k-Cone Analysis: Determining All Candidate Values for Kinetic Parameters on a Network Scale

Iman Famili,* Radhakrishnan Mahadevan,*† and Bernhard O. Palsson*
*Department of Bioengineering, University of California San Diego, La Jolla, California 92093-0412; and †Genomatica, Inc., San Diego, California 92121

ABSTRACT The absence of comprehensive measured kinetic values and the observed inconsistency in the available in vitro kinetic data has hindered the formulation of network-scale kinetic models of biochemical reaction networks. To meet this challenge we present an approach to construct a convex space, termed the *k*-cone, which contains all the allowable numerical values of the kinetic constants in large-scale biochemical networks. The definition of the *k*-cone relies on the incorporation of in vivo concentration data and a simplified approach to represent enzyme kinetics within an established constraint-based modeling approach. The *k*-cone approach was implemented to define the allowable combination of numerical values for a full kinetic model of human red blood cell metabolism and to study its correlated kinetic parameters. The *k*-cone approach can be used to determine consistency between in vitro measured kinetic values and in vivo concentration and flux measurements when used in a network-scale kinetic model. *k*-Cone analysis was successful in determining whether in vitro measured kinetic values used in the reconstruction of a kinetic-based model of *Saccharomyces cerevisiae* central metabolism could reproduce in vivo measurements. Further, the *k*-cone can be used to determine which numerical values of in vitro measured parameters are required to be changed in a kinetic model if in vivo measured values are not reproduced. *k*-Cone analysis could identify what minimum number of in vitro determined kinetic parameters needed to be adjusted in the *S. cerevisiae* model to be consistent with the in vivo data. Applying the *k*-cone analysis a priori to kinetic model development may reduce the time and effort involved in model building and parameter adjustment. With the recent developments in high-throughput profiling of metabolite concentrations at a whole-cell scale and advances in metabolomics technologies, the *k*-cone approach presented here may hold the promise for kinetic characterization of metabolic networks as well as other biological functions at a whole-cell level.

INTRODUCTION

The absence of kinetic parameters and notable inconsistency in the available in vitro kinetic data has hindered characterization of the kinetic properties in biological systems and studying biological functions using kinetic-based models. Traditionally, kinetic models are constructed using rate equations derived to describe conditions in vitro and thus rely on the use of in vitro measured kinetic parameters. However, the conditions at which in vitro experiments are performed are often vastly different from those inside the cell. In vitro kinetic measurements are generally made at a high substrate to enzyme concentration ratio, enzymes are assayed in isolation from other enzymes, and molecules that may interfere with enzyme activities in the cell are absent from the assay. Thus, in vitro kinetic rates and in vitro kinetic parameters describe enzymatic behaviors that may not truly represent the observed physiological kinetic behavior in the cell.

The ability of in vitro kinetic measurements to quantitatively predict in vivo phenotype has been examined by a number of experimental studies (Rizzi et al., 1997; Teusink at al., 2000; Vaseghi et al., 1999; Wright and Kelly, 1981). Results have shown that in vitro derived kinetic models do not adequately describe in vivo phenotypes, and thus kinetic models instead must use in vivo data to accurately describe in vivo behavior. Several methods have been developed to address this issue by incorporating in vivo measurements in constructing kinetic models (Rizzi et al., 1997; Theobald et al., 1997; Vaseghi et al., 1999). Using these methods, the response of the cell to an environmental perturbation has been studied by combining rigorous analytical tools with in vivo experimental data sets, including intracellular metabolite concentrations and reaction rates. These methods, however, require considerable mathematical efforts (Visser and Heijnen, 2003) and rely on kinetic rate equations that are derived from in vitro measurements.

Alternative computational methods have also been developed for estimating kinetic parameters in biochemical networks using measured variables (Lei and Jorgensen, 2001; Mendes and Kell, 1998; Moles et al., 2003; Segre et al., 2003). These methods utilize nonlinear optimization techniques to find the most probable set of values for the model parameters that produce the experimental behavior (Mendes and Kell, 1998 (No. 1547); Lei and Jorgensen, 2001 (No. 3334); Segre et al., 2003 (No. 3335); Moles et al., 2003 (No. 3333)) and compute sets of parameter values that result in an optimal design of the system (Mendes and Kell, 1998; Moles et al., 2003).

Submitted July 26, 2004, and accepted for publication December 22, 2004.

Address reprint requests to Bernhard O. Palsson, Dept. of Bioengineering, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0412. Tel.: 858-534-5668; Fax: 858-822-3120; E-mail: palsson@ucsd.edu.

0006-3495/05/03/1616/10 $2.00

doi: 10.1529/biophysj.104.050385

This work introduces an alternative framework for incorporating steady-state in vivo data, in particular metabolomics data, with constraint-based modeling approach to determine all candidate numerical values of kinetic constants. With the recent developments in metabolomics technologies and the success in constraint-based modeling approach at the genome scale (Price et al., 2003; Reed and Palsson, 2003), this method may hold the promise for a network-scale characterization of kinetic parameters. Here, we describe an approach for integrating experimental data and constraint-based modeling to construct a kinetic solution space called the *k*-cone, demonstrate the application of *k*-cone analysis to interpret kinetic properties of biological networks, apply this approach to evaluate how well the in vitro measured parameters reproduce observed in vivo measurements, and demonstrate the utility of *k*-cone analysis in kinetic-based modeling. We first use a kinetic model of the human red blood cell (Jamshidi et al., 2001) to benchmark the usefulness of the *k*-cone approach for determining the kinetic values. Second we implement the *k*-cone analysis to study central metabolism in *Saccharomyces cerevisiae* for which a kinetic-based model has been developed based on available in vitro kinetic parameters and measured in vivo concentrations and reaction rates (Teusink et al., 2000).

METHODS AND MATERIALS

Development of the conceptual framework

Steady-state flux space

The stoichiometric matrix, **S**, is a mapping between the flux space and the time derivatives of metabolite concentrations (Fig. 1). At steady state, this transformation describes a flux balance (Bonarius et al., 1997; Kauffman et al., 2003; Varma and Palsson, 1994),

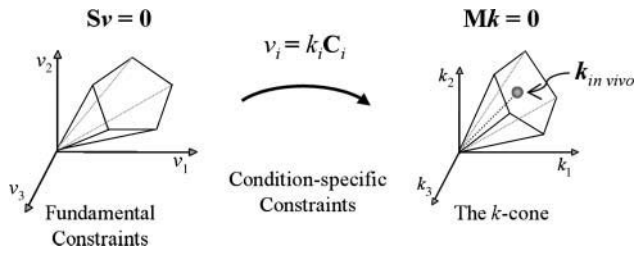

FIGURE 1 Transformation of the network-derived flux space to the condition dependent *k*-cone. Flux solution space (*left*) is constructed based on the fundamental physicochemical constraints including reaction stoichiometry, maximum enzyme capacity, and limited thermodynamics (Price et al., 2003; Reed and Palsson, 2003). Condition-specific metabolite concentrations allow for a transformation that maps the flux solution space onto a solution space containing kinetic constants, or the *k*-cone (*right*). Thus, any combination of the kinetic parameters inside the *k*-cone is attainable by the system and satisfies both physiochemical and condition-specific constraints in the network. **S**, stoichiometric matrix; $\boldsymbol{v}$, flux vector; **M**, matrix containing stoichiometric information and steady-state metabolite concentrations; $\boldsymbol{k}_{\text{in vivo}}$, vectors of in vivo kinetic values; **x**, metabolite concentrations.

