Computational and Structural Biotechnology Journal 23 (2024) 2861–2871

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Algorithm of spatial–temporal simulation for environment–strain interactions in strain–strain consortia based on resource competition mechanism

Chen Yang, Boyuan Xue, Qianqian Yuan, Shaojie Wang *, Haijia Su *

State Key Laboratory of Chemical Resource Engineering, Beijing Key Laboratory of Bioprocess, and Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, People's Republic of China

ARTICLE INFO

Keywords:
Spatial–temporal evolution
Strain–environment interactions
Strain–strain consortia
Resource competition relationship

ABSTRACT

Interaction simulation for co-culture systems is important for optimizing culture conditions and improving yields. For industrial production, the environment significantly affects the spatial–temporal microbial interactions. However, the current research on polymicrobial interactions mainly focuses on interaction patterns among strains, and neglects the environment influence. Based on the resource competition relationship between two strains, this research set up the modules of cellular physicochemical properties, nutrient uptake and metabolite release, cellular survival, cell swimming and substrate diffusion, and investigated the spatial–temporal strain–environment interactions through module coupling and data mining. Furthermore, in an *Escherichia coli–Saccharomyces cerevisiae* consortium, the total net reproduction rate decreased as glucose was consumed. *E. coli* gradually dominated favorable positions due to its higher glucose utilization capacity, reaching 100 % abundance with a competitive strength of 0.86 for glucose. Conversely, *S. cerevisiae* decreased to 0 % abundance with a competitive strength of 0.14. The simulation results of environment influence on strain competitiveness showed that inoculation ratio and dissolved oxygen strongly influenced strain competitiveness. Specifically, strain competitiveness increased with higher inoculation ratio, whereas *E. coli* competitiveness increased as dissolved oxygen increased, in contrast to *S. cerevisiae*. On the other hand, substrate diffusion condition, micronutrients and toxins had minimal influence on strain competitiveness. This method offers a straightforward procedure without featured downscaling and provides novel insights into polymicrobial interaction simulation.

1. Introduction

The rapid advances in bioinformatics technology provide a convenient platform for simulating microbial systems. Co-culture systems, as opposed to mono-cultures, can improve the substrate utilization efficiency by co-utilizing complex substrates, relieve metabolic stress by dividing metabolic tasks, and realize efficient production by effectively exploiting the diverse intracellular environment [1]. Moreover, the co-culture systems have been widely used in biosynthesis of complex and value-added products. For example, in the co-culture systems of *E. coli* and *S. cerevisiae*, the production of genistein have reached 100 mg/L which was 294 times higher than that of genistein produced by a mono-culture of *S. cerevisiae* (0.34 mg/L) [2]. In addition, Hal et al. obtained 36 mg/L resveratrol in a purely de novo fashion through the co-culture of *E. coli* and *S. cerevisiae*, eliminating the need for expensive inducers or precursors and significantly reducing costs compared to previous mono-culture system [3].

Simulation of microbial interactions for co-culture systems can provide reasonable strategies for optimizing strain ratios and culture conditions, enabling the full utilization of metabolic division and resource exchange among different strains to improve overall production. The complexity and non-uniform distribution of environmental variables in industrial production however significantly impact the accuracy of simulating spatial–temporal microbial interactions. Traditional simulation algorithms primarily focus on interactions among strains in co-culture systems, while insufficient attention is paid to environmental variables. As a result, these algorithms may not provide sufficient accuracy for industrial production processes such as fermenters and large-

* Corresponding authors.
E-mail addresses: wangshaojie@buct.edu.cn (S. Wang), suhj@mail.buct.edu.cn (H. Su).

https://doi.org/10.1016/j.csbj.2024.06.033
Received 25 January 2024; Received in revised form 26 June 2024; Accepted 26 June 2024
Available online 27 June 2024

scale bioreactors.

Among the traditional simulation algorithms for strain interactions, Joint-FBA [4,5], OptCom [5,6], SteadyCom [5,7] and cFBA [5,8] are based on the hypothesis of "overall optimization" [5]. This hypothesis assumes that the metabolic processes of strains ultimately reach maximum biomass production, thereby achieving low computational cost [5]. The Joint-FBA algorithm simulates public metabolite exchange among strains through exchange compartment and substrate concentration change [1,4]. Based on this exchange, Joint-FBA couples the intracellular fluxes for a mono strain and the objective function for maximizing overall biomass production to simulate polymicrobial fluxes [1,4]. To improve the simulation accuracy of the Joint-FBA algorithm, OptCom, SteadyCom and cFBA algorithms were developed. Among these algorithms, OptCom is capable of simulating the fitness between individual and community levels through the use of multi-level simulation and multi-objective optimization [1,6]. SteadyCom and cFBA define the objective function by measuring the properties of polymicrobial networks, thereby enabling the simulation of co-culture consortia containing more strains than Joint-FBA and OptCom [1,7]. However, these algorithms suffer from the disadvantage of "forced altruism", which results in unstable solutions in the state of strain competition for resources [1,5].

In order to eliminate the drawbacks of "forced altruism", Cai et al. proposed the NECom algorithm. NECom is based on the hypothesis of "Nash equilibrium", which assumes that a mono strain adjusts its own metabolic strategy according to the strategies of other strains to maximize growth rate and save shadow price [1,9]. NECom reflects the driving force on the community resulting from the metabolic goals of individual strains in the state of strain competition [1,9], but the algorithm is unable to simulate cases that do not reach a steady state of "Nash equilibrium".

The aforementioned hypothesis-based algorithms have made significant progress in achieving genome-scale simulation of polymicrobial metabolism while maintaining low computational costs. However, these algorithms rarely consider the influence of environmental variables on strain interactions, and their simplifying hypothesis impose significant limitations.

In addition, the algorithms that can simulate the environment influence on interactions, such as dFBA [10], closely dependent on a large amount of experimental data. For example, the parameters associated with the maximum and minimum substrate uptake rates, as well as the undissociated substrate constant, should be estimated through the dynamic parameter estimation approach [11]. Moreover, batch-specific parameters associated with substrate uptake kinetics should be estimated through the non-linear regression of experimental values for the undissociated substrate concentration and total substrate concentration [11]. Consequently, the experimental cost of dFBA is high, and the simulation of spatial–temporal dynamics is not possible.

To study the spatial–temporal influence of environmental variables, such as substrate concentration, agitation rate, inoculum amount, micronutrients, and toxins, on the interactions between two strains, this research established five modules based on the resource competition relationship. These modules encompass cellular physicochemical properties, nutrient uptake and metabolite release, cellular survival, cell swimming, and substrate diffusion. By coupling these modules and conducting data mining, the spatial–temporal interactions between two strains and the environment were simulated with a low experimental cost. Furthermore, the research explored the spatial–temporal evolution of glucose competition levels between strains in a consortium of *Escherichia coli* and *Saccharomyces cerevisiae*. The simulation took into account the influences of high-dimensional environmental and strain features on the competition level. Importantly, the simulation steps are simplified and do not require feature downscaling. The code for the analysis of this research is available at ziyuanzu/STPGMI (github.com).

2. Methods

2.1. Symbols and descriptions

The descriptions of the symbols in the methods are shown in Table 1.

