bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Simulating single-cell metabolism using a stochastic flux-balance analysis algorithm

David S. Tourigny[1*], Arthur P. Goldberg[2], and Jonathan R. Karr[2]

[1]*Columbia University Irving Medical Center, 630 West 168th Street, New York, NY 10032 USA*

[2]*Icahn Institute for Data Science and Genomic Technology, and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA*

[*]*Correspondence: dst2156@cumc.columbia.edu*

Abstract

Stochasticity from gene expression in single cells is known to drive metabolic heterogeneity at the population-level, which is understood to have important consequences for issues such as microbial drug tolerance and treatment of human diseases like cancer. Experimental methods for probing metabolism in single cells currently lag far behind advancements in single-cell genomics, transcriptomics, and proteomics, which motivates the development of computational techniques to bridge this gap in the systems approach to single-cell biology. In this paper, we present SSA-FBA (stochastic simulation algorithm with flux-balance analysis embedded) as a modelling framework for simulating the stochastic dynamics of metabolism in individual cells. SSA-FBA extends the constraint-based formalism of metabolic network modelling to the single-cell regime, providing a suitable approach to simulation when kinetic information is lacking from models. We also describe an advanced algorithm that significantly improves the efficiency of exact SSA-FBA simulations, which is necessary because of the computational costs associated with stochastic simulation and the observation that approximations can be inaccurate and numerically unstable. As a preliminary case study we apply SSA-FBA to a single-cell model of *Mycoplasma pneumoniae*, and explore the use of simulation to understand the role of stochasticity in metabolism at the single-cell level.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Introduction

Recent experimental advances are driving a data explosion in systems biology by enabling researchers to profile single cells, including their genome, transcriptome, and proteome [1, 2]. Such single-cell measurements can yield information on thousands of individual cells in a single experiment. This can provide insights on intracellular function and the role of intercellular heterogeneity in a variety of biological systems ranging from microbial populations [3, 4] to human diseases such as cancer [5, 6].

Relatively less advanced are methodologies to probe the metabolism of single cells [7, 8, 9, 11, 12]. This presents a barrier to studying metabolic reprogramming in tumour biology for example, which is now understood to be a central hallmark of cancer [13, 14]. Single-cell metabolism is challenging due to low abundances of many metabolites, compartmentalisation in eukaryotic cells, and the wide diversity of intracellular metabolites that lack regular structure [9]. Moreover, metabolism is more dynamic than many cellular processes such as DNA replication and gene expression, which means attempts to capture the metabolic state of an individual cell is susceptible to perturbation by changes in cellular behaviour and the surrounding environment. Current experimental obstacles to studying single-cell metabolism combined with its fundamental biological importance necessitates the development of computational techniques that infer the metabolism of single cells from other sources, such as single-cell transcriptomic or proteomic data and information about metabolism at the population-level [10].

While our current capacity to probe or model the metabolism of single-cells is limited, considerable attention has been devoted to the metabolism of cellular populations, where metabolic network modelling has received a great deal of success combining limited experimental data and computational simulation [15, 16, 17]. Extensions of these population-based frameworks, such as dynamic metabolism expression models [18] or dynamic enzyme-cost flux-balance analysis [19], have also been developed to incorporate the dynamics of gene expression. These deterministic approaches fail to capture effects that are relevant for metabolism in single cells however, where stochasticity is understood to play a pivotal role governing the dynamics of metabolism and growth

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

at the single-cell level [20, 21]. To date, there have only been a handful of attempts (see [22] and references therein) to extend metabolic network modelling to the single-cell regime, which have mainly focussed on integrating single-cell transcriptomics data with flux-balance analysis (FBA) in the context of cancer.

In general, the few existing approaches to stochastic modelling of single-cell metabolism fall into two distinct categories: (i) analytical treatment of the equations for expression of a single enzyme catalysing a single reaction [23, 24], or (ii) computational modelling of whole cells [25, 26] using hybrid methods involving ordinary differential equations (ODEs), particle-based stochastic simulation algorithms (SSA, [27, 28]), and dynamic FBA (DFBA, [29]). Although the shared aim of these approaches is to relate single-cell behaviour to phenotype, the two categories fall at opposite ends of a wide spectrum of increasing complexity: whole cell modelling attempts to accommodate as much detail as possible, but generates models that are a long way from analytic tractability. On the other hand, analytic methods can provide a mechanistic understanding of how the expression of a single enzyme governs fluctuations in metabolite levels, but this has less relevance outside the context of the metabolic reaction network. The goal of this paper is to introduce a stochastic extension of FBA that we call SSA-FBA (stochastic simulation algorithm with flux-balance analysis embedded), which attempts to address some limitations of previous ad-hoc treatments of metabolism in whole cell models and proves to be an appropriate tool for simulating metabolic network models of single cells.