$$\mathbf{S}\boldsymbol{v} = \mathbf{0}, \tag{1}$$

where **S** is an $m \times n$ matrix of stoichiometric coefficients and $\boldsymbol{v}$ is an $n \times 1$ vector of reaction rates. The solution to Eq. 1 resides in the null space of **S** and can be calculated using a null space operation, $N(\mathbf{S})$ (see, e.g., Strang, 1988). The columns of a matrix **P** span the null space of **S** where by definition $\mathbf{SP} = \mathbf{0}$ (Schilling et al., 1999; Schuster and Hilgetag, 1994). Defining a biologically meaningful basis for the null space of **S** has resulted in the development of a number of analytical tools such as extreme pathway analysis and elementary flux mode analysis (Papin et al., 2003; Schilling et al., 1999; Schuster and Hilgetag, 1994).

Kinetic solution space: the k*-cone formalism*

Reaction rate laws map an m-dimensional concentration vector, **x**, to the flux space via $\boldsymbol{v} = f(\mathbf{x})$, where f represents an n-dimensional nonlinear function comprising kinetic rate laws. If the concentrations are known, this function can be written in terms of the kinetic parameters as $f(\mathbf{x}) = \text{diag}(\mathbf{C})\boldsymbol{k}$ for mass action kinetic rates, where $\boldsymbol{k}$ is a vector of all elementary rate constants, and **C** is a vector of numerical product of substrate concentrations for each reaction or $C_j = \prod_{i=1}^{m} x_i^{S_{i,j}}, j = 1, \cdots, n$ (e.g., for $2x_1 + x_2 \rightarrow x_3$, $C = x_1^2 x_2$). The elements of **C** typically represent a first-order (reaction with one substrate) and bilinear (reaction with two substrates) kinetics. All reversible reactions are decoupled into a forward and reverse reaction. Mass action rates are thus typically given by either $f_i(\mathbf{x}) = k_i x_i$ (first-order reaction), or $f_i(\mathbf{x}) = k_i x_{i,1} x_{i,2}$ (bilinear reaction). If it is known that the kinetic orders deviate from unity (Schnell and Turner, 2004), the elements of **C** can be computed accordingly. Substituting $f(\mathbf{x})$ for $\boldsymbol{v}$ in Eq. 1 we obtain,

$$\mathbf{M}\boldsymbol{k} = \mathbf{0}, \tag{2}$$

where $\mathbf{M} = \mathbf{S}\,\text{diag}(\mathbf{C})$ is an $m \times n$ matrix. Note that Eq. 2 is analogous in form to Eq. 1. The set of all solutions to Eq. 2, therefore, resides in the null space of **M**, $N(\mathbf{M})$. Thus, by definition, the null space is spanned by the columns of a matrix **K**, which satisfy $\mathbf{MK} = \mathbf{0}$. We call the null space of **M**, the *k*-cone. The *k*-cone contains all the possible kinetic values that satisfy Eq. 2, which is derived from Eq. 1. Note that Eq. 1 is time and condition invariant, where as Eq. 2 can be stated for any observed steady state of a network and is thus condition dependent.

A convex basis set for the *k*-cone can be calculated directly from **M** by computing its null space for which all $k_i \geq 0$. Alternatively, it is possible to obtain a *k*-cone basis set for $N(\mathbf{M})$ by first computing a convex basis set for the null space of **S, P**, and then scaling **P** by substrate concentrations, so $\mathbf{K} = [\text{diag}(\mathbf{C})]^{-1}\mathbf{P}$. This definition follows from rewriting $\mathbf{SP} = \mathbf{0}$ as $\mathbf{S}\,[\text{diag}(\mathbf{C})][\text{diag}(\mathbf{C})]^{-1}\,\mathbf{P} = \mathbf{0}$, and because $\mathbf{M} = \mathbf{S}\,\text{diag}(\mathbf{C})$ we have that $\mathbf{K} = \text{diag}(\mathbf{C})^{-1}\,\mathbf{P}$. Note that **P** can be computed using existing extreme pathway algorithm (Schilling et al., 1999) and diag(**C**) is nonsingular.

In addition to the stoichiometric constraints and concentration measurements, incorporating equilibrium constants reduces the dimensionality of the *k*-cone by imposing a relationship between two elementary rate constants. For illustration of the *k*-cone formalism see Supplementary Material.

Note that experimental metabolite concentrations are often measured from a cell population. Incorporating such measurements in the *k*-cone analysis thus results in values that represent average kinetic estimates for the cell population. Further, a steady state of a population does not account for the inherent noise in biological networks.

Computational methods

To calculate the relationship between kinetic parameters measured in vitro and the *k*-cone, a combination of linear and nonlinear optimization methods were used.

Calculating the closest distance between vector k′ *to a* k*-cone*

To determine whether a specific point is in the k-cone as defined in Eq. 2, the following optimization problem was solved:

$$\min_{\boldsymbol{k}} \|\boldsymbol{k}-\boldsymbol{k}'\|_2 \quad \text{Subject to}: \quad \begin{matrix}\mathbf{M}\boldsymbol{k}=\mathbf{0} \\ 0\le l_i\le k_i\le u_i \quad i=1,\ldots,n\end{matrix}, \tag{3}$$

where $\boldsymbol{k}'$ is any given vector of numerical values for the kinetic constants, and l and u are the lower and upper bounds of the individual kinetic parameters k_i. Here, the objective is to minimize the Euclidean distance (2-norm) between the arbitrary point and the k-cone. This problem is a standard quadratic programming problem and was solved using MATLAB's (Mathworks, Natick, MA) quadprog function. If the Euclidean distance is found to be nonzero, then the arbitrary point lies outside the k-cone (see Example VI in the Supplementary Material).