2.2. Automated method of metabolic networks construction for strain–strain consortia

The k_{cat} values [12,13] extracted by Mao et al. was used for the *E. coli* iML1515 model, and the iMM904.xml metabolic network was downloaded from BIGG database (BiGG Models (ucsd.edu)). The AutoPACMEN toolbox moreover automatically extracted the k_{cat} values [14] for the *S. cerevisiae* iMM904 model. As there was no uniprot linkage for the enzymes in iMM904.xml to automatically extract molecular mass information, the relative masses of all enzymes were set to a uniform value of 36.8389 based on the median value of *E. coli* iML1515.

2.3. Simulation of spatial–temporal interactions between two strains and environment

In this research, five modules were designed based upon modularization to study the effects of strain metabolism, strain reproduction and death, substrate exchange, substrate concentration and agitation condition on the spatial–temporal interactions between strains and the environment for strain–strain consortia. The modules consist of the cellular physicochemical properties module, the nutrient uptake and metabolite release module, the cellular survival module, the cell swimming module and the substrate diffusion module. In addition, the spatial–temporal interactions between two strains and the environment were simulated through module coupling and data mining. This research used pyomo platform to perform the simulation of the cellular physicochemical properties module, the nutrient uptake and metabolite release module, the cellular survival module, the cell swimming module, and the substrate diffusion module.

2.3.1. Module of cellular physicochemical properties simulation

A module was designed to simulate the cellular metabolism of strain–strain consortia based on cellular physicochemical properties. In the module, the program package [13] of Yang et al. simulated the intracellular metabolic fluxes by limiting the solution space with the FBA [15] and kinetic constraints [13]. The relevant algorithms are shown in Eqs. (1)–(5). For choosing among multiple solutions of a multi-objective optimization problem, the solutions with the highest growth rate would be the candidates. Then, an approximate enzyme cost function, a sum of absolute fluxes, was defined to select the candidate solutions with the lowest enzyme cost, while the metabolite levels were limited to the physiological range.

The target functions

$$maximize v_{bio} \tag{1}$$

$$minimize v_{total} \tag{2}$$

The FBA constraint

$$0 = nf_{met} = S \bullet \vec{v} \tag{3}$$

$$LB \leq \vec{v} \leq UB$$

$$\vec{v} \geq 0$$

The kinetic constraints

$$e_i = \frac{v_i \bullet MW_i}{\sigma n k_{cat,i}} \tag{4}$$

Table 1
The symbols and descriptions in the methods.

Symbols	Descriptions	Symbols	Descriptions
i	The serial number of a reaction	$N^{str,rep}$	The number of reproducing strains
s	Refers to a substrate	N^{e}	The environmental capacity
t	the time	N^{str}	The number of strains
x	A space point on a one-dimensional space axis	$N_s^{str,de}$	The number of the death strains determined by a mono substrate
j	The serial number of a space point	$N^{str}_{j1\rightarrow j2}$	The number of microorganisms swimming from the j1st grid to the j2nd grid
l	The length of the one-dimensional space	N^{swim}	The number of the strains swimming out of a grid
Δt	The time interval between iterations	N^{grid}	The number of spatial grids
Δx	The length of the spatial grid	ΔN^{str}	The square of the relative change rate of the strain concentration in a grid
v_{bio}	The biomass yield	$V_{\max,nu}$	The maximum reaction rate of the nutrient uptake reaction (set to a constant value of twenty)
v_{total}	The sum of the fluxes	$K_{M,nu}$	The Michaelis–Menten constant of the nutrient uptake reaction (set to a constant value of two)
$\vec{v}$	The flux vector	D	The diffusion coefficient
v_i	The flux of the ith reaction	MC^{e}_{glc}	The average glucose concentration available for strains in a grid
v_s^{rel}	The release rate of substrate	*ths*	The set threshold
nf_{met}	The vector of net fluxes of metabolites	p	The parameter to determine when to stop the iteration
S	The stoichiometric matrix	MN^{str}	The average number of the mono strain
LB and UB	The vector of lower bounds and upper bounds of fluxes	cs^{str}	The competitiveness for resource of a mono strain
UB_{nu}	The upper bound of the nutrient uptake rates	r^{str}	The relative abundance of a mono strain
e_i	The enzyme cost of the ith reaction	R^2	The determination coefficient
MW_i	The enzyme molecular weight of the ith reaction	y	The true value of the sample
σ	The average saturation of all enzymes	$\widehat{y}$	The predicted value of the sample
n	The number of reactions	$\overline{y}$	The mean value of the sample
k_{cat}	The turnover number of the enzymes	y_p	The predicted value for the pth sample
p_pool	The intracellular protein	y_{base}	The baseline
e_pool	The enzyme content	$x_{p,q}$	The qth feature of the pth sample
e_pool_{eff}	The effective enzyme pool variable	$f(x_{p,q})$	The SHAP value of $x_{p,q}$, indicating the contribution of $x_{p,q}$ to the predicted value y_p. A positive $f(x_{p,q})$ indicates that the feature has a positive effect on the predicted value; conversely, it has an inverse effect
f	The enzyme proportions in proteins	C_O^h	The dissolved oxygen concentration in hydrostatic water
C_s^{intr}	The intracellular concentration of a mono substrate	P_O	The atmospheric pressure
C_s^e	The concentration of the substrate in environment	VF_o	The volume fraction of oxygen in air
C^e_{bio}	The concentration of biomass in environment	K	The Henry constant of oxygen dissolved in water
C^e_{glc}	The glucose concentration in a grid	ds_w	The density of water
ΔC_s	The concentration of the substrate consumed by cellular metabolism	MW_w	The molar mass of a water molecule
$Cycle^{str,rep}$	The reproduction cycle	C_O^{nh}	The dissolved oxygen concentration in none-hydrostatic water
rep^{str}	The strain reproduction rate	D^{nh}	The diffusion coefficient in none-hydrostatic water
de^{str}	The strain death rate	D^h	The diffusion coefficient in hydrostatic water
de_s^{str}	The strain death rate determined by a mono substrate		

$$\sigma \bullet \sum_{i=1}^{n} n \bullet e_i \le e\_pool_{eff} \tag{5}$$

2.3.2. Module of cellular survival simulation

The cellular survival module was designed to simulate the reproduction rates, as shown in Eqs. (6)–(7), and to simulate the death rates, as shown in Eqs. (8)–(10), for the strains in all spatial grids at different time points.

$$N^{str,rep}_{x,t} = \min_s\left(\min\left(N^{e}_{s,x,t}, N^{str}_{s,x,t}\right)\Delta t \Big/ Cycle^{str,rep}\right) \tag{6}$$

The reproduction rate at grid j at time point t was calculated from Eq. (6), as shown in Eq. (7).

$$rep^{str}_{x,t} = N^{str,rep}_{x,t} \Big/ N^{str}_{x,t} \tag{7}$$

$$\Delta C_{s,x,t}\left(N^{str}_{x,t} - N^{str,de}_{s,x,t}\right) = C^{e}_{s,x,t} + N^{str,de}_{s,x,t} C^{intr}_{s,x,t} \tag{8}$$

The death rate at grid j at time point t was calculated from Eq. (8), as shown in Eq. (10).

$$de^{str}_{s,x,t} = \frac{N^{str,de}_{s,x,t}}{N^{str}_{x,t}} = \frac{N^{str}_{x,t} \bullet \Delta C_{s,x,t} - C^{e}_{s,x,t}}{N^{str}_{x,t} \bullet (\Delta C_{s,x,t} + C^{intr}_{s,x,t})} \tag{9}$$

$$de^{str}_{x,t} = \max_s(de^{str}_{s,x,t}) \tag{10}$$

2.3.3. Module of nutrient uptake and metabolite release simulation

The nutrient uptake and metabolite release simulation module was designed to simulate the interaction between strain metabolism and environmental substrate concentrations based on the method of Harcombe et al. [16]. The effects of the processes of microbial nutrient uptake and metabolite release on the environmental substrate concentrations are shown in Eqs. (11)–(12), and the effect of environmental substrate concentrations on the upper bound of nutrient uptake rates is shown in Eq. (13).