The remainder of this paper is organised as follows. In Section 2 we outline the conceptual definition of SSA-FBA and relate it to a formal description of single-cell metabolism based on the chemical master equation. In Section 3 we compare exact and approximate implementations of SSA-FBA, and then introduce an advanced algorithm that significantly improves the efficiency of an exact simulation. A case study of a single *Mycoplasma pneumoniae* cell containing 176 reactions is simulated using SSA-FBA in Section 4, where we also explore the consequences of stochasticity at the single-cell level. We conclude in Section 5 with a summary of the main results and directions for future work. Further details on SSA-FBA and its advanced implementation can

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be found in the Supplementary Appendix accompanying this manuscript and all code and data are freely available at https://gitlab.com/davidtourigny/single-cell-fba.

2 SSA-FBA: stochastic simulation algorithm with flux-balance analysis embedded

Stochasticity in the metabolism and growth of single cells is generally believed to emerge because of fluctuations in enzyme expression levels [20, 21, 23, 24], where low mRNA and protein copy numbers limit the precision of gene regulation [30]. Due to the relatively high copy numbers of most metabolites, metabolism is thought to have little intrinsic stochastic variation (Figure 1a). This general observation will motivate SSA-FBA as an appropriate framework for modelling single-cell metabolism, where the dynamics of reaction fluxes internal to a metabolic network are captured deterministically by FBA while SSA is used to model changes in the copy numbers of enzyme molecules and metabolites that are produced or consumed on the periphery of the metabolic network (Figure 1b).

Prior to separating out the various sources of stochasticity, the combined single-cell network of M metabolic and enzyme expression reactions can be captured by the chemical master equation (CME) [31]

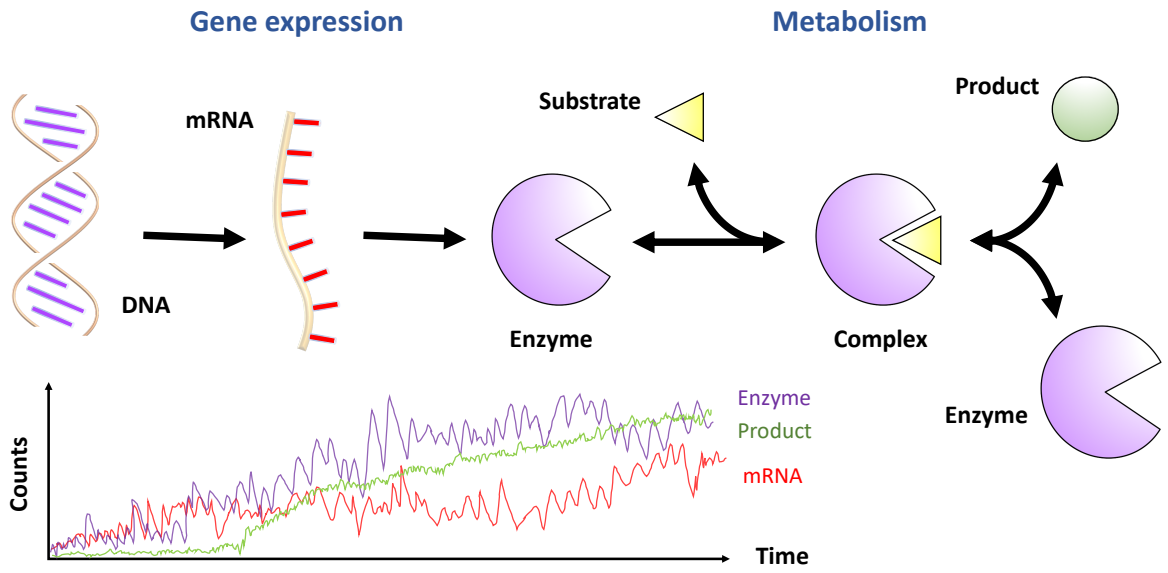
$$\frac{dP(\mathbf{n})}{dt} = \sum_{j=1}^{M} a_j(\mathbf{n}-\mathbf{S}_j)P(\mathbf{n}-\mathbf{S}_j) - a_j(\mathbf{n})P(\mathbf{n}) \tag{1}$$

where $\mathbf{n}$ is an N-dimensional vector for counts of N chemical species, $a_j(\mathbf{n})$ the propensity value of reaction j given $\mathbf{n}$, and $\mathbf{S}_j$ the stoichiometry of reaction j. The probability density function $P(\mathbf{n})$ describes the probability of the system to occupy state $\mathbf{n}$ at time t. Typically, the system of equations (1) (often infinite) can not be solved directly, and therefore Gillespie introduced SSA [27, 28] for sampling trajectories $\mathbf{n}(t)$ through state space that are exact realisations of the CME. One serious limitation of SSA is its computational cost when applied to large chemical reaction networks. Various adaptations of Gillespie's original algorithm, both exact and approximate, have