Calculating the closest distance from a vector k′ *to a* k*-cone that is constructed using experimental error*

If the error in the experimental measurements of the metabolite concentrations is considered (i.e., replace diag($\mathbf{C}$) with diag($\mathbf{C}+\delta\mathbf{C}$) in $\mathbf{S}$ diag($\mathbf{C}$)$\boldsymbol{k}=\mathbf{0}$ so $\mathbf{S}$ diag($\mathbf{C}$)$\boldsymbol{k}+\mathbf{S}$ diag($\delta\mathbf{C}$)$\boldsymbol{k}=\mathbf{0}$), the k-cone analysis becomes slightly more complicated. The distance minimization problem was formulated as follows:

$$\min_{\boldsymbol{k},\delta y} \|\boldsymbol{k}-\boldsymbol{k}'\|_2 \quad \text{Subject to}: \quad \begin{matrix}\mathbf{M}\boldsymbol{k}+\delta\mathbf{M}\mathrm{diag}(\delta\mathbf{y})\boldsymbol{k}=\mathbf{0} \\ \delta\mathbf{M}=\mathbf{S}\mathrm{diag}(\delta\mathbf{C}) \\ -1<\delta y_i<1 \\ 0\le l_i\le k_i\le u_i\end{matrix}, \tag{4}$$

where $\delta\mathbf{M}$ is defined as $\delta\mathbf{M}=\mathbf{S}\mathrm{diag}(\delta\mathbf{C})$, $\delta\mathbf{y}$ is a scaled vector representing the extent of the error in each concentration for calculating the closest distance to an arbitrary point k', and l_i and u_i are the lower and upper bounds on the values of individual kinetic parameters, k_i, respectively. In this formulation, $\delta\mathbf{y}$ vector is used as an optimization variable to identify any point within the defined experimental error and not just the extremities (see Example IV in the Supplementary Material). This formulation helps to investigate the entire space of concentrations within the experimental error and hence all possibilities for $\delta\mathbf{M}$ when calculating the distance to the k-cone with experimental error.

Calculating the minimum required number of changes in k′ *to project it into the* k*-cone*

To calculate the minimum number of parameters needed to transform an arbitrary kinetic point $\boldsymbol{k}'$ into a k-cone, the following mixed integer linear programming (MILP) problem was formulated:

$$\min_{\boldsymbol{k},\theta} \sum_{i=1}^{n}\theta_i \quad \text{Subject to}: \quad \begin{matrix}|\boldsymbol{k}-\boldsymbol{k}_{\text{in vitro}}|\le \mathrm{N}\boldsymbol{\theta} \\ \mathbf{M}\boldsymbol{k}=\mathbf{0} \\ 0\le l_i\le k_i\le u_i \quad i=1,\ldots,n\end{matrix}, \tag{5}$$

where $\boldsymbol{k}_{\text{in vitro}}$ represents the in vitro measurements of the kinetic constants, $\boldsymbol{\theta}$ is the vector of Boolean variables (i.e., θ_i is 0 or 1), n is the dimension of the kinetic space, and N is a large positive number. If a specific in vitro kinetic parameter is changed, then θ_i corresponding to that parameter must be 1, so that the inequality in Eq. 5 can be satisfied. The solution computed by MILP thus provides the minimum number of changes that needs to be made and the associated vector of kinetic constants that lies within the k-cone (see Example VI in the Supplementary Material). The MILP problem was formulated in MATLAB (Mathworks) and solved using LINDO API's MILP solver (LINDO Systems, Chicago, IL).

Metabolic networks used

Human red blood cell metabolic network

The metabolic network of the human red blood cell (RBC) was constructed based on a mechanistic model of RBC available in Mathematica (Jamshidi et al., 2001). The network contained 33 internal reactions and 36 metabolites (Fig. 2; see also Tables 3 and 4 provided in the Supplementary Material). The kinetic rate laws and kinetic parameters were adopted from the mechanistic model of RBC and steady-state metabolite concentrations under normal or increased metabolic load conditions were obtained from simulation results (Table 5 in Supplementary Material). The computed concentration values from the kinetic model of RBC have shown to agree well with measured values (Joshi and Palsson, 1990).

S. cerevisiae *central metabolic network and steady-state calculations*

The central metabolic network of *S. cerevisiae* was adopted from Teusink et al. (2000) and contained 18 internal reactions and 24 metabolites (Tables 6 and 7 as provided in the Supplementary Material). The in vivo flux measurements, as reported in Teusink et al. (2000), did not result in a complete carbon balance in the network. To obtain a balanced steady-state flux distribution, a set of flux values were estimated that best fit the experimental data using linear optimization and flux balance analysis software (Edwards et al., 1999). An associated flux error was extrapolated for each reaction using the experimental error (Table 6 as provided in the Supplementary Material). The same procedure was used to estimate the model-based flux distribution in cases where reaction rates were not reported in Teusink et al. (2000) (Table 6 as provided in the Supplementary Material).

Computing the observed kinetic values

Ideally, the reaction network should be represented by elemental rate reactions (i.e., bilinear association of two components at each step of the enzymatic reaction). However, the available kinetic models use enzymatic rate laws following simplifying assumptions applied to the corresponding reaction mechanism. To avoid dealing with the complexity of such rate laws, we instead used ''observed'' kinetic values that represent pseudoelementary rate constants (PERCs). The observed kinetic values for irreversible reactions were calculated by dividing the steady-state fluxes by the numerical product of the steady-state substrate concentrations. For reactions that contained an additional component in the model such as an inhibition or activation factor, a numerical multiplier was calculated. The multiplier was used to adjust for the relationship between metabolites and reaction rates at a given steady state. For example, the kinetic rate equation for glucose 6-phosphate isomerase in the human red blood cell (RBC) is defined in the kinetic model of RBC as $v_{\mathrm{PGI}}=(k_{\mathrm{pgi,f}}\times \mathrm{G6P}-k_{\mathrm{pgi,r}}\times \mathrm{F6P})\times \mathrm{Exp}[(13.042-4040.7/\mathit{TEMP})]$, where *TEMP* represents the internal temperature of the red cell in Kelvin (Jamshidi et al., 2001). Thus, the combined term $k_{\mathrm{pgi,f}}\times \mathrm{Exp}[(13.042–4040.7/\mathit{TEMP})]$ was taken to be the observed kinetic parameter corresponding to $v_{\mathrm{PGI,f}}$.

RESULTS

The k-cone formalism was used to determine the candidate kinetic values for: 1), the human red blood cell metabolic network for which a kinetic-based model is available (Jamshidi

et al., 2001), and 2), the central metabolism in *S. cerevisiae* for which in vivo concentrations and reaction rates have been measured and a kinetic-based model has been previously developed using in vitro kinetic values (Teusink et al., 2000).

Human red blood cell metabolism

Values for the pseudoelementary rate constants

To compute the numerical values of PERCs for the red blood cell metabolism under given conditions, the forward and reverse kinetic values of the red blood cell model and equilibrium constants for the reversible reactions were obtained from the RBC model (Table 8, columns 1–3, as provided in the Supplementary Material). By decomposing the reversible reactions into a forward and reverse reaction, the total number of components in the nominal kinetic vector included 50 kinetic values (2×17 reversible reactions + 16 irreversible reactions) and 17 exchange fluxes (Table 1). Thus, $\boldsymbol{k}$ is a 67-dimensional vector.

