Stochastic method for isotope labeling systems

Stochastic simulation algorithm for isotope labeling metabolic networks

Quentin Thommen,¹ Julien Hurbain,² and Benjamin Pfeuty²

¹Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, UMR9020-U1277
- CANTHER - Cancer Heterogeneity Plasticity and Resistance to Therapies,
F-59000
²Univ. Lille, CNRS, UMR 8523 - PhLAM - Physique des Lasers Atomes et Molécules,
F-59000 Lille, France

(*Electronic mail: quentin.thommen@univ-lille.fr)

(Dated: 4 January 2022)

Carbon isotope labeling method is a standard metabolic engineering tool for flux quantification in living cells. To cope with the high dimensionality of isotope labeling systems, diverse algorithms have been developed to reduce the number of variables or operations in metabolic flux analysis (13C-MFA), but lacks generalizability to dynamic metabolic responses (13C-DMFA).

In this study, we present a stochastic simulation algorithm (SSA) derived from the chemical master equation of the isotope labeling system. This algorithm allows to compute the time evolution of isotopomer and metabolite concentrations in non-steady conditions where the computational time only scales with the number of reactions of the metabolic system, not the number of isotopomers. The computational efficiency of the algorithm is benchmarked with two metabolic networks of different sizes and topologies. SSA method for the forward simulation of the isotope labeling system can be easily combined with Monte Carlo Markov Chain method for the inverse problem of dynamic flux estimation.
I. INTRODUCTION

In a network of (bio)chemical reactions, the temporal evolution of the state probabilities is described by the chemical master equation (CME) through a general formalism\(^1\). Deterministic kinetic rate equations, on the one hand, can be derived from the first moments of the probability distribution and allow for a thorough analysis of the network dynamics by various analytical techniques\(^2\). The probabilistic features of the dynamics such as bimodal distributions or coefficients of variation, on the other hand, can be investigated with stochastic implementation of the CME through well-established stochastic simulation algorithms\(^3,4\). Many biochemical reaction networks of interest, such as metabolic networks, have a sufficiently large number of chemical species to justify the prevalence and relevance of deterministic approaches. In case of metabolic networks, a wide range of theoretical tools based on balance rate equation framework (flux balance analysis Orth, Thiele, and Palsson\(^5\), metabolic control theory Kell and Westerhoff\(^6\), metabolic flux analysis Zamboni et al.\(^7\)) have therefore been developed over the last decades to ease a comprehensive analysis of metabolic processes. In contrast, stochastic modeling is mostly useful to study fluctuations in metabolic pathways arising from a small number of some molecular species (e.g., enzymes)\(^8,9\) but is prone to be computationally inefficient because of high species concentration and different reaction timescales. Here, we explore an alternative proposal that stochastic simulation algorithm can be more efficient than deterministic approaches to study some averaged properties of isotope labeling metabolic systems.

Metabolic flux analysis (MFA) is an example of metabolic engineering techniques developed essentially within a deterministic modeling framework. Carbon isotope labeling methods have motivated a set of refined MFA methods involving rate equations of isotopomers, cumomers or elementary metabolite units\(^11,13\) and adapted to whether labeling or/and flux patterns are stationary or dynamics. While 13C-MFA performs metabolic and isotopic steady-state flux analysis, non-stationary 13C-MFA (13C-NMFA) incorporates transient 13C labeling data, but still assumes constant metabolic fluxes and constant metabolite pool sizes for intracellular metabolites\(^14,15\). Several dynamic metabolic flux analysis (13C-DMFA) methods have further been proposed to experimental contexts where fluxes are expected to vary in time\(^16,18\) though “rigorous 13 C-DMFA methods for analyzing systems where changes in fluxes and changes in 13 C-labeling occur at similar time-scales are still relatively underdeveloped.”\(^19\). Indeed, several powerful specialized software packages have been developed for 13C-MFA, but not yet so far for 13C-DMFA though a gen-
eralized computational framework using EMU-based simulation and B-splines has been recently proposed[18].

In contrast to such trend towards more complex and powerful deterministic algorithms, we propose here a Stochastic Simulation Algorithm (SSA) specifically designed for simulating the dynamics of the isotope labeling patterns (i.e., forward problem) that could be combined with standard methods of bayesian inference for the estimation of concentration and fluxes from isotope labeling patterns (i.e., inverse problem). The proposed algorithm is both simple to implement, generic and computationally efficient. The forward problem is first formulated in the framework of the chemical master equation. The internal state of the isotopomer network can be typically simulated by two numerical schemes: a deterministic schemes in the limit of high number of metabolite species and a stochastic Monte Carlo scheme based on the discreteness of enzymatic reaction processes. A comparative analysis of the two algorithms is performed for two different metabolic networks: the first network mimics a simple and tunable propagation scheme of carbon atoms while the second network consists in the pentose phosphate pathways featured with a complex propagation scheme of carbon atoms. The stochastic scheme (SSA) is shown to become much more efficient for large number of isotopomers per metabolite species.

II. CHEMICAL MASTER EQUATION MODEL FOR ISOTOPE LABELING NETWORKS

A. Isotopic labeling networks

An isotope labeling network contains information to track the propagation of labeled atoms in a chemical reaction network. The labeling state associated to isotopic changes is represented by an ordered sequence of atoms which defines an (positional) isotopomer. The distribution of labeling state evolves over time as the atoms propagate in the network. Let us illustrate this with a particular Bi Bi reaction of the non-oxidative branch of the pentose phosphate pathway mediated by the enzyme transaldolase (Figure 1a). The sequences of letters depict the permutation of carbon atoms during the reaction. The transaldolase enzyme breaks Sedoheptulose 7-phosphate (S7P) into two subunits and reassembles them with glyceraldehyde 3-phosphate (GAP) into two new molecules: erythrose 4-phosphate (E4P) and fructose 6-phosphate (F6P)[20]. Thus, the fourth carbon of S7P becomes the first carbon of E4P while the first carbon of GAP becomes the fourth
FIG. 1. Atom rearrangements in metabolic reactions. (a) Rearrangement of the carbon sequence by the enzyme transaldolase. The carbons of the two reactants, conventionally represented by letter sequences in the writing of the chemical reaction, or schematically here by colored squares are reorganized by transaldolase to form the products of the reaction. (b) The upper glycolytic pathways supplemented by the pentose phosphate pathway provides an example of isotope labeling network. Emphasis is put on the propagation of carbons which are represented by colored squares. Chemical reactions rearrange carbons, often reversibly, thus allowing complex propagation.

carbon of F6P.