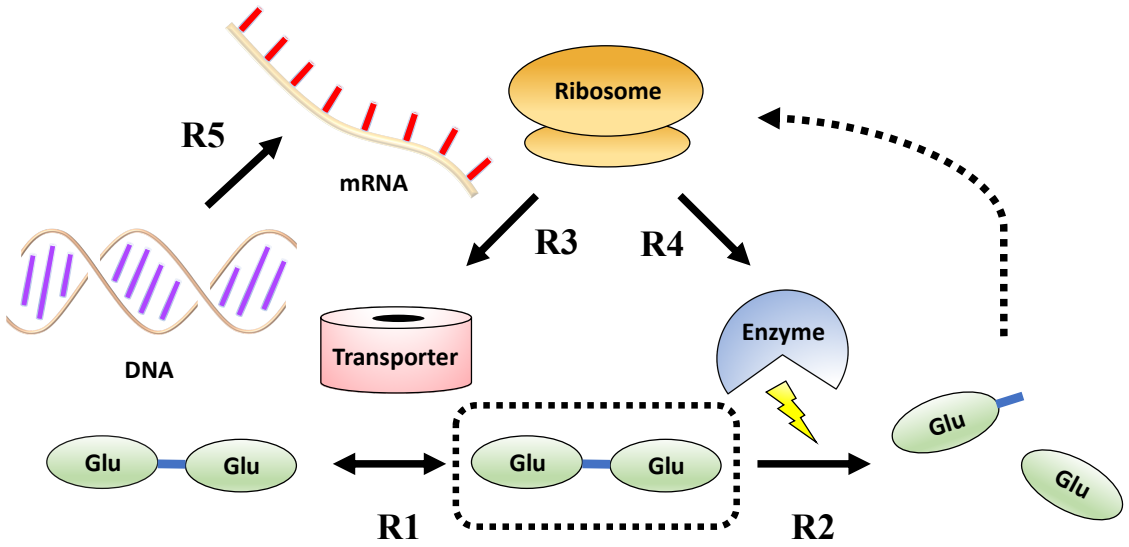
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1: Cartoon illustrations of single-cell metabolism and SSA-FBA.

(a) Extrinsic stochastic contributions come from fluctuations in the expression levels of enzymes catalysing metabolic reactions. Intrinsic stochastic contributions to metabolism are less by comparison.



(b) Simple SSA-FBA model where intracellular glutamylglutamic acid (enclosed by dotted box) is the only species internal to the metabolic reaction network, and therefore transport and cleavage reactions R1, R2 form the subset of SSA-FBA reactions with gene expression reactions R3, R4, R5 making up the SSA-only reactions. Reactions R3, R4, R5 thus have their propensity values calculated by a rate law that depends on the counts of external species (e.g., glutamic acid, ribosome, mRNA) while propensity values for R1, R2 are obtained by solving FBA. Since the rates of reactions R1 and R2 depend on the number of transporter and enzyme molecules, respectively, counts of these external species also serve to constrain SSA-FBA propensity values in the embedded LP problem.



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

therefore been proposed to improve the runtime of SSA, which can be particularly restricting for metabolism due to the multi-scale nature of enzymatic reactions. Since the average counts of enzymes are many times lower than those of metabolites, and metabolic reactions operate on a much faster time-scale than those of gene expression, the propensity values of these reactions can differ by several orders of magnitude. In order to deal with the computational challenges of multiple time-scales, several SSA methods that use a *stochastic quasi-steady state assumption* (a stochastic generalisation of the quasi-steady state assumption in DFBA [29]) have been introduced to simulate trajectories corresponding to a reduction of the CME (1) on the basis of time-scale separation (see [32, 33, 34] and Supplementary Appendix S1).

While the above methods for accommodating time-scale separation in SSA could in principle be used to simulate single-cell metabolism, in practice it is rarely the case that rate equations or parameters on which these methods depend are known for the majority of reactions in a metabolic network model. Consequently, our proposal is to borrow from the solution to this same problem in the constraint-based formalism of (D)FBA [15, 16, 29] where linear programming (LP) is used to identify a numerical solution to the deterministic quasi-steady state conditions of the metabolic reaction network. In SSA-FBA, reactions are separated into three mutually disjoint subsets based on whether the participating species are defined to be internal or external to the metabolic reaction network (see Figure 1b for illustration). The three subsets of reactions in an SSA-FBA model are:

- **FBA-only reactions**. Responsible for interconverting among internal species
- **SSA-only reactions**. Responsible for interconverting among external species
- **SSA-FBA reactions**. Responsible for interconverting between internal and external species

For example, the SSA-only reactions may correspond to reactions involved in gene expression, while the FBA-only and SSA-FBA reactions correspond to reactions involved in metabolism. Furthermore, since external species such as enzymes have the ability to determine the propensity values of reactions in the metabolic network, the constraints or bounds on those of the FBA-only and SSA-FBA reactions are allowed to vary as functions of external species. The LP problem

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

representing the metabolic reaction network therefore takes the form

$$\begin{aligned} &\text{maximise}: \quad z = \mathbf{c} \cdot \mathbf{a}_{FBA} \\ &\text{subject to}: \quad \mathbf{S} \cdot \mathbf{a}_{FBA} = 0, \quad \mathbf{l}(\mathbf{n}) \leq \mathbf{a}_{FBA} \leq \mathbf{u}(\mathbf{n}) \end{aligned} \tag{2}$$

where $\mathbf{a}_{FBA}$ is a vector containing the propensity values of FBA-only and SSA-FBA reactions, $\mathbf{S}$ is the stoichiometry matrix of the metabolic reaction network, and $\mathbf{l}(\mathbf{n}), \mathbf{u}(\mathbf{n})$ are bounds that depend on the counts of external species (see Supplementary Appendix S1 for extended discussion). The coefficient vector $\mathbf{c}$ is understood to be chosen in order for the cell to fulfil some biologically-relevant objective.