Red blood cell k*-cone calculation*

The *k*-cone for the human red blood cell, following the framework presented earlier and using the extreme pathway algorithm (Schilling et al., 1999), was calculated to have 56 nonnegative convex basis vectors that are in 30-dimensional null space of $\mathbf{M}_{\mathrm{RBC}}$. Adding the equilibrium constants (Table 8 as provided in the Supplementary Material) into $\mathbf{M}_{\mathrm{RBC}}$, reduced the number of spanning convex vectors from 56 to 24 (equivalently reducing the linear 30-dimensional null space of $\mathbf{M}_{\mathrm{RBC}}$ into a 13-dimentional null space for $\mathbf{M}_{\mathrm{RBC\&Keq}}$). In both cases, the nominal kinetic vector (Table 1) satisfied Eq. 2 and was verified by Eq. 3 to be inside the *k*-cone.

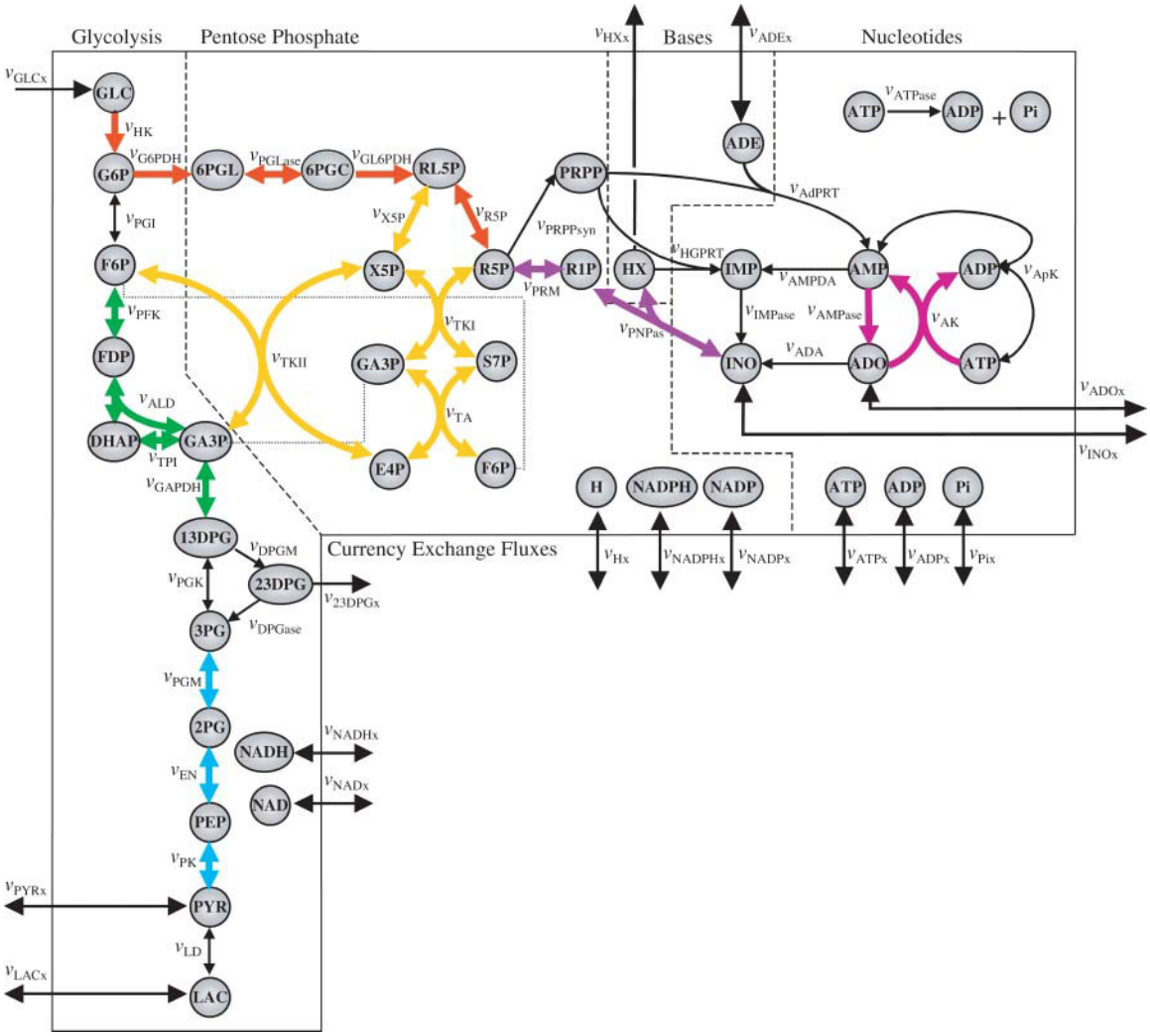

FIGURE 2 Metabolic network in the human red blood cell. The metabolite and reaction abbreviations are listed in Tables 3 and 4, as provided in Supplementary Material (adopted from Wiback and Palsson, 2002). The highlighted color-coded sets of reactions are kinetically correlated as detailed in the Results section.

TABLE 1 Pseudoelementary rate constants (PERCs) for the no-load condition in the red blood cell model with no metabolic load at steady-state (*k*_nom, nominal PERCs values)

PERCs	k-$_{nom}$	Units	PERCs	k-$_{nom}$	Units
kHK	1.044 E − 1	1/[mM/h]	kR5P-lf	2.156 E + 3	[1/h]
kPGlf	6.074 E + 4	[1/h]	kR5P-lr	8.391 E + 2	[1/h]
kPGlr	1.482 E + 5	[1/h]	kXu5P-lf	2.354 E + 4	[1/h]
kPFK	1.893 E + 1	1/[mM/h]	kXu5P-lr	7.848 E + 3	[1/h]
kALDf	7.412 E + 2	[1/h]	kTKlf	1.924 E + 3	1/[mM/h]
kALDr	9.150 E + 3	1/[mM/h]	kTKlr	1.603 E + 3	1/[mM/h]
kTPlf	3. 110 E + 4	[1/h]	kTKllf	1.792 E + 3	1/[mM/h]
kTPlr	6.842 E + 5	[1/h]	kTKllr	1.740 E + 2	1/[mM/h]
kGAPDHf	1.166 E + 5	1/[mM^2/h]	kTAf	3.535 E + 3	1/[mM/h]
kGAPDHr	1.032 E + 11	1/[mM^2/h]	kTAr	3.367 E + 3	1/[mM/h]
kPGKf	8.586 E + 6	1/[mM/h]	kAMPase	1.625 E + 0	[1/h]
kPGKr	4.770 E + 3	1/[mM/h]	kADA	3.999 E + 2	[1/h]
kDPGM	2.183 E + 3	[1/h]	kAK	1.879 E + 3	1/[mM/h]
kDPGase	1.074 E − 1	[1/h]	kApKf	5.607 E + 1	1/[mM/h]
kPGMf	5.073 E + 3	[1/h]	kApKr	3.772 E + 1	1/[mM/h]
kPGMr	3.449 E + 4	[1/h]	kAMPDA	1.086 E − 2	[1/h]
kENf	5.153 E + 4	[1/h]	kAdPRT	6.490 E + 2	1/[mM/h]
kENr	3.016 E + 4	[1/h]	kIMPase	9.184 E − 2	[1/h]
kPK	2.520 E + 2	1/[mM/h]	kPNPasef	1.993 E + 3	1/[mM/h]
kLDf	3.149 E + 5	1/[mM/h]	kPNPaser	2.215 E + 4	1/[mM/h]
kLDr	7.055 E + 3	1/[mM/h]	kPRMf	4.371 E + 2	[1/h]
kG6PDH	8.293 E + 3	1/[mM/h]	kPRMr	3.287 E + 1	[1/h]
kPGLase	1.260 E + 2	[1/h]	kPRPPsyn	3.066 E + 1	1/[mM/h]
kGL6PDHf	1.734 E + 4	1/[mM/h]	kHGPRT	5.002 E + 1	1/[mM/h]
kGL6PDHr	1.021 E + 2	1/[mM^2/h]	kATPase	4.116 E − 1	[1/h]

