Discretised flux balance analysis for reaction-diffusion simulation of single-cell metabolism

Yin Hoon Chew and Fabian Spill

School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, England, United Kingdom.

*Corresponding author(s). E-mail(s): y.h.chew@bham.ac.uk; f.spill@bham.ac.uk;

Abstract

Metabolites have to diffuse within the sub-cellular compartments they occupy to specific locations where enzymes are, so reactions could occur. Conventional flux balance analysis (FBA), a method based on linear programming that is commonly used to model metabolism, implicitly assumes that all enzymatic reactions are not diffusion-limited though that may not always be the case. In this work, we have developed a spatial method that implements FBA on a grid-based system, to enable the exploration of diffusion effects on metabolism. Specifically, the method discretises a living cell into a two-dimensional grid, represents the metabolic reactions in each grid element as well as the diffusion of metabolites to and from neighbouring elements, and simulates the system as a single linear programming problem. We varied the number of rows and columns in the grid to simulate different cell shapes, and the method was able to capture diffusion effects at different shapes. We then used the method to simulate heterogeneous enzyme distribution, which suggested a theoretical effect on variability at the population level. We propose the use of this method, and its future extensions, to explore how spatiotemporal organisation of sub-cellular compartments and the molecules within could affect cell behaviour.

Keywords: flux balance analysis, grid-based, metabolism, reaction-diffusion, spatial method, cell shape

1 Introduction

Cellular metabolism is a set of biochemical reactions occurring inside living cells that supports growth and maintenance. Nutrient metabolites are transported into cells, then converted through a series of reactions into energy and building blocks of macromolecular components that make up cellular biomass. Reactions generally occur at regions where enzymes catalysing the reactions are localised. Therefore, metabolites have to diffuse to where enzymes are located to participate in reactions.

In cells, diffusion occurs as molecules collide constantly with one another in a crowded environment. In time, these collisions generate a net movement for each molecular species, from regions where its concentration is high to regions where its concentration is low. Therefore, factors such as the mass and size of the molecular species, and how crowded the environment is, affect the rate of diffusion (Minton, 2001; Gershon et al, 1985; Luby-Phelps et al, 1987). Metabolites are relatively small so they tend to have higher rates of diffusion compared to large molecules like enzymes. However, the effect of metabolite diffusion on the overall rate of reaction depends on diffusion rate relative to enzyme catalytic rate (Schnell and Turner, 2004). Consider the example of a simple single-substrate enzyme-catalysed reaction. Substrate (metabolite) diffusion and catalysis occurs in serial, so the overall rate of reaction is limited by the slower of the two events. When diffusion occurs much faster than enzyme catalysis, the reaction is said to operate at the reaction-limited regime. On the other hand, the reaction is said to operate at the diffusion-limited regime if

diffusion is much slower than enzyme catalysis. Which regime any particular reaction operates at may be dynamic because the sub-cellular environment in cells can vary a lot among different cell types and change as the cells grow. Interestingly, several recent studies suggested that some cells may have evolved to achieve optimal growth rates by adopting a molecular composition that provides an optimal (macro)molecular crowding; so as to balance a concentrated environment that supports high catalytic rates and room for movement that enables diffusion (Vazquez, 2010; Oldewurtel et al, 2021; Pang and Lercher, 2023; Dourado et al, 2021).

Modelling provides a useful tool to study complex reaction network like metabolism. One established method for modelling metabolic reactions is ordinary differential equations (ODEs). However, this method requires enzyme kinetic parameters whose values are only available for a small fraction of metabolic reactions, such as those in the central carbon metabolism (Teusink et al, 2000; Peskov et al, 2012; Jahan et al, 2016). This limits the use of ODEs for modelling large metabolic network. To circumvent the lack of kinetic measurements, flux balance analysis (FBA) has been commonly used especially for modelling metabolism at the genome scale (Orth et al, 2010). FBA uses linear programming to solve for the rates of reactions by optimising an objective function (that represents an assumed biological objective in each particular study), constrained by measurements such as nutrient uptake, gene expression and/or enzyme concentration. Due to the way linear programming works, reaction rates in the solution may not be unique or obey kinetic rate laws, but analysis of the solution space can provide useful information about the system. While the above are generic methods that can be adopted to model detailed events in metabolism, many metabolic modelling using these methods focus only on the reaction events. Some models consider multiple events and lump them into

simple effective functions to represent the overall rate but they also implicitly ignore effects of diffusion. For example, a common rate law used in ODE metabolic models is the Michaelis-Menten equation and its parameters are usually measured *in vitro*, which does not represent the crowded environment in cells where metabolite diffusion is slowed down (Chang et al, 2021; Wittig et al, 2018; Srinivasan, 2022). Genome-scale FBA models, on the other hand, are mostly concerned with (enzyme) capacity that constrains the rate instead of the exact rate because kinetic parameters are not easily available (Blazier and Papin, 2012; Machado and Herrgard, 2014). Estimation from various studies, including analysing the catalytic efficiency of >1000 different enzymes, suggested that ignoring diffusion effects may be valid in most cases because enzymatic reactions tend to occur at much lower rates compared to the diffusion of metabolites (Sweetlove and Fernie, 2018; Bar-Even et al, 2015). Thus, it is believed that most metabolic reactions are not diffusion-limited. However, there exists a small number of highly efficient enzymes, where reactions occur very fast such that metabolite diffusion becomes limiting (Pettersson, 1992; Sweetlove and Fernie, 2018). Diffusion-limited effects are exacerbated if the enzymes are in low abundance and heterogeneously distributed within particular cell shapes that require metabolites to travel a long distance or a narrow path (Kinsey et al, 2011; Szymanska et al, 2015).

Various approaches have been used to consider diffusion effects, directly and indirectly, in the modelling of metabolism. These include creating separate rate laws for diffusion-limited and reaction-limited regimes, adjusting the Michaelis-Menten constant to incorporate effects of molecular crowding, and developing a general theory of growth rate optimisation that balances mass conservation, reaction kinetics and limits on cellular density (that can affect diffusion) (Vazquez, 2010; Pang and Lercher, 2023; Dourado and Lercher, 2020). These

approaches made theoretical predictions of protein content and cellular density that would be optimal for growth, and it was found that the predicted values fell in the range of experimental measurements in bacterial cells such as *Escherichia coli*. This suggests that the principle of optimality that is assumed in methods such as FBA may be valid, at least under certain growth conditions. Additionally, the approaches above provide an improved representation of metabolism that occur in a crowded environment in cells. They have been useful for studying the quantitative relationship between molecular composition (hence diffusion indirectly) and growth at the cell scale. However, they are not suitable for exploring diffusion effects brought about by spatial properties such as cell shape and heterogeneous distribution of molecular species in the cell. Studies suggested that these effects are not always negligible (Kinsey et al, 2011). A common method for spatial modelling that accounts for both reaction and diffusion is partial differential equations (PDEs). Reaction is usually modelled using the law of mass action that relates reaction rate to substrate concentration, while diffusion is usually modelled using Fick's first law that relates diffusion rate to concentration gradient. Thus like ODEs, this method requires parameters such as reaction rate constants and diffusion coefficients. Acquiring their values is still a non-trivial problem. A recent work that used PDEs to model the energy metabolism in glial cells proposed a new method to enable simulation of complex cell shapes, and they intentionally left out parameter identification for this very reason (Farina et al, 2021).

To help circumvent the difficulty in obtaining parameter values, we present a method for spatial modelling of reaction-diffusion in single-cell metabolism that has low parameter requirements. Specifically, we expand FBA spatially, add representation of intracellular metabolite diffusion, and use linear programming to obtain rates of reaction and diffusion that would yield an optimal

value for an objective function. Note that our method is not an alternative to PDE or similar spatial modelling method, but an option for theoretical exploration when rate laws and data such as concentration and kinetic measurements are scarce. In the next section, we provide a description of our method. Then, we describe a simple set of toy metabolic reactions and the simulation settings that we used to test and demonstrate the utility of the method; followed by a discussion of the simulation results and their implications. We also suggest potential extension to this method for future work.