Figure 1b illustrates a typical example of metabolic isotope labeling network comprising the upper part of glycolysis and the pentose phosphate pathway\cite{20}. Although this network represents a restricted set of metabolic pathways of a cell, its complexity comes from the large number of carbon isotopomer (e.g., 128 for S7P), the reversibility of reaction and the specific rearrangement of carbon during Bi Bi reactions. Such entanglement and reversibility of the chemical reactions pro-
duce a complex propagation pattern of the carbon atoms and, thus, a complex dynamics of carbon isotopomer distribution. The choice of this particular network builds upon recent and systematic experimental investigations by isotopic labeling methods substituting one or more 12C by a 13C. Mass spectroscopy techniques allow to measure at each time increment the number of carbon 13 of each species, but not the position. Following the import of labeled species, the propagation of the label is typically assessed by time measurements of mass isotopomer distribution which provide information on the network fluxes.

B. Isotopomer index and addressing operators

A chemical reaction network such as the one depicted in Figure 1 is defined by \( K \) reactions between \( M \) chemical species whose concentrations are denoted by \( S_m \) with \( m \in [1,M] \). The labeling states of the species \( S_m \) is an ordered sequence \( (s_{m,1}, s_{m,2}, \ldots, s_{m,l_m}) \) of length denoted \( l_m \) made of elements \( s_{m,i} \in [0,q-1] \). The species \( S_m \) has therefore \( L_m = q^{l_m} \) different labeling states or positional isotopomer indexed by \( n_m = \sum_{i=1}^{l_m} s_{m,i} q^{i-1} \), also noted \( n_m = (s_{m,1}, s_{m,2}, \ldots, s_{m,l_m})_q \) and called the isotopomer index. A similar approach restricted to \( q = 2 \) has already been introduced to describe isotopomer distribution vectors. In the case 13C-labeling, each carbon may be in two different states (i.e., \( q = 2 \)) and the sequence (0, 0, 1, 0, 0, 0, 0) for S7P indicates that 13C label is in third carbon position and corresponds to the labeling state number 4, the S7P species has therefore \( 2^7 \) different labeling states.

If the permutation rule is known, one can define addressing operators that compute the isotopomer index of the product from the isotopomer index of the reactant, and vice versa for each reaction of the network. The addressing operator forms an alternative to atom mapping matrices defined by Zupke et al. In the case of the reaction mediated by transaldolase (Figure 1a), the addressing operators

\[
\sigma_{F6P}(n_{S7P}, n_{GAP}) = (s_{a,1}, s_{a,2}, s_{a,3}, s_{b,1}, s_{b,2}, s_{b,3})_q \tag{1}
\]

\[
\sigma_{E4P}(n_{S7P}, n_{GAP}) = (s_{a,4}, s_{a,5}, s_{a,6}, s_{a,7})_q, \tag{2}
\]

compute the product index from reactant index \( n_{S7P} = (s_{a,1}, \ldots, s_{a,7})_q \) and \( n_{GAP} = (s_{b,1}, s_{b,2}, s_{b,3})_q \). In the same manner,

\[
\sigma_{S7P}(n_{F6P}, n_{E4P}) = (s_{c,1}, s_{c,2}, s_{c,3}, s_{d,1}, s_{d,2}, s_{d,3}, s_{d,4})_q \tag{3}
\]

\[
\sigma_{GAP}(n_{F6P}, n_{E4P}) = (s_{c,4}, s_{c,5}, s_{c,6})_q \tag{4}
\]
Stochastic method for isotope labeling systems

compute the reactant index from product index \( n_{F6P} = (s_{c,1}, \ldots, s_{c,6})_q \) and \( n_{E4P} = (s_{d,1}, \ldots, s_{d,4})_q \).

Therefore, in the context of a 13C labeling \((q = 2)\), the reaction between a doubly labeled S7P (1000100) and a simply labeled GAP (001) – i.e. \( n_{S7P} = 17 \) and \( n_{GAP} = 4 \) – produces an F6P (100001) and an E4P (0100) – i.e. \( n_{F6P} = 33 \) and \( n_{E4P} = 2 \).

C. Chemical master equation description

The chemical master equation (CME) is a general and accurate formalism to describe the stochastic dynamics in (bio)chemical reaction networks.\(^\text{[25]}\) This formalism can be easily extended to also describe the stochastic dynamics of labeling states of chemical species. With above notations for isotopomer index and addressing operators, the probabilistic dynamics in isotope labeling network can naturally be described in the CME framework. The chemical species with the largest number of isotopomers determines the size \( N \) of the state space. In the case of the network in Figure [1b] with a two-state labeling, the size of the state space is determined by the number of isotopomer of S7P that is \( N = 2^7 \). The state of the whole system is therefore described by a \( M \times N \) integer matrix \( \omega \): \( \omega_{m,n} \) indicating the number of \( m \)th species in the \( n \)th labeling state. The total number of the \( m \)th species is noted \( \Omega S_m = \sum_n \omega_{m,n} \) where \( \Omega \) is a volume (involved as a scaling factor) and \( S_m \) is a concentration. The probability that the internal sequence of the \( m \)th species corresponds to the \( n \)th isotopomer is denoted \( \rho_{m,n} = \omega_{m,n} / \Omega S_m \).

The formalism of the CME describes the temporal evolution of the probability of the system to be in the state \( \omega \), noted \( \mathcal{P}_\omega(t) \).