Correlated kinetic variables

To determine the allowable range and distribution of the kinetic values, the *k*-cone was uniformly sampled using a randomized Monte Carlo approach (Price et al., 2004b). This approach has proven to be efficient for characterizing flux solution space in metabolic networks (Almaas et al., 2004; Price et al., 2004a; Wiback et al., 2004) and was utilized here to determine the content of the *k*-cone in normal and under metabolic load conditions. The *k*-cone for the human red blood cell was randomly sampled and a histogram for each kinetic variable was generated (Fig. 3, *black line*). Each histogram is a one-dimensional projection of the *k*-cone onto an axis that represents a kinetic parameter. Thus, the height of each histogram indicates how ''deep'' the space is for that parameter at a given value.

Many of the kinetic parameters exhibited similar histogram projections (e.g., k_{PGIf} and k_{PGIr}), suggesting that kinetic values may display correlated functions. A pairwise cross-correlation analysis of the randomly selected points in the *k*-cone showed that the candidate values of the rate constants are highly correlated. For a correlation coefficient of >0.9, the kinetic parameters fell into 10 correlated sets of kinetic constants. All the reversible reactions appeared in the 10 correlated groups. The pairwise correlation between the forward and reverse reactions is a result of the imposed equilibrium constraints and the ratio between their values was equal exactly to the K_{eq}, as expected. Four of the 10 correlated groups contain only the forward and reverse reactions: glucose-6-phosphate isomerase (PGI), phosphoglycerate kinase (PGK), lactate dehydrogenase (LD), and adenylate kinase (ApK) (Fig. 2). The remaining six correlated groups are composed of (see Fig. 2):

1. Hexokinase (HK), oxidative branch of pentose phosphate pathway, and ribose-5-phosphate isomerase (R5P),
2. Nonoxidative branch of pentose phosphate pathway excluding ribose-5-phosphate isomerase,
3. Phosphofructokinase (PFK), fructose-bisphosphate aldolase (ALD), triosephosphate isomerase (TPI), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH),
4. Phosphoglycerate mutase (PGM), enolase (EN), and pyruvate kinase (PK),
5. Purine nucleoside phosphorylase (PNPase), phosphoribomutase (PRM), and phosphoribosyl pyrophosphate synthetase (PRPPsyn),
6. Adenosine monophosphate phosphohydrolase (AMPDA) and inosine monophosphtase (IMPase).

Note that the coordinated function of kinetic parameters results from the network topology and can be predicted from extreme pathway analysis and flux coupling analysis (Schilling et al., 1999). The correlation within a group of kinetic parameters implies that their values change coordinately, and thus determining the value of one parameter sets the value for the rest of the parameters in the group. Thus,

