



# Estimating Relative Changes of Metabolic Fluxes

Lei Huang<sup>1</sup>, Dongsung Kim<sup>2</sup>, Xiaojing Liu<sup>3</sup>, Christopher R. Myers<sup>4,5</sup>, Jason W. Locasale<sup>1,2,3,5\*</sup>

**1** Graduate Field of Computational Biology, Cornell University, Ithaca, New York, United States of America, **2** Graduate Field of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York, United States of America, **3** Division of Nutritional Sciences, Cornell University, Ithaca, New York, United States of America, **4** Laboratory of Atomic and Solid State Physics, and Institute of Biotechnology, Cornell University, Ithaca, New York, United States of America, **5** Field of Computational Biology and Medicine, Cornell University, Ithaca, New York, United States of America

## Abstract

Fluxes are the central trait of metabolism and Kinetic Flux Profiling (KFP) is an effective method of measuring them. To generalize its applicability, we present an extension of the method that estimates the relative changes of fluxes using only relative quantitation of <sup>13</sup>C-labeled metabolites. Such features are directly tailored to the more common experiment that performs only relative quantitation and compares fluxes between two conditions. We call our extension rKFP. Moreover, we examine the effects of common missing data and common modeling assumptions on (r)KFP, and provide practical suggestions. We also investigate the selection of measuring times for (r)KFP and provide a simple recipe. We then apply rKFP to <sup>13</sup>C-labeled glucose time series data collected from cells under normal and glucose-deprived conditions, estimating the relative flux changes of glycolysis and its branching pathways. We identify an adaptive response in which *de novo* serine biosynthesis is compromised to maintain the glycolytic flux backbone. Together, these results greatly expand the capabilities of KFP and are suitable for broad biological applications.

**Citation:** Huang L, Kim D, Liu X, Myers CR, Locasale JW (2014) Estimating Relative Changes of Metabolic Fluxes. *PLoS Comput Biol* 10(11): e1003958. doi:10.1371/journal.pcbi.1003958

**Editor:** Christos A. Ouzounis, Institute of Agrobiotechnology, Greece

**Received:** May 18, 2014; **Accepted:** October 1, 2014; **Published:** November 20, 2014

**Copyright:** © 2014 Huang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors acknowledge support from the sources NIH R00 CA168997 and R01 AI110613 (JWL) and NSF IOS-1127017 (CRM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: locasale@cornell.edu

This is a *PLOS Computational Biology Methods* article.

## Introduction

In recent years there has been renewed interest in metabolism resulting from discoveries of its connections to gene regulation [1], epigenetics [2], immunity [3], and pathogenesis of diseases such as cancer [4–6]. Independently, technological advances in metabolomics promise great improvement of our capabilities in metabolism studies and drug development [7–10]. However, despite the surge of interest and technological advances, quantitative systems-level characterization of the central trait of metabolism, metabolic flux, has been scarce and challenging. This is in part due to the mathematical nature of flux: rather than the amount of something that is experimentally measurable, it is defined as the rate of change in that amount and has to be inferred through modeling. Several modeling frameworks exist for the purpose. First, the century-old enzyme kinetics [11] and its systems analog, kinetic models of metabolic networks, offer a natural bridge from amount to flux, but unfortunately suffer from the “parameter problem” [12,13] of depending on many and usually poorly-characterized kinetic parameters. Second, structural models such as Flux Balance Analysis ambitiously aim to predict global distributions of fluxes with minimal data, but the prediction accuracy is still at a stage where validation against more direct estimation results is necessary. Third, isotope-based methods exploit the elegant and powerful experimental design of isotopes, and are the workhorse for reliable flux estimations.

Among the isotope-based methods, Kinetic Flux Profiling (KFP) [14,15] has been proven to be powerful [16–19], with a good balance between experimental ease, model simplicity, and prediction accuracy. In many ways complementary to Metabolic Flux Analysis (MFA) [20], another major isotope-based method which typically uses stationary isotopomer distribution data and is good at estimating relative flux distributions at branch points, KFP uses kinetic isotopomer distribution data and is good at estimating absolute flux scales along linear pathways.

The basic idea of KFP can be illustrated using a toy metabolic network. Consider a system of only one metabolite *A* connected to the environment by an influx  $J_1$  and an outflux  $J_2$ ; the system is at steady state so  $J_1 = J_2 = J$  (Figure 1a). KFP works by switching the system from a <sup>12</sup>C-labeled environment to a <sup>13</sup>C-labeled one at time  $t=0$ , measuring the concentrations of <sup>13</sup>C-labeled *A* (termed  $A^*$ ) at several time points thereafter, and estimating  $J$  from the time series data of  $A^*$ . After the switching of environment,  $A^*$  will gradually infiltrate the pool of *A* as a result of  $A^*$ -carrying influx, with the dynamics described by

$$\frac{dA^*}{dt} = J - J \frac{A^*}{A}, \text{ with the initial condition } A^*(0) = 0. \quad (1)$$

The two terms  $J$  and  $J \frac{A^*}{A}$  in the right-hand side respectively describe the infiltration of  $A^*$  into the *A* pool by the influx and the

## Author Summary

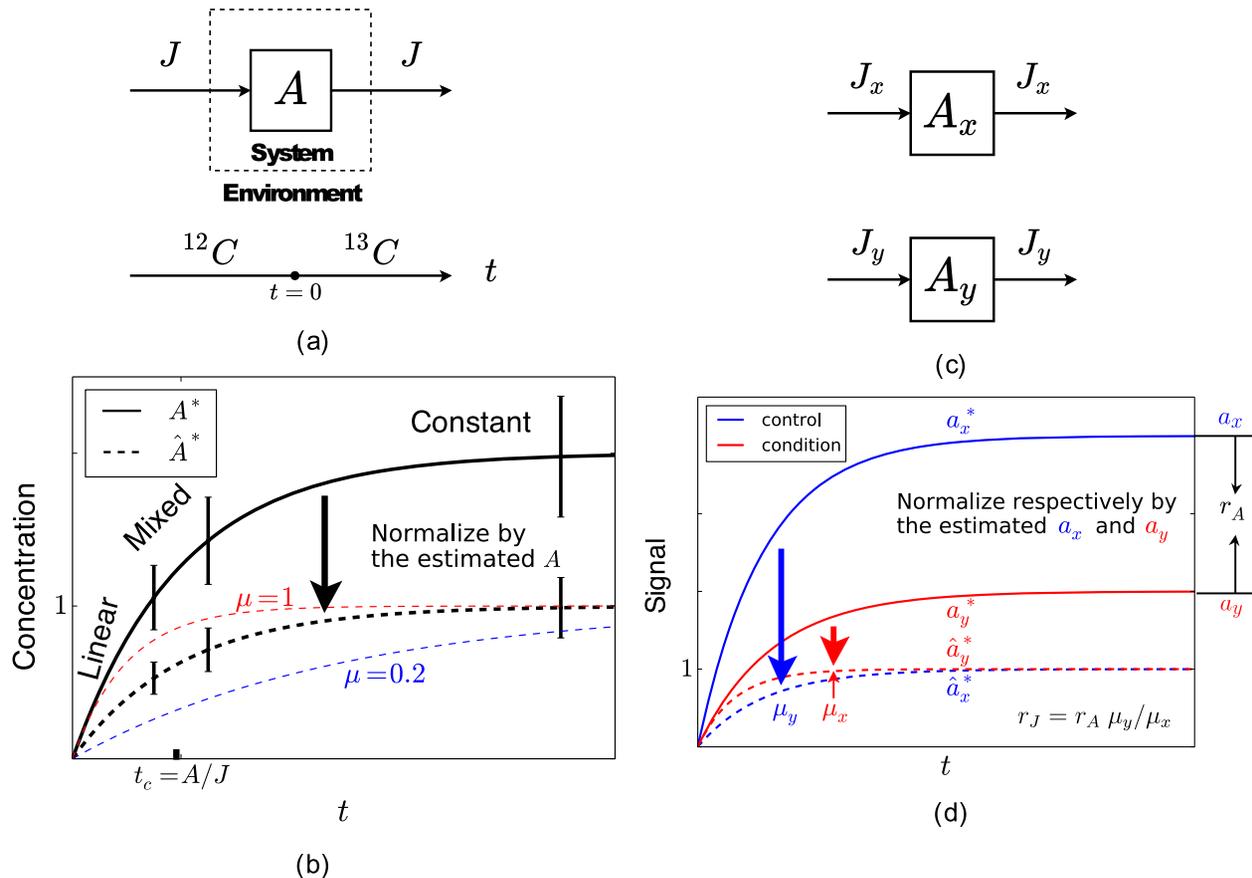
Metabolism underlies all biological processes, and its quantitative study is crucial for our understanding. The central trait of metabolism, metabolic fluxes, cannot be directly measured and are estimated usually through modeling. Existing modeling methods, however, are limited by poorly-characterized parameters, crude precision, or labor-intensiveness. Motivated by these limitations, and recognizing a most common goal in the field of comparing the fluxes between two conditions, we develop an extension of an existing method that takes in time-series relative-quantitation data of isotope-labeled metabolites (a kind of data that modern metabolomic technologies readily generate), and outputs the relative changes of fluxes in the metabolic networks of interest. We also carefully examine some issues on model construction and experimental design, and improve the reliability and strength of the method. We apply our method to data collected from cells in normal and glucose-deprived conditions, demonstrate the efficacy of the method and arrive at new biological insight.

opposing depletion by the outflux, and their net difference describes the rate of change of  $A^*$ . The equation is a first-order linear ordinary differential equation (ODE), and can be solved using standard techniques (see Text S1). Its solution is the simple exponential approach function

$$A^*(t) = A(1 - e^{-\frac{t}{A}}), \quad (2)$$

and geometrically corresponds to a family of curves parameterized by  $A$  and  $J$  (Figure 1b).