$$C^{e,nume}_{bio,x,t+\Delta t} = C^{e,nume}_{bio,x,t} + N^{str}_{x,t} \bullet v_{bio,x,t} \bullet \Delta t \tag{11}$$

$$C^{e,nume}_{s,x,t+\Delta t} = C^{e,nume}_{s,x,t} + v^{rel}_{s,x,t} \bullet N^{str}_{x,t} \bullet \Delta t \tag{12}$$

$$UB_{nu,x,t} = \frac{V_{\max,nu\bullet} \quad C^{e,nume}_{s,x,t}}{C^{e,nume}_{s,x,t} + K_{M,nu}} \tag{13}$$

2.3.4. Module of substrate diffusion simulation

The distribution of substrate concentrations can be calculated based on Fick's second law [17] (as shown in Eq. (14)), the substrate diffusion module to simulate the substrate concentrations at different time points is shown in Eqs. (15)–(20). The module transformed the first-order derivative of substrate concentrations on time scale into the second-order derivative on spatial scale.

$$\frac{\partial C^{e,dif}_s}{\partial t} = D\frac{\partial^2 C^{e,dif}_s}{\partial x^2} \tag{14}$$

The Eq. (14) was expressed in a differential form for iterative calculation, as shown in Eqs. (15)-(17).

$$\frac{C^{e,dif}_{s,\ x+\Delta x,t+\Delta t} - C^{e,dif}_{s,\ x+\Delta x,t}}{\Delta t} = D\frac{C^{e,dif}_{s,\ x+2\Delta x,t} - 2C^{e,dif}_{s,\ x+\Delta x,t} + C^{e,dif}_{s,\ x,t}}{\Delta x^2} \quad (0 \le x \le l-2\Delta x) \tag{15}$$

$$\frac{C^{e,dif}_{s,\ 0,t+\Delta t} - C^{e,dif}_{s,\ 0,t}}{\Delta t} = \frac{C^{e,dif}_{s,\ \Delta x,t+\Delta t} - C^{e,dif}_{s,\ \Delta x,t}}{\Delta t} \tag{16}$$

$$\frac{C^{e,dif}_{s,\ l,t+\Delta t} - C^{e,dif}_{s,\ l,t}}{\Delta t} = \frac{C^{e,dif}_{s,\ l-\Delta x,t+\Delta t} - C^{e,dif}_{s,l-\quad \Delta x,t}}{\Delta t} \tag{17}$$

According to Eqs. (15)-(17), the substrate concentration $C^{e,dif}_{s,t+\Delta t}$ at the next moment was calculated from $C^{e,dif}_{s,t}$ at the previous moment, as shown in equation Eqs. (18)-(20).

$$C^{e,dif}_{s,\ x+\Delta x,t+\Delta t} = C^{e,dif}_{s,\ x+\Delta x,t} + D\frac{C^{e,dif}_{s,\ x+2\Delta x,t} - 2C^{e,dif}_{s,\ x+\Delta x,t} + C^{e,dif}_{s,\ x,t}}{\Delta x^2}\Delta t \quad (0 \le x \le l-2\Delta x) \tag{18}$$

$$C^{e,dif}_{s,\ 0,t+\Delta t} = C^{e,dif}_{s,\ 0,t} + C^{e,dif}_{s,\ \Delta x,t+\Delta t} - C^{e,dif}_{s,\ \Delta x,t} \tag{19}$$

$$C^{e,dif}_{s,\ l,t+\Delta t} = C^{e,dif}_{s,\ l,t} + C^{e,dif}_{s,\ l-\Delta x,t+\Delta t} - C^{e,dif}_{s,l-\quad \Delta x,t} \tag{20}$$

2.3.5. Module of cell swimming simulation

The cell swimming module was designed to simulate the random swimming of strains to obtain nutrients. When the average nutrient concentration in the adjacent grid is greater than that of the current grid, the swimming proportion is determined based on the relative difference in average concentrations and a set threshold, as shown in Eqs. (21)–(26):

$$MC^e_{glc,x} = C^e_{glc,x} \Big/ N^{str}_x \tag{21}$$

$$N^{str}_{x\to x+\Delta x} = \min\left(ths, \frac{MC^e_{glc,x+\Delta x} - MC^e_{glc,x}}{MC^e_{glc,x}}\right) \bullet N^{str}_x \quad (MC^e_{glc,x-\Delta x} \le MC^e_{glc,x} < MC^e_{glc,x+\Delta x}, \quad \Delta x \le x \le l-\Delta x \ ; \ or MC^e_{glc,x} < MC^e_{glc,x+\Delta x} \ , \ 0 \le x < \Delta x) \tag{22}$$

$$N^{str}_{x\to x-\Delta x} = \min\left(ths, \frac{MC^e_{glc,x-\Delta x} - MC^e_{glc,x}}{MC^e_{glc,x}}\right) \bullet N^{str}_x \quad (MC^e_{glc,x-\Delta x} > MC^e_{glc,x} \ge MC^e_{glc,x+\Delta x}, \quad \Delta x \le x \le l-\Delta x \ ; \ or MC^e_{glc,x-\Delta x} > MC^e_{glc,x} \ , \ l-\Delta x < x \le l) \tag{23}$$

$$N^{swim}_x = \min\left(ths, \frac{MC^e_{glc,x+\Delta x} + MC^e_{glc,x-\Delta x} - 2MC^e_{glc,x}}{MC^e_{glc,x}}\right) \bullet N^{str}_x \tag{24}$$

$$N^{str}_{x\to x+\Delta x} = \frac{MC^e_{glc,x+\Delta x} - MC^e_{glc,x}}{MC^e_{glc,x+\Delta x} + MC^e_{glc,x-\Delta x} - 2MC^e_{glc,x}} \bullet N^{swim}_x \quad (MC^e_{glc,x-\Delta x} > MC^e_{glc,x} < MC^e_{glc,x+\Delta x}, \quad \Delta x \le x \le l-\Delta x) \tag{25}$$

$$N^{str}_{x\to x-\Delta x} = \frac{MC^e_{glc,x-\Delta x} - MC^e_{glc,x}}{MC^e_{glc,x+\Delta x} + MC^e_{glc,x-\Delta x} - 2MC^e_{glc,x}} \bullet N^{swim}_x \quad (MC^e_{glc,x-\Delta x} > MC^e_{glc,x} < MC^e_{glc,x+\Delta x}, \quad \Delta x \le x \le l-\Delta x) \tag{26}$$

2.3.6. Coupling of the modules

The modules were coupled through the substrate concentration distribution and strain concentration distribution. When the two concentration distributions at the initial moment were defined, the strain interactions could be simulated through iteration over the entire time scale.

Firstly, at each time point, the effect of the swimming strain concentrations in the cell swimming simulation module on the strain concentration distribution was simulated based on the initial strain concentration distribution and initial substrate concentration distribution at the time point. Then, in the nutrient uptake and metabolite release simulation module, the upper bounds of the substrate uptake rates were calculated from the substrate concentration so that the metabolism fluxes could be simulated by the cellular physicochemical properties simulation module. At the same time, in this nutrient uptake and metabolite release simulation module, the substrate uptake rate, the metabolite release rate and the nutrient requirements of the strains were calculated through the simulated fluxes.