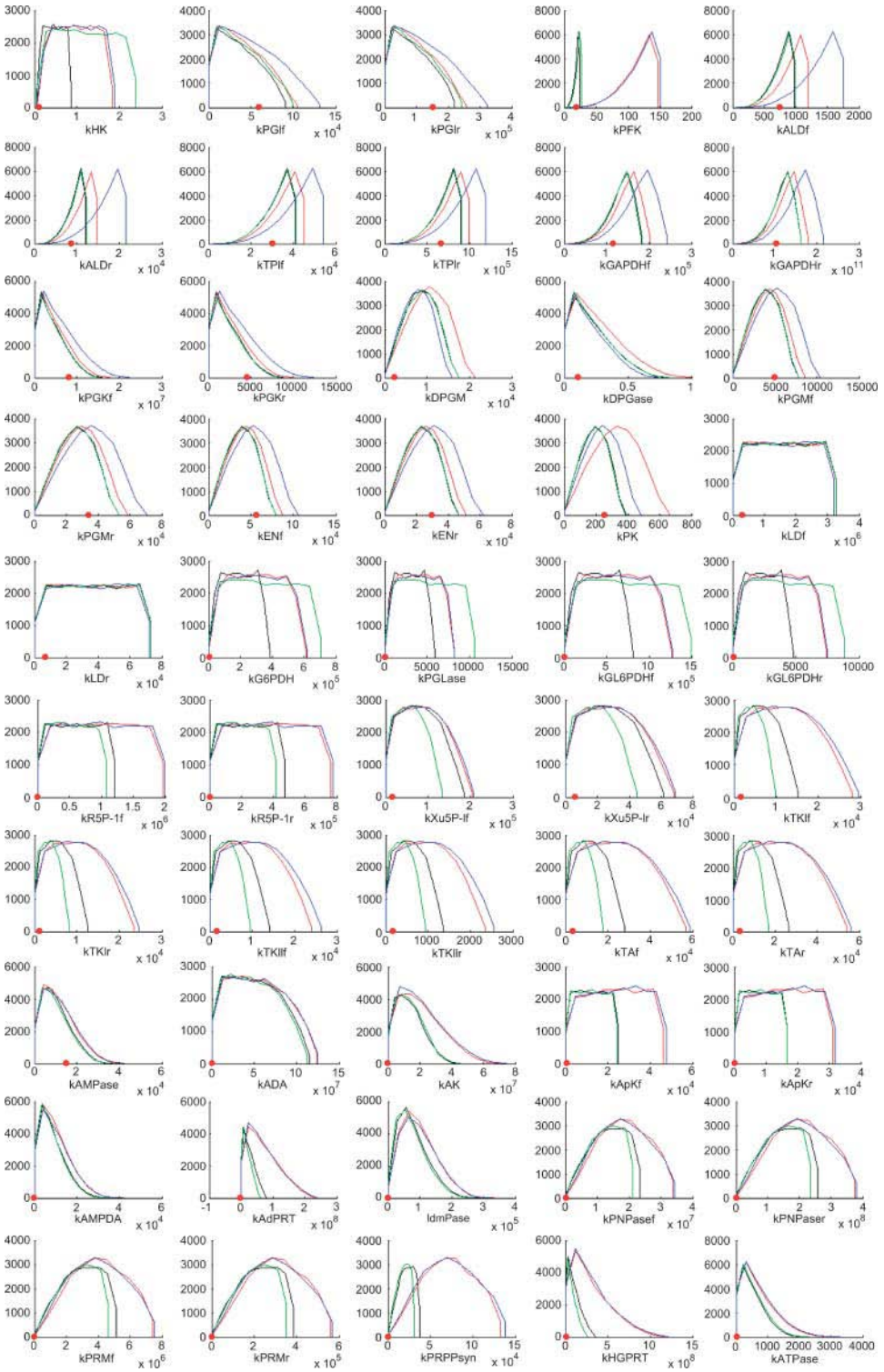

FIGURE 3 Randomized sample point histograms of the red blood cell *k*-cone under normal or no load (*black*), ATP load (*red*), 2,3-DPG load (*green*), and NADH load (*blue*). Randomized points (20,000) were sampled from the red blood cell *k*-cone. The histogram plot of the sample points illustrates the allowable range of kinetic variables and its height indicates the depth of the solution space at different variable values. The red dots indicate the pseudoelementary rate constants for the no-load condition.

the steady-state kinetic parameters are not independent and are fundamentally constrained by Eq. 1 and condition-specific constraints of Eq. 2, which represent how the measured concentrations lead to determination of the numerical values for the pseudoelementary rate constants.

Shrinking the k*-cone*

A network under different steady-state conditions may have concentration values that change significantly and, consequently, they result in different condition-dependent *k*-cones. Assuming that the enzyme concentrations do not change over

time, the actual $\boldsymbol{k}$ must then be able to satisfy all conditions simultaneously and as a result it must reside at the intersection of all *k*-cones (Fig. 4 *A*). By calculating the *k*-cone, under various experimental conditions and by eliminating the kinetic values that cannot satisfy the network demands under all alternative conditions, we can thus define a much narrower range for $\boldsymbol{k}$ (see Example V in the Supplementary Material).

To define a more precise range of values for kinetic parameters in the human red blood cell network, RBC *k*-cone was computed under four physiological conditions (Jamshidi et al., 2002): a), the normal or no-load condition; b), ATP load where ATP_{PUMP} (i.e., ATP + 2 Na_{in} +3 K_{out} → ADP + Pi + 2 Na_{out} + 3 K_{in}) was set to a constant value of 0.37 mM/h; c), 2,3-DPG load where 2,3-DPG was siphoned off from the network at a constant rate of 0.45 mM/h; and d), NADH load with a constant drain of 1.7 mM/h. The cell faces such conditions during osmotic imbalance (ATP load), at high altitude (2,3-DPG load), or when the cell must reduce methemoglobin (NADH load).

The histogram projections of kinetic variables were generated as described previously (Fig. 3). The intersection between the four condition-dependent *k*-cones contains $\boldsymbol{k}$ vectors that satisfy all conditions. Thus, the kinetic range common to all four conditions was used to represent the reduced solution space. The most constraining range was imposed by the no-load condition in k_{PFK} compared to ATP and NADH loads. k_{PFK} histogram also showed that under ATP and NADH load, the space is narrow in the range between 0 and 25 mM^{-1} h^{-1}, as indicated by the height of the histogram. The most constraining condition was 2,3-DPG load, reducing the kinetic values in 33 out of 50 cases. As expected, the observed kinetic vectors for all conditions fell within the most restricted kinetic set or inside the intersection of all *k*-cones (data not shown). Thus, constraining the range of allowable combinations of kinetic parameters in a network by examining various experimental conditions results in a set that is biologically more accurate.

The histogram projection of the observed kinetic parameter for lactate dehydrogenase (*k*LD) remained the same under all conditions (Fig. 3). The phosphofructokinase (*k*PFK) histogram, however, moved to the right under ATP and NADH loads and covers a wider range of parametric values. The observed change in *k*PFK histogram clearly demonstrates that the solution space shifts under ATP and NADH load conditions and a different set of kinetic values become available.

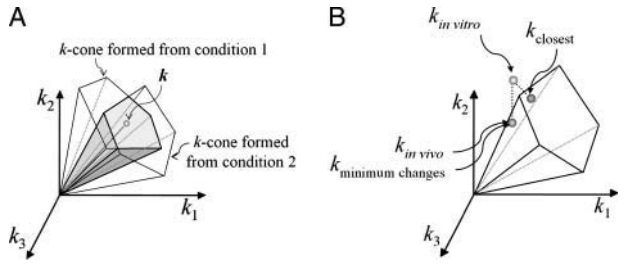

FIGURE 4 Schematic representation of the *k*-cone formation from different experimental conditions and for the central metabolic model of *S. cerevisiae*. The solution space and the kinetic values are shown here only for illustration. (*A*) The steady-state metabolite concentrations under different experimental conditions may result in the formation of distinct *k*-cones. The true $\boldsymbol{k}$ however must always fall within the intersection of all *k*-cones (i.e., the *shaded* area) to satisfy all conditions simultaneously. (*B*) $k_{in\ vitro}$ represents the observed in vitro measured kinetic parameters. $k_{closest}$ represents the closest point from $k_{in\ vitro}$ to the *k*-cone, which requires 24 out of 25 values to be changed. $k_{in\ vivo}$ represents the in vivo and adjusted kinetic set. $k_{minimum\ changes}$ is obtained from MILP calculation and is identical to $k_{in\ vivo}$ (only 12 out of 25 parameters needed to be changed).

S. cerevisiae central metabolism

K-cone analysis was performed for the central metabolism in *S. cerevisiae* using in vivo measurements, in vitro kinetic parameters, and previously published kinetic-based model as was reported by Teusink and colleagues (Teusink et al., 2000). In this study, in vivo experimental data and in vitro kinetic measurements were obtained under identical conditions and from the same yeast source (Teusink et al., 2000). Teusink and colleagues have shown that the kinetic-based model of central metabolism using the in vitro kinetic measurements (''unadjusted model'') was unable to reproduce the in vivo experimental data sets, and to correctly reproduce the experimental values, kinetic parameters of half of the enzymes had to be changed (''adjusted model'') (Teusink et al., 2000).

To demonstrate the utility of *k*-cone analysis in kinetic-based modeling, the methods introduced earlier were applied to the central metabolism of *S. cerevisiae* to determine whether *k*-cone analysis can identify what in vitro kinetic parameters do not reproduce in vivo data and to use this analysis to identify a set of candidate values that agree with experimental measurements.

