards this task and its corresponding efficiency was stored. To find the next priority
427 task, the experimentally measured flux through the previous task was enforced as a lower bound
428 in the metabolic model and that task was removed from the list of metabolic tasks to be
429 evaluated. Following this, the task efficiency calculation steps were repeated to identify the next
430 highest priority task. This loop was repeated until all metabolic tasks were ordered. The list of
431 state-specific metabolic tasks and their corresponding task efficiencies are reported in
432 Supplementary Table ST4.

433 **4.4. Training the cell state distribution predictor**

434 The cell state distribution predictor is a machine-learning model that predicts cell state based on
435 bioreactor conditions. Using bioreactor media concentrations, partial pressure of oxygen, and the
436 temperature of the bioreactor as inputs, the phase classifier model is trained using the three-step
437 process depicted in Figure 3 and predicts the fraction of cell population in the production state.
438 For each condition considered in this study, time points were classified into either growth state,
439 production state, or mixed populations based on whether the fraction of cells in the production
440 state were less than 20%, greater than 80%, or somewhere in between, respectively.
441 Concentrations of all metabolites were grouped into these three classes and plotted to identify
442 metabolites correlated with phase shifts. Candidate metabolites were chosen such that their (a)
443 median concentrations changed drastically between the growth and production states, and (b)
444 they were depleted, or close to depleted in the production state. Oxygen and temperature were
445 included to account for premature state shifts arising from hypoxic (Zeh et al., 2021) and

446 hypothermic shifts (Wulhfard et al., 2008). The second step is to reduce the dimensionality of the
447 data such that the growth and production state data are separated into distinct clusters. We used
448 Linear Discriminant Analysis (LDA) to achieve this. The projected concentration w is related to
449 feature i (metabolite concentration, partial pressure of oxygen, or temperature) via a weighted
450 linear combination using weights k_i using Equation (1):

$$w = \sum_i^N k_i * feature_i \quad (1)$$

451 Following this, projected concentrations were computed for all three classes and logistic
452 regression was performed to compute the parameters a and b , representing the steepness of the
453 transition and the bias, respectively and model the transition from growth to production state
454 using Equation (2):

$$p(w) = \frac{1}{1 + e^{a*w-b}} \quad (2)$$

455

456 **4.5.Simulating metabolite concentration profiles and culture parameters using COSMIC-** 457 **dFBA**

458 COSMIC-dFBA simulates bioreactor concentration profiles by solving the following initial value
459 problem (IVP) for cell density ($X(t)$), cell size ($S(t)$), and concentration of metabolite i ($C_i(t)$)
460 from time t_0 to t_f :

$$\frac{dX(t)}{dt} = \mu_{eff}X(t) \quad (3)$$

$$\frac{dS(t)}{dt} = v_{bmac}S(t) \quad (4)$$

$$\frac{dC_i(t)}{dt} = v_{i,eff}X(t) \quad (5)$$

$$X(t_0) = X_0 \quad (6)$$

$$S(t_0) = S_0 \quad (7)$$

$$C_i(t_0) = C_{i,0} \quad (8)$$

461

462 In Equations (3), (4), and (5), μ_{eff} , $v_{bmass,eff}$, and $v_{i,eff}$ represent the effective growth rate,
463 effective biomass accumulation rate, and the effective uptake/secretion rate of metabolite i , and
464 are related to the growth and production phase fluxes via the population fraction parameter $p(t)$
465 computed using Equations (1) and (2):

$$\mu_{eff}(t) = (1 - p(t))\mu_{growth} + p(t)\mu_{production} \quad (9)$$

$$v_{bmass,eff}(t) = (1 - p(t))v_{bmass,growth} + p(t)v_{bmass,production} \quad (10)$$

$$v_{i,eff}(t) = (1 - p(t))v_{i,growth} + p(t)v_{i,production} \quad (11)$$

466

467 The above IVP is solved using the Bulirsch-Stoer algorithm (Bulirsch and Stoer, 1966) with
468 adaptive step-size control (Deuflhard, 1983) to reduce the number of times the metabolic model
469 must be solved without loss of accuracy. COSMIC-dFBA is encoded and executed in
470 MATLABTM. The source code is provided as a zip file in the supplementary material.