Later, in the cellular survival simulation module, the reproduction strain concentration and death strain concentration were calculated based on the nutrient requirements and the strain concentration. Next, based on the substrate exchange rates, the reproduction strain concentration and the death strain concentration, the effect of the strains on the substrate concentration distribution was calculated. Next, in the substrate diffusion simulation module, the effect of substrate diffusion on substrate concentration distribution was calculated based on the initial substrate concentration distribution of each time point. Thus, the final substrate concentration distribution was calculated based on the effect of the cell and the substrate diffusion. Lastly, the final strain concentration distributions were calculated based on the effect of strain swimming, the reproduction strain concentration and the death strain concentration. The calculation of the next time point was started and iterated until the end.

2.3.7. The initial condition of the strain concentrations and substrate concentrations

At the initial time, the strain concentrations were obtained by randomly sampling for two strains in each space grid. Furthermore, the initial substrate concentrations were set to a dimensionless number in each space grid.

2.3.8. The calculation method for dissolved oxygen concentration

The dissolved oxygen concentration was calculated by Henry's law [18]. Firstly, for the conditions in hydrostatic water, the dissolved oxygen concentration was calculated by the Henry coefficient, as shown in Eq. (27).

$$C^h_O = P_O * VF_o / K * ds_w / MW_w \tag{27}$$

Then, the dissolved oxygen concentration under non-hydrostatic water was calculated by the diffusion coefficient, as shown in Eq. (28).

Table 2
The characterizations of the coupled modules.

Modules	Characterizations
The cellular physicochemical properties module	Simulate the metabolic fluxes and the concentration of intermediate metabolites in the cells under a steady-state condition based on flux balance analysis and enzyme kinetic constraints. Calculate the substrate requirements for normal growth and reproduction, as well as the release of intracellular metabolites at death.
The cellular survival module	Simulate the reproduction rate and death rate of strains at any time point based on nutrient and strain concentrations.
The nutrient uptake and metabolite release module	Calculate the fluxes of substrate exchange reactions between strains and the environment to simulate the nutrient uptake and metabolite release processes of strains.
The substrate diffusion module	Simulate the effect of substrate diffusion processes on strain interactions in a concise form of a linear equation.
The cell swimming module	Simulate the feeding behavior of two strains. Together with the first four modules, it simulates the resource competition level of strains under specific conditions.

$$C_O^{nh} = C_O^h * D^{nh}/D^h \tag{28}$$

2.3.9. Determination condition for the spatial–temporal steady state of strain–environment interaction relationships

The grids were divided in a one-dimensional space. The parameter p•Δt was assumed to be the threshold of the sum of the relative change rates of strain concentrations in all grids in each iteration relative to the previous iteration, as shown in Eqs. (29)–(30). When p is zero for five consecutive iterations, the consortium is considered to have reached a steady state and the iteration is stopped.

$$\Delta N_{j,t}^{str} = \begin{cases} \left(\left(N_{j,t}^{str} - N_{j,t-1}^{str}\right)/N_{j,t-1}^{str}\right)^2 & \left(t \geq 1, \quad N_{j,t-1}^{str} > 0\right) \\ \left(N_{j,t}^{str}\right)^2 & \left(t \geq 1, \quad N_{j,t-1}^{str} = 0\right) \end{cases} \tag{29}$$

$$\sum_{j=1}^{N^{grid}} \sum_{str=}^{2} {}_1 \Delta N_{j,t}^{str} \leq p \bullet \Delta t \tag{30}$$

2.3.10. Quantification of the competitiveness of a mono strain

The relative abundance calculation achieved the quantification of a mono strain competitiveness for resource in a strain–strain consortium, as shown in Eqs. (31)–(33):

$$MN_t^{str} = \sum_{j=1}^{N^{grid}} N_{j,\ t}^{str} / N^{grid} \tag{31}$$

$$r_t^{str} = MN_t^{str} \Big/ \sum_{str=1}^{2} MN_t^{str} \quad t \geq 1 \tag{32}$$

$$cs^{str} = average\left(r_t^{str}\right) if MN_t^1 + MN_t^2 > 0 \tag{33}$$

2.3.11. Data mining approach to strain–environment interactions for strain–strain consortia

The FLAML automated machine learning method [19] was used to simulate the influence of various features on the strain–environment interaction relationships for strain–strain consortia. The fit of the machine learning method was assessed by the determination coefficient, as shown in Eq. (34).

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{34}$$

The SHAP feature importance analysis [20] was performed to analyze the rules extracted by machine learning method. The SHAP values are shown in Eq. (35):

$$y_p = y_{base} + f(x_{p,1}) + f(x_{p,2}) + \cdots + f(x_{p,q}) \tag{35}$$

3. Results and discussion

3.1. Design and characterization of the algorithm simulating the spatial–temporal interactions between two strains and environment

Global simulation of interactions between strains and environment in co-culture consortia is important for understanding strain proportions, distribution, and consortia evolution. This, in turn, provides reasonable guidance for optimizing production conditions. Currently, existing algorithms frequently neglect the influence of environmental variables on microbial interactions, which limits the investigation of spatial–temporal dynamics and potentially reduce simulation accuracy. To enhance the simulation of strain–environment interactions, a combination of module coupling and data mining techniques was used to explore the spatial–temporal interactions facilitated by resource competition mechanisms.

The characterizations of the coupled modules are listed in Table 2. Importantly, the coupling of the modules provided an in-depth study of the impact of cell survival and metabolic demands on the spatial–temporal evolution of resource competition in industrial production. Moreover, the computation partition of the modules reduced the algorithm redundancy and improved the algorithm readability. Furthermore, the data mining of strain–environment interactions provided insights into the mechanisms of high-dimensional environmental features on strain competitiveness for resource.

3.2. The calculated dissolved oxygen concentration

In this research, 0.02 cm^2/h was considered to be the diffusion coefficient in hydrostatic water. The empirical Henry constants of dissolved oxygen in 25 ℃, 30 ℃ and 35 ℃ are shown in Table 3.

For the agitation conditions, the calculated dissolved oxygen concentrations at different diffusion and temperature conditions are shown in Table 4.

The calculated dissolved oxygen concentrations were used in the upper bounds of oxygen uptake rates for strains. At a same temperature, the dissolved oxygen concentration increased with the substrate diffusion coefficient increased. To analyze the effect of a mono feature on the strain–strain consortia, the strain competitiveness level at the same dissolved oxygen concentration and different substrate diffusion coefficients were calculated (see Supplemental_Table S1). Moreover, the same dissolved oxygen concentration at different substrate diffusion

Table 3
The empirical Henry constants of dissolved oxygen in 25 ℃, 30 ℃ and 35 ℃.

Temperature (℃)	Henry constant (E × 10^{-6} kPa)
25	4.44
30	4.81
35	5.14

Table 4
The calculated dissolved oxygen concentrations at different diffusion and temperature conditions.

Temperature (℃)	Diffusion coefficient (cm^2/h)	Dissolved oxygen concentration (mM)
25	0.02	2.65
25	0.06	7.95
25	0.1	13.25
30	0.02	2.45
30	0.06	7.35
30	0.1	12.25
35	0.02	2.29
35	0.06	6.87
35	0.1	11.45

coefficients could be maintained by aeration control, temperature control, and other operations.