Numerical values for the PERCs

The pseudoelementary rate constants for in vivo measurements, unadjusted kinetic model, and adjusted kinetic model were computed as described previously (Table 2). The PERCs for hexokinase and phosphoglycerate kinase were not calculated because the experimental values for glucose and 1,3-diphosphoglycerate concentrations were not reported (Teusink et al., 2000). Hexokinase and phosphoglycerate kinase were instead represented as flux variables in the observed kinetic vector (see Table 2). Note that the reversible reactions were not decomposed into forward and reverse reactions for simplicity.

The k*-cone*

The convex basis for the *k*-cone was calculated as described above and it contained four vectors. The observed (or PERC) in vivo $\boldsymbol{k}$ was shown to reside in the *k*-cone, as expected (see

the schematic representation shown in Fig. 4 *B*). In comparison, the unadjusted $\boldsymbol{k}$ was found to be outside the *k*-cone ($k_{\text{in vitro}}$ in (Fig. 4 *B*). The adjusted $\boldsymbol{k}$ was also shown to be inside the solution space.

Incorporating experimental error in k*-cone calculation*

To determine if incorporating the experimental error would account for discrepancy of the in vitro $\boldsymbol{k}$ with the kinetic solution space, an optimization procedure was formulated (see Computational methods) by which the experimental error was integrated into *k*-cone calculation. Even when experimental error was taken into account, the unadjusted $\boldsymbol{k}$ was shown to be outside the solution space and could not reproduce the observed in vivo experimental results.

Parameter adjustments using the k*-cone approach*

To use *k*-cone analysis for determining the required changes in the unadjusted kinetic parameters, two approaches were utilized. First, the closest *k*-cone solution point to the unadjusted *k* parameters was calculated using linear optimization (see Computational methods). The *k*-cone solution that resulted from this calculation required that 24 of 25 parameters in the unadjusted $\boldsymbol{k}$ must be changed (Table 2). Next, the minimum number of changes to convert the unadjusted $\boldsymbol{k}$ into a *k*-cone solution point was calculated using an MILP approach (see also Computational methods). In this case, only 12 of 25 parameters were changed for the unadjusted $\boldsymbol{k}$ and the resulting parameter set matched closely the adjusted values that Teusink and colleagues generated by manually altering the model parameters (Table 2).

The MILP procedure identified a solution that is identical to the changes required in the kinetic model of *S. cerevisiae*. Although other sets of parameter changes exist for which the unadjusted $\boldsymbol{k}$ can be modified to satisfy the in vivo measurements, the minimum MILP solution is unique (i.e., there is only one solution with the minimum of 12 changes). The smallest set following the minimal set of 12 adjustments required 17 parameters to be changed.

DISCUSSION

The absence of a comprehensive kinetic parameter set and the observed inconsistency in the available kinetic data have hindered characterization of the kinetic properties in biological networks and studying biological functions using kinetic-based models. This work introduces a new approach for incorporating in vivo data with constraint-based modeling approach to construct a kinetic solution space (the *k*-cone) for large-scale biological networks and to estimate numerical values for the kinetic parameters at steady state. Here, we have shown that: 1), the *k*-cone can be defined and computed using constraint-based modeling approach and steady-state metabolite concentrations; 2), the *k*-cone contains all allowable combinations of kinetic values that satisfy a measured steady-state concentration and exchange fluxes; 3), the *k*-cone can be

TABLE 2 Observed in vivo, unadjusted, adjusted, closest, and MILP-obtained kinetic vectors for the kinetic model of core *S. cerevisiae* metabolism; PERCs, pseudoelementary rate constants

PERCs	In vivo	Unadjusted model	Adjusted model	Closest point	MILP solution	Units
vHK	108	88	88	384	88	mmol/min/L-cytosol
kPGIR	39.67	72.15	31.51	157	32	1/min
kPFK	62.21	279.61	49.41	246	49	L-cytosol/mmol/min
kALDR	17.64	128.67	14.01	70	14	1/min
kTPIR	97.53	79.73	72.84	475	73	1/min
kGAPDH	978.89	2638.77	756.67	4271	757	L-cytosol/mmol/min
vPGK	176.2	136.2	136.2	769	136	mmol/min/L-cytosol
kPGMR	195.78	378.33	151.33	854	151	1/min
kENOR	1468.33	3405.00	1135.00	6406	1135	1/min
kPYK	1906.93	1508.31	1474.03	8319	1474	L-cytosol/mmol/min
kPDC	95.24	15.99	73.62	416	74	1/min
kADH	2547.81	18958.82	1944.49	11594	1944	L-cytosol/mmol/min
kAKR	0	0	0	0	0	L -cytosol/mmol/min
kG3PDH	57.61	614.86	57.61	0	58	L-cytosol/mmol/min
kATPase	52.63	36.91	36.76	305	37	1/min
kGLGsyn	2.45	2.23	2.45	0	2	L-cytosol/mmol/min
kTREsyn	1.96	1.67	1.96	0	2	L-cytosol^2/mmol^2/min
kSUCCsyn	1.81	0.85	1.81	0	2	L-cytosol^8/mmol^8/min
vGLCxt	−108	−88	−88	−384	−88	mmol/min/L-cytosol
vETHxt	168.92	128.92	128.92	769	129	mmol/min/L-cytosol
vCO2xt	176.2	136.20	136.20	769	136	mmol/min/L-cytosol
vGLGsyn	6	6	6	0	6	mmol/min/L-cytosol
vTREsyn	4.8	4.80	4.80	0	5	mmol/min/L-cytosol
vGLYxt	18.2	18.20	18.20	0	18	mmol/min/L-cytosol
vSUCCxt	3.64	3.64	3.64	0	4	mmol/min/L-cytosol

reduced by imposing additional constraints such as equilibrium constants; 4), *k*-cone analysis allows for identification of sets of correlated kinetic parameters; 5), the range of feasible numerical values for the kinetic constants determined from the *k*-cone analysis can be reduced by computing the *k*-cone under different conditions; 6), consistency in in vitro parameters to reproduce in vivo measurements can be investigated using *k*-cone analysis and independent from a kinetic model; and 7), the minimum required in vitro parameters that must be adjusted to reproduce in vivo measurements can be computed using this approach and mixed integer linear programming techniques.

In addition to characterizing a feasible solution space for kinetic parameters in biological networks, *k*-cone analysis may be used to develop kinetic-based models by defining an allowable range of kinetic values that are consistent with large-scale data sets. In vitro obtained kinetic parameters, as was shown for central metabolic network of *S. cerevisiae*, may fall outside the values that satisfy the network phenotype. Computing the kinetic solution space a priori to kinetic model development may reduce the time and effort involved in model building and parameter adjustment.