Propensity values for SSA-FBA reactions obtained from an optimal solution to the LP problem (2) are combined with the propensity values of SSA-only reactions, calculated from counts of external species directly, to determine the next reaction event according to Gillespie's original algorithm (although this step can also be replaced by more advanced methods). Since FBA-only reactions do not effect the counts of external species, their propensity values are not required at this stage. On the other hand, execution of either an SSA-only or SSA-FBA reaction updates the counts of external species, which in turn update the bounds of both FBA-only and SSA-FBA propensity values in (2). This bidirectional coupling implies that the associated FBA problem is "embedded" in SSA in much the same way an LP problem is embedded within an ODE in DFBA [29]. The relative scale of SSA-only to SSA-FBA propensity values obtained from the embedded FBA problem (2) serves as an additional parameter of an SSA-FBA model, as are the stoichiometry values of the reactions these correspond to. The choice of this relative scaling factor proves to be important for simulation efficiency and is related to the multi-scale nature of single-cell metabolism, because numerical exploration of the state space will become challenging if propensity values differ by several orders of magnitude.

It is important to highlight that, unlike various multi-scale versions of SSA attempting to best-approximate trajectories of the CME (1) [32, 33, 34], the goal of SSA-FBA is to serve partly as a lumping framework to build and simulate single-cell metabolic models involving large re-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

action networks lacking kinetic information [15, 16]. Thus, some degree of freedom is afforded to the model builder when making their choices for partitioning reactions into the three subsets (FBA-only, SSA-only, or SSA-FBA) and there is further flexibility regarding the relative scaling of SSA-only to SSA-FBA propensity values. Our provided guidelines are that this scale factor should be inversely correlated with the stoichiometry of SSA-FBA reactions (i.e., stoichiometry of SSA-FBA reactions should increase as their propensity values scale down relative to SSA-only propensity values) and distinguishing between internal and external species (and hence partitioning of reactions) should be loosely based on the understanding that time-scales of internal species are relatively fast compared to those of external species. Further considerations for SSA-FBA are outlined in Supplementary Appendix S1.

3 Implementation of SSA-FBA

The definition of SSA-FBA provided in Section 2 involves obtaining expected propensity values for SSA-FBA reactions from an embedded LP problem, whose constraints depend on the counts of species external to the metabolic reaction network. The resulting optimal expected propensity values in turn serve to determine the random selection of a reaction to execute during the next time interval of the SSA simulation. Since both SSA-only and SSA-FBA propensity values are taken into consideration during selection of this reaction, and because execution of a reaction from either set changes the counts of external species, implementing SSA-FBA following this exact approach appears to imply that runtime will scale in the number of reaction execution events in a way that depends on the complexity of the embedded LP problem.

A related issue arises in the simplest implementation of DFBA [29], where the Euler method is used to integrate the ODE over a time interval across which an optimal flux distribution does not to change. When the end of the interval is reached, the constraints of the embedded LP problem are updated using the current values of kinetic variables and re-solved for a new optimal flux distribution used to parameterise integration over the next interval. As the size of the time inter-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

val used for integration increases, fewer calls to the LP solver are required to complete a DFBA simulation, which can improve overall runtime considerably. We first explored a similar approach to approximating SSA-FBA simulations, also observing that a similar concept has been used to simulate whole cell models [25, 26]. The resulting simulations were no longer exact in the sense defined in Section 2, but we hypothesised that for a suitable approximation we might be able to improve overall runtime without overly compromising the accuracy of SSA-FBA. For demonstration purposes, we designed a toy model that contained only two SSA-only reactions (R0 and R1), one SSA-FBA reaction (R2, growth rate of *Mycoplasma genitalium*) and one variable FBA bound (oxygen transport reaction). The toy model contained three species S_0, S_1, S_2 and is represented by the reaction schema

$$S_0 \underset{\text{R1}}{\overset{\text{R0}}{\rightleftarrows}} S_1 \overset{\text{R2}}{\rightarrow} S_2, \tag{3}$$

where it is understood that the propensity value of reaction R2 depends indirectly on counts of species S_1, which bounds the maximal rate of the oxygen uptake reaction in the metabolic network model of *M. genitalium* [25] based on the functional form $u(S_1) = S_1/10$.