3.3. Spatial–temporal evolutionary patterns of strain–strain consortia based on resource competition mechanism

In order to deeply investigate the influence of cell survival and metabolic demands on strain resource competition in industrial production, and to accurately predict the dominant strain, the spatial–temporal evolution processes of the strain concentration, relative abundance, and glucose consumption in an *E. coli–S. cerevisiae* consortium [21] were simulated. Furthermore, the simulation of the spatial–temporal evolution was hoping to guide the product biosynthesis process of the consortium. Based on the coupling of the five modules, the spatial–temporal evolution of strain reproduction, death, and distribution in resource competition was explored, particularly under conditions of initial strain distribution heterogeneity.

Since *E. coli* and *S. cerevisiae* reproduced every 20 and 120 min, respectively, the $\Delta t/Cycle_{breed}$ of 1/5 and 1/30 was set for *E. coli* and *S. cerevisiae* respectively in each spatial grid and each iteration (interval four minutes). For the simulated physical domain, 60 grids were divided in a one dimensional space with 30 cm length. The initial concentrations of all common metabolites in the iML1515 and iMM904 models were set to a dimensionless number of three hundred. Because there was no nutrient supplementation, both strains underwent a decay period after a platform period with the concentrations of both strains approximated to zero. The substrate diffusion coefficient D in the substrate diffusion module was set to 0.02 cm^2/h. The initial metabolite concentration was set to 300 mM, and the initial strain concentration was a random number between 1 mM and 20 mM in each space grid. The spatial–temporal changes of the total strain concentration and the glucose concentration, the net increase rate of the total strain concentration, and the changes of the average glucose concentration at 35 ℃ are shown in Fig. 1.

From Fig. 1a and b, it is observed that the spatial distributions of glucose concentrations and strain concentrations tend to be homogeneous as the time goes on. These indicate that substrate diffusion process and cell swimming process for nutrients play important roles in homogenizing substrate distribution and strain distribution. The substrate diffusion caused the increasingly homogeneous distribution of glucose concentrations, and the cell swimming caused the strain concentrations gradually to be homogeneous with the glucose concentrations.

Furthermore, it is observed that starting from the fourth time point, the distribution of the total strain concentration closely resembled the distribution of glucose concentration at the previous time point. This observation indicates that cell swimming plays a significant role in driving the competition for glucose between strains. Both strains tend to compete for grids with high glucose concentrations while moving away from grids with low glucose concentrations.

From Fig. 1b and c, it can be found that in the first thirteen time points when glucose was sufficient, the demand for environment nutrients in the cellular physicochemical properties module was satisfied, the total strain concentration was rising. The total reproduction concentration of N^{rep} calculated by the cellular survival module was more than the total death concentration of N^{death}. As the glucose concentration decreased, the total net reproduction rate, $r_t^{\text{reproduction}} - r_t^{death}$ decreased from 10.0 % at the time point two to 3.74 % at the time point thirteen. After that the total strain concentration had a decreasing trend, with $N^{reproduction} < N^{death}$. As glucose almost depleted, the total net reproduction rate $r_t^{\text{reproduction}} - r_t^{death}$ decreased from − 11.3 % at time point fourteen to − 66.7 % at time point twenty-seven, after which the total strain concentration remained at a very low level. This suggests that the strains respond to substrate concentration by reproduction and death.

In order to accurately analyze the competition level for glucose of a mono strain and thus reasonably predict the dominant strain, the spatial–temporal changes of relative abundance of $r_{j,t}^{E.coli}$ and $r_{j,t}^{S.cerevisiae}$ and total relative abundance of $r_t^{E.coli}$ and $r_t^{S.cerevisiae}$ among the spatial grids were simulated, as shown in Fig. 2:

It can be found from Fig. 2a and b that $r_{j,t}^{E.coli}$ calculated by the cellular survival module and the cell swimming module was greater than $r_{j,t}^{S.cerevisiae}$ at most of the spatial grids with high glucose concentration at the previous time point. From Fig. 2c, it can be found that $r_t^{E.coli}$ gradually increased from 58.6 % to 100 %, and the *E. coli* competitiveness of $cs^{E.coli}$ was quantified to be 0.86. On the other hand, $r_t^{S.cerevisiae}$ gradually decreased from 41.40 % to 0 %, and the *S. cerevisiae* competitiveness of $cs^{S.cerevisiae}$ was 0.14. These indicate that the *E. coli* competitiveness is significantly higher than the *S. cerevisiae* competitiveness and the *E. coli* become the dominant strain.

The phenomenon validates the conclusions of Barber et al., who experimentally investigated the evolution of an *E. coli–S. cerevisiae* consortium. They found that in the first 420 generations, there were only two groups of the two strains coexisting, and the wild *E. coli* was significantly dominant in the remaining fifty-eight groups [21]. The reason for this phenomenon may be that the generation time of *E. coli* is only 1/6 of that of *S. cerevisiae*, and the strong ability to reproduce increases the *E. coli* competitiveness. Moreover, *E. coli* is a prokaryotic strain, which has a simple cellular structure, flexible metabolic regulation and greater adaptability to environment changes than the eukaryotic strain of *S. cerevisiae*. Qualitatively, the simulation method for spatial–temporal interactions between these two strains and the environment has a reasonable level of accuracy.

3.4. The influence of industrial production environment on the strain competitiveness

To analyze the impact of the industrial production environment on the competitive dynamics between different strains within consortia and to offer insights for optimizing the industrial production process, an exploration of the influence mechanisms of five key environmental factors on strain competitiveness was conducted. Specifically, the potential effects of substrate diffusion, dissolved oxygen concentration, inoculation ratio, micronutrient availability, and toxin presence on strain competitiveness were investigated.

The competitive dynamics within strain–strain consortia are significantly influenced by the industrial production environment, with substrate diffusion and dissolved oxygen concentration being particularly affected by operational parameters such as agitation rate, temperature, and aeration. The inoculation ratio of strains also critically shapes the trajectory of strain abundance, determined by their initial relative proportions. In this study, the intricate relationships between these environmental factors and strain competitiveness are examined, with a focus on the role of micronutrient uptake in strain performance. Micronutrients are integral to the growth process and are essential for sustaining normal metabolic functions in strains.