In case of human red blood cell, implementation of *k*-cone analysis in determining the range of kinetic parameters and defining the intersection of the *k*-cones was particularly applicable. Red cell is enucleated upon maturation and consequently lacks transcriptional regulation. The enzyme concentration is also assumed not to change over time and thus the changes that are observed in the metabolite concentrations in the red cell are due solely to the imposed demand on the cell and the control mechanism by the kinetic regulation. Thus, determining the intersecting cone in the RBC metabolic network is appropriate. For other cell types, however, changes in enzyme concentrations and transcriptional regulation can affect the observed concentration values by changing the total rate of enzymatic conversion. In such cases, other aspects of cellular metabolism including enzyme concentrations must be taken into account in the network reconstruction. Such analyses could thus be used to integrate proteomic and transcriptomic data for enzymes in addition to metabolomic and fluxomic data in a chemically and genetically structured setting.

Shrinking the solution space as described here can also be used for networks that exhibit multiple steady states. Although multiple steady-state solutions were not encountered in the red blood cell kinetic model, the set of concentration profiles that result from multiple steady states can also be used as additional criteria for constraining the *k*-cone.

If all the network interactions are captured correctly, the intersection of the *k*-cones obtained from different conditions or multiple steady states will provide the most concise representation of what kinetic values are allowed in the system. There may be, however, cases in which the *k*-cones do not intersect. Such cases may arise if the network is incomplete and some network components are missing. In combination with the *k*-cone analysis, a hypothesis-driven iterative approach should lead to the identification of the missing components and refinement of the network structure.

As a proof of concept and also for simplicity, the treatment of the rate equations in this study was centered around linearization of enzyme-saturated rate curves. To take into account the nonlinear nature of kinetic reactions, necessary modification must be applied that may include incorporating enzyme concentrations, or utilizing nonlinear optimization techniques to define and study the allowable solution space. The usefulness of *k*-cone analysis even at this simplified stage is clearly evident for kinetic model development and assessment of in vitro measured kinetic values. For cases where the kinetic mechanism of biochemical reactions is not available, the use of pseudoelementary constants may be the most efficient way to apply the *k*-cone analysis to genome-scale biological network.

For visualization purposes, the *k*-cone analysis was introduced here using extreme pathway analysis to compute a nonnegative convex *k*-cone. The *k*-cone formalism, however, does not require a nonnegative convex basis and other bases sets including orthonormal bases generated by singular value decomposition can be used to calculate the allowable kinetic solution space. Other computational methods used here to generate the results including the randomized sampling, minimum Euclidian distance, and MILP calculations can also be combined equally well with other types of bases to generate similar results. The *k*-cone analysis is thus not restricted by the computational limitation of extreme pathway calculation and can be applied to large-scale biological networks.

With the recent developments in high-throughput profiling of metabolic concentration at a whole-cell scale and advances in metabolomics technologies, this approach may hold the promise for kinetic characterization of metabolic networks as well as other biological functions at a whole-cell level.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

We thank Bas Teusink for valuable discussions and his help with the kinetic model of central metabolism for *S. cerevisiae* and Sharon Wiback for her input and help with the red blood cell kinetic model.

We thank the Whitaker Foundation for its support through the Graduate Fellowships in Biomedical Engineering to I.F., the National Science Foundation (BES 03-31342), and the National Institutes of Health (R01 GM68837). The authors disclose a potential financial conflict of interest.

REFERENCES

Almaas, E., B. Kovacs, T. Vicsek, Z. N. Oltvai, and A. L. Barabasi. 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*. 427:839–843.

Bonarius, H. P. J., G. Schmid, and J. Tramper. 1997. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* 15:308–314.

Edwards, J. S., R. Ramakrishna, C. H. Schilling, and B. O. Palsson. 1999. Metabolic flux balance analysis. *In* Metabolic Engineering. S. Y. Lee and E. T. Papoutsakis, editors. Marcel Dekker, New York

Jamshidi, N., J. S. Edwards, T. Fahland, G. M. Church, and B. O. Palsson. 2001. Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics.* 17:286–287.

Jamshidi, N., S. J. Wiback, and B. O. Palsson. 2002. *In silico* model-driven assessment of the effects of single nucleotide polymorphisms (SNPs) on human red blood cell metabolism. *Genome Res.* 12:1687–1692.

Joshi, A., and B. O. Palsson. 1990. Metabolic dynamics in the human red cell. Part IV. Data prediction and some model computations. *J. Theor. Biol.* 142:69–85.

Kauffman, K. J., P. Prakash, and J. S. Edwards. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14:491–496.

Lei, F., and S. B. Jorgensen. 2001. Estimation of kinetic parameters in a structured yeast model using regularisation. *J. Biotechnol.* 88:223–237.

Mendes, P., and D. Kell. 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics.* 14:869–883.

Moles, C. G., P. Mendes, and J. R. Banga. 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13:2467–2474.

Papin, J. A., N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson. 2003. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* 28:250–258.

Price, N. D., J. A. Papin, C. H. Schilling, and B. Palsson. 2003. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol.* 21:162–169.

Price, N. D., J. L. Reed, and B. O. Palsson. 2004a. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2:886–897.

Price, N. D., J. Schellenberger, and B. Ø Palsson. 2004b. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys. J.* 87:2172–2186.

Reed, J. L., and B. O. Palsson. 2003. Thirteen years of building constraint-based in silico models of *Escherichia coli*. *J. Bacteriol.* 185:2692–2699.

Rizzi, M., M. Baltes, U. Theobald, and M. Reuss. 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*. 2. Mathematical model. *Biotechnol. Bioeng.* 55:592–608.

Schilling, C. H., S. Schuster, B. O. Palsson, and R. Heinrich. 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15:296–303.

Schnell, S., and T. E. Turner. 2004. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. Biophys. Mol. Biol.* 85:235–260.

Schuster, S., and C. Hilgetag. 1994. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* 2:165–182.

Segre, D., J. Zucker, J. Katz, X. Lin, P. D'Haeseleer, W. P. Rindone, P. Kharchenko, D. H. Nguyen, M. A. Wright, and G. M. Church. 2003. From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS.* 7:301–316.

Strang, G. 1988. Linear Algebra and its Applications. Saunders College Publishing, Fort Worth, TX.

Teusink, B., J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep .2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* 267:5313–5329.

Theobald, U., W. Mailinger, M. Baltcs, M. Rizzi, and M. Reuss. 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*. I. Experimental observations. *Biotechnol. Bioeng.* 55:305–316.

Varma, A., and B. O. Palsson. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology.* 12:994–998.

Vaseghi, S., A. Baumeister, M. Rizzi, and M. Reuss. 1999. *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metab. Eng.* 1:128–140.

Visser, D., and J. J. Heijnen. 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab. Eng.* 5:164–176.

Wiback, S. J., I. Famili, H. J. Greenberg, and B. O. Palsson. 2004. Monte Carlo sampling can be used to determine the size and shape of the steady state flux Space. *J. Theor. Biol.* 228:437–447.

Wiback, S. J., and B. O. Palsson. 2002. Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.* 83:808–818.

Wright, B. E., and P. J. Kelly. 1981. Kinetic models of metabolism in intact cells, tissues, and organisms. *Curr. Top. Cell. Regul.* 19:103–158.