With some measurements of  $A^*$  along the curve, parameters  $A$  and  $J$  can be estimated in a standard way: a least-squares fitting algorithm gives the best fit, and sensitivity analysis or Monte Carlo simulations give the uncertainties. However, it helps to understand *why* KFP should work in this case. First, it is easy to see from Eq. 2 that parameter  $A$  determines the saturation level of  $A^*$  and  $J/A$  determines the rate at which the saturation level is approached; in other words,  $A$  determines the *scale* and  $J/A$  determines the *rate*. To highlight this, we define a rate parameter,  $\mu \equiv J/A$ ; its inverse,  $t_c \equiv 1/\mu = A/J$ , is conventionally called the characteristic time-scale



**Figure 1. Understanding KFP and rKFP.** (a) A schematic diagram of KFP applied to a toy metabolic network. At  $t=0$ , the system is switched from a  $^{12}\text{C}$ -labeled environment to  $^{13}\text{C}$ -labeled one, and  $A^*$  is measured at a few time points thereafter. (b) For a given trajectory of  $A^*(t)$  (the black solid curve), the three time regimes (linear, mixed and constant) are marked and three measurements are made (two in the mixed regime and one in the constant). Normalizing it gives  $\hat{A}^*(t)$  between 0 and 1 (the black dashed curve), parameterized by a single parameter  $\mu$ , which can be estimated by comparing the normalized measurements to  $\hat{A}^*$ 's of different  $\mu$ 's (the red and blue dashed curves). (c) A schematic diagram of rKFP applied to the same network in (a). Relative quantitation is performed on  $A^*$  in two conditions (with subscripts  $x$  and  $y$  respectively) with the goal of estimating  $r_J = J_y/J_x$ . (d) The ratio in  $\mu$  between  $\hat{a}_x^*(t)$  and  $\hat{a}_y^*(t)$  is  $r_J/r_A$  (Eq. 6), and since  $\mu$ 's and  $r_A$  are identifiable from relative quantitation, so is  $r_J$ . doi:10.1371/journal.pcbi.1003958.g001

and numerically corresponds to the time needed to go from the initial condition to  $1 - \frac{1}{e}$  ( $\approx 0.63$ ) of the saturation level. Second, like the familiar Michaelis-Menten hyperbolic curves, the exponential approach curves can also be thought as having three regimes, defined with respect to the characteristic time-scale: the linear regime when  $t \ll t_c$ , the constant regime when  $t \gg t_c$ , and the mixed regime in between when  $t \sim t_c$  (Figure 1b). Measurements of  $A^*$  in the linear regime are informative of only  $J$  (for the slope of that regime is  $J$ ), the constant regime of only  $A$  (for  $A^*$  in that regime is constantly  $A$ ), and the mixed regime of both. In practice, however, usually only the constant and mixed regimes are measured due to their experimental accessibility. Finally, after  $A$  is estimated from measurements in the constant regime, the estimation of  $J$  from measurements in the mixed regime can be understood in the following way: normalize all measurements by the estimated  $A$  to describe the normalized variable  $\hat{A}^* \equiv A^*/A = 1 - e^{-\mu t}$ ; parameterized only by  $\mu$  and now increasing from 0 to 1,  $\hat{A}^*$  changes from a sharply rising curve to a gently rising one as  $\mu$  decreases; the normalized measurements in the mixed regime nail down the specific  $\hat{A}^*$  within the family of curves, together with  $\mu$  and  $J$  (Figure 1b). The discussion above can be succinctly summarized as “parameters  $A$  and  $J$  are *identifiable* when applying KFP to the system in Figure 1a”. Later in the paper we will see how the reasoning described here in terms of normalization and rate can be used again to understand the estimation of relative changes of fluxes and the ratios of pool sizes, and the selection of measuring times.

Applying KFP to systems of arbitrary size and network topology and with multiple influxes is less straightforward and requires care [17]. When a reaction involves more than one substrate, the labeling states are no longer just labeled or unlabeled as in KFP, and tracing the origins and fates of  $^{13}\text{C}$  labels requires the knowledge of Carbon Transition Map [21] of the reactions. The assumption of irreversibility can eliminate this complication for decomposition reactions, but is only valid for far-from-equilibrium ones. Also, multiple influxes to a system complicate modeling the dynamics of  $^{13}\text{C}$ -labeled metabolites in the same way as multi-substrate reactions do. For this reason, KFP works best for systems consisting of mono-molecular reactions, and works for general systems only through gathering additional information, making assumptions, or using only part of the data that is more amenable to model. The last strategy corresponds to the idea used in an extension of KFP called extended KFP (or eKFP) [17], and is relevant in our later discussion on the capacity of KFP and our extension of it in studying metabolic cycles.

Two considerations motivate us to extend KFP beyond its current scope. First, KFP requires *absolute quantitation* of metabolites, meaning that their absolute concentrations have to be measured, while many experimental techniques such as mass spectrometry can only perform *relative quantitation* readily, meaning that the measurement output is scaled from the absolute concentration by a metabolite-specific unknown constant; going from relative quantitation to absolute quantitation typically requires performing relative quantitation on some reference samples whose absolute concentrations are known, which can often be a challenge due to the increased effort of both additional experiments and procurement of reference samples. Second, often it is the relative changes of fluxes (or biological quantities in general) between two conditions that are of interest or biological relevance (*e.g.*, wildtype *vs.* mutant, control *vs.* drug-treated; [16,22]), and estimating the absolute fluxes of the two conditions only to get their ratios is inefficient (the information regarding their scales is eventually

discarded) and roundabout (three rounds of estimation are carried out, one for each condition and one for their ratio).

In this paper, we report an extension of KFP that can estimate the relative changes of fluxes using only relative quantitation, which we call rKFP, hence addressing the two considerations above. To improve the reliability and strength of KFP and rKFP, we examine some issues in the application of the methods, on both setting up models and selecting measuring times. Finally, we apply rKFP to experimental data collected in normal and glucose-deprived conditions, estimating the relative flux changes in glycolysis and its branching pathways and arriving at new biological insight.

## Results/Discussion

### Extending KFP to Estimate Relative Flux Changes

Consider again the toy system in Figure 1a, now in two different conditions; the same experimental procedures of switching environment at  $t=0$  and subsequent measurements of  $A^*$  apply, but only with relative quantitation (Figures 1c and 1d). The aim is to estimate the relative change of  $J$  between the two conditions.

We start by writing down the relationship between the relative quantitation measurements (which we call signals) and the absolute quantitation measurements (concentrations) for the two conditions:

$$a_i^* = p_A A_i^*, \quad i = x, y, \quad (3)$$

where an upper-case letter denotes concentration and lower-case signal,  $p$  their ratio, superscript  $*$  a labeled quantity, and subscripts  $x$  and  $y$  quantities for the two conditions respectively (a list of notation can be found in Table 1). Since now we can only perform relative quantitation and hence  $a_i^*$  becomes the measurable, we establish its dynamics by plugging in the dynamics of  $A_i^*$  which we have solved in studying KFP:

$$a_i^*(t) = p_A A_i (1 - e^{-J_i t/A_i}), \quad i = x, y. \quad (4)$$

The above two equations highlight a simple but important fact:  $a_i^*(t)$  is simply  $A_i^*(t)$  scaled by an unknown constant  $p_A$ , and they share the same intrinsic dynamics. Therefore, the reasoning of normalization and rate described in the discussions of KFP in the introduction appears even *more* natural in this situation: although relative quantitation leaves us oblivious to the scale, we can still normalize the measurements to uncover the rate.

$$\hat{a}_i^*(t) \equiv \frac{a_i^*}{a_i} = 1 - e^{-J_i t/A_i} = 1 - e^{-\mu_i t}, \quad i = x, y. \quad (5)$$

Defining the relative changes of pool size  $r_A \equiv A_y/A_x$  and flux  $r_J \equiv J_y/J_x$ , the relative change of  $\mu$  can then be expressed in terms of  $r_A$  and  $r_J$ :

$$\mu_y = \frac{J_y}{A_y} = \frac{r_J J_x}{r_A A_x} = \frac{r_J}{r_A} \mu_x. \quad (6)$$

Since  $\mu_x$  and  $\mu_y$  are identifiable using normalized measurements in the same way as  $\mu$  described in the introduction, and  $r_A$  is obviously identifiable (Figure 1d), so is  $r_J$ . This concludes our explanation of why rKFP should work for the toy system in Figure 1c.