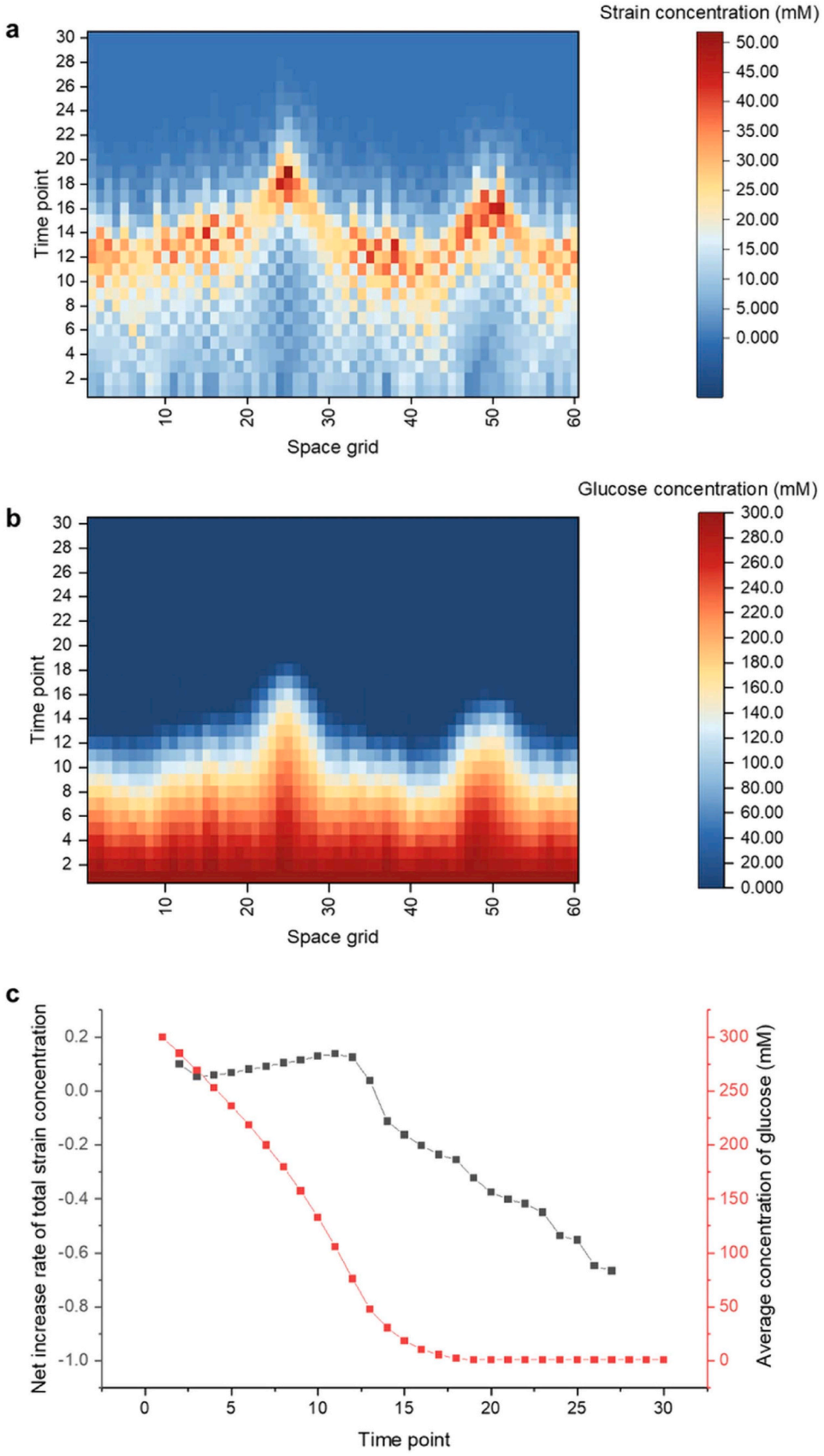

Fig. 1. Heatmaps of spatial–temporal changes of the total strain concentration and the glucose concentration, and scatterplots of the net increase rate of the total strain concentration and the change of the average glucose concentration (a– the spatial–temporal changes of the total strain concentration, b– the spatial–temporal changes of the glucose concentration, c– the net increase rate of the total strain concentration and the change of the average glucose concentration).

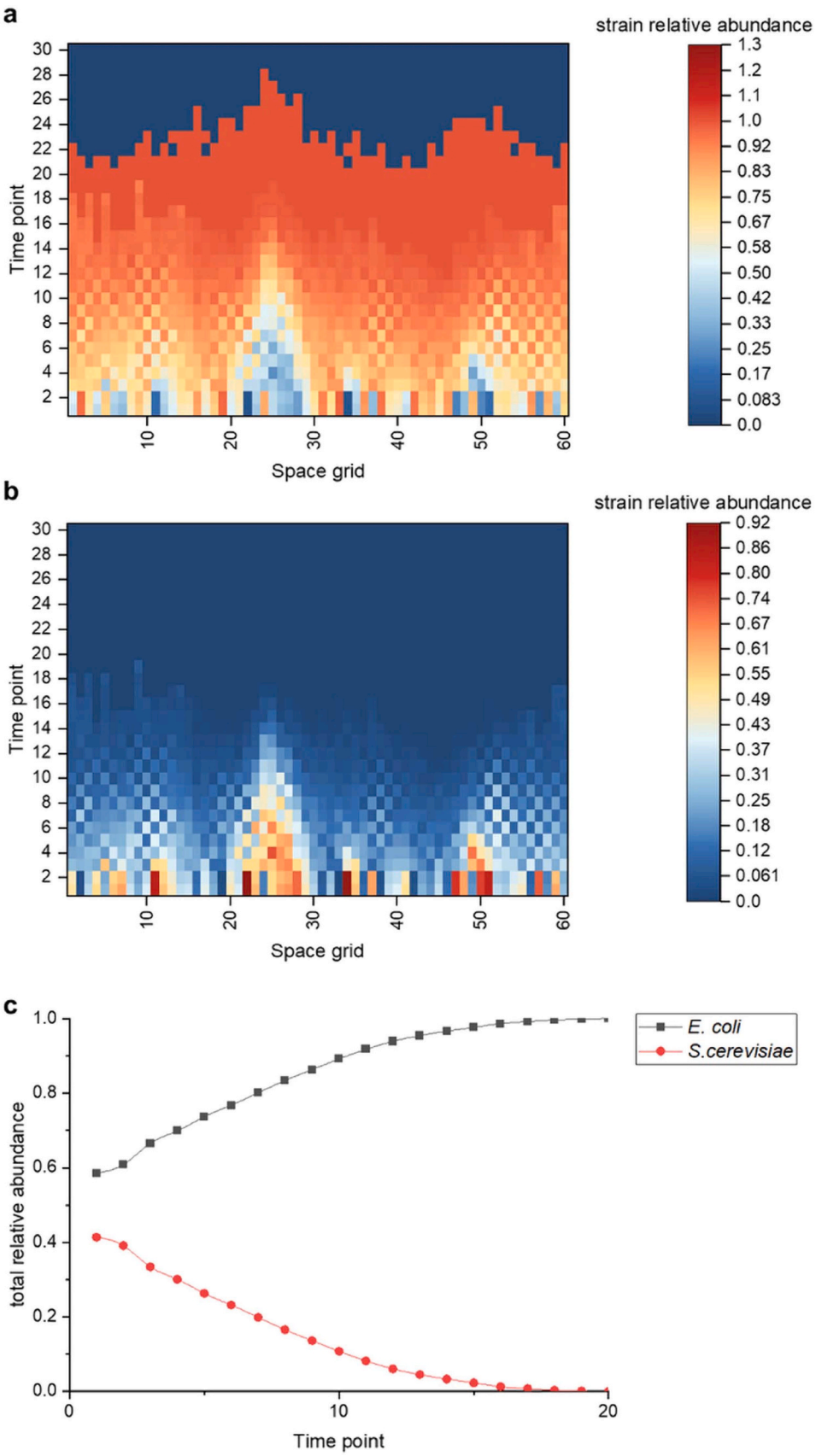

Fig. 2. Heatmaps of spatial–temporal changes of the spatial–temporal changes of relative abundance and scatterplots of the total relative abundance among the spatial grids for *E. coli* and *S. cerevisiae* (a– the spatial–temporal changes of relative abundance for *E. coli*, b– the spatial–temporal changes of relative abundance for *S. cerevisiae*, c– the total relative abundance among the spatial grids for *E. coli* and *S. cerevisiae*).

Micronutrients, such as iron, play a crucial role in the growth and metabolic processes of strains. Iron, an essential nutrient involved in vital pathways including respiration, nitrogen fixation, and DNA biosynthesis [22], is predominantly absorbed by strains in the soluble Fe^{2+} form [22]. It has been hypothesized that an increased availability of TonB protein for iron uptake may enhance the growth rate of *E. coli* [21].