The chemical reactions that define the network are characterized by both a concentration-dependent flux of reagents \( v_k \) with \( k \in [0, K] \) and a permutation rule between the position of labeled atoms of reactants and products. Reactions are distinguished depending on their input, output, or internal position in the network. For instance, the network depicted in Figure [1b] has one input reaction, 3 output reactions and 11 internal reactions. As seen latter on, input reactions always require a particular consideration since the reactant is not modified. For keeping notations simple, we restrict to Bi Bi reactions of the type \( A + B \rightarrow C + D \) where \( A, B, C, D \) are either chemical species or empty sets. In this case, the CME reads,
Stochastic method for isotope labeling systems

\[
\frac{d}{dt} \rho_{(\omega)}(t) = \sum_{k=1}^{K_\omega} \sum_{n,n'} \left[ \frac{E_{a_k,n} + E_{b_k,n'}}{\Omega S_m} - 1 \right] \rho_{a_k,n} \rho_{b_k,n'} \rho_{(\omega)}(t)
\]

where \( E_{m,n} \) is a matrix of the canonical base (only the element at the intersection of row \( m \) and column \( n \) is non-zero and is unity), and \( \delta_{i,j} \) the Kronecker symbol (unity if indexes are equal, zero either). The \( K_n \) first reactions concern internal and output reactions whereas the remaining \( K_I \) concerns input reactions (\( \emptyset \rightarrow C \)). In this latter case, \( I_{c_k,n} \) is the fixed probability to have an isotopomer \( n \) of the input species \( c_k \).

### III. DERIVED SIMULATION ALGORITHMS

#### A. Deterministic Simulation Algorithm (DSA)

The CME (Equation 5) can be approximated in the large size limit \( \Omega \rightarrow \infty \) by deterministic rate equation dynamics. The probability of each internal sequence is denoted by \( \rho_{m,n} \) such that \( S_{m,n} = S_m \rho_{m,n} \) describes the concentration of \( n \)-th isotopomer of the \( m \)-th species. The time evolution of isotopomer concentrations is governed by the distribution rules specific to each reaction and is formalized mathematically by a permutation of the concatenated internal sequence between the reagents and the products. The deterministic system evolves according to the ordinary differential equations,

\[
\frac{d}{dt} S_{m,n} = \sum_{k=1}^{K} N_{m,k} v_k(S) \Phi_{m,n}^k(\rho),
\]

where \( N \) denotes the stoichiometry matrix, \( v_k \in [0,K] \) the concentration-dependent flux of reagents, and \( \Phi_{m,n}^k(\rho) \) the flux fraction describing the permutation rules of chemical reaction satisfying \( \sum_n \Phi_{m,n}^k(\rho) = 1 \). Algorithm to simulate **Equation 7** with standard Runge-Kutta-Fehlberg method of order 5 with adaptive step is called Deterministic Simulation Algorithm (DSA).
Let us first consider the internal reactions restricted to Bi Bi reactions of the form \( A + B \rightarrow C + D \). The reaction is characterized by the reordering of atom position defining addressing operations of the products according to the indices of the reagents. The operator \( \sigma_A(n_c, n_d) \) gives the isotopomer index of the \( A \) species that produce \( C \) and \( D \) of isotopomer index \( n_c \) and \( n_d \), respectively. In that case, the reaction index \( k \) is omitted and the flux fraction reads,

\[
\Phi_{a,n_a}(\rho) = \rho_{a,n_a} \\
\Phi_{b,n_b}(\rho) = \rho_{b,n_b} \\
\Phi_{c,n_c}(\rho) = \sum_{n_d=0}^{L_d-1} \rho_{a,\sigma_A(n_c,n_d)} \rho_{b,\sigma_B(n_c,n_d)} \\
\Phi_{d,n_d}(\rho) = \sum_{n_c=0}^{L_c-1} \rho_{a,\sigma_A(n_c,n_d)} \rho_{b,\sigma_B(n_c,n_d)}
\]

where \( L_c \) and \( L_d \) are the number of isotopomers of \( C \) and \( D \) species. If \( B \) is an empty set then \( \rho_{b,x} = 1 \), and if \( D \) is an empty set Equation (11) is useless. If the reaction is an output reaction, then \( C \) and \( D \) are empty sets, \( \Phi_{a,n_a}(\rho) \) and \( \Phi_{b,n_b}(\rho) \) are computed with the same rules as internal reactions. As mentioned, the input reactions must be treated separately since the reactants are not variables but parameters. We consider here input reactions of simple form \( \emptyset \rightarrow C \) and we note \( I_n \) the probability of synthesis of the species \( C \) in the state \( n \), thus \( \Phi_{c,n_c}(\rho) = I_n \) in this case.

Alternatively, the dynamical system (Equation 7) can also be written as

\[
\frac{d}{dt} S_m = \sum_{k=1}^{K} N_{m,k} v_k(S) \quad m \in [0,M] \\
S_m \frac{d}{dt} \rho_{m,n} = \sum_{k=1}^{K} N_{m,k} v_k(S) \left( \Phi_{m,n}(\rho) - \rho_{m,n} \right)
\]

The first equation describes the time evolution of species concentrations while the second equation describes the time evolution of the fraction of different isotopomers. This additional equation highlights the key role of the concentrations \( S_m \) in the timescale of changes in isotopomer distribution: higher concentration values lead to slower evolution of isotopomer distributions.

The permutation rules defined in \( \Phi \) may be easily extended to more complex reactions. In the framework developed here, they only depend on the permutation relations and not on the mathematical forms of concentration-dependent flux, because we assume that the internal modification does not impact the reaction rate. If the construction rules are simple to establish and to implement in a numerical code, the computation time of the flux vector of the dynamic system (the right-hand
Stochastic method for isotope labeling systems

side term of Equation 7 increases significantly with the length of the sequences and the number of isotopomers, because of the many summations of terms. Moreover, this implementation computes the evolution of all possible isotopomers while the experimental labeling used nowadays generates only a small subset of the possible isotopomers. The deterministic system therefore requires a large number of unnecessary calculations even with an optimized implementation. It nevertheless serves as a useful benchmark to check the relative accuracy and efficiency of other methods.