**Table 1.** Definitions of variables and their identifiabilities in rKFP.

Symbol	Meaning	Identifiable in rKFP
$A$ (uppercase)	metabolite $A$ or its pool size (absolute quantitation) <sup>1</sup>	No
$a$ (lowercase)	signal of the pool size (relative quantitation)	Yes
$A^*$	concentration of $^{13}\text{C}$ -labeled $A$ (absolute quantitation)	No
$a^*$	signal of $^{13}\text{C}$ -labeled $A$ (relative quantitation)	Yes
$\hat{A}^*$	the fraction of $A^*$ in $A$ , $\hat{A}^* \equiv A^*/A$	Yes
$\hat{a}^*$	the fraction of $a^*$ in $a$ , $\hat{a}^* \equiv a^*/a$	Yes
$J$	flux	No
$\mu$	rate, $\mu \equiv J/A$	Yes
$p_A$	proportionality constant between signal (relative quantitation) and concentration (absolute quantitation) of $A$ , $p_A \equiv a/A = a^*/A^*$	No
$\bullet_x$ (subscript $x$ )	a quantity of control	-
$\bullet_y$ (subscript $y$ )	a quantity of condition	-
$r_A$	relative change of pool size, $r_A \equiv \dot{A}_y/A_x$	Yes
$r_J$	relative change of flux, $r_J \equiv \dot{J}_y/J_x$	Yes
$\rho_{A_i, A_j}$	ratio of pool size, $\rho_{A_i, A_j} \equiv A_j/A_i$	Yes

1: Which one of the two possible meanings is meant should be inferable from the context when not specified; it is a slight abuse of notation customary in the field (e.g., [46]).

doi:10.1371/journal.pcbi.1003958.t001

To summarize the mechanics of rKFP for the system, one sets up the following two equations:

$$\begin{cases} \frac{da_x^*}{dt} = p_A(J_x - J_x \frac{a_x^*}{a_x}) = p_A(J_x - J_x \frac{a_x^*}{A_x p_A}), \\ \frac{da_y^*}{dt} = p_A(J_y - J_y \frac{a_y^*}{a_y}) = p_A(J_x r_J - J_x r_J \frac{a_y^*}{A_x r_A p_A}), \end{cases} \quad (7)$$

which form a model that is parameterized by  $\theta = (A_x, J_x, p_A, r_A, r_J)$ , and predicts  $a_x^*(t)$  and  $a_y^*(t)$  (the solutions are Eq. 4); measurements of  $a_x^*(t)$  and  $a_y^*(t)$  allow for estimating  $r_J$  (and  $r_A$ ) with precision (identifiable). Two remarks follow: first, while parameters  $A_x$ ,  $J_x$  and  $p_A$  are obviously non-identifiable, two of their functions,  $a_x = p_A A_x$  and  $\mu_x = J_x/A_x$ , are, amounting to two constraints on  $(A_x, J_x, p_A)$ ; second, in light of the constraints, the model can be parameterized in other natural ways: for example, one can replace  $A_x$  with  $a_x$ , or  $J_x$  with  $\mu_x$ , and the resulting new parameterization would be as interpretable.

For larger networks consisting of single-substrate reactions arranged in linear pathways and branch points, rKFP proceeds in a similar way: equations like Eq. 7 are constructed for each metabolite in the network, and relative quantitation measurements of the metabolites allow for estimating the relative changes of all the independent fluxes with precision. However, for metabolic networks involving reactions of multiple substrates, especially cycles (multi-substrate reactions are usually present in cycles as part of the network design; see supplementary Text S1), rKFP in its above form cannot handle the situation. Fortunately, a variant of KFP, termed extended KFP (or eKFP), has been developed to overcome the problem [17], and its rKFP version, which we term reKFP, can estimate the relative flux changes for cycles without having to deal with the complications of modeling carbon transitions that arise in multi-substrate reactions [23]. The essential idea of eKFP is that while there can be multiple labeled states for the reactants in multi-substrate reactions and keeping

track of their transitions can be complicated, there is always only one unlabeled state and modeling its dynamics is relatively simple and stays within the KFP framework; the downsides are that only a fraction of the information in the data is used and that the measurements of the unlabeled metabolites have to be accurate which sometimes can be nontrivial to achieve due to media contamination. A detailed discussion of how reKFP can be applied to cycles is contained in supplementary Text S1.

Before we conclude the discussion of rKFP, we mention one more nontrivial quantity identifiable from relative quantitation, the ratio of pool sizes. As an illustrating example, consider a metabolic pathway of two metabolites with relative quantitation; we again normalize the measurements to uncover the intrinsic dynamics. The idea is that the further the second metabolite lags behind the first one, the more abundant it is compared to the first one (Figure S1a).

Formally, one can plug  $\rho_{A_1, A_2} \equiv \rho \equiv A_2/A_1$  and  $\mu_1 \equiv J/A_1$  into the normalized  $\hat{A}_2^*(t)$  (Text S1 contains a derivation of  $\hat{A}_2^*(t)$ ):

$$\begin{aligned} \hat{A}_2^*(t) &= 1 - \left( \frac{A_1}{A_1 - A_2} e^{-\frac{\mu_1 t}{A_1}} + \frac{-A_2}{A_1 - A_2} e^{-\frac{\mu_1 t}{A_2}} \right) \\ &= 1 - \left( \frac{1}{1 - \rho} e^{-\mu_1 t} + \frac{-\rho}{1 - \rho} e^{-\frac{\mu_1 t}{\rho}} \right). \end{aligned}$$

Since  $\mu_1$  is identifiable from relative quantitation of  $A_1^*$ , the single parameter that is left,  $\rho$ , determines how much  $\hat{A}_2^*$  lags behind  $\hat{A}_1^*$  and is identifiable from the comparison (Figure S1b). In the introduction we discussed the general challenge of absolute quantitation, but here we show that if absolute quantitation can be performed on, or good prior knowledge exists for, some metabolites (e.g., the cellular glucose concentration is known to be about 5 mM [24]), this information of scale can be propagated across the network to other metabolites through the estimates of

pool size ratios, estimating absolute concentrations from relative quantitation.

### Missing Data: Effects and Pitfalls

Despite the great advances in metabolomic technologies, it is nevertheless common to have missing data. It is a result of the chemical properties of metabolites: some are too unstable to be accurately measured, and some isomers are too similar to each other to be distinguishable. To set up a computational model for KFP or rKFP under missing data, one in principle has the following options: (1) use a reduced model where the network components corresponding to the missing data are removed; (2) use a full model where all components are kept and the part of the model corresponding to missing data represents additional degrees of freedom unconstrained by data; (3) use the full model but incorporate prior information for the part of model uncovered by data; (4) spend additional effort to collect all data and use the full model. We observe that a common practice in applying KFP is to choose the first option and use reduced models when there is missing data (e.g., [16]). We acknowledge that this option is tempting and reduced models do offer many advantages such as conceptual simplicity and computational manageability (after all, “all models are wrong; some are useful”); however, because of the potential bias reduced models might introduce to parameter estimation and model prediction, we believe that their use would be better justified after a careful consideration of their effects.

Consider the following example. Figure 2a provides a cartoon of a typical situation in applying KFP: in a linear pathway of three metabolites,  $A_2$  is hard to measure and therefore constitutes missing data. The above four options concretize to the followings: (1) use the reduced model consisting of only  $A_1$  and  $A_3$ ; (2) use the full model consisting of all three metabolites with  $A_2$  uncovered by data; (3) use the full model but put a prior distribution (in the Bayesian sense) on the  $A_2$  pool size based on previous knowledge; (4) try to collect  $A_2$  to complete the data and use the full model. We note that the desirability of option (3) depends on what prior distribution is available and how close it is to the true value, i.e., how well one *a priori* knows the missing piece. Hence it has to be judged on a case-by-case basis and cannot be discussed generally. We therefore exclude option (3) in our subsequent analysis, and only note that in the limit of a correct tight prior the case converges to option (4) of completing the data, and in the limit of a loose prior the case converges to option (2) of effectively having no prior information.

We call this scenario of missing data in Figure 2a *missing metabolite*, and the corresponding procedure of constructing reduced models *metabolite removal*. We identify and name three additional common scenarios of missing data with their associated reduction procedures: first, *missing pathway*, where a branching pathway has poor data coverage with most metabolites unmeasured, associated with *pathway removal*, where the branching pathway is removed from the model; second, *undistinguished metabolites*, where the individual identity of a set of metabolites in a measurement cannot be resolved, associated with *lumping*, where the undistinguished metabolites are lumped into a single pool; third, *unknown reversibility*, where the extent of reversibility of a reaction is unknown, associated with *assuming irreversibility*, where potentially reversible reactions are modeled as irreversible for simplicity. In the remainder of this section, we describe in details the effects of metabolite removal on both KFP and rKFP, for this case is the simplest and hence most illustrative, and also for this case turns out to be important (see below); after that we briefly describe the results on pathway removal and lumping, leaving the

details to the supplementary text; results on assuming irreversibility are discussed in the next section.