2 Discretised flux balance analysis

In its basic form, FBA assumes that metabolite concentrations are at steady state, which is generally valid because metabolism operates at a much faster time scale compared to other cellular processes such as transcription and translation (Orth et al, 2010). With this assumption, one can write the mass balance for each metabolite m as the following linear equation:

$$\sum_{r \in R} S_{m,r} v_r = 0 \tag{1}$$

where,

$$m \in M$$

M is the set of all metabolites and R is the set of all reactions in the model. $S_{m,r}$ represents the stoichiometric coefficient, which indicates the number of molecules of metabolite m that participate in reaction r relative to other participants. The stoichiometric coefficient takes a negative sign if the metabolite is a substrate in the reaction, and a positive sign if it is a product. v_r is the flux, i.e. the conversion rate of a substrate to product metabolites, of reaction r. Equation (1) can also be represented mathematically in matrix form as

follows:

$$S\boldsymbol{v} = 0 \tag{2}$$

where S is a matrix of stoichiometric coefficients with each row representing a metabolite and each column representing a reaction (Fig. 1c). $\boldsymbol{v}$ is then a column vector with each element representing a reaction flux variable. Like ODEs, FBA can be used to solve compartmental models by treating the same metabolite as a distinct species for each of the different compartments (or organelles) that the metabolite occupies; and incorporating transport event between compartments. Therefore, R includes not only biochemical reactions, but also metabolite transport; from extracellular compartment into the cell through the plasma membrane, or from one sub-cellular compartment to another. In this case, transport flux variable is the rate of metabolite movement between compartments, akin to "converting" a metabolite in one compartment to the same metabolite in another compartment. Thus, the stoichiometric coefficient takes the value of one, with a positive sign for metabolite in the target compartment and a negative sign for the origin compartment. When modelling single-cell metabolism, flux is in units of particle number (usually moles) per unit of time per cell.

In this work, we expand FBA to enable spatial modelling within single cells and include intracellular metabolite diffusion. To spatially represent a cell, we first simplify the representation of cell shape by assuming that each living cell is a rectangle. We discretise a cell into a two-dimensional I x J grid, where I (number of rows) represents cell width and J (number of columns) represents cell length (Fig. 1a). We term each square element on the grid a cell region k, which has the properties i and j that denote its coordinate on the grid. Similar to compartmental models, we treat the same metabolite as a distinct species for each of the cell regions. Thus, each cell region contains

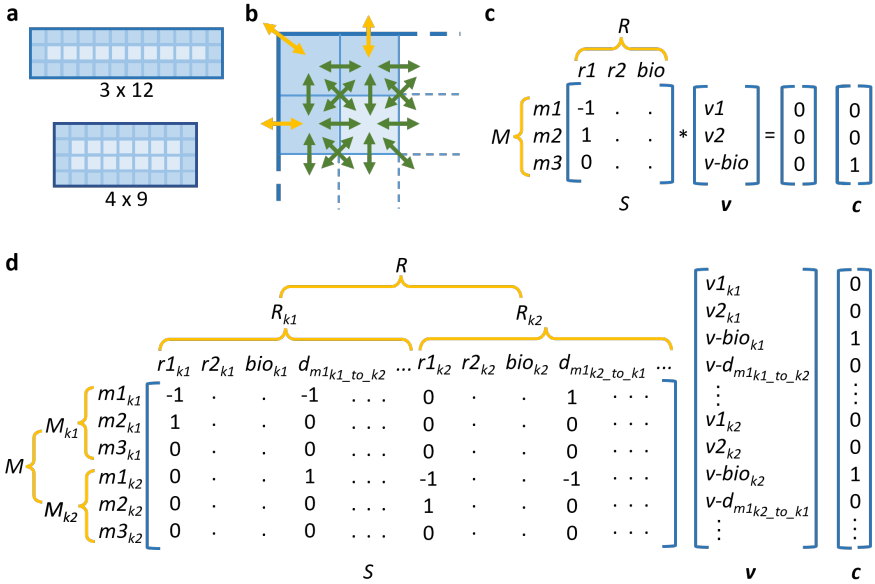

Fig. 1 a) In discretised flux balance analysis (FBA), a cell is discretised into a two-dimensional I x J grid, where I represents cell width and J represents cell length. Each square element on the grid is a cell region k. Cell regions can be grouped into layers, from cell boundary to the centre of the cell; b) Metabolites can diffuse in and out of each cell region to its neighbouring regions (green arrows). For regions at the boundary of a cell, metabolites can also be imported and exported to the extracellular environment (yellow arrows); c) In FBA, the metabolic network is represented as an M x R matrix of stoichiometric coefficients S, where each row represents a metabolite and each column represents a reaction. $\boldsymbol{v}$ is a column vector representing reaction flux variables. FBA assumes that the network is at steady state, and solves for the variables by optimising an objective function. $\boldsymbol{c}$ is a vector of weight for each variable in the objective function. In this example, *v-bio* is the objective function to be optimised; d) Discretised FBA expands conventional FBA by creating regional metabolites M_k and reactions R_k that also include metabolite diffusion d between regions. $\boldsymbol{v}$ and $\boldsymbol{c}$ are also expanded accordingly.

its own sets of metabolites M_k and reactions R_k. Conventional FBA explicitly represents transport of metabolites between compartments by including these events as part of the reaction set. In our discretised cell, intracellular diffusion can be considered a form of passive transport between cell regions. We can therefore add representation of intracellular metabolite diffusion as a natural extension to FBA. We do this in the following manner: 1) for every cell region k, diffusion of metabolites to and from its neighbouring regions are created and added to R_k; 2) if a cell region is located at the boundary between

cell and the extracellular compartment, transport of metabolites between compartments are also included in R_k (Fig. 1b). Similar to a transport event, the stoichiometric coefficient for a diffusion event takes the absolute value of one, and diffusion flux variable is the rate of net metabolite movement between cell regions. The new sets of metabolites M and reactions R for the whole cell are now the union of all cell-region sets M_k and R_k, respectively. Our method therefore expands the number of elements in the matrix of the original FBA problem, by decomposing each metabolite and reaction flux at the whole-cell scale into their constituents at the cell-region scale, and incorporating diffusion fluxes between regions (Fig. 1d). Flux in this case is in units of particle number (or moles) per unit of time per cell region.

Metabolic network usually contains a higher number of reactions than the number of metabolites. This creates an under-determined system where there are more unknown fluxes than there are equations. FBA uses linear programming to determine the unknown fluxes by adding inequality constraints and optimising an objective function. The linear programming problem can be formulated as follows:

$$\text{Maximise (or minimise):}$$
$$Z = \boldsymbol{c}^T \boldsymbol{v}$$
$$\text{Subject to:}$$
$$S\boldsymbol{v} = 0$$
$$l_r \leq v_r \leq u_r$$

Here, $\boldsymbol{c}$ is a vector of weights, or objective function coefficients, that is used to indicate the level of contribution of each flux to the objective function Z. A common objective function in modelling genome-scale metabolic network is the biomass reaction, which is a pseudo-reaction that represents cell growth (see

example in Section 3.1). This can be indicated by setting the weight of biomass reaction to one and all others to zero. As our method decomposes cell-scale flux variables into regional-scale variables, we need to re-compose the latter to generate the equivalent objective function. Using biomass reaction as an example, our method will set the weight of biomass reaction in every cell region to one, so that their fluxes will be added back to give whole-cell biomass growth (Figs. 1c and 1d). The optimised value of the objective function depends on the inequality constraints. Each flux variable v_r is constrained between a lower bound l_r and an upper bound u_r. Their values can be set based on experimental measurements or knowledge about the variable. For example, in the case of irreversible reactions that can only occur in the forward direction, one can set their lower bounds to zero. If no data or knowledge is available, variables can be left unbounded, i.e. the lower and upper bounds are set to negative infinity and positive infinity, respectively. For diffusion flux variables, their lower bounds are automatically set to zero because our method creates a separate variable for each direction between any two cell regions. Note that FBA does not have any notion of metabolite concentration; flux values in the optimal solution depend on the constraints set in each particular simulation. These constraints eliminate infeasible states to help narrow down the solution space. Within the feasible region, it is possible to have more than one set of solution that yield the same optimal value for the objective function. Depending on the accuracy of the constraints, some of these solutions may not be biologically feasible. Nevertheless, in the absence of parameters such as rate constants and diffusion coefficients, setting the objective function and constraints for flux variables in different ways allows one to theoretically explore the solution space and gain insights into the system (Loghmani et al, 2022).

Note that in general, FBA does not have explicit representation of mass, volume and surface area. The scale of the system to be modelled is indicated by the units of the constraints. Therefore, changing the number of cell regions, i.e. grid elements, does not necessarily indicate a change in cell size in our method. For example, one can increase the number of grid elements but sets the units of the constraints such that the sum over all regions will recover the same cell-scale units and values. Doing this does not change cell size but increases the spatial resolution of the simulation. One advantage of having spatial resolution is that we can use different constraints for the same reaction in different regions. This can be useful for studying local effects and how they influence the whole system.

Our method has been implemented as a Python API. Codes for the method, as well as the following toy model and simulations, are available on GitHub at https://github.com/YinHoon/discretised_FBA. Our API uses COBRApy, a constraint-based modeling package that is commonly used for running conventional FBA (Ebrahim et al, 2013).