Simulations began with initial species counts $S_0 = 1000$, $S_1 = 0$, $S_2 = 0$, and the toy model was simulated across 3000 reaction execution events using either the exact (embedded LP problem updated and re-solved after the execution of every reaction event) or an approximate SSA-FBA method. The approximate SSA-FBA method only updated the constraints and re-solved the embedded LP problem after a fixed number, Δ_{event}, of reaction events had executed. As a general trend we found that when Δ_{event} was small the variances of the approximate SSA-FBA trajectories were comparable to those of exact SSA-FBA (Figures 2a and 2b), while increasing Δ_{event} increased the variance of approximate SSA-FBA trajectories correspondingly (Figure 2c). A further shortcoming of approximate SSA-FBA was that trajectories occasionally (particularly for larger Δ_{event} as visible in Figure 2c, but observed for all Δ_{event} tested) displayed numerical instability and diverged from realistic values towards the latter end of a simulation. This type of numerical instability also arises during direct integration of many DFBA models [35], but is particularly problematic for SSA-FBA because its stochastic nature means the pathology can not be conclusively ruled-out on

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

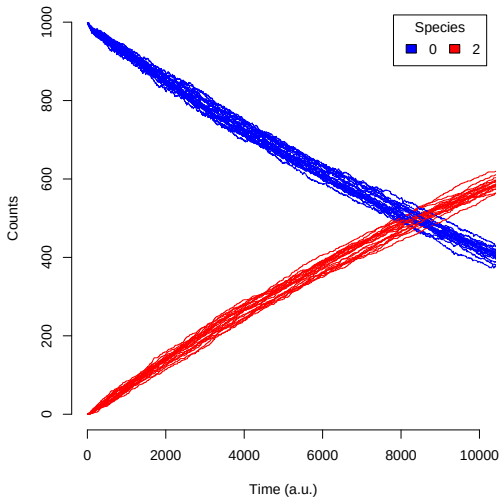
the basis of trial simulations (for example, to establish optimal size of Δ_{event}). It relates to the fact that the embedded LP can become infeasible within an integration interval, which induces a closed domain of definition for the dynamic system.

We therefore searched for a strategy to reduce the computational cost of exact SSA-FBA rather than risk the loss of accuracy and numerical stability associated with approximation, and were motivated by a recent method for numerical integration of ODEs with embedded LP problems [35] to derive an exact optimal basis SSA-FBA algorithm (Supplementary Appendix S2). The new simulation algorithm is guaranteed to be exact in the sense of Section 2, but is distinguished from the direct implementation of exact SSA-FBA because it improves efficiency and reduces the frequency of calls to an LP solver by using an optimal basis to calculate SSA-FBA and FBA-only propensity values. We evaluated the performance of our exact optimal basis SSA-FBA algorithm using the *M. genitalium* metabolic network model by selecting a random sequence of K execution events, each event corresponding to a selection of SSA-FBA or FBA-only reactions (between 1 and K_{max}) whose upper bounds are known in the original model (i.e., not zero or infinity). Execution of K events then involved updating the bounds of the reactions selected for the ith event by multiplying their previous value by some $0 < \sigma < 1$ so that the cumulative effect was to slowly drive the growth rate of *M. genitalium* to zero over the course of simulation. After execution of each event, SSA-FBA and FBA-only propensity values were calculated either by directly solving the LP problem or using the efficient optimal basis algorithm. Employing this approach rather than evaluating a full SSA-FBA simulation allowed us to confidently compare the performance of direct and efficient implementations, while confirming that calculated propensity values remained identical.

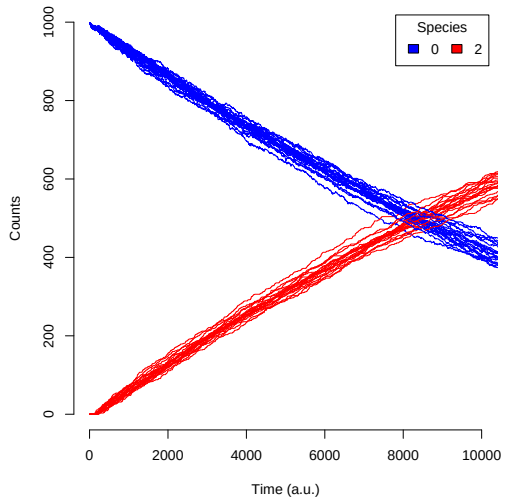
Figure 2d displays representative results of the comparisons between the efficient optimal basis algorithm and direct implementation of exact SSA-FBA, showing that the former improves runtime by an order of magnitude over the latter. The efficient optimal basis algorithm also scales better in the number of reaction execution events than the direct implementation (slopes of linear regressions using data from Figure 2d are 19.0 vs 1.7 for direct vs exact, respectively). We therefore conclude that our efficient optimal basis SSA-FBA algorithm can be used to significantly

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

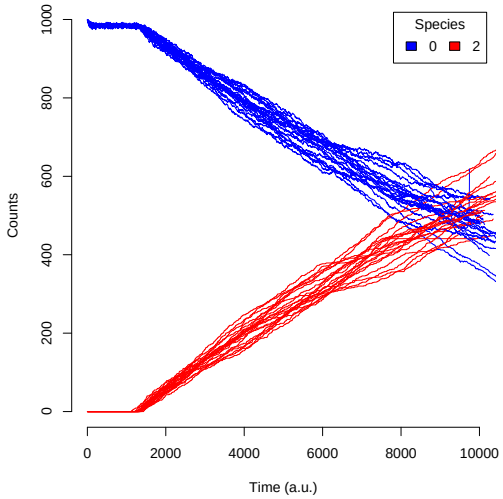
Figure 2: Comparison of exact and approximate SSA-FBA implementations using the *M. genitalium* metabolic network model. First three panels each display trajectories from 20 representative simulations, each consisting of 3000 reaction execution events. Error bars in fourth panel are standard deviations for runtimes across four simulations with parameters $\sigma = 0.9$, $K_{max} = 5$ and $K = 50000$, 100000, or 200000 described in the main text.