Back to Figure 2a, intuitively, using the reduced model with  $A_2$  removed would *underestimate*  $J$ , for two reasons. First, since the influx  $J$  is  $^{13}\text{C}$ -saturated and constant with time, after time  $t$  one would expect  $Jt$  amount of  $^{13}\text{C}$  in the system, distributed across different metabolite pools; removing  $A_2$  from the network excludes the  $^{13}\text{C}$  in that pool, causing an underestimation of the  $^{13}\text{C}$  in the system and hence an underestimated  $J$ . Second, the presence of any metabolite pool slows down the infiltration process of  $^{13}\text{C}$  along the network, and if the slow-down of the infiltration by the  $A_2$  pool is concealed by removing  $A_2$  from the network, the slowed infiltration would be attributed to a lower  $J$ . Both factors become more pronounced as the pool size of  $A_2$  increases, and so should be the underestimation.

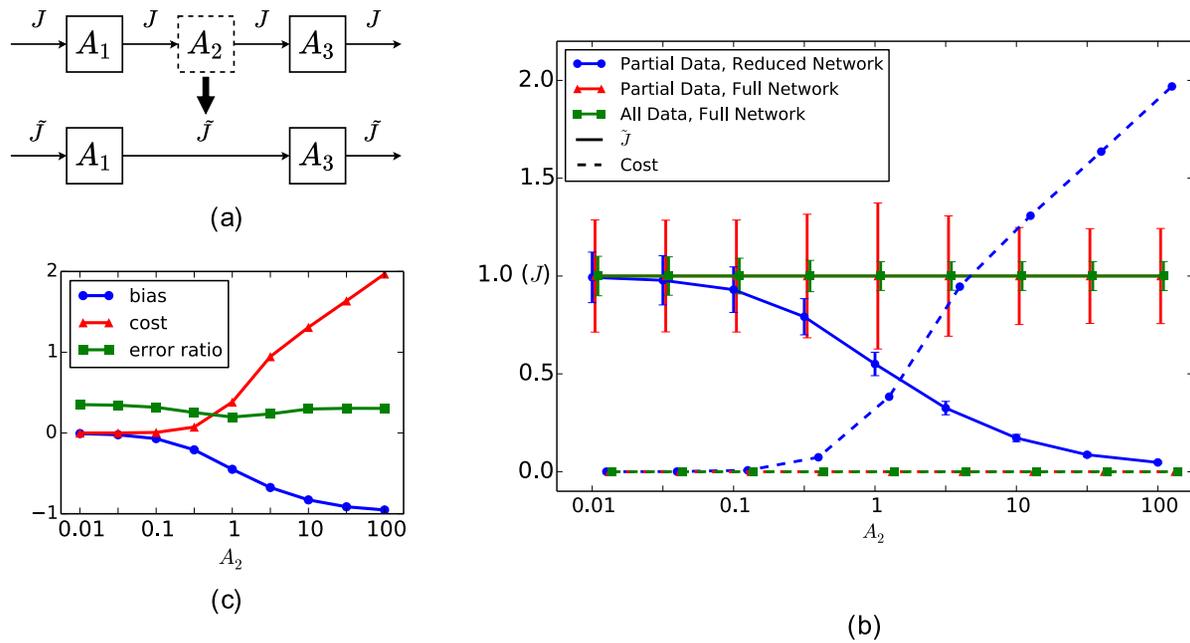
Figure 2b shows the results for the three options. First, using the reduced model indeed underestimates  $J$ , and it worsens as  $A_2$  increases (the blue solid curve), confirming our intuition. Second, using the reduced model decreases the goodness-of-fit between the model and the data, quantified by the cost of fitting, and it increases with  $A_2$  as well (the blue dashed curve). Third, using full models causes no bias or cost (red/green and solid/dashed curves). Fourth, the extra hard work of collecting the data of  $A_2$  pays off by shrinking the uncertainties of the estimated  $J$  (the red *vs.* the green error bars). These observations suggest the following three summary statistics:

- bias, defined as  $(\tilde{\theta} - \theta)/\theta$ , where  $\tilde{\theta}$  and  $\theta$  are respectively the estimated and true value of a parameter ( $J$  in this case).
- cost, the cost of fitting using the reduced model.
- error ratio, defined as  $\sigma_c/\sigma_p$  where  $\sigma$  is the standard error of the parameter estimate and subscripts  $c$  and  $p$  refer respectively to the cases of complete and partial data.

These statistics summarize the effects and pitfalls of missing data: error ratio quantifies their adverse effects on parameter estimation (or, looking optimistically, the reward of completing them), and bias and cost quantify the harm of using reduced models. Figure 2c re-plots the results in Figure 2b in terms of the three summary statistics. From these plots we can see that KFP is rather sensitive to metabolite removal: in the case of the toy model, a missing metabolite of pool size equal to others causes roughly a 50% bias. The same should hold for general pathways: if the total pool size of the missing metabolites in a pathway is relatively large, the bias should be considerable. For this reason, we give the general suggestion that, unless the missing metabolites are known *a priori* to have small pool sizes, one should use full models in KFP.

We next examine how metabolite removal affects rKFP (Figure 3a). We have concluded that removing  $A_2$  in KFP underestimates  $J$  and the underestimation worsens as  $A_2$  increases. Since  $r_J$  is the ratio between  $J_y$  and  $J_x$ , we expect that the bias in  $r_J$  depends on  $A_2$  in the two conditions,  $A_{2x}$  and  $A_{2y}$ : if  $A_{2x}$  is large and  $A_{2y}$  small,  $J_x$  is implicitly more underestimated than  $J_y$  (implicitly as rKFP does not explicitly estimate  $J$ ), giving an overestimated  $r_J$ , and in the same way a small  $A_{2x}$  and a large  $A_{2y}$  give an underestimated  $r_J$ . Like KFP, we expect the bias to increase with the pool size difference of  $A_2$  between two conditions.

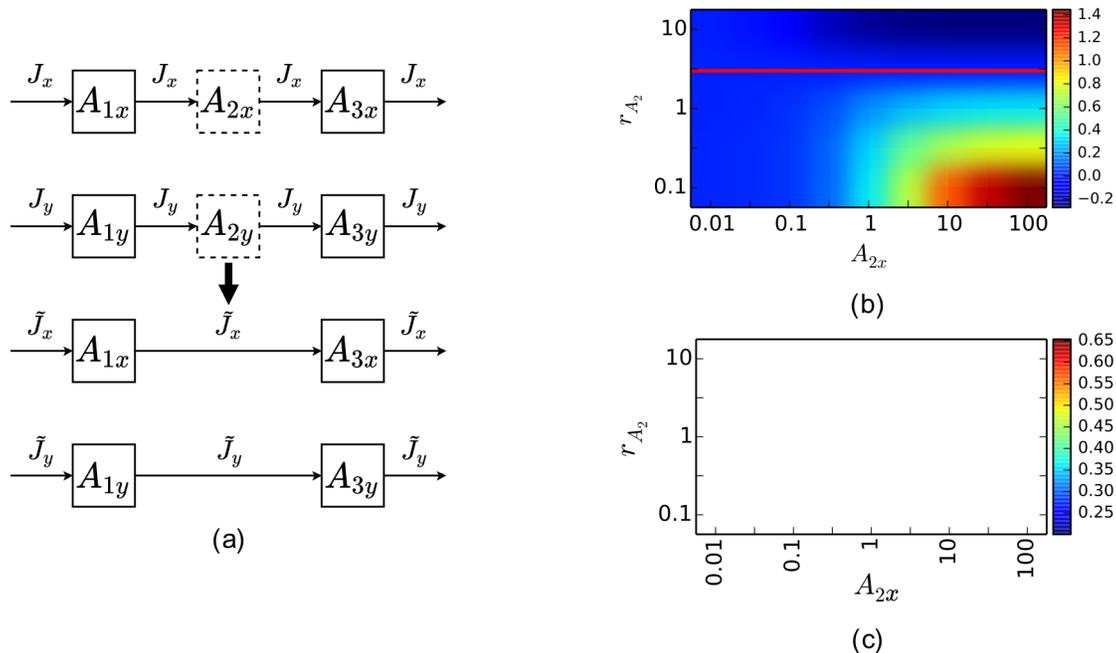
Figure 3b plots the bias of  $r_J$ , and confirms our intuition. However, an important feature distinguishes the case in rKFP from KFP. Figure 3b shows a red line where the relative change of  $A_2$  is the same as  $A_1$  and  $A_3$ , and above the red line  $r_J$  is underestimated and below overestimated; in other words, when the relative change of  $A_2$  pool size exactly matches the others in