3 Model description

In this section, we describe the toy metabolic model that we created specifically to test our method. We first present the model at the cell scale, i.e. the set of reactions and parameters at the organism level that is normally represented when using conventional FBA. We then describe how our method transforms cell-scale representations into regional-scale representations that include region-specific reactions and metabolite diffusion.

3.1 Cell-scale representation

The toy metabolic network consists of only a few reactions, as follows:

$$\longrightarrow \mathrm{A[e]} \tag{R1}$$

$$\mathrm{A[e]} \xrightarrow{\mathrm{E1}} \mathrm{A[c]} \tag{R2}$$

$$\mathrm{A[c]} \xrightarrow{\mathrm{E2}} \mathrm{B[c]} \tag{R3}$$

$$\mathrm{B[c]} \xrightarrow{\mathrm{E3}} \mathrm{Biomass} \tag{R4}$$

$$\mathrm{Biomass} \longrightarrow \tag{R5}$$

The model consists of two compartments (as denoted in square brackets): an extracellular compartment "e" and a cellular compartment "c". In reaction R2, metabolite A is transported from the extracellular compartment into the cellular compartment via a transporter E1. Metabolite A is then converted into B in the cell, catalysed by enzyme E2. Finally, B is converted into cell biomass, catalysed by enzyme E3. R1 and R5 are exchange reactions that act as the source and sink to balance out all other reactions in the model. R5 is the biomass reaction, a proxy of cell growth, that will be set as the objective function and maximised during simulation. All reactions are irreversible.

Constraints are required to simulate the model. Common types of data for constraining conventional FBA models include nutrient uptake rates, fluxomic, transcriptomic and proteomic data (Blazier and Papin, 2012; Machado and Herrgard, 2014; Cook and Nielsen, 2017). When using transcriptomic and proteomic data, an assumption made is that the level of gene expression indicates maximum enzyme capacity which limits the rate of reaction. For our simulation, we chose to constrain the fluxes of reactions R2, R3 and R4 whose rates are limited by the amount and efficiency of E1, E2 and E3, respectively. We picked these reactions so that we could demonstrate the utility of spatial

resolution in our method (see next section for details). We constrained the reactions by setting their upper bounds as follows:

$$u_r = k_{cat}[E] \tag{3}$$

where k_{cat} is the catalytic constant and $[E]$ is the concentration of enzyme (transporter) E that catalyses (facilitates) reaction r. We employed this equation because it is commonly used in enzyme kinetic studies to represent maximum enzyme capacity (Davidi et al, 2016; Heckmann et al, 2018, 2020). We set the total cellular concentration of each enzyme (or transporter) to 100 and their k_{cat} values to 1.0. We further set the upper bound of exchange reaction R1 to 100, as a proxy of available nutrient in the media. We chose this set of parameter values so that the maximum feasible growth per cell is 100 and no specific enzyme is limiting growth. Note that we are not assigning specific units in this toy example, but the parameter values are to be taken as being consistent in units.

3.2 Regional-scale representation

Given a cell-scale representation such as the one described above, our Python API converts it into a regional-scale representation using the following steps (see Appendix A for more details):

1. Create a two-dimensional I x J grid;
2. Create the metabolite objects in every cell region;
3. Create the reaction objects in every cell region;
4. Create an objective function that sums the cell-scale objective function over all cell regions;

5. Create the reaction objects that represent transport processes in every cell region located at the boundary between cell and the extracellular compartment;
6. Create the reaction objects that represent diffusion processes in every cell region;
7. Set the bounds of each reaction object.

In the regional-scale representation, the same intracellular reaction is represented in each cell region as a distinct reaction object with its own upper and lower bounds. This means that we can simulate the same reaction as having different local rates, which is a more realistic representation of what happens in cells given how heterogeneous the sub-cellular environment is. One of the factors that can give rise to different local reaction rates is the heterogeneous distribution of enzymes that catalyse the reaction. To enable simulation of this, we have included additional functions in our method to distribute enzymes and calculate the local upper bounds of enzyme-catalysed reactions using equation (3). Once the regional bounds have been calculated, they can be used in step (7) above to set the bounds of corresponding reaction objects (see Appendix B for more details).

Our method provides four options for distributing enzymes in the cell: 1) uniformly; 2) on a positive gradient (increasing concentration from the boundary to the centre of the cell); 3) on a negative gradient or; 4) randomly. For non-uniform enzyme distribution, our method uses the following to determine the fraction $[E]_k$ that each cell region will receive out of the total cellular concentration $[E]$:

$$[E]_k = \frac{n_k}{\sum_{k \in K_E} n_k}[E] \tag{4}$$

where K_E is the set of cell regions where enzymes will be distributed to and n_k is a non-negative value assigned to cell region k. For random distribution, n_k

is sampled from a random number generator. For positive (negative) gradient distribution, cell regions in the outermost (innermost) layer of the cell will be assigned a value of one and increase by a user-defined gradient parameter value for each layer into (away from) the cell centre (Fig. 1a).

4 Simulations

We conducted a series of simulations using the model above, with the aim of testing our method and its utility. First, we wanted to check that our method has the ability to reproduce the expected effects of diffusion. Even though we have expanded conventional FBA to explicitly represent events of intracellular metabolite diffusion, their fluxes are solved using linear programming that depends on the objective function and set of constraints used. There is no guarantee that the solution(s) will capture the diffusion phenomenon at all. To test this, we ran simulations of different cell shapes, with intracellular reactions operating at either reaction-limited or diffusion-limited regime. To simulate a reaction-limited regime ($v_{reaction} \ll v_{substrate-diffusion}$), we left the diffusion flux of substrate in the reaction unbounded, i.e. the substrate diffuses instantaneously. To simulate a diffusion-limited regime, ($v_{reaction} \gg v_{substrate-diffusion}$), we omitted diffusion from the set of reaction objects, i.e. the substrate cannot diffuse. Second, we wanted to test the utility of the method for simulating spatial heterogeneity within a single cell. For this test, we simulated different types of enzyme distribution as previously described. Altogether, the simulations comprised combinations of the following parameters:

- **Cell shape:** 1x36, 2x18, 3x12, 4x9, 6x6 (note that the total number of cell regions is the same for every cell shape so they all have the same spatial resolution)

- **Rate-limiting regime:** no diffusion (both reactions R3 and R4 are diffusion-limited), metabolite A diffuse instantaneously while metabolite B cannot diffuse (R3 is reaction-limited, R4 is diffusion-limited), metabolite B diffuse instantaneously while metabolite A cannot diffuse (R3 is diffusion-limited, R4 is reaction-limited), both metabolites A and B diffuse instantaneously (both R3 and R4 are reaction-limited)
- **Distribution pattern for transporter E1:** uniform, random (we only distributed this enzyme among cell regions in the outermost layer of the cell)
- **Distribution pattern for enzyme E2:** uniform, positive gradient, negative gradient, random (we distributed this enzyme among all cell regions)
- **Distribution pattern for enzyme E3:** uniform, positive gradient, negative gradient, random (we distributed this enzyme among all cell regions)

For positive and negative gradients, we set the gradient parameter to 1 and -1, respectively. For combinations that contain random enzyme distribution(s), simulation is no longer deterministic so it may yield a different result in every simulation. Therefore, we ran 100 simulations for each of these combinations to determine population-level behaviour. Note that we did not seed the random number generator for these simulations. We repeated the simulations multiple times, and our results were robust and showed consistent behaviour. Scripts for running these simulations can be found on the method's GitHub page. All were steady-state simulations, an inherent feature of FBA. Simulation results are presented next.

4.1 Method captures the effects of diffusion

We first tested whether our method can reproduce the effects of intracellular metabolite diffusion, by comparing simulations across different cell shapes. For this test, we focused only on deterministic simulations, i.e. parameter

combinations that do not involve any random enzyme distribution. To allow a quantitative comparison across cell shapes, we used the perimeter-to-area (PA) ratio to represent each shape. We defined this ratio as the number of cell regions in the outermost layer divided by the total number of cell regions. This ratio was chosen over a commonly used aspect ratio because it is the two-dimensional equivalent of surface-area-to-volume ratio, which is known to correlate positively with the rate of diffusion (Okie, 2013). If the effects of diffusion are captured by the method, we would expect to see differences in optimised cell growth at different PA ratios when diffusion becomes limiting. Note that in our simulations, transporter E1 was distributed only to regions in the outermost layer regardless of cell shape. We can therefore rule out cellular capacity for nutrient uptake as the cause of any differences in simulated growth if observed. As seen in Fig. 2, when either or both metabolites A and B have limited ability to diffuse, there is a general trend of reduced growth at low PA ratios. On the other hand, in the control where both metabolites can diffuse instantaneously, i.e. all intracellular reactions are reaction-limited as implicitly assumed in conventional FBA, cell growth can achieve the maximal value of 100 regardless of the ratio. We observed these relations in all combinations of non-random enzyme distributions. It is worth noting that both 1x36 and 2x18 cells achieve the same growth in our simulations despite having different elongated shapes. This is because at this spatial resolution, they have the same PA ratio where all cell regions are in the outermost layer. Each cell region can receive nutrients from the extracellular compartment via transporters instead of being dependent on diffusion from neighbouring regions. A higher spatial resolution would be needed to discriminate between the two shapes, but we included both in our simulations as controls to check that the reduced growth effects we observed in other shapes were indeed due to intracellular diffusion.