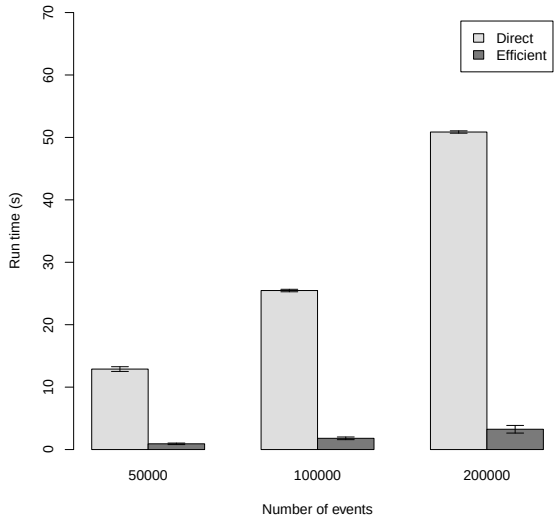
(a) Exact SSA-FBA simulation of toy model (equivalent to $\Delta_{event} = 0$).



(b) Approximate SSA-FBA simulation of toy model with $\Delta_{event} = 50$.



(c) Approximate SSA-FBA simulation of toy model with $\Delta_{event} = 500$.



(d) Comparison of runtimes for exact SSA-FBA implementations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

improve the runtime of exact SSA-FBA simulations.

4 Case study: *M. pneumoniae*

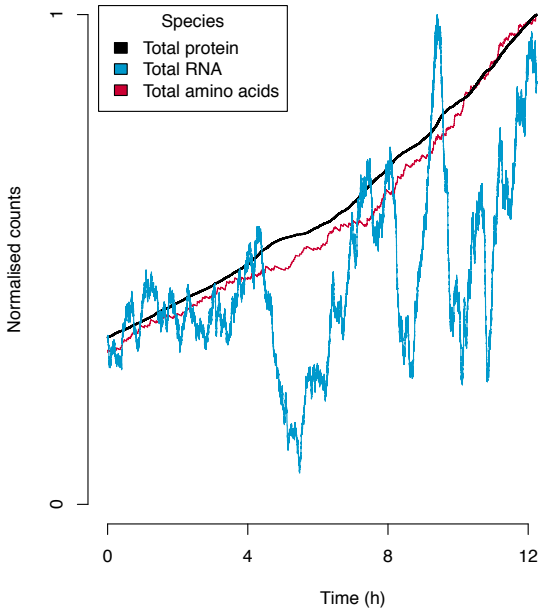
In this section we present preliminary results from using SSA-FBA to simulate the metabolism of a single *M. pneumoniae* cell. *Mycoplasma* are among the smallest bacterial cells yet discovered and *M. pneumoniae* the best known and examined [36], making it an excellent case study for understanding single-cell metabolism in a relatively simple organism. Its close relative *M. genitalium* was previously used to build a whole cell model [25].

We constructed a reduced metabolic reaction network of *M. pneumoniae* consisting of 86 metabolic reactions, including a metabolism production reaction that constitutes the objective function of the embedded FBA problem (see Supplementary Model file accompanying this manuscript). The full SSA-FBA model also has 30 protein species that are either enzymes or transporters regulating flux through the FBA reactions, or serve a direct role in gene expression (RNA polymerase, ribosomal protein, or RNAse). Furthermore, each protein species is associated with an mRNA molecule and three reactions corresponding to RNA transcription, RNA degradation, and protein synthesis, making 90 SSA-only reactions (176 reactions in total). Species on the periphery of the metabolic reaction network that appear as reactants in the Michaelis-Menten-like rate laws for SSA-only reactions (nucleoside triphosphates and amino acids) were designated as external species together with extracellular substrates, and therefore reactions of the embedded FBA model involved in their production or consumption were assigned to the SSA-FBA reaction subset. The remaining reactions of the embedded FBA model made up the FBA-only subset. The full list of parameter values and species are provided in the Supplementary Model file and associated code, but note that species internal to the metabolic reaction network, although recorded and tracked by the model, are not relevant to the SSA-FBA formalism.