**Figure 2. Metabolite removal in KFP.** (a) A schematic diagram of getting the reduced model through metabolite removal in KFP. Dashed squares represent metabolites removed in the reduced model; thick dark arrow represents reduction;  $\tilde{J}$  represents the estimated  $J$  (potentially biased). (b) The estimation results for the three options. The solid curves represent  $\tilde{J}$ , the dashed curves represent the cost of fitting (normalized by the number of data points to be comparable across options), and three colors represent the three options. Parameter values used for generating the simulated data:  $A_1 = A_3 = J = 1$  (overall patterns independent of the choice here). (c) The estimation results in (b) in terms of the three summary statistics. doi:10.1371/journal.pcbi.1003958.g002

the pathway,  $r_J$  is estimated without bias. This hints at another important advantage of rKFP over KFP: bias can be introduced to the two conditions in such a similar way that some of it is canceled

out. The question remains of how likely the pool size of a metabolite changes in a similar way to all others in a pathway. Figure S3 plots the fitting of experimental data of glycolysis in two



**Figure 3. Metabolite removal in rKFP.** (a) A schematic diagram of getting the reduced model through metabolite removal in rKFP. The same figure scheme as in Figure 2a applies here. Parameter values used in generating the simulated data:  $A_{1x} = A_{3x} = J_x = 1$ ,  $A_{1y} = A_{3y} = 3$ , and  $J_y = 2$ . (b) Dependence of bias on the pool size of the missing metabolite in two conditions. (c) Dependence of error ratio on the pool size of the missing metabolite in two conditions. doi:10.1371/journal.pcbi.1003958.g003

conditions, and shows that the relative changes in pool size of all metabolites in the pathway fall within a factor of five. From an enzyme kinetic perspective, this says that the reactions along the pathway are in similar *elasticity* regimes [25], which is also consistent with the consensus that reactions in a pathway in general share the *flux control* [26]. We hence believe that in rKFP some of the bias is indeed canceled out which makes it *more robust* to metabolite removal than KFP. However, we also note in Figure 3b that the bias worsens quickly as the relative change in pool size of different metabolites diverge: when it is off by a factor of three, say,  $r_{A_2} = 1$  (while  $r_{A_1} = r_{A_3} = 3$ ), the bias becomes about 40%. This implies that in a typical situation despite the canceling the remaining bias is still too large to justify using the reduced model, and hence we again suggest using full models in rKFP for missing metabolites. The ideas of canceling and metabolites along a pathway changing pool sizes similarly, however, find their use in the next section where they serve to justify using reduced models for unknown reversibility.

To summarize, we have demonstrated above the effects of missing metabolites on (r)KFP and the pitfalls of removing them; we suggest that unless one has good prior knowledge about the missing pool size or its relative change between two conditions that ensures a small bias, full models should be used. The approach we use for the demonstration is an intuitive and computational one, and in supplementary Text S1 we outline a complementary and analytical method that allows for a more precise description and deeper insight into the dependence of bias on relevant parameters. Also included there are some detailed studies of how two other reduction scenarios, pathway removal and lumping, might affect KFP and rKFP. We summarize the results in the following: first, both KFP and rKFP are rather robust to the removal of a minor branching pathway (that is, a branching pathway that does not carry the majority of the flux), while using the full model greatly inflates the confidence interval of the parameter estimates due to the additional degrees of freedom; second, since lumping is usually applied to isomers which are often close to chemical equilibrium, it is generally rather innocuous in light of the results in the next section. In a later section on data analysis, we see that the results in this section are verified by comparing the estimation results of different models.

## Modeling Reversible Reactions

Biochemical reactions often operate close to equilibrium [27] (for example, it is conventionally thought that this applies to seven out of the ten reactions in glycolysis [28]), and a reaction's distance to equilibrium is commonly quantified by the difference in Gibbs free energy,  $\Delta G$ , between substrates and products: the closer  $\Delta G$  is to zero, the closer the reaction is to equilibrium. One of the properties of a close-to-equilibrium reaction is that both of its forward and backward fluxes are large and most of them are canceled out to give a much smaller net flux; quantitatively this is described by the *flux-force relationship*:  $J^+ / J^- = e^{-\frac{\Delta G}{RT}}$ , where  $J^+$  and  $J^-$  are the forward and backward flux respectively,  $R$  the gas constant and  $T$  the temperature [29]. Quantity  $\frac{J^+ + J^-}{J^+ - J^-}$ , the total flux over the net flux, therefore describes how much “futile” work the reaction does compared to “useful” work, and predicts from the flux-force relationship that as the reaction approaches equilibrium, exponentially more fluxes simply go back and forth compared to the net flux (Figure S2a). This has an important implication in the context of (r)KFP: when the reversible fluxes are large compared to the net flux, the time-scale of the mixing of substrate and product pools is much shorter than that of the

infiltration of the  $^{13}\text{C}$ -labeled metabolites in each pool, making the two pools effectively one as far as (r)KFP is concerned (Figure S2b).

This suggests another potential issue in setting up models for KFP or rKFP: ignoring the reversible fluxes when the reaction is close to equilibrium, as is commonly done, might introduce bias. To check this, we again use a toy system (Figure 4a) and, similar to the case of missing metabolite, conceive three options for modeling a reversible reaction: (a) model it as irreversible; (b) model it as reversible and let  $\Delta G$  be a free parameter; (c) model it as reversible and measure  $\Delta G$  to some precision (which typically requires absolute quantitation). Figure 4b plots the summary statistics for KFP, which shows that there is a small bias and a moderate cost when a highly reversible reaction is assumed irreversible. It suggests that KFP might be robust to assuming irreversibility; however, even the small bias can be avoided since in KFP  $\Delta G$  can be calculated and is not missing information: KFP uses concentration data, which can be used to calculate  $\Delta G$  through its definition (see below); therefore one can explicitly incorporate in the model  $J^+$  and  $J^-$  which depend on the parameter  $J$  through

the relationships  $J^+ = \frac{J}{e^{-\frac{\Delta G}{RT}} - 1}$  and  $J^- = \frac{J e^{-\frac{\Delta G}{RT}}}{e^{-\frac{\Delta G}{RT}} - 1}$ . Also plotted

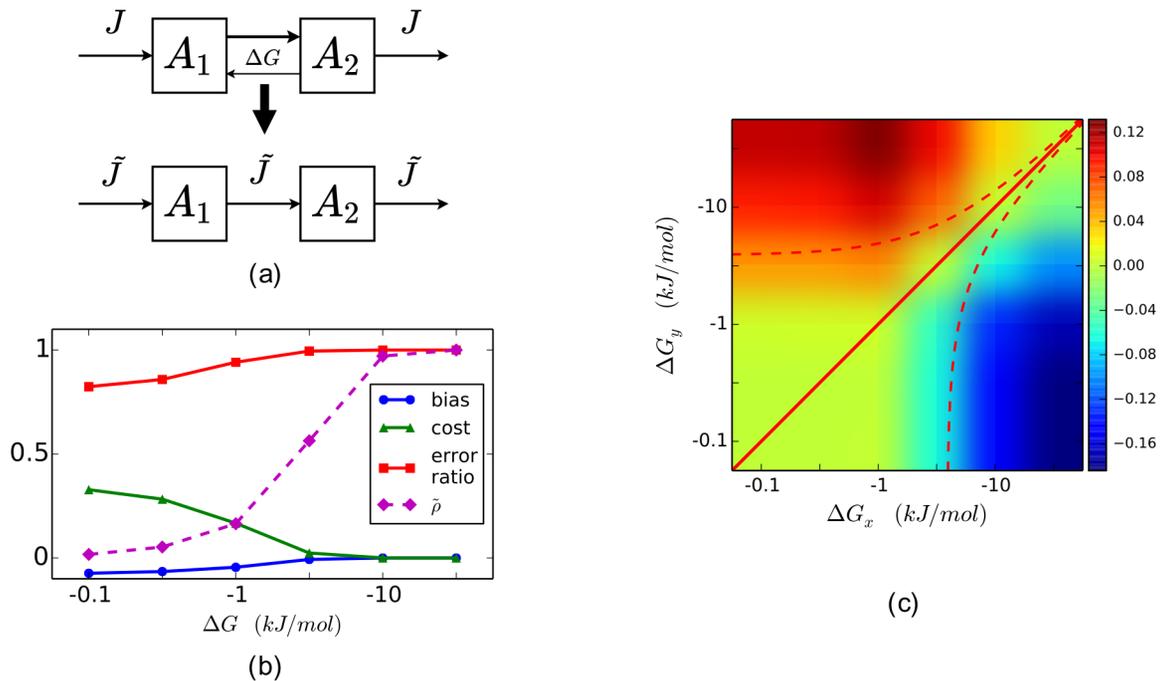
is the estimated pool size ratio between  $A_1$  and  $A_2$  with one being the true value (the purple dashed line), which shows a vast underestimation when the reaction is close to equilibrium; this can be explained by the following: as the reaction becomes more reversible,  $A_1^*$  and  $A_2^*$  share the dynamics to a greater extent (Figure S2b), and  $A_2^*$  closely following behind  $A_1^*$  is interpreted by the algorithm as resulting from  $A_2 \ll A_1$  (Figure S1a).