B. Stochastic Simulation Algorithm (SSA)

The CME is in fact a continuous-time approximation of discrete time stochastic processes. Stochastic algorithms are often used to simulate the molecular dynamics in chemical reaction networks and capture the statistical and temporal features of fluctuations. In the case of the above CME (Equation 5), time evolution of isotopomer distribution can also be simulated by a stochastic Monte-Carlo algorithm based on the next reaction methods, here called Stochastic Simulation Algorithm (SSA). Each chemical species is represented by a finite sample of isotopomers (Figure 2) where the sampling size is proportional to the concentration of the corresponding chemical species. The sample size of the variable $m$ is $\Omega S_m$ where $\Omega$ is a volume. The occurrence of a chemical reaction is determined by the standard next reaction methods that we have adapted, SSA is summarized as:

Init: Compute the reaction time for all reactions

$$
\tau_k = \frac{1}{\Omega v_k(S)} \log \left( \frac{1}{U_k} \right)
$$

where $U_k$’s are independent uniform random deviates in $[0, 1]$;

Step 1: Find the smallest reaction time $\tau_{k'} = \min(\tau_k)$ and do reaction $k'$ by randomly picking the reagents from their samples and synthesizing the products following the permutation rule of the reaction, as schematically represented in the Figure 2;

Step 2: Increment time $t$ by $\tau_{k'}$ and compute a next time for reaction $k$;

Step 3: Adjust the set of reaction times to account for sample size variation induced by reaction $k'$
Stochastic method for isotope labeling systems

FIG. 2. Monte Carlo method for time evolution of isotopomer distribution. The Bi Bi reaction in the non-oxidative branch of the pentose phosphate pathway mediated by the Transaldolase. Each chemical species is represented by a set of molecules in a given labeling state. When the reaction $k$ occurs at a time $\tau_k$, two reactants are randomly picked and removed from the corresponding set (S7P and GAP here), rearrangement is performed, and the products are added in the corresponding set (F6P and GAP here).

and iterate to Step 1

In this sequential process, each stochastic occurrence of a chemical reaction induces discrete changes in the number of species and of isotopomers associated to each chemical species, which results in stochastic fluctuations of both species concentrations and isotopomer distribution.

As if, fluctuations come from two distinct mechanisms: on the one hand the fluctuation of the reaction times $\tau_k$ and, on the other hand, the random choices of the reactants in the lists. In the context of metabolic networks, fluctuation of the reaction times $\tau_k$ are rarely relevant. Because of the high copy number of metabolites, numerous reactions occur and fluctuations of the reaction
times $\tau_k$ do not induce much concentration fluctuations. Here this can be obviously obtained by increasing the volume $\Omega$ such that the copy number in the model correspond to the real metabolite copy number. However, one can also modify the algorithm to cancel fluctuations in reaction times $\tau_k$ by replacing the random variables $U_k$ by constant value that keeps $E(\log(1/U_k))$ unchanged ($U_k = e^{-1}$ $\forall k$ in this case). In this way, $\Omega$ parameter value may be much smaller and represent the typical sample size used for isotopomers. Only the relative abundances of the isotopomers display stochastic fluctuations in this case. If $\Omega = 100 \times 10^6/N_A$ L, with $N_A$ the Avogadro constant, then a concentration of $1 \mu$M corresponds to 100 copies in the SSA, it is therefore more convenient to express $\Omega$ in unit of the number of copies per microMolar ($nc/\mu$M).

IV. A COMPARATIVE ANALYSIS BETWEEN DETERMINISTIC AND STOCHASTIC ALGORITHMS

A. An archetypical network for algorithms benchmarking

The archetypical network schematically presented on the Figure 3 establishes a simple framework to benchmark the proposed algorithms. This network comprises 8 chemical species ($S_m$) made of two subunits which are reordered through 10 chemical reactions including two input reactions and two output reactions. One of the input reactions may introduce a labeling of the first $S_1$ subunit (asterisk on Figure 3). The labeling is transmitted to the output species $S_2$ through the upper network path ($S_1 \rightarrow S_6 \rightarrow S_8 \rightarrow S_2$) and to the output species $S_4$ through the lower path $S_1 \rightarrow S_5 \rightarrow S_7 \rightarrow S_2$. The upper and the lower paths define the elementary mode of the flux balance analysis.

In the simple case of this archetypical network, the deterministic dynamical system that model the isotopomers kinetics is straightforward to derive, due to the simplicity of the network, and serves as a control of the algorithm numerical implementation. For simplicity, the law of mass action is used to investigate the network in non-equilibrium condition and the values of the kinetic constants are set to unity except the zero-order kinetic constants driving $S_2$ production rate that is used as a control parameter (see table 2 of the supplementary information ).
Stochastic method for isotope labeling systems

FIG. 3. Elementary isotope labeling network used to illustrate the Monte-Carlo algorithm for the simulation of labeling dynamics. Each $S_i$ species contains two atoms that can be marked and represented by colored squares. The colors make it possible to follow the propagation of labeling along the reaction network by applying the permutation rules of the various chemical reactions. $v_{ij}$ denote the chemical reaction rates. The black star indicates the labeling of the input species $S_1$ and its propagation through the network. The output species $S_2$ is labeled only by the flow through the upper path.