471

472 Acknowledgements

473 This work was supported by a generous grant from Amgen.

474 **References**

- 475 Ahn-Horst, T.A., Mille, L.S., Sun, G., Morrison, J.H., and Covert, M.W. (2022). An expanded whole-cell
476 model of *E. coli* links cellular physiology with mechanisms of growth rate control. *NPJ Syst Biol Appl* 8,
477 30.
- 478 Ay, A., and Arnosti, D.N. (2011). Mathematical modeling of gene expression: a guide for the perplexed
479 biologist. *Crit Rev Biochem Mol Biol* 46, 137-151.
- 480 Ben Yahia, B., Gourevitch, B., Malphettes, L., and Heinzle, E. (2017). Segmented linear modeling of CHO
481 fed-batch culture and its application to large scale production. *Biotechnol Bioeng* 114, 785-797.
- 482 Ben Yahia, B., Malphettes, L., and Heinzle, E. (2021). Predictive macroscopic modeling of cell growth,
483 metabolism and monoclonal antibody production: Case study of a CHO fed-batch production. *Metab Eng*
484 66, 204-216.
- 485 Bulirsch, R., and Stoer, J. (1966). Numerical treatment of ordinary differential equations by extrapolation
486 methods. *Numerische Mathematik* 8, 1-13.
- 487 Chen, Y., McConnell, B.O., Gayatri Dhara, V., Mukesh Naik, H., Li, C.T., Antoniewicz, M.R., and
488 Betenbaugh, M.J. (2019). An unconventional uptake rate objective function approach enhances
489 applicability of genome-scale models for mammalian cells. *NPJ Syst Biol Appl* 5, 25.
- 490 Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.Y., Moody, D.B., Murray,
491 M., and Galagan, J.E. (2009). Interpreting expression data with metabolic flux models: predicting
492 *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol* 5, e1000489.
- 493 de Oliveira, R.D., Le Roux, G.A.C., and Mahadevan, R. (2023). Nonlinear programming reformulation of
494 dynamic flux balance analysis models. *Computers & Chemical Engineering* 170, 108101.
- 495 Dean, J., and Reddy, P. (2013). Metabolic analysis of antibody producing CHO cells in fed-batch
496 production. *Biotechnol Bioeng* 110, 1735-1747.
- 497 Deuflhard, P. (1983). Order and stepsize control in extrapolation methods. *Numerische Mathematik* 41,
498 399-422.
- 499 Galleguillos, S.N., Ruckerbauer, D., Gerstl, M.P., Borth, N., Hanscho, M., and Zanghellini, J. (2017). What
500 can mathematical modelling say about CHO metabolism and protein glycosylation? *Computational and*
501 *Structural Biotechnology Journal* 15, 212-221.
- 502 Garcia Sanchez, C.E., and Torres Saez, R.G. (2014). Comparison and analysis of objective functions in flux
503 balance analysis. *Biotechnol Prog* 30, 985-991.
- 504 Gustafsson, J., Anton, M., Roshanzamir, F., Jornsten, R., Kerkhoven, E.J., Robinson, J.L., and Nielsen, J.
505 (2023). Generation and analysis of context-specific genome-scale metabolic models derived from single-
506 cell RNA-Seq data. *Proc Natl Acad Sci U S A* 120, e2217868120.
- 507 Gutierrez, J.M., Feizi, A., Li, S., Kallehauge, T.B., Hefzi, H., Grav, L.M., Ley, D., Baycin Hizal, D.,
508 Betenbaugh, M.J., Voldborg, B., et al. (2020). Genome-scale reconstructions of the mammalian secretory
509 pathway predict metabolic costs and limitations of protein secretion. *Nat Commun* 11, 68.
- 510 Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-
511 Hizal, D., Huang, Y., Ley, D., et al. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster
512 Ovary Cell Metabolism. *Cell Syst* 3, 434-443 e438.
- 513 Helen, O.M., Chih-Chung, K., Magdalena, M., Magnus, L., Åsa, S., Anna, B., Hanna, T., Sophia, H.,
514 Mathias, U., Luigi, G., et al. (2022). Deciphering the determinants of recombinant protein yield across
515 the human secretome. *bioRxiv*, 2022.2012.2012.520152.
- 516 Jerby, L., Shlomi, T., and Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic
517 models: application to human liver metabolism. *Mol Syst Biol* 6, 401.

518 Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N.,
519 Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from
520 genotype. *Cell* *150*, 389-401.

521 Kasemiire, A., Avohou, H.T., De Bleye, C., Sacre, P.Y., Dumont, E., Hubert, P., and Ziemons, E. (2021).
522 Design of experiments and design space approaches in the pharmaceutical bioprocess optimization. *Eur*
523 *J Pharm Biopharm* *166*, 144-154.

524 Laakso, K., Koskeniemi, K., Koponen, J., Kankainen, M., Surakka, A., Salusjarvi, T., Auvinen, P., Savijoki,
525 K., Nyman, T.A., Kalkkinen, N., et al. (2011). Growth phase-associated changes in the proteome and
526 transcriptome of *Lactobacillus rhamnosus* GG in industrial-type whey medium. *Microb Biotechnol* *4*,
527 746-766.

528 Lin, Y., Yan, S., Chang, X., Qi, X., and Chi, X. (2022). The global integrative network: integration of
529 signaling and metabolic pathways. *aBIOTECH* *3*, 281-291.

530 López-Meza, J., Araíz-Hernández, D., Carrillo-Cocom, L.M., López-Pacheco, F., Rocha-Pizaña Mdel, R., and
531 Alvarez, M.M. (2016). Using simple models to describe the kinetics of growth, glucose consumption, and
532 monoclonal antibody formation in naive and infliximab producer CHO cells. *Cytotechnology* *68*, 1287-
533 1300.