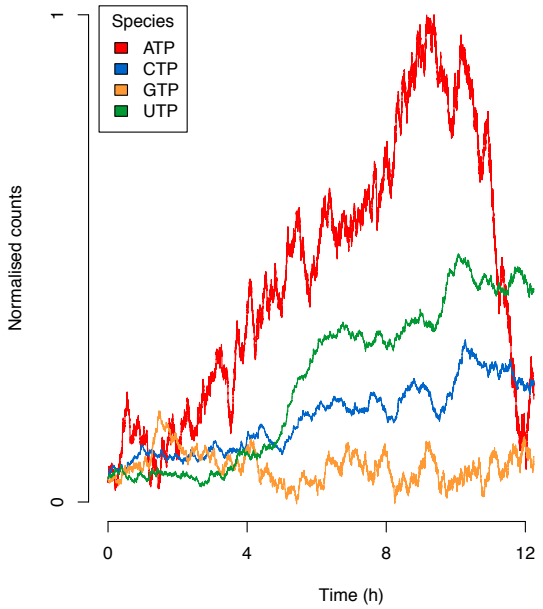
For the parameter values we evaluated, simulating one million reaction execution events took just over one minute on a personal laptop computer with a 2.8 GHz Intel Core i7 processor (76.86

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3: Example simulation of the *M. pneumoniae* SSA-FBA model used as a case study.



(a) Total counts of proteins, mRNA and amino acids normalised by their maximum over time.



(b) Counts of nucleoside triphosphates normalised by their maximum over time.

seconds with standard deviation of 1.4 seconds across four simulations). We experimented with the relative scaling of SSA-FBA and SSA-only propensity values, and identified a range of stoichiometry and scaling factors yielding simulations that produced biologically-realistic trajectories after a runtime of just 7-9 seconds on the same machine (Figure 3). The SSA-FBA model generally displayed a wide range of different behaviours, where alternative groups of metabolite or protein species would tend to accumulate more than others in each simulation instance, supporting the notion that stochasticity at the single-cell level is a driver for metabolic heterogeneity [10, 21, 24]. One serious limitation of the current model is that no explicit method is in place to prevent the counts of some metabolites from becoming negative, which occurred on occasion during individual simulations. Ongoing work aims to address this issue, obtain more biologically-realistic parameter values and rate laws, and appropriately quantify the stochastic nature of SSA-FBA and its relevance for metabolic heterogeneity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 Conclusion

In this paper we have presented what we believe to be the first comprehensive framework for simulating single-cell metabolism in a way that captures effects of stochasticity in the context of large metabolic reaction networks. SSA-FBA involves a hybrid method for embedding FBA in SSA, which is well-suited to using metabolic models for which kinetic information is lacking. We also have proposed an advanced optimal basis algorithm for exact simulation of SSA-FBA models, which can otherwise become computationally expensive due to the nature of embedding an LP problem in SSA with multiple time-scales. A preliminary case study using our algorithm to simulate the metabolism of a single *M. pneumoniae* demonstrates that SSA-FBA has the potential to reveal how stochasticity at the single-cell level impacts metabolic heterogeneity at the level of the population.

Future extensions of our work will involve applying SSA-FBA to larger, more realistic models of single cells, in addition to reconsidering how standard constraint-based formulations of metabolite pools and an objective function should be adapted to suit the biological nature of single-cell metabolism. More complex models of single-cell metabolism could include mechanisms that impart regulatory control of stochasticity [37], and pave the way for combining insights from simulation with experimental advances in microfluidics [38] or real-time quantification of RNA translation events within individual cells [39]. From an algorithmic perspective, SSA-FBA could be extended to incorporate additional time-scales governed by continuous stochastic or deterministic processes (e.g., [40]), and its computational efficiency perhaps further enhanced through parallelisation methods [41].

Acknowledgements

DST is a Simons Foundation Fellow of the Life Sciences Research Foundation. The work of JRK is supported by National Science Foundation award 1649014 and National Institutes of Health award R35GM119771. We would like to extend our thanks to ME Beber and J Carrasco Muriel for their

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

close collaboration with the first author on a related project that provided some of the motivation for the current study.

References

[1] Angerer P, Simon L, Tritschler S, Wolf FA, Fischer D, Theis FJ (2017) Single cells make big data: new challenges and opportunities in transcriptomics. *Curr. Opin. Syst. Biol.* 4: 85-91.

[2] Stuart T, Satija R (2019) Integrative single-cell analysis. *Nat. Rev. Genet.* 20: 257-272.

[3] Ackermann M (2015) A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* 13: 497-508.

[4] Schreiber F, Ackermann M (2020) Environmental drivers of metabolic heterogeneity in clonal microbial populations. *Curr. Opin. Biotechnol.* 62: 202-211.

[5] Dagogo-Jack I, Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15: 81-94.

[6] Xiao Z, Locasale JW, Dai Z (2020) Metabolism in the tumor microenvironment: insights from single-cell analysis. *Oncoimmunology* 9: 1726556.

[7] Rubakhin SS, Lanni EJ, Sweedler JV (2013) Progress toward single cell metabolomics. *Curr. Opin. Biotechnol.* 24: 95-104.

[8] Fessenden M (2016) Metabolomics: small molecules, single cells. *Nature* 540: 153-155.

[9] Emara S, Amer S, Ali A, Abouleila Y, Oga A, Masujima T (2017) Single-Cell Metabolomics. In: Sussulini A. (eds) *Metabolomics: From Fundamentals to Clinical Applications. Advances in Experimental Medicine and Biology*, vol 965. Springer, Cham.

