In contrast to stoichiometric-based models, the development of large-scale kinetic models of metabolism has been hindered by the challenge of identifying kinetic parameter values and kinetic rate laws applicable to a wide range of environmental and/or genetic perturbations. The recently introduced ensemble modeling (EM) procedure provides a promising remedy to address these challenges by decomposing metabolic reactions into elementary reaction steps and incorporating all phenotypic observations, upon perturbation, in its model parameterization scheme. Here, we present a kinetic model of *Escherichia coli* core metabolism that satisfies the fluxomic data for wild-type and seven mutant strains by making use of the EM concepts. This model encompasses 138 reactions, 93 metabolites and 60 substrate-level regulatory interactions accounting for glycolysis/gluconeogenesis, pentose phosphate pathway, TCA cycle, major pyruvate metabolism, anaerobic reactions and a number of reactions in other parts of the metabolism. Parameterization is performed using a formal optimization approach that minimizes the discrepancies between model predictions and flux measurements. The predicted fluxes by the model are within the uncertainty range of experimental flux data for 78% of the reactions (with measured fluxes) for both the wild-type and seven mutant strains. The remaining flux predictions are mostly within three standard deviations of reported ranges. Converting the EM-based parameters into a Michaelis-Menten equivalent formalism revealed that 35% of $K_m$ and 77% of $k_{cat}$ parameters are within uncertainty range of the literature-reported values. The predicted metabolite concentrations by the model are also within uncertainty ranges of metabolomic data for 68% of the metabolites. A leave-one-out cross-validation test to evaluate the flux prediction performance of the model showed that metabolic fluxes for the mutants located in the proximity of mutations used for training the model can be predicted more accurately. The constructed model and the parameterization procedure presented in this study pave the way for the construction of larger-scale kinetic models with more narrowly distributed parameter values as new metabolomic/fluxomic data sets are becoming available for *E. coli* and other organisms.

1. Introduction

Over the past decade stoichiometric based genome-scale metabolic models have been derived for a variety of organisms (Kim et al., 2012). The global nature of these models enables the assessment of theoretical limits of metabolic performance (Reed and Palsson, 2003) and the identification of plausible engineering strategies (Xu et al., 2011). These predictions can be significantly sharpened through the introduction of regulatory constraints (Shlomi et al., 2007; Covert et al., 2001) and restrictions implied by thermodynamic information (Henry et al., 2007). Stoichiometric models have recently been used to model microbial communities as well (Klitgord and Segrè, 2010; Salimi et al., 2010; Stolyar et al., 2007; Zhuang et al., 2011; Zhuang et al., 2012; Zomorrodi et al., 2014; Zomorrodi and Maranas, 2012). Despite these advances, stoichiometric models alone cannot quantitatively capture the effect of concentration levels and enzyme saturation on reaction throughput and regulation. Kinetic models have the potential to capture these interdependencies. However, in contrast to stoichiometric models, the construction of large-scale kinetic models has been plagued by a number of challenges among which are the need for selection of kinetic rate laws for each reaction and estimation of individual kinetic parameters as well as the paucity of relevant experimental data to support unambiguous kinetic model parameterization.
A. Khodayari et al. / Metabolic Engineering 25 (2014) 50–62

A. Khodayari et al. / Metabolic Engineering 25 (2014) 50–62

A challenge associated with determination of parameters of kinetic models using the sampling methods is an extremely large feasible search space (Schellenberger and Palsson, 2009). Ensemble modeling (EM) of metabolic networks (Tran et al., 2008) was developed towards addressing this challenge through successively reducing the size of parameter space by using phenotypic data such as experimental flux and/or concentration measurements while incorporating thermodynamic constraints (Rizk and Liao, 2009). In this platform, any type of reaction mechanism able to describe enzyme saturation phenomenon can be employed to construct the model. For each reaction, if the functional form of the enzyme kinetic is already determined, it can be used to frame the model, otherwise, each reaction is decomposed into elementary reaction steps using mass action kinetics, which is able to conserve the mechanistic details of enzymatic reactions and capture saturation behavior and substrate-level regulation (Tan and Liao, 2012). The EM procedure starts with constructing an initial set of kinetic models all predicting the same phenotype (e.g., flux distribution) for the wild-type strain, but with different dynamic behaviors (Tan and Liao, 2012). Next, the impact of defined perturbations (Alper et al., 2005; Kitagawa et al., 2005) is simulated by all models, where each model may predict a different phenotype with a dynamic response. These newly obtained phenotypic data provide a basis for screening models against experimental data. Each perturbation and screening cycles train the model using a new set of experimental measurements and are repeated until a minimal set of kinetic models all of which can predict the wild-type and mutant phenotypes with a reasonable accuracy is identified. This approach has been successfully applied to study the production of lysine (Contador et al., 2009), fatty acid oxidation (Dean et al., 2010), aromatic production (Rizk and Liao, 2009) and modeling cancer cells (Khazaee et al., 2012). Furthermore, in an earlier effort, a systematic procedure was introduced for identifying gene/enzyme perturbations leading to a maximum reduction in the number of models retained in the ensemble after each round of perturbation (Zomorrodi et al., 2013).