Moreover, toxins released by strains can influence metabolic decisions, thereby affecting strain interaction patterns and abundance. It was hypothesized by Barber et al. that the *E. coli* reduced its uptake of bacteriocins secreted by *S. cerevisiae* by altering its own metabolic decisions, thereby enhancing its environmental toxin tolerance and growth levels [21]. However, this hypothesis was not supported by experimental data. In this research, the bacteriocin resistance of *E. coli* was analyzed to explore the inference mechanism of environmental toxins on strain competitiveness.

In this research, a comprehensive set of input features was defined to assess their impact on strain competitiveness within the consortium, which included the substrate diffusion coefficient, dissolved oxygen, the inoculation ratio of *S. cerevisiae*, the upper bound of Fe^{2+} uptake rates, the upper bound of Fe^{3+} uptake rates, and the bacteriocin resistance. The strain competitiveness strengths of $cs^{E.coli}$ and $cs^{S.cerevisiae}$ were set as the output features. The initial metabolite concentration was set to 300 mM, and the initial strain concentration is a random number between 1 mM and 20 mM in each space grid. A total of 138 pieces of data were simulated through the coupling of the five modules (see Supplemental_Table S1). The FLAML automated machine learning method mined the influences of the input features on the output features. It was found that RandomForestRegressor [23] had performed the best simulation and the determination coefficient between the predicted and simulated values in the validation set was 0.680, which indicated that RandomForestRegressor had extracted some rules. However, these rules require further translation into interpretable formats to facilitate a deeper understanding. Consequently, the study proceeded to analyze feature importance using the SHAP method, which quantified and elucidated the influence of each feature on the competitiveness of *E. coli* and *S. cerevisiae*, as depicted in Fig. 3.

The feature importance for $cs^{E.coli}$ and $cs^{S.cerevisiae}$ was found in the following order: the inoculation ratio of *S. cerevisiae* > dissolved oxygen > bacteriocin resistance > substrate diffusion coefficient > upper bound of Fe^{2+} uptake rates> upper bound of Fe^{3+} uptake rates. The influence of the environmental features of strain inoculation ratio and dissolved oxygen on the strain competitiveness was much larger than the influence of the strain metabolic properties. This indicates that the environmental features are crucial for the simulation accuracy of the strain competitiveness. Based on the range of SHAP values for each feature from Fig. 3, inoculation ratio of *S. cerevisiae* and dissolved oxygen were classified as strong influencing features, while the substrate diffusion coefficient, the upper bound of Fe^{3+} uptake rates, the upper bound of Fe^{2+} uptake rates, and the bacteriocin resistance were classified as weak influencing features.

The inoculation ratio of *S. cerevisiae* ($r_0^{S.cerevisiae}$) was the environmental feature with the greatest influence on the strain competitiveness. As $r_0^{S.cerevisiae}$ increased, $cs^{E.coli}$ gradually decreased and $cs^{S.cerevisiae}$ gradually increased. It shows that the optimization of strain inoculation ratio is able to significantly regulate the strain competition level in industrial production.

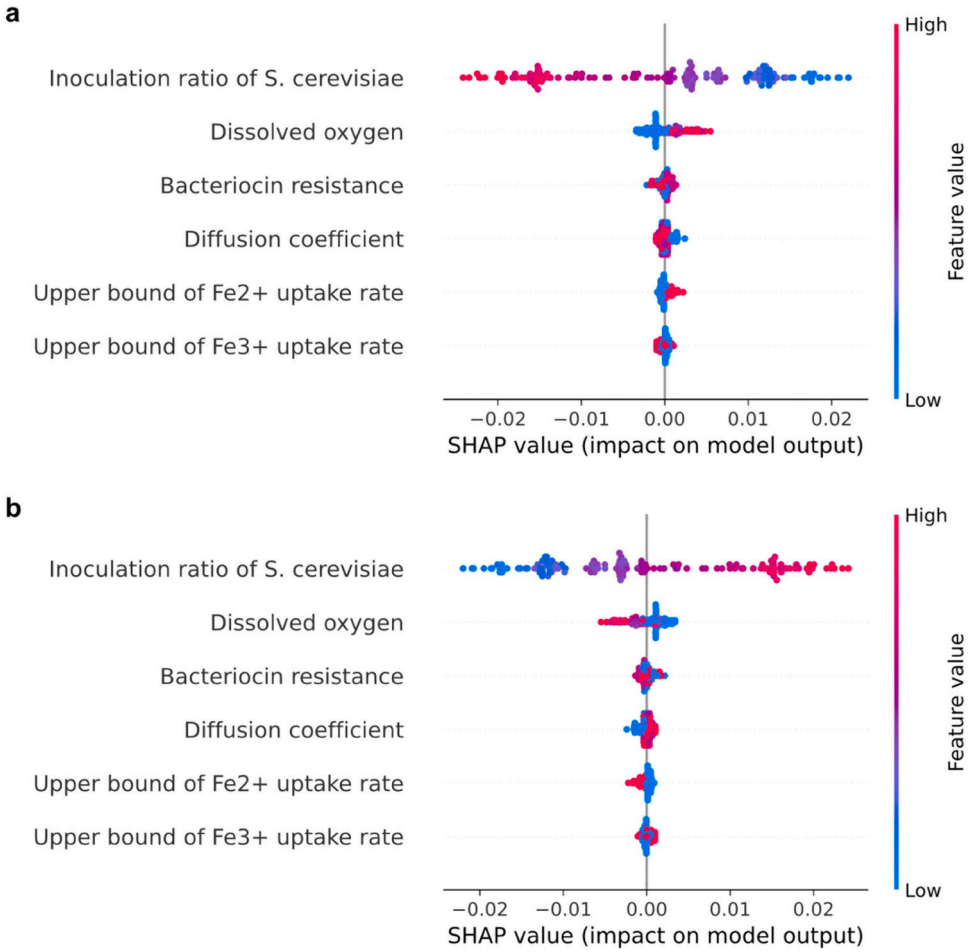

Fig. 3. Distribution of SHAP values for each feature for the competitiveness of the strains (a– *E. coli*, b– *S. cerevisiae*).

Dissolved oxygen had a positive effect on $cs^{E.coli}$, and had a negative effect on $cs^{S.cerevisiae}$. Interestingly, *E. coli* and *S. cerevisiae* are facultative aerobes, and oxygen can promote both of their growth. The disparity in their responses to oxygen levels may be attributed to the differences in their metabolic complexity. *E. coli*, being a prokaryotic organism, possesses a relatively simpler metabolic pathway, which can make it more sensitive to changes in oxygen availability. Consequently, oxygen may have a more pronounced effect on the energy production efficiency and growth rate of *E. coli* compared to the eukaryotic *S. cerevisiae*, which exhibits a more stable and complex metabolic profile. These findings highlight the significant impact of modulating dissolved oxygen concentrations through environmental control strategies, such as agitation, temperature, and aeration, on the competitiveness of strains in industrial fermentation processes.

The substrate diffusion coefficient (D) within the substrate diffusion module was identified as a weak influencing feature, yet its effects on *E. coli* ($cs^{E.coli}$) and *S. cerevisiae* ($cs^{S.cerevisiae}$) were distinct, with a negative impact on the former and a positive impact on the latter. This difference suggests that, within the operational parameters of industrial bioreactors, an increased agitation rate leads to a more uniform distribution of substrate concentration, and fewer strains swim due to glucose. Therefore, the less *S. cerevisiae* is affected by *E. coli*, and the competitiveness of *E. coli* and *S. cerevisiae* is weaker and greater, respectively.