[10] Evers TMJ, Hochane M, Tans SJ, Heeren RMA, Semrau S, Nemes P, Mashaghi A (2019) Deciphering metabolic heterogeneity by single-cell analysis. *Anal. Chem.* 91: 13314-13323.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[11] Duncan KD, Fyrestam J, Lanekoff I (2019) Advances in mass spectrometry based single-cell metabolomics. *Analyst.* 144: 782-793.

[12] Damiani C, Gaglio D, Sacco E, Alberghina L, Vanoni M (2020) Systems metabolomics: from metabolomic snapshots to design principles. *Curr. Opin. Biotechnol.* 63: 190-199.

[13] Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell.* 144: 646-674.

[14] Pavlova NN, Thompson CB (2016) The emerging hallmarks of cancer metabolism. *Cell Metab.* 12: 27-47.

[15] Edwards JS, Covert M, Palsson B (2002) Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol. 4: 133-140.

[16] O'Brien EJ, Monk JM, Palsson BO (2015) Using genome-scale models to predict biological capabilities. *Cell* 161: 971-987.

[17] Angione C (2019) Human systems biology and metabolic modelling: a review- from disease metabolism to precision medicine. *Biomed Res. Int.* 2019: 8304260.

[18] Yang L, Ebrahim A, Lloyd CJ, Saunders MA, Palsson BO (2019) DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression. *BMC Syst Biol.* 13: 2.

[19] Waldherr S, Oyarzún DA, Bockmayr A (2015) Dynamic optimization of metabolic networks coupled with gene expression. *J. Theor. Biol.* 365: 469-485.

[20] Kiviet DJ, Nghe P, Walker N, Boulineau S, Sunderlikova V, Tans SJ (2014) Stochasticity of metabolism and growth at the single-cell level. *Nature* 514: 376-379.

[21] Wehrens M, Büke F, Nghe P, Tans SJ (2018) Stochasticity in cellular metabolism and growth: approaches and consequences. *Curr. Opin. Syst. Biol.* 8: 131-136.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[22] Damiani C, Maspero D, Di Filippo M, Colombo R, Pescini D, Graudenzi A, Westerhoff HV, Alberghina L, Vanoni M, Mauri G (2019) Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. *PLoS Comput. Biol.* 15: e1006733.

[23] Levine E, Hwa T (2007) Stochastic fluctuations in metabolic pathways. *Proc. Natl. Acad. Sci. USA*. 104: 9224-9229.

[24] Tonn MK, Thomas P, Barahona M, Oyarzún DA (2019) Stochastic modelling reveals mechanisms of metabolic heterogeneity. *Commun. Biol.* 2: 108.

[25] Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr., Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150: 389-401.

[26] Karr JR, Takahashi K, Funahashi A (2015) The principles of whole-cell modeling. *Curr. Opin. Microbiol.* 27: 18-24.

[27] Gillespie DT (1976) General method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22: 403-434.

[28] Gillespie DT (1977) Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* 81: 2340-2361.

[29] Mahadevan R, Edwards JS, Doyle FJ 3rd (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83: 1331-1340.

[30] Elowitz MB, Levine AJ, Siggia ED (2002) Stochastic gene expression in a single cell. *Science* 279: 1183-1186.

[31] Gillespie DT (1992) A rigorous derivation of the chemical master equation. *Physica A* 188: 404-425.

[32] Rao CV, Arkin AP (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.* 118: 4999-5010.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.110577; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[33] Cao Y, Gillespie DT, Petzold LR (2005) The slow-scale stochastic simulation algorithm. *J. Chem. Phys.* 122: 014116.

[34] Cao Y, Gillespie DT, Petzold LR (2005) Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comput. Phys.* 206: 395-411.

[35] Harwood SM, Höffner K, Barton PI (2016) Efficient solution of ordinary differential equations with a parametric lexicographic linear program embedded. *Numer. Math.* 133: 623-653.

[36] Waites KB, Talkington DF (2004) *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin. Microbiol. Rev.* 17: 697-728.

[37] Borri A, Palumbo P, Singh A (2016) Impact of negative feedback in metabolic noise propagation. *IET Syst. Biol.* 10: 179-186.

[38] Leygeber M, Lindemann D, Sachs CC, Kaganovitch E, Wiechert W, Nöh K, Kohlheyer D (2019) Analyzing microbial population heterogeneity-expanding the toolbox of microfluidic single-sell cultivations. *J. Mol. Biol.* 431: 4569-4588.

[39] Morisaki T, Lyon K, DeLuca KF, DeLuca JG, English BP, Zhang Z, Lavis LD, Grimm JB, Viswanathan S, Looger LL, Lionnet T, Stasevich TJ (2016) Real-time quantification of single RNA translation dynamics in living cells. *Science* 352: 1425-1429.

[40] Haseltine EL, Rawlings JB (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* 117: 6959.

[41] Goldberg AP, Jefferson DR, Sekar JAP, Karr JR (2020) Exact parallelization of the stochastic simulation algorithm for scalable simulation of large biochemical networks. *arXiv preprint* https://arxiv.org/abs/2005.05295