These results demonstrated that our method can spatially simulate reaction-diffusion in single-cell metabolism. It can capture the theoretical effects of both reaction-limited and diffusion-limited regimes on growth, as well as the spatial effect of cell shape.

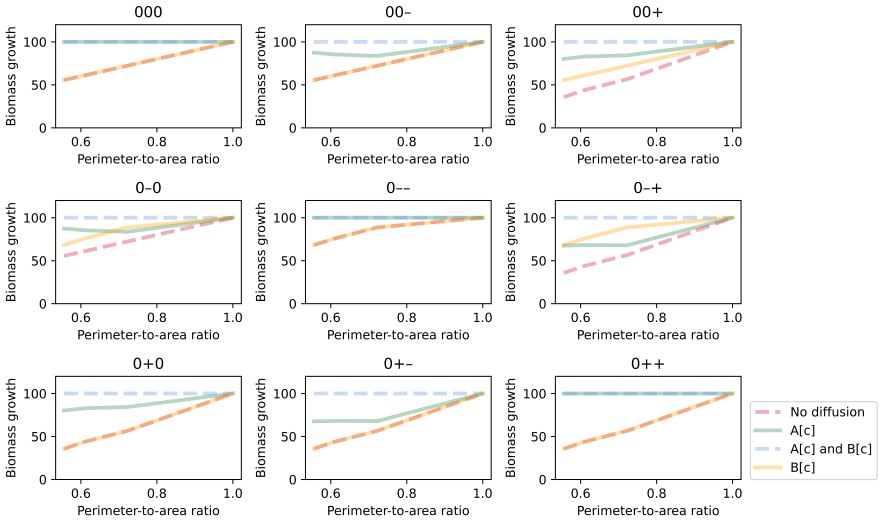

Fig. 2 Simulated biomass growth plotted against perimeter-to-area ratio for different rate-limiting regimes (no diffusion, metabolite A can diffuse instantaneously, metabolite B can diffuse instantaneously, both metabolites can diffuse instantaneously). Each panel illustrates plots for a different combination of enzyme distribution patterns, as indicated by three characters at the top of the panel ("0": uniform distribution; "+": positive gradient; "-": negative gradient) that represent the distribution of transporter E1, enzyme E2 and enzyme E3, respectively. Opacity of the lines was adjusted so that overlapping lines appear as a blend of the overlapping colours.

The extent to which optimal growth is reduced with PA ratio, i.e. the slope of lines in Fig. 2, depends on the metabolite(s) that is not diffusible. In general, the effect of cell shape is strongest when both metabolites A and B cannot diffuse. This is because all intracellular reactions are diffusion-limited, so a reduction in diffusion in rounder cells increases the severity of the limitation. The effect is weaker if metabolite A can diffuse instantaneously, since only one reaction is diffusion-limited. However, combination of enzyme distribution also plays a role. We explored this next.

4.2 Method enables study of enzyme distribution effect

To inspect how enzyme distribution influences the effect of diffusion on growth, we further analysed simulated growth of a 6x6 cell where diffusion-limiting effect is most severe. Again, we only focused on deterministic simulations where transporter E1 has a uniform distribution, while enzymes E2 and E3 can be distributed uniformly or on a (positive/negative) gradient. We found that for each rate-limiting regime, combinations of enzyme distribution can be grouped based on their effective maximal growth rates (Fig. 3).

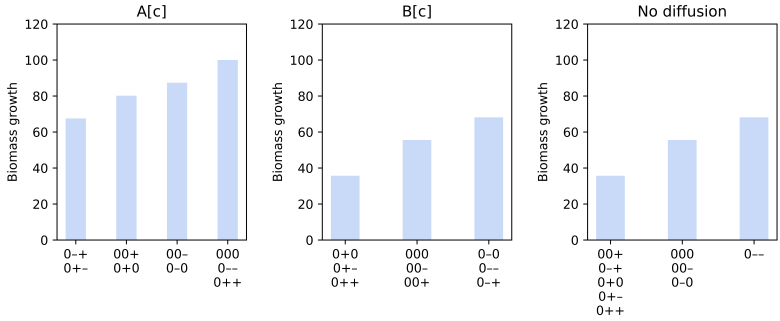

Fig. 3 Simulated biomass growth for a 6x6 cell for different rate-limiting regimes: only metabolite A can diffuse instantaneously (left panel); only metabolite B can diffuse instantaneously (middle panel); and both metabolites are restricted (right panel). The combinations of enzyme distribution patterns that produce each growth level are indicated at the bottom of each bar. Each combination of three characters ("0": uniform distribution; "+": positive gradient; "-": negative gradient) represents the distribution of transporter E1, enzyme E2 and enzyme E3, respectively.

For the regime where only metabolite A can diffuse instantaneously (left panel in Fig. 3), a maximal growth of 100 can still be achieved as long as enzymes E2 and E3 have the same distribution pattern, i.e. both enzymes have the same concentration in every cell region. This is because A can move freely and instantaneously so it is optimally channelled in the simulation to occupy each cell region at a rate proportional to regional concentration of enzyme E2. This enables enzyme E2 in every region to operate at maximal

capacity, without being limited by diffusion, to convert A to B. If in addition there is proportional amount of regional enzyme E3 to convert B to biomass on site, growth is not limited by the diffusion of B either so the maximal growth is achieved. On the contrary, if the enzymes are distributed in opposing gradient to each other, the growth in every cell region is limited by either of the enzymes. Even if A can diffuse accordingly to enable maximal conversion to B in every region, some regions do not have sufficient E3 to fully support on-site conversion of B to biomass, and any excess B cannot diffuse to other regions with higher E3 to make up for it. Thus, overall growth is lowest for combinations with opposing gradients. In the middle range are cases where one enzyme is uniformly distributed while the other has a gradient distribution. In these cases, a negative gradient is more favourable for growth. This can be explained by the fact that the number of cell regions in each layer decreases from the outer layers to the inner layers (Fig. 1a). With a negative gradient where concentration of an enzyme is higher in the outer layers, more cell regions are allocated with high amount of the enzyme thereby generating higher overall growth. Again, this highlights how PA ratio, metabolite diffusion and enzyme distribution interact to give rise to growth behaviour.

Our analysis also shows that there are only three growth levels when metabolite A cannot diffuse, regardless of whether B can diffuse (middle and right panels in Fig. 3). However, combinations of enzyme distribution are grouped differently depending on the ability of B to diffuse. If B can instantaneously diffuse, combinations are grouped based only on the distribution of enzyme E2. A negative gradient is most favourable, followed by a uniform distribution and lastly a positive gradient. The reason is quite simple. Since metabolite A cannot diffuse, they stay in cell regions in the outermost layer after being transported into cell. A negative gradient means there is a higher

concentration of enzyme E2 in the outermost layer to maximise conversion of A to B. Hence, there is more B that can be converted to biomass either immediately in this layer or after diffusing to the inner layers, to produce a higher growth rate. If both metabolites A and B cannot diffuse, reactions and growth only occur in cell regions in the outermost layer. Thus, a high growth rate can only be achieved when both enzymes E2 and E3 have high concentrations in this layer. In accordance with this, growth is lowest when any enzyme has a positive gradient distribution, i.e. a low concentration in the outermost layer.

In the optimal solutions above, for cases with reaction-limited regime, metabolites diffuse in the direction and rate that maximises growth. Each metabolite tends to move towards regions with high enzyme capacity, where it will be consumed at a higher rate and stays at a lower level. If we assume that volume exclusion by other molecules is the same in each region, then the optimal solutions have simulated metabolite diffusion in the direction from regions with high concentrations to regions with low concentrations, without explicitly representing concentration gradient. Note however that our simple reaction network only consists of single-substrate linear pathway. For more complex reaction network, the optimal solution may not necessarily favour diffusion that follows the concentration gradient. Additional constraints may be needed to obtain more accurate solutions.