Figure 4c illustrates how ignoring reaction reversibilities biases rKFP, which can be understood in a similar way as Figure 3b in using reduced models. First, the direction of the bias depends on  $\Delta G$ 's of the two conditions; since ignoring reaction reversibilities underestimates  $J$  in KFP (Figure 4b), if the reaction is more reversible in control (x) than in condition (y), then  $J$  would be more underestimated in control, giving an overestimated  $r_J$  (the red region), and *vice versa*. Second, if  $\Delta G_x = \Delta G_y$ , then  $r_J$  is estimated without bias (the red line).

We naturally wonder how likely the change of  $\Delta G$  falls around the red line, and find that its definition offers the clue. Since  $\Delta G \equiv \Delta G^o + RT \ln Q$ , where  $\Delta G^o$  is the standard Gibbs free energy change and  $Q$  the reaction quotient defined as the product concentrations divided by the substrate concentrations. For a mono-molecular reaction with  $S$  the substrate and  $P$  the product,  $\Delta G_x$  and  $\Delta G_y$  can be related in the following way:

$$\begin{aligned} \Delta G_y - \Delta G_x &= RT(\ln Q_y - \ln Q_x) = RT\left(\ln \frac{P_y}{S_y} - \ln \frac{P_x}{S_x}\right) \\ &= RT \ln \frac{P_y S_x}{P_x S_y} = RT \ln \frac{r_P}{r_S}. \end{aligned} \quad (8)$$

In other words, the change in  $\Delta G$  depends on the relative changes in pool size of  $P$  and  $S$ . In the previous section we conclude that the relative changes in pool size of the metabolites along a pathway are likely to be similar, and Figure 4c shows that, *unlike* the case of metabolite removal, the bias increases *slowly* as the relative pool size changes of different metabolites diverge (the red dashed curves correspond to a five-fold difference in the relative pool size changes, a range of variation we observe in our glucose-deprivation data, and delineate a region surrounding the diagonal line of very small bias). We therefore conclude rKFP should be robust to assuming irreversibility and reduced models can be used.



**Figure 4. Modeling reversible reactions in KFP and rKFP.** (a) A schematic diagram of the toy system considered here; for rKFP two copies of each network are similarly made as in Figure 3a. Parameter values used in generating the simulated data:  $A_1 = A_2 = A_{1x} = A_{2x} = J = J_x = 1$ ,  $A_{1y} = A_{2y} = 3$ , and  $J_y = 2$ . (b) Dependence of the summary statistics on  $\Delta G$  in KFP. (c) Dependence of the bias of  $r_j$  on  $\Delta G$  of the two conditions in rKFP. The red solid diagonal line corresponds to  $\Delta G_x = \Delta G_y$ , where there is no bias. The red dashed curves correspond to a five-fold difference in the relative pool size changes between the substrate and product, a range we observe in our data. doi:10.1371/journal.pcbi.1003958.g004

We last note that, like the case of pathway removal discussed in the supplementary text, using the full model in this case and incorporating  $\Delta G$ 's as free parameters would leave the model underconstrained by data and lead to huge uncertainties in parameter estimates, since one additional parameter would be accumulated for each reaction modeled as reversible. Incorporating prior information on  $\Delta G$ 's would help, but at a systems level the information is scarce as it requires accurate absolute quantitation [27] and predictive computational methods are still being developed [30]. It is also our experience that rKFP models constructed this way with additional  $\Delta G$  parameters are prone to numerical problems in data fitting and parameter sampling, since for each reaction modeled as reversible two copies of  $\Delta G$  ( $\Delta G_x$  and  $\Delta G_y$ ) need to be made and they are entangled with parameters of relative pool size changes through Eq. 8. In light of all these challenges, the robustness of rKFP to assuming irreversibility is advantageous.

### Selecting Measuring Times

In addition to the issues on setting up models as discussed in the previous two sections, we have also investigated the experimental design issue of selecting measuring times for a metabolic pathway in (r)KFP. We list below the main conclusions, and leave the derivation details to the supplementary text.

First, we note that the problem can be formulated as a classical experimental design one [31]: select a set of measuring times such that the errors of the estimated parameters ( $J$ 's in KFP and  $r_j$ 's in rKFP) are the smallest. However, such an approach suffers from computational intractability, physical uninterpretability and un-realisticness (Text S1).

We choose to adopt an alternative approach for which there is a physical intuition: since the dynamics of the metabolites in a pathway are linear combinations of exponential functions (Text

S1), and in the introduction we show that for a single exponential function, there is a well-defined characteristic time-scale, we wonder if we can make use of this special mathematical structure and target the measuring times on the different time-scales.

We show that for the first metabolite in a pathway, if a late time has been measured so that its (relative) scale is estimated, the optimal measuring time would be around  $t_c = A_1/J$ , the characteristic time-scale. This highlights another significance of  $t_c$ : if the curve of  $A^*(t)$  in Figure 1b is held fixed at two ends and the middle left wiggling as  $\mu$  changes, placing a pin at the characteristic time-scale leaves the least wiggle room.

We also show by using a similar reasoning that for the second metabolite in a pathway the optimal measuring time is around  $(A_1 + A_2)/J$ , and similarly for the  $k$ -th metabolite around  $\sum_{i=1}^k A_i/J$ . That is, the measuring time for the  $k$ -th metabolite should be around the sum of the characteristic time-scales of all the metabolites up to the  $k$ -th one; this makes intuitive sense, given that the dynamics of  $A_k^*$  is a mixture of  $k$  time-scales due to the actions of the  $k$  pools. We therefore provide the following practical suggestion: for a metabolic pathway of  $n$  metabolites with a rough prior guess of the flux (and pool sizes for rKFP), after measuring at a late time so that the (relative) scales of all metabolites can be estimated, other measurements should go to  $\sum_{i=1}^k A_i/J$ ,  $k = 1, \dots, n$ . Given this scheme of selecting measuring times, it is calculated through simulation and shown in the supplementary text how the precision of estimated parameters depends on the number of data points and how those data points should be optimally distributed across different metabolites and times.

We last note that a typical metabolic network exhibits strong separation of concentration scales. For the example of glycolysis, one source reports that the most abundant metabolite (glucose) is

about 35 times more than the second most abundant one and the full range of variation spans over three orders of magnitude [28]. This has an important implication: the most abundant metabolite dominates the numerator of  $\sum_{i=1}^k A_i/J$ , and hence the sampling time. If the dominant metabolite happens to be the first one, as is suggested for glycolysis in [28], then the whole pathway effectively shares one dynamics and measuring time.

### Applying rKFP to Glucose Deprivation Data

In this last section, we apply rKFP to experimental data collected on glycolysis and its two main branching pathways (pentose-phosphate pathway, or PPP, and serine synthesis pathway) from cells in normal (5 mM glucose) and glucose-deprived (0.5 mM glucose) media conditions, and estimate the relative changes in the fluxes of the three pathways.

Figure 5a shows the diagram of the network, where the upper pathway branching off glycolysis at G6P is (part of) the PPP, and the lower pathway branching off glycolysis at 3PG is (part of) the serine synthesis pathway; two additional branching pathways, glycogen synthesis pathway and glycerol synthesis pathway, emanating from G6P and DHAP, respectively, are also portrayed (the dashed arrows). Further described in the network diagram is our data coverage, which contains all types of missing data we have considered: data of metabolites BPG and a few others in branching pathways are missing, data of the glycogen and glycerol synthesis pathways are missing, isomers such as G6P/F6P, GAP/DHAP and 3PG/2PG are not distinguished, and the extent of reversibility of the reactions is not known.

We set up three computational models for rKFP, corresponding to three different choices of treating missing data as discussed in earlier sections, and the estimation results of all three models are shown in Figure 5c: dark blue histograms correspond to the choice recommended in the earlier sections, namely, using the full model only for missing metabolite, light green the choice of using the reduced model for all types of missing data, and light red the choice of using the full model for both missing metabolite and pathway. We make two observations from the comparison of the histograms: the green histograms are tighter than the blue but shifted, and the red histograms have similar averages to the blue (except for PPP) but much larger variances.

The observations support our recommended choice. The shifted green histograms relative to the blue suggest that metabolite removal introduces significant bias, likely because the pool sizes of the missing metabolites change differently from the other metabolites. The flattened red histograms relative to the blue suggest that keeping in the model the missing pathways with no data leaves the estimation greatly underconstrained; on the other hand, their similar averages suggest that the bias introduced by pathway removal is small in this case, likely because the missing pathways have much lower fluxes than the main glycolysis pathway (see the supplementary text) which is consistent with some previous studies (*e.g.*, [16,32]; also see Figure S3) and that the cell line in our study exhibits the Warburg effect [4] with most of the glycolytic flux going to lactate fermentation. In summary, through the comparison of different models for both toy and real networks, we have chosen the following choice: in the face of missing data, we keep the model components that matter (metabolites) and leave out those that do not and would make the estimation underconstrained or numerically unstable (pathways, distinct pools and reversibilities). For future applications of rKFP, we suggest users either follow our recommended choice, or, better yet, set up different models and compare the results to verify the choice as is done here.