B. Forward simulations of labeling patterns in non-equilibrium flux condition

For comparison, DSA and SSA are both implemented for both networks depicted in Figure 1b and 3. The law of mass action is used and the values of the kinetic constants are set to unity for both network (see table 1 and 2 of the supplementary information). The control parameters are two zero-order kinetic constants driving : (1) the G6P production rate in the realistic network Figure 1b named $v_{G6P}$ in the following, and (2) $S_2$ production rate in archetypical network Figure 3 named $v_2$. At initial time, networks are in steady state without labeling ($v_{G6P} = 1$ and $v_2 = 0.1$, then a sudden increase of control parameters ($v_{G6P} = 1 \rightarrow 2$ and $v_2 = 0.1 \rightarrow 0.5$) and the labeling are simultaneously applied (Figure 4a-b). The used labeling concerns the first atom of G6P and $S_1$. Both deterministic (solid lines) or stochastic (dots) methods show a similar temporal evolution of the species concentration and isotopomer fractions following the increase of the controls parameter, demonstrating the consistency between the two methods. In stochastic simulations, the fluctuations come solely from the random choices of the reactants in the lists and note from the reaction times $\tau_k$. The forward simulations of labeling patterns in equilibrium flux condition is investigated in the supplementary information.
FIG. 4. DSA and SSA in non-equilibrium flux conditions. (a-b) Samples of concentrations and isotopomer fractions obtained by numerical implementation of the DSA (solid lines) or the SSA (dots). Algorithm are run with $\varepsilon = 10^{-4}$ for DSA and $\Omega = 100 \text{ nc/\mu M}$ for SSA. (a) network shown in figure 3) is used. The plots display the $S_{1/2}$ concentration (black) as well as the unlabeled m+0 (green) and one labeled m+1 (red) concentrations. (b) Network shown in figure 1b is used. The plots display the G6P and Ru5 concentrations (black) as well as the unlabeled m+0 (green) and one labeled m+1 (red) concentrations. (c-d) Number of basic operations and computational time for the simulation shown in (a-b) for the DSA (c) and SSA (d), for the metabolic network of Figure 3 (filled stars) and for the simple metabolic network of Figure 1b (filled circles). $\varepsilon$: relative precision of the Runge-Kutta-Fehlberg method; CT: CPU time; #RHS: number of right-hand side evaluation; #reac: number of performed reaction; $\Omega$ sample size.

C. Computational performance

Figure 4c-d compare the computational performance of DSA and SSA as function of the algorithm parameters $\varepsilon$ and $\Omega$ for both networks depicted in Figure 1b and 3 and in the same
Stochastic method for isotope labeling systems

FIG. 5. Computation performance. Computation time as function of the number of labeled atoms in the metabolic network of Figure 3 for the DSA (orange filled circles) and for the SSA (green filled stars). $\varepsilon = 10^{-4}$ for the DSA and $\Omega = 100 \text{ nc/\mu M}$ for the SSA.

framework as in Figure 4a-b. Both methods were implemented with the same highest level of optimization, using low-level bit-manipulation tools to implement addressing operators in a fortran code (compiled with gfortran and optimisation flag “-O3”). The computational cost is measured as the computational time (CT) on an Intel(R) Core(TM) i5-6300U CPU at 2.40GHz for such simulations. In SSA, one reaction is typically performed in 75 CPU clock.

In the DSA, the number of evaluations of the right-hand side of Equation 8 increases with the relative precision $\varepsilon$ of the integration scheme in a similar way for the two networks (Figure 4c), but the computational time is much higher for the isotope labeling network of much larger isotope number due to the increased size of RHS vector. Indeed, the time of one right-hand side (RHS) evaluation is multiplied by 144 when the number of possible isotope per species increases from 4 to 128 for the isotope metabolic networks of Figure 3 and Figure 1b, respectively.

In the SSA, the number of chemical reaction occurrences can be approximated by the product $T v \Omega n$ where $n$ is the number of reactions in the network, $\Omega$ the volume, $v$ the typical flux values and $T$ the time interval, irrespective of the number of isotope per molecular species. Accordingly, the computational time, that is proportional to the number of occurring chemical reactions, increases proportionally to $\Omega$, but does not seem to differ significantly between the two networks in spite of their significant difference in the number of isotope Figure 4d.

Such qualitative difference between computational performance of DSA and SSA is illustrated in Figure 5 where the number of atoms per molecules is increased in the isotope labeling network of Figure 3. Since the computational time of SSA does not increase with number of isotope per species, SSA can become much more efficient from several order of magnitude for high enough number of isotopeomers. SSA was tested with values of $\Omega = 100 \text{ nc/\mu M}$. 

14
In the present study, we compare deterministic and stochastic simulation methods to simulate the dynamics of isotope labeling metabolic networks and illustrates their respective characteristics with a simple metabolic network. Both methods are derived from the same chemical master equation framework and use addressing operators that efficiently define product configurations from the reactant configurations using the permutation rule defining each chemical reaction. The only requirement is to define the network-specific addressing operator that can be built from the low-level bit manipulation functions in the most common case of a two-state labeling, which makes it more efficient than the atom mapping matrixes. Such common CME-based framework eases systematic comparison between the dynamical and computational properties of deterministic versus stochastic implementations of the CME.

The deterministic simulation algorithm (DSA) shares many similitudes to isotopomer methods involving mapping matrixes and therefore shows the same limitations related to the numerical computation of the equation system (Equation 7) whose dimensionality increases with the number of isotopomers. Also, more complex reactions than Bi Bi reactions would require much more complicated mathematical formulation and computational implementations. In contrast to DSA, the stochastic simulation algorithm (SSA) is relatively simple to implement and to extend to more complex reactions with integer stoichiometric coefficients or with more than two reactants and products. More importantly, the computational time in SSA depends only on the number of occurring reactions from which isotopomer distribution is updated in discrete manner, and thus does not depends on the number of isotopomers as in the DSA. This feature can provide a significant comparative advantage in isotopic systems with very high isotopomer number associated with advanced isotopic methods, combining 13C and 2H for instance. The common computational limitation of SSA algorithms is associated to small time steps in presence of large numbers of molecules and high rate constants, which have motivated various approaches such as hybrid modeling, paralleling or other CME approximations to simulate the stochastic dynamics of biochemical systems. In the proposed SSA algorithms, the number of metabolite species related to volume is an algorithm parameter, not a biological parameter, and the high number of reactions can be largely compensated by the fact that the number of operation (e.g., computational time) does not depend on isotopomer number per species. Only the parameter need to be set to a minimal value (typically 100 – 1000) to optimize the tradeoff between precision (below experimen-
Stochastic method for isotope labeling systems

...mental uncertainty) and computational efficiency. Moreover, temporal averaging (over time below the timescale of flux changes) can be performed to effectively reduce fluctuation of isotopomer concentrations and obtain more precise and statistically significant values of isotopomer fractions, in the perspective of the inverse problem.