4.3 Effect on cell-to-cell variability

In this section, we focus on simulations that involve random enzyme distribution. Fig. 4 shows the growth distribution of 100 stochastic simulations for each rate-limiting regime and combination of enzyme distribution for a 6x6 cell. As previously observed, the maximal growth of 100 is always achieved when metabolites A and B can diffuse instantaneously. For some rate-limiting

regimes, cells can switch between deterministic and stochastic growth, depending on the combination of enzyme distribution. Interestingly, several growth distributions have lower standard deviation than others. To explore this further, we plotted the coefficient of variation (CV) against the PA ratio for each case (Fig. 5). We used CV instead of standard deviation because the former is normalised to the population mean so it is comparable across cases.

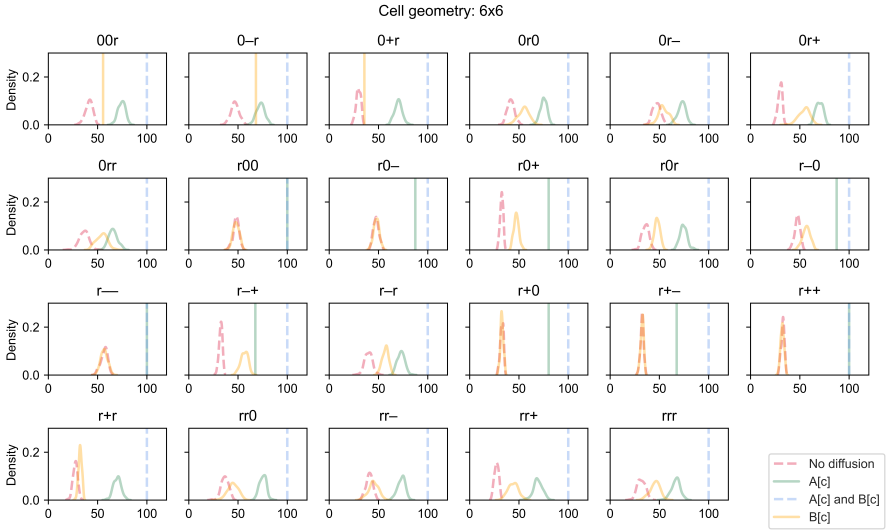

Fig. 4 Kernel density plots of simulated biomass growth for different rate-limiting regimes (no diffusion, metabolite A can diffuse instantaneously, metabolite B can diffuse instantaneously, both metabolites can diffuse instantaneously). Each panel illustrates plots for a different combination of enzyme distribution patterns, as indicated by three characters at the top of the panel ("r": random distribution; "0": uniform distribution; "+": positive gradient; "-": negative gradient) that represent the distribution of transporter E1, enzyme E2 and enzyme E3, respectively. Each plot describes the population distribution of 100 simulations. A vertical line indicates that all 100 simulations produce the same value. Opacity of the lines was adjusted so that overlapping lines appear as a blend of the overlapping colours.

In the cases where metabolites A and B diffuse instantaneously, the CV is always zero as expected. For most other rate-limiting regimes, the CV does not generally change much with the PA ratio, and there are overlaps among regimes for some combinations of enzyme distribution. However, CV tends to be slightly sensitive to the PA ratio when both metabolites cannot diffuse.

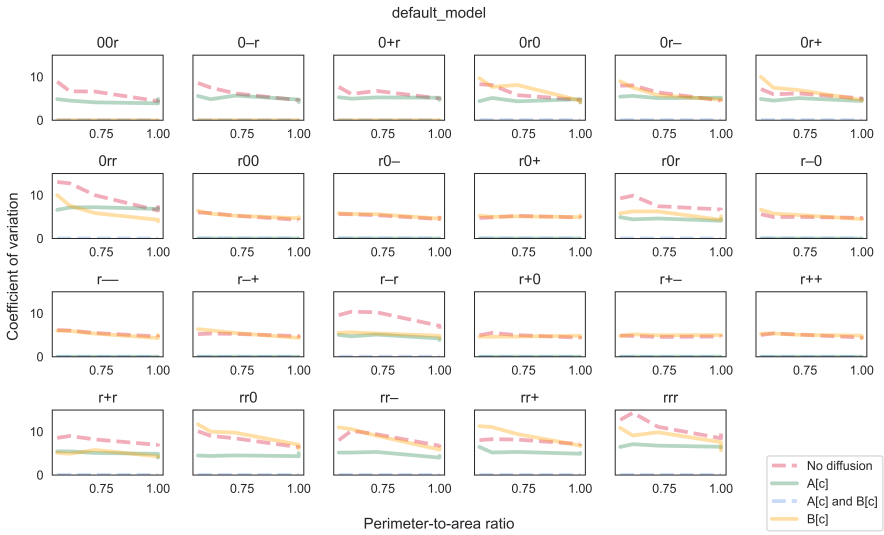

Fig. 5 The coefficient of variation (CV) of biomass growth simulated using the "default model". CV is plotted against perimeter-to-area ratio for different rate-limiting regimes (no diffusion, metabolite A can diffuse instantaneously, metabolite B can diffuse instantaneously, both metabolites can diffuse instantaneously). Each panel illustrates plots for a different combination of enzyme distribution patterns, as indicated by three characters at the top of the panel ("r": random distribution; "0": uniform distribution; "+": positive gradient; "-": negative gradient) that represent the distribution of transporter E1, enzyme E2 and enzyme E3, respectively. Opacity of the lines was adjusted so that overlapping lines appear as a blend of the overlapping colours.

Specifically, CV increases at decreasing PA ratio. As previously discussed, growth only occurs in regions in the outermost layer when both metabolites cannot diffuse. At low PA ratios, only a small number of cell regions actually contribute to overall growth. The contribution of each region depends on its enzyme concentrations, some of which are sampled randomly. Since variance is usually high when sample size is small, the variance of regional growth among a small number of contributing regions tends to be high (Hajagos and Steiner, 2001). Consequently, cell growth which is the sum of these highly variable regional growth can take a wide range of values; thus CV tends to be higher at low PA ratios.

In the simulations so far, the reaction network and parameter values were designed such that no specific enzyme would be limiting growth if simulated

as a cell-scale representation. In other words, there are just enough of each enzyme to support a specific cell growth level assuming that all reactions are not diffusion-limited. We called this the "default model". When the "default model" is simulated as a regional-scale representation, some regions may have disproportional enzyme contents depending on how each enzyme is distributed. As diffusion becomes limiting, there are regions where some enzymes are in excess or unused. Consequently, the overall cell growth is lower than the theoretical maximum (Fig. 3). How much growth is reduced thus indicates the degree to which enzymes are not fully utilised. At the population level, a high CV indicates that a large proportion of cells under-utilise their enzymes and grow much lower than the maximum and mean levels. Therefore, the CV measures the extent of under-utilisation within the population.

We also explored alternative models where some enzymes were in excess even when diffusion is not limiting, by changing the stoichiometry in the reaction network while maintaining total cellular amount of each enzyme. We chose this over changing enzyme amount directly so that enzyme distribution patterns were consistent and comparable across models. In the first alternative which we called the "B-limiting model", we increased the stoichiometric coefficient of metabolite B in reaction R4 as follows:

$$2\ \mathrm{B[c]} \xrightarrow{\mathrm{E3}} \mathrm{Biomass} \tag{R4a}$$

Note that we are not indicating metabolite B is a growth-limiting nutrient. With this change, E3 becomes an excessive enzyme. Compared to the "default model", there is not much difference in CV sensitivity to the PA ratio (Fig. 6). The most noticeable difference is a decrease in CV for several combinations of enzyme distribution in the regime where only metabolite A can diffuse

instantaneously. We also tested a second alternative which we called the "A-limiting model", where we increased the stoichiometric coefficient of metabolite A in reaction R3 instead, as follows:

$$2\ \mathrm{A[c]} \xrightarrow{\mathrm{E2}} \mathrm{B[c]} \tag{R3a}$$

In the "A-limiting model", both enzymes E2 and E3 become excessive. CV becomes zero in all cases where metabolite A can diffuse instantaneously, regardless of whether B can diffuse (Fig. 7). More interestingly, the sensitivity of CV to the PA ratio increases for the other rate-limiting regimes, i.e. CV is high at low ratios and low at high ratios. In a number of cases, CV even approaches zero at the ratio of one, going from highly stochastic to almost deterministic just by changing the PA ratio. Together, our simulations suggest that enzyme distribution patterns within individual cells could affect not only single-cell behaviour but also variability at the population level. Besides rate-limiting regime and cell shape, effect at the population level also depends on the overall relative abundance of enzymes.