After changing the upper bound of Fe^{2+} uptake rate of *E. coli* to 1.2 times of the original bound, there was a slight enhancement of $cs^{E.coli}$ and a slight decrease of $cs^{S.cerevisiae}$. This may be due to that the water-soluble nature of Fe^{2+} makes it less energy demand. Thus, the competitiveness is enhanced with the increase of the Fe^{2+} uptake capacity for *E. coli*. These findings imply that the competitiveness of strains can be strategically improved by genetically enhancing their Fe^{2+} uptake capabilities. Combined with the influence of Fe^{3+}, it also indicates that strains tend to be more competitive and use energy more efficiently when they absorb more soluble micronutrients, which provides a strategic direction for the optimization of engineered strains in industrial applications.

The effect directions of the upper bound of Fe^{3+} uptake rate and the bacteriocin resistance for *E. coli* on strain competitiveness are no clear. The reason about the upper bound of Fe^{3+} uptake rate may be that Fe^{3+} is extremely difficult to dissolve in water, and the strain metabolisms has low requirements for Fe^{3+}. In addition, the effect of bacteriocin resistance for *E. coli* on strain competitiveness may corroborate the experimental result that no bacteriocins were detected in the *E. coli–S. cerevisiae* consortium of Barber et al. [21].

4. Conclusions

To deeply understand the influence of industrial environment on co-culture consortia, the spatial–temporal interaction between strains and environment were simulated and data-mined. A novel framework was proposed, encompassing cellular physicochemical properties, nutrient uptake and metabolite release, cellular survival, cell swimming, and substrate diffusion within an *E. coli–S. cerevisiae* consortium. The results showed that the strains responded to substrate concentration changes by reproduction and death. As the substrate was consumed, the total net reproduction rate of the two strains gradually decreased from 10 % at the beginning to − 66.7 %, after which the total strain concentration remained at a very low level. The swimming caused the strains to continuously compete for the grid with high glucose concentration. The total relative abundance of *E. coli* increased to 100 % and gradually occupied most of the spatial grids with high glucose concentration, with a strain competitiveness strength of 0.86, while the total relative abundance of *S. cerevisiae* decreased to 0 % with a strain competitiveness strength of 0.14. Furthermore, data mining was employed to simulate the influence of the industrial production environment on strain competitiveness for resource. The results showed that inoculation ratio and dissolved oxygen content were strong features influencing the competitiveness. As the relative inoculum ratio increased, the strain competitiveness increased. As the dissolved oxygen content increased, the competitiveness of *E. coli* gradually increased and the competitiveness of *S. cerevisiae* gradually decreased. On the other hand, the substrate diffusion condition, the micronutrients and toxins were weak influencing features. This method provides new ideas for the simulation of spatial–temporal interactions and the optimization of production processes in industrial environment with two or more strains.

CRediT authorship contribution statement

Shaojie Wang: Writing – review & editing, Validation. **Haijia Su:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Chen Yang:** Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Boyuan Xue:** Writing – original draft, Methodology. **Qianqian Yuan:** Validation, Software.

Declaration of Competing Interest

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.

Acknowledgments

We express our thanks for funding support from National Key R&D Program of China (2021YFC2100700), National Natural Science Foundation of China (21838001, 31961133018), Fundamental Research Funds for the Central Universities (JD2428), and TIB BioDesign Center.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.06.033.

References

[1] Stubbendieck RM, Vargas-Bautista C, Straight PD. Bacterial communities: interactions to scale. Front Microbiol 2016;7:1234.
[2] Lou H, Hu L, Lu H, Wei T, Chen Q. Metabolic engineering of microbial cell factories for biosynthesis of flavonoids: a review. Molecules 2021;26:4522.
[3] Yuan S-F, Yi X, Johnston TG, Alper HS. De novo resveratrol production through modular engineering of an Escherichia coli–saccharomyces cerevisiae co-culture. Microb Cell Factor 2020;19:143.
[4] Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E. Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun 2011;2:589.
[5] Cai J, Tan T, Joshua Chan SH. Predicting Nash equilibria for microbial metabolic interactions. Bioinformatics 2020;36:5649–55.
[6] Zomorrodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 2012;8:e1002363.
[7] Chan SHJ, Simons MN, Maranas CD. SteadyCom: predicting microbial abundances while ensuring community stability. PLoS Comput Biol 2017;13:e1005539.
[8] Khandelwal RA, Olivier BG, Roling WF, Teusink B, Bruggeman FJ. Community flux balance analysis for microbial consortia at balanced growth. PLoS One 2013;8:e64567.
[9] Cai J, Tan T, Joshua Chan SH. Bridging evolutionary game theory and metabolic models for predicting microbial metabolic interactions. bioRxiv 2020.
[10] Henson MA, Hanly TJ. Dynamic flux balance analysis for synthetic microbial communities. IET Syst Biol 2014;8:214–29.
[11] Özcan E, Seven M, Şirin B, Çakır T, Nikerel E, Teusink B, Toksoy Öner E. Dynamic co-culture metabolic models reveal the fermentation dynamics, metabolic capacities and interplays of cheese starter cultures. Biotechnol Bioeng 2020;118:223–37.
[12] Mao Z, Zhao X, Yang X, Zhang P, Du J, Yuan Q, Ma H. ECMpy, a simplified workflow for constructing enzymatic constrained metabolic network model. Biomolecules 2022;12:65.
[13] Yang X, Mao Z, Zhao X, Wang R, Zhang P, Cai J, Xue C, Ma H. Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. Metab Eng 2021;67:133–44.
[14] Bekiaris PS, Klamt S. Automatic construction of metabolic models with enzyme constraints. BMC Bioinforma 2020;21:19.

[15] Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol 2010; 28:245–8.
[16] Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N, Mehta P, Marx CJ, Segre D. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Cell Rep 2014;7:1104–15.
[17] Lu Y, Li M. Simultaneous rapid determination of the solubility and diffusion coefficients of a poorly water-soluble drug based on a novel UV imaging system. J Pharm Sci 2016;105:131–8.
[18] Qi X, Sun X, Li J, Hu T, Fang J, Zhou L, Yu B, Sun L, Liu W, Cai X, Ding Y, Xie Y, Zuo G. Determination of Henry's law coefficient of oxygen in LAB for JUNO. J Instrum 2024;19:P03011.
[19] Wang C, Wu Q, Weimer M, Zhu E. Flaml: a fast and lightweight automl library. Proc 4th Mach Learn Syst Conf 2021:434–47.
[20] Lundberg S.M., Lee S.-I. (2017) A Unified Approach to Interpreting Model Predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA.
[21] Barber JN, Sezmis AL, Woods LC, Anderson TD, Voss JM, McDonald MJ. The evolution of coexistence from competition in experimental co-cultures of Escherichia coli and Saccharomyces cerevisiae. ISME J 2021;15:746–61.
[22] Jordi MP, José María M-N, José Manuel F-RJNrE. The role of iron in host–microbiota crosstalk and its effects on systemic glucose metabolism. Nat Rev Endocrinol 2022;18:683–98.
[23] Breiman L. Random forests. In: Schapire RE, editor. Machine Learning. Netherlands: Kluwer Academic Publishers; 2001. p. 5–32.