Besides its computational efficiency for large isotope labeling systems, a SSA method has also the advantage to be a direct method that does not rely on any steady-state assumptions and that can be implemented without advanced knowledge about decomposition techniques or specialized softwares. Sticking close to the usual description of chemical reaction network also eases direct interpretation of the results and eases algorithm embedding inside a complex chain of procedures involving parametric estimation. For the inverse problem, SSA can be interfaced with stochastic sampling techniques such as Monte Carlo Markov Chain algorithm which is nowadays the state-of-the-art technique for estimating flux parameters with their uncertainty. In conclusion, SSA-based method seems to meet the main objectives of 13C-DMFA and could therefore be used in advanced study of metabolic networks.

ACKNOWLEDGEMENTS

The authors thank Darka Labavic for careful reading the manuscript.

FUNDING

This work has been supported by the LABEX CEMPI (ANR-11-LABX-0007) and by the Ministry of Higher Education and Research, Hauts de France council and European Regional Development Fund (ERDF) through the Contrat de Projets Etat-Region (CPER Photonics for Society P4S).

REFERENCES

1D. T. Gillespie, “A rigorous derivation of the chemical master equation,” Physica A: Statistical Mechanics and its Applications 188, 404–425 (1992).

2J. M. T. Thompson and H. B. Stewart, *Nonlinear dynamics and chaos* (John Wiley & Sons, 2002).
Stochastic method for isotope labeling systems

3D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” The journal of physical chemistry 81, 2340–2361 (1977).
4D. T. Gillespie, “Approximate accelerated stochastic simulation of chemically reacting systems,” The Journal of chemical physics 115, 1716–1733 (2001).
5J. D. Orth, I. Thiele, and B. Ø. Palsson, “What is flux balance analysis?” Nature biotechnology 28, 245–248 (2010).
6D. B. Kell and H. V. Westerhoff, “Metabolic control theory: its role in microbiology and biotechnology,” FEMS Microbiology Reviews 2, 305–320 (1986).
7N. Zamboni, S.-M. Fendt, M. Rühl, and U. Sauer, “13C-based metabolic flux analysis,” Nature protocols 4, 878–892 (2009).
8E. Levine and T. Hwa, “Stochastic fluctuations in metabolic pathways,” Proceedings of the National Academy of Sciences 104, 9224–9229 (2007).
9D. S. Tourigny, A. Goldberg, and J. R. Karr, “Simulating single-cell metabolism using a stochastic flux-balance analysis algorithm,” bioRxiv (2020).
10E. J. Clement, T. T. Schulze, G. A. Soliman, B. J. Wysocki, P. H. Davis, and T. A. Wysocki, “Stochastic simulation of cellular metabolism,” IEEE Access 8, 79734–79744 (2020).
11K. Schmidt, M. Carlsen, J. Nielsen, and J. Villadsen, “Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices,” Biotechnology and bioengineering 55, 831–840 (1997).
12W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, and A. A. de Graaf, “Bidirectional reaction steps in metabolic networks: III. explicit solution and analysis of isotopomer labeling systems,” Biotechnology and bioengineering 66, 69–85 (1999).
13M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, “Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions,” Metabolic engineering 9, 68–86 (2007).
14W. Wiechert and K. Nöh, “From stationary to instationary metabolic flux analysis,” Technology transfer in biotechnology , 145–172 (2005).
15J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, “An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis,” Biotechnology and bioengineering 99, 686–699 (2008).
16M. R. Antoniewicz, D. F. Kraynie, L. A. Laffend, J. González-Lergier, J. K. Kelleher, and G. Stephanopoulos, “Metabolic flux analysis in a nonstationary system: fed-batch fermentation
Stochastic method for isotope labeling systems

of a high yielding strain of E. coli producing 1, 3-propanediol,” Metabolic engineering 9, 277–292 (2007).

17R. W. Leighty and M. R. Antoniewicz, “Dynamic metabolic flux analysis (DMFA): a framework for determining fluxes at metabolic non-steady state,” Metabolic engineering 13, 745–755 (2011).

18L.-E. Quek, J. R. Krycer, S. Ohno, K. Yugi, D. J. Fazakerley, R. Scalzo, S. D. Elkington, Z. Dai, A. Hirayama, S. Ikeda, et al., “Dynamic 13C flux analysis captures the reorganization of adipocyte glucose metabolism in response to insulin,” Iscience 23, 100855 (2020).

19M. R. Antoniewicz, “A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications,” Metabolic Engineering (2020).

20T. McKee, Biochemistry: the molecular basis of life (Oxford University Press, USA, 2003) Chap. chapter 8: Carbohydrate Metabolism.

21C. M. Metallo, J. L. Walthier, and G. Stephanopoulos, “Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells,” Journal of biotechnology 144, 167–174 (2009).

22A. Kuehne, H. Emmert, J. Soehle, M. Winnefeld, F. Fischer, H. Wenck, S. Gallinat, L. Terstegen, R. Lucius, J. Hildebrand, et al., “Acute activation of oxidative pentose phosphate pathway as first-line response to oxidative stress in human skin cells,” Molecular cell 59, 359–371 (2015).

23M. R. Antoniewicz, “Methods and advances in metabolic flux analysis: a mini-review,” Journal of industrial microbiology and biotechnology 42, 317–325 (2015).

24C. Zupke and G. Stephanopoulos, “Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices,” Biotechnology Progress 10, 489–498 (1994).

25D. T. Gillespie, “The chemical langevin equation,” The Journal of Chemical Physics 113, 297–306 (2000).

26N. G. Van Kampen, Stochastic processes in physics and chemistry, Vol. 1 (Elsevier, 1992).

27M. A. Gibson and J. Bruck, “Efficient exact stochastic simulation of chemical systems with many species and many channels,” The journal of physical chemistry A 104, 1876–1889 (2000).

28T. B. Jacobson, P. A. Adamczyk, D. M. Stevenson, M. Regner, J. Ralph, J. L. Reed, and D. Amador-Noguez, “2H and 13C metabolic flux analysis elucidates in vivo thermodynamics of the ed pathway in zymomonas mobilis,” Metabolic engineering 54, 301–316 (2019).