5 Discussion

In this paper, we present discretised FBA, a spatial modelling method and algorithms that extend conventional FBA to include representation of metabolite diffusion inside cells. Implemented on a grid-based system, our method allows us to spatially explore metabolism and cell behaviour such as growth when diffusion becomes a limiting factor in enzymatic reactions. Coupling FBA with spatial representation has been done previously, for example to describe the interactions and spatial dynamics among microbial communities (Henson, 2015; Borer et al, 2019; Karimian and Motamedian, 2020; Dukovski et al,

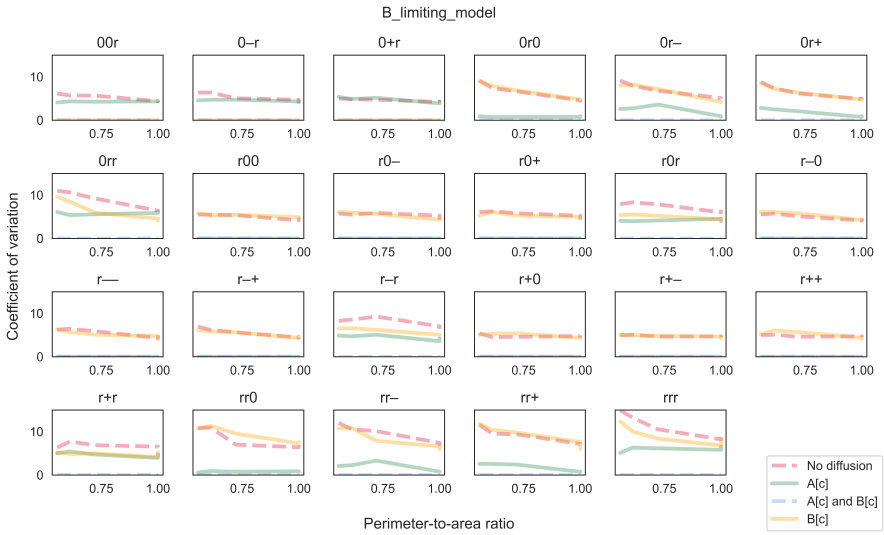

Fig. 6 The coefficient of variation (CV) of biomass growth simulated using the "B-limiting model". CV is plotted against perimeter-to-area ratio for different rate-limiting regimes (no diffusion, metabolite A can diffuse instantaneously, metabolite B can diffuse instantaneously, both metabolites can diffuse instantaneously). Each panel illustrates plots for a different combination of enzyme distribution patterns, as indicated by three characters at the top of the panel ("r": random distribution; "0": uniform distribution; "+": positive gradient; "-": negative gradient) that represent the distribution of transporter E1, enzyme E2 and enzyme E3, respectively. Opacity of the lines was adjusted so that overlapping lines appear as a blend of the overlapping colours.

2021). To the best of our knowledge, our work is the first spatially informed FBA method that looks into reaction-diffusion within a single cell. Compared to existing spatial modelling methods such as PDE, our method requires very few parameters.

We simulated diffusion-limited and reaction-limited regimes using extreme settings where metabolites are either restricted or able to diffuse freely and instantaneously. This allowed us to compare the maximum extent of their potential impact on behaviour such as growth. In real systems, metabolites are generally small so they can move around quite freely and rapidly within the same compartment, but can be impeded by surrounding macromolecules. Even so, it is unlikely that metabolites will be fully restricted to the extreme level of diffusion limitation in our simulations. Besides, highly efficient enzymes

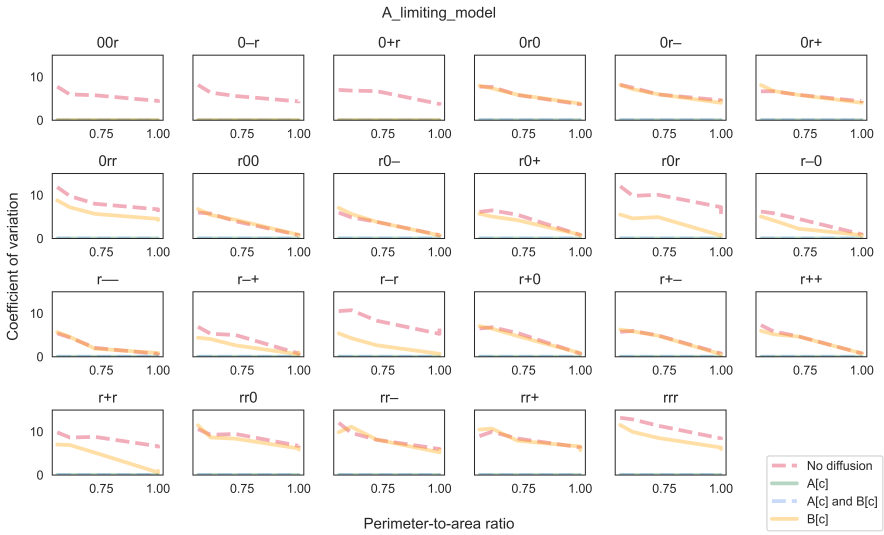

Fig. 7 The coefficient of variation (CV) of biomass growth simulated using the "A-limiting model". CV is plotted against perimeter-to-area ratio for different rate-limiting regimes (no diffusion, metabolite A can diffuse instantaneously, metabolite B can diffuse instantaneously, both metabolites can diffuse instantaneously). Each panel illustrates plots for a different combination of enzyme distribution patterns, as indicated by three characters at the top of the panel ("r": random distribution; "0": uniform distribution; "+": positive gradient; "-": negative gradient) that represent the distribution of transporter E1, enzyme E2 and enzyme E3, respectively. Opacity of the lines was adjusted so that overlapping lines appear as a blend of the overlapping colours.

that are diffusion-limited are rare (Sweetlove and Fernie, 2018). The majority of enzymes have not evolved to the "perfect" level of catalytic efficiency, so most reactions likely operate under a reaction-limited regime even when the diffusion rates of substrates are low (Bar-Even et al, 2015). Additionally, metabolism consists of reaction networks that are much more complex than our toy models, so we expect the overall impact is intricate but much less severe. Some studies have proposed that diffusion limitation can be overcome through metabolite channelling, where intermediate products in a series of reactions are directly transferred from one enzyme to the next in the pathway to shorten diffusion distance (Obata, 2020). All these seem to indicate a negligible effect of metabolite diffusion on single-cell metabolism. However, several studies have noticed that the density of some bacterial cells seem to be optimal

for achieving cellular efficiency, with a tradeoff between high volume density that can improve biochemical efficiency and molecular crowding that can slow down diffusion (Pang and Lercher, 2023; Vazquez, 2010). Since the sub-cellular environment is not static, spatial and structural organisation in cells could provide another dimension of regulation. Our method provides a theoretical means to study the relationship between local reaction/diffusion limitations and optimal growth (or any other objective function). This relationship could be useful, for example to study cancer cells that are known to have abnormal growth and display drug resistance predictable by cell shape (Longden et al, 2021). Another potential use is to facilitate the design of synthetic systems, such as protocells where structure is a major consideration (Gözen et al, 2022).

We also demonstrated another utility of our spatial method by presenting simulations of different enzyme distribution patterns. Experimental observations have shown that most cellular molecules are heterogeneously distributed. Such heterogeneity could arise as the direct effect of regulation and the indirect effect of cell shape (Halatek et al, 2018; Ovádi and Saks, 2004). Since enzymes are usually large and metabolism is a relatively fast process, we assumed that enzymes do not move during the time-scale of our steady-state simulations. Our results suggest that some distribution patterns are more favourable for growth than others. Growth favours distributions where enzymes that act in series have similar range of concentration or capacity within the same region. This supports the notion that cells have organelles and compartments, among other theories, to concentrate enzymes and substrates that participate in the same pathways to increase biochemical efficiency (Diekmann and Pereira-Leal, 2013; Conrado et al, 2007).

Our results also suggest that enzyme distribution within individual cells can affect variability at the population level, if some reactions are diffusion-limited. Enzymes are required for cell growth but their production is costly (Buttgereit and Brand, 1995; Russell and Cook, 1995). In our simulations, the same investment, i.e. the same amount of total cellular enzyme, could yield different returns in growth if cells change their shape and enzyme distribution pattern. At the population level, a low CV indicates a trade-off between overall return (population growth) and low cell-to-cell variability. On the other hand, a high CV means that a small number of cells in the population yield unusually high (or low) growth, which could confer a selective advantage under fluctuating environments (Balaban et al, 2004; Acar et al, 2008). Most studies on variability have focused on stochasticity in gene expression (Thattai and van Oudenaarden, 2004; Schwabe et al, 2011; Evans and Zhang, 2020); our study suggests that other aspects such as cell shape and spatial organisation may have effects under certain conditions.