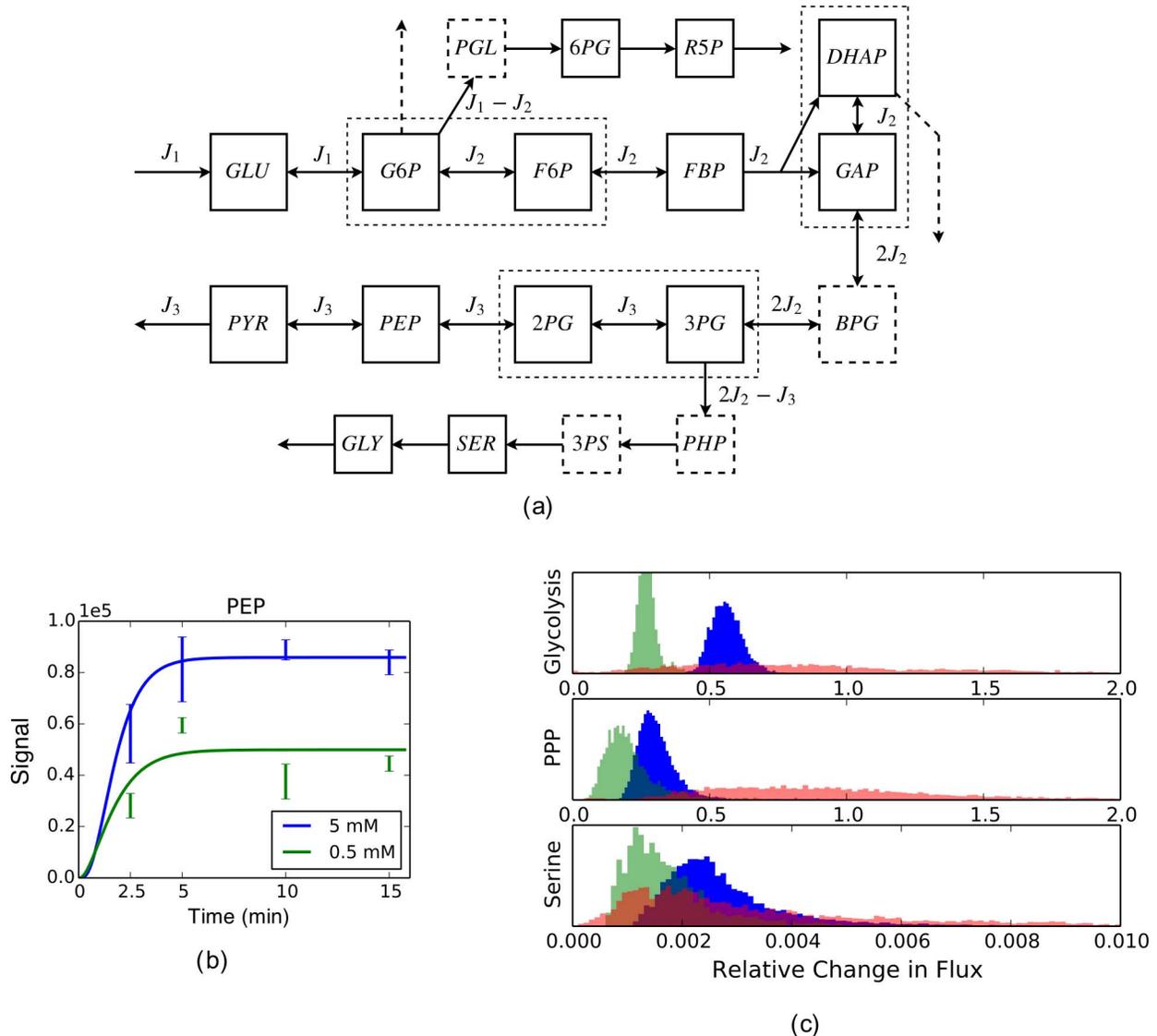
From the parameter distributions under our recommended model choice, we observe that, despite a 10-fold decrease in the media glucose concentration, glycolysis and PPP fluxes are reduced by about 40% and 60% respectively, while the serine pathway is basically incapacitated by the glucose deprivation. We make the following interpretation of the results. As it is generally believed that tumor cell proliferation sometimes requires metabolic adaptation to a microenvironment deprived of glucose, the effect of glucose deprivation on metabolism has been of interest [33–35]. Also of interest is the activity of serine biosynthesis pathway for its implication in tumorigenesis [36]. Here we show that when the external glucose source is depleted, the cells adapt their metabolism by largely maintaining the backbone of glycolysis flux at the expense of serine biosynthesis flux. This suggests the possibility of additional contextual requirements of PHGDH, the enzyme that diverts flux from glycolysis into *de novo* serine biosynthesis, and the possibility of other mechanisms for serine utilization in the condition of low glucose availability.

## Materials and Methods

### Computation

Individual steps of the computation in this study are listed and explained below using KFP as the example (many of them use *SloppyCell*, a Python package originally developed for analyzing biochemical networks [37]). Relevant Python codes are deposited at <http://github.com/leihuang/rkfp>.

- Encoding models: models are encoded in an SBML-compliant format [38] in *SloppyCell* and mathematically correspond to systems of ODEs of  $\mathbf{A}^*$ , the concentration vector of labeled metabolites.
- Simulating models: *daskr*, an algebraic-differential equation solver [39] used in *SloppyCell* numerically integrates the models.
- Generating simulated data: given the integrated trajectories of  $\mathbf{A}^*$ , simulated data are generated by a selection of measuring times following the suggestions in the section on selecting measuring times, and the associated noise is generated by choosing a constant value when deriving analytical results for its tractability and assuming to be proportional to data when carrying out simulations for its realism (a proportionality constant of 0.2 is used).
- Estimating parameters: letting  $Y_{ik}$  be the measurement of the model prediction  $A_i^*(t_k)$  and  $\sigma_{ik}$  be the noise associated with  $Y_{ik}$ , one defines the cost of fitting as a function of parameter  $\theta$ :
 
$$C(\theta) \equiv \sum_{i,k} \left( \frac{A_i^*(t_k, \theta) - Y_{ik}}{\sigma_{ik}} \right)^2$$
 (also known as  $\chi^2$ ); *SloppyCell* uses the Levenberg-Marquardt algorithm [40] to minimize  $C(\theta)$  and find the parameter estimate  $\tilde{\theta}$ . In our analysis of experimental data where the model is large and the data is noisy, we find that having a good initial guess is crucial [41] and having a large L-M parameter  $\lambda$  and a small trust region helps.
- Integrating sensitivity curves: *SloppyCell* calculates  $\frac{\partial A_i^*}{\partial \theta_\alpha}$  as a function of  $t$  in an accurate way (more details in Text S1), which is important for calculating Jacobian and errors (below);
- Estimating errors: (1) construct the Jacobian of the model,
 
$$\mathbf{D} = \left( \frac{\partial A_i^*}{\partial \theta_\alpha}(t_k) \right)_{ik,\alpha}$$
 by filling a matrix with the sensitivities at the assumed measuring times along the sensitivity curves; (2) normalize each entry by  $\sigma_{ik}$ ; (3) perform the singular value



**Figure 5. Analysis of experimental data.** (a) The diagram of glycolysis and its two branching pathways used in the analysis, where dashed squares (e.g., BPG) represent missing metabolites, dashed arrows represent missing pathways, and dashed rectangles containing solid squares represent undistinguished metabolites. Abbreviations for metabolites: GLU: glucose; G6P: glucose-6-phosphate; F6P: fructose-6-phosphate; FBP: fructose-1,6-bisphosphate; DHAP: dihydroxyacetone phosphate; GAP: glyceraldehyde-3-phosphate; BPG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; PYR: pyruvate; PGL: 6-phosphogluconolactone; 6PG: phosphogluconate; R5P: the pool of ribose 5-phosphate, ribulose 5-phosphate and xylulose 5-phosphate; PHP: phosphohydroxypyruvate; 3PS: 3-phosphoserine; SER: serine; GLY: glycine. (b) An exemplary plot of the data of a metabolite and its fit. Plots of all metabolites and their fits can be found in Figure S3. (c) Histograms of  $r_j$ 's as generated by sampling the corresponding posterior distributions in a way detailed in the Methods. Glycolysis flux refers to  $J_3$  in the diagram, PPP flux  $J_1 - J_2$ , and serine synthesis flux  $2J_2 - J_3$ . The three histograms for each flux correspond to three different modeling choices described in the text.

doi:10.1371/journal.pcbi.1003958.g005

decomposition  $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ; (4) the square roots of the diagonal entries of the matrix  $\mathbf{V}(\mathbf{\Sigma}\mathbf{\Sigma})^{-1}\mathbf{V}^T$  give the estimates of errors.