29J. Pahle, “Biochemical simulations: stochastic, approximate stochastic and hybrid approaches,” Briefings in bioinformatics 10, 53–64 (2009).
Stochastic method for isotope labeling systems

30 G. Klingbeil, R. Erban, M. Giles, and P. K. Maini, “Stochsimgpu: parallel stochastic simulation for the systems biology toolbox 2 for matlab,” Bioinformatics 27, 1170–1171 (2011).
31 A. Singh and R. Grima, “The linear-noise approximation and moment-closure approximations for stochastic chemical kinetics,” arXiv preprint arXiv:1711.07383 (2017).
32 A. Gábor and J. R. Banga, “Robust and efficient parameter estimation in dynamic models of biological systems,” BMC systems biology 9, 1–25 (2015).
33 G. I. Valderrama-Bahamóndez and H. Fröhlich, “Mcmc techniques for parameter estimation of ode based models in systems biology,” Frontiers in Applied Mathematics and Statistics 5, 55 (2019).
Stochastic method for isotope labeling systems

Appendix A: Supplementary information

1. Used chemical reaction networks

Table 1 lists the chemical reactions and carbon rearrangements of the upper part of the increased glucolysis of the pentose phosphate pathway (Figure 1b). To illustrate the dynamics of the propagation of the labeled carbons and for the sake of simplicity, the reaction rate used corresponds to the mass action law with a kinetic parameter of unit value. Figure 6 completes the description of the dynamics presented in Figure 1b. To be complete, Table 2 presents similar information as Table 1 but for the archetypical network (Figure 3).
| reac nb | chemical reaction | reaction speed | parameter values |
|---------|-------------------|----------------|------------------|
| 1       | GLU (abcde) → G6P (abcde) | $v_{G6P} = k_1$ | $k_1 = 1 \rightarrow 2 \mu M/s$ |
| 2       | G6P (abcdef) → F6P (abcdef) | $v_2 = k_2[G6P]$ | $k_2 = 1/s$ |
| 3       | F6P (abcdef) → FBP (abcdef) | $v_3 = k_3^+ [F6P] - k_3^- [FBP]$ | $k_3^+ = k_3^- = 1 /s$ |
| 4       | FBP (abcdef) → DHAP (cba) + GAP (def) | $v_4 = k_4^+ [FBP] - k_4^- [DHAP][GAP]$ | $k_4^+ = 1/s; k_4^- = 1/\mu M/s$ |
| 5       | DHAP (abc) → GAP (abc) | $v_5 = k_5^+ [DHAP] - k_5^- [GAP]$ | $k_5^+ = k_5^- = 1 /s$ |
| 6       | G6P (abcdef) → 6PG (abcdef) | $v_6 = k_6[G6P]$ | $k_6 = 1/s$ |
| 7       | 6PG (abcdef) → CO2 (a) + Ru5 (bcdef) | $v_7 = k_7[6PG]$ | $k_7 = 1/s$ |
| 8       | Ru5 (abcde) → R5P (abcde) | $v_8 = k_8^+ [Ru5] - k_8^- [R5P]$ | $k_8^+ = k_8^- = 1 /s$ |
| 9       | Ru5 (abcde) → X5P (abcde) | $v_9 = k_9^+ [Ru5] - k_9^- [X5P]$ | $k_9^+ = k_9^- = 1 /s$ |
| 10      | X5P (abcdef) + R5P (ABCDE) → S7P (abABCDE) + GAP (cde) | $v_{10} = k_{10}^+ [X5P][R5P] - k_{10}^- [S7P][GAP]$ | $k_{10}^+ = k_{10}^- = 1 /\mu M/s$ |
| 11      | S7P (abcdefg) + GAP (ABC) → F6P (abcABC) + E4P (defg) | $v_{11} = k_{11}^+ [S7P][GAP] - k_{11}^- [F6P][E4P]$ | $k_{11}^+ = k_{11}^- = 1 /\mu M/s$ |
| 12      | X5P (abcdef) + E4P (ABCD) → F6P (abABCD) + GAP (cde) | $v_{12} = k_{12}^+ [X5P][E4P] - k_{12}^- [F6P][GAP]$ | $k_{12}^+ = k_{12}^- = 1 /\mu M/s$ |
| 13      | GAP (abc) → | $v_{13} = k_{13}[GAP]$ | $k_{13} = 1 /s$ |
| 14      | R5P (abcde) → | $v_{14} = k_{14}[R5P]$ | $k_{14} = 1 /s$ |
| 15      | CO2 (a) → | $v_{15} = k_{15}[CO2]$ | $k_{15} = 1 /s$ |

**TABLE I.** List of chemical reactions and reaction speed used to simulate the 13C propagation through the upper glycolytic pathways supplemented by the pentose phosphate pathway displayed in Figure 1b.
Stochastic method for isotope labeling systems

FIG. 6. DSA and SSA in non-equilibrium flux conditions. Samples of concentrations and isotopomer fractions obtained by numerical implementation of the DSA (solid lines) or the SSA (dots). Algorithm are run with $\varepsilon = 10^{-4}$ for DSA and $\Omega = 100$ nc/µM for SSA. Network shown in figure 1b is used. The plots display the concentrations (black) as well as the unlabeled m+0 (red), single labeled m+1 (green), double labeled m+2 (blue), and triply labeled m+2 (cyan) concentrations (in µM). Same parameters and intial conditions as in Figure 4b.
Stochastic method for isotope labeling systems