6 Conclusions and future work

In conclusion, our method expands the widely used FBA and increases its spatial resolution. FBA is widely used because it has low parameter requirements yet provides flexibility for investigation. For example, one can use thermodynamics principles and gene expression data to constrain reaction fluxes and direction when kinetic parameters are lacking at the genome scale, or test various hypotheses by changing the objective function (Blazier and Papin, 2012; Machado and Herrgard, 2014; Raman and Chandra, 2009). We recommend similar use of our method. In the absence of data about diffusion

coefficients and concentrations, spatial information and measurements of factors that affect diffusivity can be used to constrain diffusion fluxes, to gain insight into the system.

Currently, our method is implemented to represent regular two-dimensional shapes. We have demonstrated the use of this method on a simple toy reaction network in a single-compartment cell. Future extensions could include capabilities to represent more complex cell shapes and organelles, as well as in three dimensions. At present, each cell region or groups of neighbouring regions can effectively be a distinct sub-cellular compartment or organelle, but all movements of metabolites in and out of a region are treated as diffusion. To represent membrane-bound organelles, extra features are needed to identify cell regions located at the boundary of an organelle, and treat metabolite movements between these regions and neighbouring regions that are not part of the organelle as cross-membrane (active/passive) transport reactions. Additional functions could facilitate the representation of organelle-specific reaction networks. These features will enable us to simulate more realistic cells and reaction networks, so that we can explore how spatiotemporal organisation of sub-cellular structure affects metabolism and cell behaviour such as growth.

Acknowledgments. We thank Alexandra-Anamaria Sorinca who tested the idea behind this method as a summer undergraduate project funded by the London Mathematical Society (URB-2022-19). This work was funded by the United Kingdom Research Innovation Future Leaders Fellowship (MR/T043571/1).

Appendix A

The following details how our Python API converts a cell-scale representation of metabolic network into a regional-scale representation (see Algorithms 1–5):

1. Create a two-dimensional I x J grid, where I (cell width) and J (cell length) are defined by the user;
2. Create in every cell region a metabolite object for each intracellular metabolite. For metabolites in the extracellular compartment, only one object per metabolite will be created. The user has to provide input of metabolites as two separate lists; one for intracellular and another for extracellular metabolites;
3. Create in every cell region a reaction object for each intracellular biochemical reaction. For reactions in the extracellular compartment including exchange reactions, only one object per reaction will be created. Again, the user has to provide the two groups of reactions separately;
4. Create an objective function that sums over all cell regions, the variable which represents the flux of a user-specified intracellular reaction, e.g. the biomass reaction;
5. Create in every boundary cell region a reaction object that represents the transport of each intracellular metabolite to/from the extracellular compartment. Boundary cell regions are the regions located in the outermost layer of the cell, i.e. the boundary between cell and the extracellular compartment. For this step, the user has to provide the list of intracellular metabolites that are transported to/from the extracellular compartment, and the list of import/export reactions;
6. For each cell region:
    (a) identify its neighbouring regions;
    (b) for each neighbouring region, create a reaction object for every intracellular metabolite to represent diffusion from own region to the neighbouring region;

7. Set the bounds of each reaction object, based on user input. If no bounds are specified for a reaction, it will be left unbounded.

Algorithm 1 Create cell regions

Input: Cell width, I
Cell length, J
Output: Array of cell regions, K

```
create empty array K
for i = 0, ..., I − 1 do
    for j = 0, ..., J − 1 do
        create cell region k
        K(i, j) ⇐ k
    end for
end for
```

Algorithm 2 Create regional metabolites

Input: Array of cell regions, K
List of intracellular metabolites, M_{intra}
List of extracellular metabolites, M_{extra}
Output: Lists of metabolites M

```
create empty list M
for m ∈ M_intra do
    for k ∈ K do
        create and add metabolite m_k to M
    end for
end for
for m ∈ M_extra do
    create and add metabolite m to M
end for
```

Appendix B

To simulate different enzyme distribution patterns, we provide additional functions in the API that do the following:

1. Set the concentration of a user-specified enzyme in each cell region based on the total cellular concentration and distribution pattern provided by users.

Algorithm 3 Create regional reactions

Input: Array of cell regions, K
List of intracellular reactions, R_{intra}
List of extracellular reactions, R_{extra}
Objective reaction, r_{obj}
List of import/export reactions, R_{trans}
Output: Lists of reactions R

```
create empty list R
for r ∈ R_intra do
    for k ∈ K do
        create and add reaction r_k to R
    end for
end for
for r ∈ R_extra do
    create and add reaction r to R
end for
for r ∈ R do
    if r is r_obj then
        set weight of r in objective function to 1
    else
        set weight of r in objective function to 0
    end if
end for
identify the set of boundary cell regions B
for k ∈ B do
    for r ∈ R_trans do
        create and add reaction r_k to R
    end for
end for
```

Algorithm 4 Add metabolite diffusion to list of reactions

Input: Array of cell regions, K
List of reactions, R
List of intracellular metabolites, M_{intra}
Output: Metabolite diffusion objects

```
for k ∈ K do
    identify the set of neighbouring regions N
    for n ∈ N do
        for m ∈ M_intra do
            create and add diffusion d_{m_{k→n}} to R
        end for
    end for
end for
```

Algorithm 5 Set reaction bounds

Input: Array of cell regions, K
List of all reactions, R
User-defined bounds

Output: Constrained regional-scale model

```
for r ∈ R do
    identify cell region k where r belongs
    if r has user-defined bounds then
        l_r ⇐ user-defined lower bound
        u_r ⇐ user-defined upper bound
    else
        l_r ⇐ −∞
        u_r ⇐ ∞
    end if
end for
```

The function provides the option of distributing enzymes in the cell uniformly, on a positive gradient (increasing concentration from the boundary to the centre of the cell), on a negative gradient, or randomly. Expected user inputs to determine the distribution pattern include whether the distribution is random, if distributing to all cell regions or only the outermost layer, the gradient parameter and an optional seed for the random number generator (see Algorithm 6);

2. Calculate the upper bound of the reaction catalysed by the enzyme using equation (3), based on a user-specified k_{cat} value.

References

Acar M, Mettetal J, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. Nature Genetics 40:471–475. https://doi.org/10.1038/ng.110

Balaban NQ, Merrin J, Chait R, et al (2004) Bacterial persistence as a phenotypic switch. Science 305(5690):1622–1625. https://doi.org/10.1126/science.1099390

Bar-Even A, Milo R, Noor E, et al (2015) The moderately efficient enzyme: Futile encounters and enzyme floppiness. Biochemistry 54(32):4969–4977. https://doi.org/10.1021/acs.biochem.5b00621

Blazier AS, Papin JA (2012) Integration of expression data in genome-scale metabolic network reconstructions. Frontiers in Physiology 3(299). https://doi.org/10.3389/fphys.2012.00299

Borer B, Ataman M, Hatzimanikatis V, et al (2019) Modeling metabolic networks of individual bacterial agents in heterogeneous and dynamic soil habitats (indimesh). PLOS Computational Biology 15(6):e1007,127. https://doi.org/10.1371/journal.pcbi.1007127

Buttgereit F, Brand MD (1995) A hierarchy of atp-consuming processes in mammalian cells. Biochemical Journal 312:163–167. https://doi.org/10.1042/bj3120163

Chang A, Jeske L, Ulbrich S, et al (2021) Brenda, the elixir core data resource in 2021: new developments and updates. Nucleic Acids Research 49(D1):D498–D508. https://doi.org/10.1093/nar/gkaa1025

Conrado RJ, Mansell TJ, Varner JD, et al (2007) Stochastic reaction–diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. Metabolic Engineering 9(4):355–363. https://doi.org/10.1016/j.ymben.2007.05.002

Cook DJ, Nielsen J (2017) Genome-scale metabolic models applied to human health and disease. WIREs Systems Biology and Medicine 9:e1393. https://doi.org/10.1002/wsbm.1393

Davidi D, Noor E, Liebermeister W, et al (2016) Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. Proceedings of the National Academy of Sciences 113(12):3401–3406. https://doi.org/10.1073/pnas.15142401

Diekmann Y, Pereira-Leal JB (2013) Evolution of intracellular compartmentalization. Biochemical Journal 449(2):319–331. https://doi.org/10.1042/BJ20120957

Dourado H, Lercher MJ (2020) An analytical theory of balanced cellular growth. Nature Communications 11:1226. https://doi.org/10.1038/s41467-020-14751-w

Dourado H, Mori M, Hwa T, et al (2021) On the optimality of the enzyme–substrate relationship in bacteria. PLoS Biology 19(10):e3001,416. https://doi.org/10.1371/journal.pbio.3001416

Dukovski I, Bajić D, Chacón J, et al (2021) A metabolic modeling platform for the computation of microbial ecosystems in time and space (comets). Nature Protocols 16:5030–5082. https://doi.org/10.1038/s41596-021-00593-3