- Generating posterior distribution: assuming Gaussian-distributed measurement noise, the measurement of  $A_i^*(t_k)$  is also Gaussian distributed:  $Y_{ik} \sim \mathcal{N}(A_i^*(t_k, \theta), \sigma_{ik})$ ; treating the parameter estimation problem in the Bayesian framework and assuming an uninformative prior,  $\theta$  has a posterior distribution with a probability density proportional to that of observing  $Y_{ik}$  in  $\mathcal{N}(A_i^*(t_k, \theta), \sigma_{ik})$ , following the Bayes' rule:  $p(\theta|\mathbf{Y}) \propto p_{\mathcal{N}}(\mathbf{Y}|\theta)$ ; Metropolis algorithm [42] is used to sample the

posterior distribution in SloppyCell, and parameters of 100,000 steps, 1% burn-in and 50-step thinning interval [43] are used for generating the distributions in Figure 5c.

Note that the last two points constitute the two standard ways of estimating parameter uncertainties: the first one also goes by the name of *sensitivity analysis* or *delta method*, and is computationally cheap but less accurate; the second one is also known as *ensemble method* [44], and is computationally expensive but more accurate. In this study, simulations intended to illustrate basic principles use the first method, and data analyses intended to draw realistic conclusions use the second.

## Experiments

Experimental procedures of cell culture, metabolite extraction, mass spectrometry, liquid chromatography and data processing follow those in [45]. On the procedures specific to rKFP, HCT116 human colon cancer cells cultured in 6 well dishes with RPMI 1640 were washed with PBS and transferred to two media with  $^{12}\text{C}$ -glucose of concentrations 5 mM and 500  $\mu\text{M}$  respectively, where they were incubated for 2.5 hours before switching to media with  $^{13}\text{C}$ -glucose of the same concentrations; relative quantitation of triplicates were then performed on the cells at 0, 2.5, 5, 10 and 15 minutes after the switching.

## Supporting Information

**Text S1** A file containing three supplementary figures, a derivation of the solution for KFP applied to a linear metabolic pathway, a discussion on applying (r)KFP to metabolic cycles, the description of an analytical approach to studying the effects of removing metabolites and pathways, and effects of pathway

## References

- McKnight SL (2010) On getting there from here. *Science* 330: 1338–1339.
- Katada S, Imhof A, Sassone-Corsi P (2012) Connecting threads: epigenetics and metabolism. *Cell* 148: 24–28.
- Pearce EL, Poffenberger MC, Chang CH, Jones RG (2013) Fueling immunity: Insights into metabolism and lymphocyte function. *Science* 342: 1242–1245.
- Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324: 1029–1033.
- Vander Heiden MG (2011) Targeting cancer metabolism: a therapeutic window opens. *Nature Reviews Drug Discovery* 10: 671–684.
- Locasale JW (2012) The consequences of enhanced cell-autonomous glucose metabolism. *Trends in Endocrinology and Metabolism* 23: 545–551.
- Patti GJ, Yanes O, Siuzdak G (2012) Metabolomics: the apogee of the omic trilogy. *Nature Reviews Molecular Cell Biology* 13: 263–269.
- Rabinowitz JD, Purdy JG, Vastag L, Shenk T, Koyuncu E (2011) Metabolomics in drug target discovery. Cold Spring Harbor symposia on quantitative biology 76: 235–246.
- Benjamin DI, Cravatt BF, Nomura DK (2012) Global profiling strategies for mapping dysregulated metabolic pathways in cancer. *Cell Metabolism* 16: 565–577.
- Noack S, Wiechert W (2014) Quantitative metabolomics: a phantom? *Trends in Biotechnology* 32: 238–244.
- Michaelis L, Menten MML (2013) The kinetics of invertin action: Translated by T.R.C. Boyde submitted 4 February 1913. *FEBS Letters* 587: 2712–2720.
- Gunawardena J (2010) Models in systems biology: The parameter problem and the meanings of robustness. In: Lodhi HM, Muggleton SH, editors, *Elements of Computational Systems Biology*, John Wiley & Sons, Inc. pp.19–47.
- Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, et al. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3: e189.
- Yuan J, Fowler WU, Kimball E, Lu W, Rabinowitz JD (2006) Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. *Nature Chemical Biology* 2: 529–530.
- Yuan J, Bennett BD, Rabinowitz JD (2008) Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature Protocols* 3: 1328–1340.
- Munger J, Bennett BD, Parikh A, Feng XJ, McArdle J, et al. (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nature Biotechnology* 26: 1179–1186.
- Szczowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, et al. (2013) Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell* 25: 694–714.
- Ruhl M, Coq DL, Aymerich S, Sauer U (2012)  $^{13}\text{C}$ -flux analysis reveals NADPH-balancing transhydrogenation cycles in stationary phase of nitrogen-starving *Bacillus subtilis*. *Journal of Biological Chemistry* 287: 27959–27970.
- Cobbold SA, Vaughan AM, Lewis IA, Painter HJ, Camargo N, et al. (2013) Kinetic flux profiling elucidates two independent acetyl-CoA biosynthetic pathways in *Plasmodium falciparum*. *Journal of Biological Chemistry* 288: 36338–36350.
- Wiechert W (2001)  $^{13}\text{C}$  metabolic flux analysis. *Metabolic Engineering* 3: 195–206.
- Mu F, Williams RF, Unkefer CJ, Unkefer PJ, Faeder JR, et al. (2007) Carbonate maps for metabolic reactions. *Bioinformatics* 23: 3193–3199.
- Goentoro L, Kirschner MW (2009) Evidence that fold-change, and not absolute level, of  $\beta$ -catenin dictates Wnt signaling. *Molecular Cell* 36: 872–884.
- Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic Engineering* 9: 68–86.
- Voet D, Voet JG, Pratt CW (2011) *Fundamentals of Biochemistry: Life at the Molecular Level*. John Wiley & Sons, Inc
- Cornish-Bowden A (2012) *Fundamentals of Enzyme Kinetics*. Weinheim, Germany: Wiley-Blackwell, 4 edition.
- Fell D (1996) *Understanding the Control of Metabolism*. Portland Press, 1 edition.
- Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology* 5: 593–599.
- Garrett R, Grisham C (2012) *Biochemistry*. Cengage Learning, 5 edition.
- Beard DA, Qian H (2010) *Chemical Biophysics: Quantitative Analysis of Cellular Systems*. Cambridge University Press.
- Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, et al. (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* 10: e1003483.
- Atkinson AC, Donev AN (1992) *Optimum Experimental Designs*. Oxford: Oxford University Press.
- Metallo CM, Walther JL, Stephanopoulos G (2009) Evaluation of  $^{13}\text{C}$  isotopic tracers for metabolic flux analysis in mammalian cells. *Journal of Biotechnology* 144: 167–174.
- Yun J, Rago C, Cheong I, Pagliarini R, Angenendt P, et al. (2009) Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells. *Science* 325: 1555–1559.
- Komurov K, Tseng JT, Muller M, Seviour EG, Moss TJ, et al. (2012) The glucose-deprivation network counteracts lapatinib-induced toxicity in resistant ErbB2-positive breast cancer cells. *Molecular Systems Biology* 8: 596.
- Birsoy K, Possemato R, Lorbeer FK, Bayraktar EC, Thiru P, et al. (2014) Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides. *Nature* 508: 108–112.
- Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, et al. (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genetics* 43: 869–874.
- Myers CR, Gutenkunst RN, Sethna JP (2007) Python unleashed on systems biology. *Computing in Science & Engineering* 9: 34–37.
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524–531.
- Brown PN, Hindmarsh AC, Petzold LR (1994) Using Krylov methods in the solution of large-scale differential-algebraic systems. *SIAM Journal on Scientific Computing* 15: 1467–1488.
- Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11: 431–441.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition.
- Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49: 327–335.
- Gamerman D, Lopes HF (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press.
- Brown KS, Sethna JP (2003) Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E* 68: 021904.
- Liu X, Ser Z, Locasale JW (2014) Development and quantitative evaluation of a high-resolution metabolomics technology. *Analytical Chemistry* 86: 2175–2184.
- Gunawardena J (2012) Theory and mathematical methods. In: Egelman EH, editor, *Comprehensive Biophysics*, Amsterdam: Elsevier. pp. 243–267.

removal in KFP and rKFP, and some detailed results on the effects of missing data and selecting measuring times. (PDF)

**Dataset S1**  $^{13}\text{C}$ -labeled relative-quantitation data collected from cells in normal and glucose-deprived media used for the analysis. (CSV)

## Acknowledgments

The authors would like to thank John Guckenheimer, Eli Bogart, Oleg Kogan, Greg Ezra, James Sethna, Jacob Bien, Brandon Barker and members of the Locasale lab for helpful discussions, Zheng Ser for the help with data processing, and Yiping Wang for helpful comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: LH JW. Performed the experiments: DK XL. Analyzed the data: LH. Wrote the paper: LH JW CRM.