534 Mahadevan, R., Edwards, J.S., and Doyle, F.J., 3rd (2002). Dynamic flux balance analysis of diauxic
535 growth in *Escherichia coli*. *Biophys J* *83*, 1331-1340.

536 Masson, H.O., Borland, D., Reilly, J., Telleria, A., Shrivastava, S., Watson, M., Bustillos, L., Li, Z., Capps, L.,
537 Kellman, B.P., et al. (2023). ImmCellFie: A user-friendly web-based platform to infer metabolic function
538 from omics data. *STAR Protoc* *4*, 102069.

539 Nolan, R.P., and Lee, K. (2011). Dynamic model of CHO cell metabolism. *Metab Eng* *13*, 108-124.

540 Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D.C., and Lewis, N.E. (2017). A Systematic Evaluation
541 of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst* *4*, 318-329 e316.

542 Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? *Nature Biotechnology* *28*,
543 245-248.

544 Pio, G., Mignone, P., Magazzu, G., Zampieri, G., Ceci, M., and Angione, C. (2022). Integrating genome-
545 scale metabolic modelling and transfer learning for human gene regulatory network reconstruction.
546 *Bioinformatics* *38*, 487-493.

547 Richelle, A., Kellman, B.P., Wenzel, A.T., Chiang, A.W.T., Reagan, T., Gutierrez, J.M., Joshi, C., Li, S., Liu,
548 J.K., Masson, H., et al. (2021). Model-based assessment of mammalian cell metabolic functionalities
549 using omics data. *Cell Rep Methods* *1*.

550 Samaga, R., and Klamt, S. (2013). Modeling approaches for qualitative and semi-quantitative analysis of
551 cellular signaling networks. *Cell Commun Signal* *11*, 43.

552 Sanchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.J., Kerkhoven, E.J., and Nielsen, J. (2017). Improving the
553 phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic
554 constraints. *Mol Syst Biol* *13*, 935.

555 Savinell, J.M., and Palsson, B.O. (1992). Network analysis of intermediary metabolism using linear
556 optimization. II. Interpretation of hybridoma cell metabolism. *J Theor Biol* *154*, 455-473.

557 Sidoli, F.R., Mantalaris, A., and Asprey, S.P. (2004). Modelling of Mammalian cells and cell culture
558 processes. *Cytotechnology* *44*, 27-46.

559 Sompairac, N., Modamio, J., Barillot, E., Fleming, R.M.T., Zinovyev, A., and Kuperstein, I. (2019).
560 Metabolic and signalling network maps integration: application to cross-talk studies and omics data
561 analysis in cancer. *BMC Bioinformatics* *20*, 140.

562 St John, P.C., Crowley, M.F., and Bomble, Y.J. (2017). Efficient estimation of the maximum metabolic
563 productivity of batch systems. *Biotechnol Biofuels* *10*, 28.

564 Sunley, K., Tharmalingam, T., and Butler, M. (2008). CHO cells adapted to hypothermic growth produce
565 high yields of recombinant beta-interferon. *Biotechnol Prog* *24*, 898-906.

566 Templeton, N., Dean, J., Reddy, P., and Young, J.D. (2013). Peak antibody production is associated with
567 increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnol Bioeng*
568 *110*, 2013-2024.

569 Tian, M., and Reed, J.L. (2018). Integrating proteomic or transcriptomic data into metabolic models using
570 linear bound flux balance analysis. *Bioinformatics* *34*, 3882-3888.

571 Torres, M., Altamirano, C., and Dickson, A.J. (2018). Process and metabolic engineering perspectives of
572 lactate production in mammalian cell cultures. *Current Opinion in Chemical Engineering* *22*, 184-190.

573 Varma, A., and Palsson, B.O. (1994). Stoichiometric flux balance models quantitatively predict growth
574 and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* *60*,
575 3724-3731.

576 Wulhfard, S., Tissot, S., Bouchet, S., Cevey, J., de Jesus, M., Hacker, D.L., and Wurm, F.M. (2008). Mild
577 Hypothermia Improves Transient Gene Expression Yields Several Fold in Chinese Hamster Ovary Cells.
578 *Biotechnology Progress* *24*, 458-465.

579 Wynn, M.L., Consul, N., Merajver, S.D., and Schnell, S. (2012). Logic-based models in systems biology: a
580 predictive and parameter-free network analysis method. *Integr Biol (Camb)* *4*, 1323-1337.

581 Yeo, H.C., Hong, J., Lakshmanan, M., and Lee, D.Y. (2020). Enzyme capacity-based genome scale
582 modelling of CHO cells. *Metab Eng* *60*, 138-147.

583 Zeh, N., Schlossbauer, P., Raab, N., Klingler, F., Handrick, R., and Otte, K. (2021). Cell line development
584 for continuous high cell density biomanufacturing: Exploiting hypoxia for improved productivity. *Metab*
585 *Eng Commun* *13*, e00181.

586 Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., and Lovley, D.R. (2011).
587 Genome-scale dynamic modeling of the competition between *Rhodospirillum rubrum* and *Geobacter* in anoxic
588 subsurface environments. *The ISME Journal* *5*, 305-316.

589