Ebrahim A, Lerman JA, Palsson BO, et al (2013) Cobrapy: Constraints-based reconstruction and analysis for python. BMC Systems Biology 7(74). https://doi.org/10.1186/1752-0509-7-74

Evans TD, Zhang F (2020) Bacterial metabolic heterogeneity: origins and applications in engineering and infectious disease. Current Opinion in Biotechnology 64:183–189. https://doi.org/10.1016/j.copbio.2020.04.007

Farina S, Claus S, Hale JS, et al (2021) A cut finite element method for spatially resolved energy metabolism models in complex neuro-cell morphologies

with minimal remeshing. Advanced Modeling and Simulation in Engineering Sciences 8(5). https://doi.org/10.1186/s40323-021-00191-8

Gershon ND, Porter KR, Trus BL (1985) The cytoplasmic matrix: Its volume and surface area and the diffusion of molecules through it. Proceedings of the National Academy of Sciences 82(15):5030–5034. https://doi.org/10.1073/pnas.82.15.503

Gözen I, Köksal ES, Põldsalu I, et al (2022) Protocells: Milestones and recent advances. Small 18(18):2106,624. https://doi.org/10.1002/smll.202106624

Hajagos B, Steiner F (2001) Fulfilment of the law of large numbers in case of variance determinations. Acta Geodaetica et Geophysica Hungarica 36:163–174. https://doi.org/10.1556/AGeod.36.2001.2.4

Halatek J, Brauns F, Frey E (2018) Self-organization principles of intracellular pattern formation. Philosophical Transactions of the Royal Society B 373(1747):20170,107. https://doi.org/10.1098/rstb.2017.0107

Heckmann D, Lloyd CJ, Mih N, et al (2018) Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. Nature Communications 9(5252). https://doi.org/10.1038/s41467-018-07652-6

Heckmann D, Campeau A, Lloyd CJ, et al (2020) Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. Proceedings of the National Academy of Sciences 117(37):23,182–23,190. https://doi.org/10.1073/pnas.2001562117

Henson MA (2015) Genome-scale modelling of microbial metabolism with temporal and spatial resolution. Biochemical Society Transactions 43(6):1164–1171. https://doi.org/10.1042/BST20150146

Jahan N, Maeda K, Matsuoka Y, et al (2016) Development of an accurate kinetic model for the central carbon metabolism of *Escherichia coli*. Microbial Cell Factories 15(112). https://doi.org/10.1186/s12934-016-0511-x

Karimian E, Motamedian E (2020) Acbm: An integrated agent and constraint based modeling framework for simulation of microbial communities. Scientific Reports 10(8695). https://doi.org/10.1038/s41598-020-65659-w

Kinsey ST, Locke BR, Dillaman RM (2011) Molecules in motion: influences of diffusion on metabolic structure and function in skeletal muscle. Journal of Experimental Biology 214(2):263–274. https://doi.org/10.1242/jeb.047985

Loghmani SB, Veith N, Sahle S, et al (2022) Inspecting the solution space of genome-scale metabolic models. Metabolites 12(1):43. https://doi.org/10.3390/metabo12010043

Longden J, Robin X, Engel M, et al (2021) Deep neural networks identify signaling mechanisms of erbb-family drug resistance from a continuous cell morphology space. Cell Reports 34(3):108,657. https://doi.org/10.1016/j.celrep.2020.108657

Luby-Phelps K, Castle PE, Taylor DL, et al (1987) Hindered diffusion of inert tracer particles in the cytoplasm of mouse 3t3 cells. Proceedings of the National Academy of Sciences 84(14):4910–4913. https://doi.org/10.1073/pnas.84.14.4910

Machado D, Herrgard M (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLOS Computational Biology 10(4):e1003,580. https://doi.org/10.1371/journal.pcbi.1003580

Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. Journal of Biological Chemistry 276(14):P10,577–10,580. https://doi.org/10.1074/jbc.R100005200

Obata T (2020) Toward an evaluation of metabolite channeling *in vivo*. Current Opinion in Biotechnology 64:55–61. https://doi.org/10.1016/j.copbio.2019.09.013

Okie JG (2013) General models for the spectra of surface area scaling strategies of cells and organisms: Fractality, geometric dissimilitude, and internalization. The American Naturalist 181(3). https://doi.org/10.1086/669150

Oldewurtel ER, Kitahara Y, van Teeffelen S (2021) Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. Proceedings of the National Academy of Sciences 118(32):e2021416,118. https://doi.org/10.1073/pnas.2021416118

Orth JD, Thiele I, Palsson B (2010) What is flux balance analysis? Nature Biotechnology 28:245–248. https://doi.org/10.1038/nbt.1614

Ovádi J, Saks V (2004) On the origin of intracellular compartmentation and organized metabolic systems. Molecular and Cellular Biochemistry 256:5–12. https://doi.org/10.1023/B:MCBI.0000009855.14648.2c

Pang TY, Lercher MJ (2023) Optimal density of bacterial cells. PLoS Computational Biology 19(6):e1011,177. https://doi.org/10.1371/journal.pcbi.1011177

Peskov K, Mogilevskaya E, Demin O (2012) Kinetic modelling of central carbon metabolism in *Escherichia coli*. The FEBS Journal 279(18):3374–3385. https://doi.org/10.1111/j.1742-4658.2012.08719.x

Pettersson G (1992) Evolutionary optimization of the catalytic efficiency of enzymes. European Journal of Biochemistry 206(1):289–295. https://doi.org/10.1111/j.1432-1033.1992.tb16927.x

Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Briefings in Bioinformatics 10(4):435–449. https://doi.org/10.1093/bib/bbp011

Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiological Reviews 59(1):48–62. https://doi.org/10.1128/mr.59.1.48-62.1995

Schnell S, Turner TE (2004) Reaction kinetics in intracellular environments with macromolecular crowding: Simulations and rate laws. Progress in Biophysics and Molecular Biology 85(2–3):235–260. https://doi.org/10.1016/j.pbiomolbio.2004.01.012

Schwabe A, Dobrzyński M, Rybakova K, et al (2011) Origins of stochastic intracellular processes and consequences for cell-to-cell variability and cellular survival strategies. In: Jameson D, Verma M, Westerhoff HV (eds) Methods in Enzymology, vol 500. Academic Press, chap 28, p 597–625, https://doi.org/10.1016/B978-0-12-385118-5.00028-1

Srinivasan B (2022) A guide to the michaelis–menten equation: steady state and beyond. The FEBS Journal 289(20):6086–6098. https://doi.org/10.1111/febs.16124

Sweetlove LJ, Fernie AR (2018) The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. Nature Communications 9:2136. https://doi.org/10.1038/s41467-018-04543-8

Szymanska P, Kochanczyk M, Miekisz J, et al (2015) Effective reaction rates in diffusion-limited phosphorylation-dephosphorylation cycles. Physical Review E 91(2):022,702. https://doi.org/10.1103/PhysRevE.91.022702

Teusink B, Passarge J, Reijenga CA, et al (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry. European Journal of Biochemistry 267:5313–5329. https://doi.org/10.1046/j.1432-1327.2000.01527.x

Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. Genetics 167(1):523–530. https://doi.org/10.1534/genetics.167.1.523

Vazquez A (2010) Optimal cytoplasmatic density and flux balance model under macromolecular crowding effects. Journal of Theoretical Biology 264(2):356–359. https://doi.org/10.1016/j.jtbi.2010.02.024

Wittig U, Rey M, Weidemann A, et al (2018) Sabio-rk: an updated resource for manually curated biochemical reaction kinetic. Nucleic Acids Research 46(D1):D656–D660. https://doi.org/10.1093/nar/gkx1065

Algorithm 6 Distribute enzyme

Input: Boolean; if distribution is random
Boolean; if distributing to all cell regions
Gradient parameter, p
Array of cell regions, K
Cellular concentration of enzyme, $[E]$

Output: Enzyme concentration in each cell region, $[E]_k$

```
create empty array K_E
if distributing to all cell regions then
    K_E ⇐ K
else
    K_E ⇐ cell regions in outermost layer
end if
if distribution is random then
    for k ∈ K_E do
        n_k ⇐ random number
    end for
else
    if p is 0 then                                  ▷ uniform distribution
        for k ∈ K_E do
            n_k ⇐ 1
        end for
    else
        group cell regions into layers L
        if p > 0 then                               ▷ positive gradient
            order L from outermost layer to cell centre
        else if p < 0 then                          ▷ negative gradient
            order L from cell centre to outermost layer
        end if
        for l = 1, ..., count(L) do
            for k ∈ L(l − 0) do
                n_k ⇐ l
            end for
        end for
    end if
end if
for k ∈ K_E do
    [E]_k ⇐ n_k / (Σ_{k∈K_E} n_k) [E]
end for
```