| reac nb | chemical reaction | reaction speed | parameter values |
|---------|-------------------|----------------|------------------|
| 1       | input (ab) ⚫ S1 (ab) | $v_1 = k_1$ | $k_1 = 1 \ \mu M/s$ |
| 2       | input (ab) ⚫ S3 (ab) | $v_2 = k_2$ | $k_2 = 0.1 \rightarrow 0.5 \ \mu M/s$ |
| 3       | S1 (ab) ⚫ S6 (ab) | $v_3 = k_3^+ [S1] - k_3^- [S6]$ | $k_3^+ = k_3^- = 1 / s$ |
| 4       | S6 (ab) ⚫ S8 (ab) | $v_4 = k_4 [S6]$ | $k_4 = 1 / s$ |
| 5       | S8 (ab) ⚫ S2 (ab) | $v_5 = k_5 [S8]$ | $k_5 = 1 / s$ |
| 6       | S1 (ab) ⚫ S5 (ab) | $v_6 = k_6^+ [S1] - k_6^- [S5]$ | $k_6^+ = k_6^- = 1 / s$ |
| 7       | S5 (ab) + S3 (AB) ⚫ S7 (Ab) + S4 (aB) | $v_7 = k_7 [S5][S3]$ | $k_7 = 1 / \mu M/s$ |
| 8       | S7 (ab) ⚫ S2 (ab) | $v_8 = k_8 [S7]$ | $k_8 = 1 / s$ |
| 9       | S2 (ab) ⚫ | $v_9 = k_9 [S2]$ | $k_9 = 1 / s$ |
| 10      | S4 (ab) ⚫ | $v_{10} = k_{10} [S4]$ | $k_{10} = 1 / s$ |

TABLE II. list of chemical reactions and reaction speed used to simulate the labeling propagation through the archetypical network displayed in Figure 3
Stochastic method for isotope labeling systems

2. The special case of networks at steady flux state

The proposed algorithm requires the knowledge of the mathematical law defining the reaction rate as a function of the concentrations of the species, thus allowing to update these reaction rates and occurrence times as the concentrations evolve in time. It can nevertheless be easily adapted in the case where the mathematical laws defining the reaction rates are unknown if we assume steady state of fluxes and concentrations such that only isotopomer distribution evolves over time. In this case, the reaction rates $v_k$ are constant parameters whose values are related to each other by the flux balance condition. Both the flux and concentration states of the chemical reaction network are stationary.

It is here required that $U_k = e^{-1}$ $\forall k$ in the SSA. Otherwise, if $U_k$ are random numbers and all reaction rates are constant, errors occur because the chemical reaction sequence become unrelated to the amount of chemical species, a reaction can be requested even if the reactant vanishes. This cancels fluctuations in reaction times $\tau_k$ and guarantees the steadiness of sample sizes (up to the unit change), the chemical reaction occurs periodically in this case. Only the relative abundances of the isotopomers display stochastic fluctuations. The sample sizes are, however, not necessarily identical, they define the variance of the temporal fluctuations of the isotopomer fractions.

As an example, let consider the archetypical network at steady flux state. The labeling is transmitted to the output species $S_2$ through the upper network path ($S_1 \rightarrow S_6 \rightarrow S_8 \rightarrow S_2$) and to the output species $S_4$ through the lower path $S_1 \rightarrow S_5 \rightarrow S_7 \rightarrow S_2$. The upper and the lower paths define the elementary mode of the flux balance analysis. In the steady state, their flux are respectively denoted $v_u$ and $v_d$. Following the proposed labeling, the fraction of labeled $S_2$ corresponds straightforwardly to the flux fraction in the upper path, and therefore a measurement of $S_2$ labeled fraction fully determines the $\frac{v_u}{v_u+v_d}$ ratio. At steady flux state, $(v_1 = v_9 = v_u + v_d; v_3 = v_4 = v_5 = v_u$ and $v_2 = v_6 = v_7 = v_8 = v_4 = v_d)$. We consider in the simple case where $v_d = v_u$, and the labeling is introduced at the initial time. Half of the output species $S_2$ ends up being labeled after a transient whose duration is here characterized by the half-time value denoted $t_{1/2}$ (Figure 7a). The numerical simulations in Figure 7 are obtained by choosing $v_u = v_d = 0.5 \mu M/s$ and $\Omega = 100 nc/\mu M$. The stochastic fluctuations of isotopomer concentrations are only due to the random choice of reactants (as the reactions occur periodically). The fluctuations of the stochastic dynamics quantified by the variation coefficient decrease in $1/\sqrt{\Omega}$ as expected (Figure 7b) where the limit $\Omega \rightarrow \infty$ corresponds to deterministic dynamics.
Stochastic method for isotope labeling systems

![Graph](image)

**FIG. 7.** Statistical and dynamical features of SSA. Numerical simulations of the internal dynamics of the network with the labeling indicated in figure 3. The simulations consider flux equipartition through the two network branches $v_u = v_d = 0.5 \text{ s}^{-1}$ and no subunit is labeled in the initial condition. (a) Time evolution of the $S_2$ concentrations computed from the DSA (solid line) or the SSA with $\Omega = 100 \text{ nc/mM}$ and $S_X = 1 \mu\text{M}$ (dots). The color code is indicated on the top. $t_{1/2}$ quantifies the transient duration by the time need to reach half of maximum value (25% labeling here). (b) In the stochastic simulations, the coefficient of variation of $S_2$ values (black dots) decrease as $1/\sqrt{\Omega}$. The continuous gray line displays the regression line of slope $-1/2$. (c) The transient duration $t_{1/2}$, as defined in (a), increases with the concentration ($S_6$ here). Monte-Carlo simulations (black circles) are obtained for $\Omega = 1000 \text{ nc/mM}$, only $S_6$ varies, the other being 1 $\mu\text{M}$. Similar results are obtained with the deterministic model (continuous gray line). $t^{*}_{1/2}$ is the reference value of $t_{1/2}$ for $S_6 = 1 \mu\text{M}$.

Neither the duration of the transient response nor the duration of the steady state depends on the overall size of $\Omega$ chosen for the samples; the $\Omega$ value only changes the time interval of each reaction. On the other hand, the relative abundance of different species can have a significant impact. An intermediate species in the reaction chain, such as $S_6$ for example, will reach its internal equilibrium more slowly the more abundant it is, and a long transient for $S_6$ will affect here the following species $S_8$ and $S_2$. This is obtained in a similar way with both deterministic and Monte Carlo algorithms (Figure 7c). Therefore, in the backward problem which consist in the steady flux determination from experimental time serie, the asymptotic value is enough to determine
Stochastic method for isotope labeling systems

the steady fluxes, whereas the transient gives insight to the relative species concentrations.  Al-
ternatively, steady fluxes may be determine from transient dynamic solely, provided the relative concentration of species are known.