Despite the availability of comprehensive in vivo metabolomics and fluxomic datasets for wild-type and mutant strains of well-studied microorganisms such as E. coli (Ishii et al., 2007) and Saccharomyces cerevisiae (Wisse et al., 2010; Blank et al., 2005), they have not been utilized in a systematic way for kinetic model reconstruction. This study aims to make use of these resources for the systematic parameterization of large-scale kinetic models of metabolism by combining efficient optimization-based approaches and the EM formalism. Using multiple omics (i.e., fluxomic) data available for E. coli (Ishii et al., 2007), we construct a kinetic model of the core metabolism of E. coli satisfying the in vivo flux measurements for wild-type and seven mutant strains under aerobic condition with glucose as the carbon source. This model contains 138 reactions and 93 metabolites and includes all the reactions used in the previously published kinetic models (Chassagnole et al., 2002; Kadir et al., 2010; Peskov et al., 2012). The model accounts for 60 substrate-level regulatory interactions extracted from online databases (Karp et al., 2000; Schomburg et al., 2013). An optimization-driven parameter identification method is proposed to elucidate the kinetic parameters in the form of elementary mechanisms for a given metabolic network. A leave-one-out cross-validation test was also performed to assess the flux prediction capability of the model. In addition, 35% and 77% of identified \( K_m \) and \( K_{ass} \) parameters, respectively, are consistent with the previously reported kinetic parameter values and also the predicted concentrations for 68% of metabolite concentrations are in good agreement with experimental measurements.

2. Materials and methods

2.1. Model scope and experimental data

A metabolic model composed of 138 reactions and 93 metabolites representing E. coli core metabolism (The Core E. coli model) was used to simplify the mathematical analysis. In general, however, the prediction quality deteriorates as fluxes move away from the reference state (Heijnjen, 2005; Liebermeister and Klipp, 2006). For example, approximate enzyme kinetic equations (Chakrabarti et al., 2013; Hatzimanikatis and Bailey, 1996; Hatzimanikatis et al., 1998; Hatzimanikatis et al., 1998; Heijnjen et al., 2005; Nielsen, 1997; Pozo et al., 2011; Savageau, 1972; Smallbone et al., 2013; Sorribas et al., 2007; Stanford et al., 2013; Visser and Heijnjen, 2003) or a combination of in vitro derived lumped and approximate rate equations (Drager et al., 2009; Costa et al., 2010) are frequently used to simplify the mathematical analysis. In general, however, the prediction quality deteriorates as fluxes move away from the reference state (Heijnjen, 2005; Liebermeister and Klipp, 2006).

Decomposable kinetic nonlinear formalisms have recently emerged as alternatives to lumped or approximate rate equations (Miskovic and Hatzimanikatis, 2011). These approaches allow for incorporating both metabolite and enzyme concentrations/activities and implementing any type of reaction including regulatory interactions. While these models hold promise to describe metabolism by combining efficient optimization-based approaches and the EM formalism. Using multiple omics (i.e., fluxomic) data available for E. coli (Ishii et al., 2007), we construct a kinetic model of the core metabolism of E. coli satisfying the in vivo flux measurements for wild-type and seven mutant strains under aerobic condition with glucose as the carbon source. This model contains 138 reactions and 93 metabolites and includes all the reactions used in the previously published kinetic models (Chassagnole et al., 2002; Kadir et al., 2010; Peskov et al., 2012). The model accounts for 60 substrate-level regulatory interactions extracted from online databases (Karp et al., 2000; Schomburg et al., 2013). An optimization-driven parameter identification method is proposed to elucidate the kinetic parameters in the form of elementary mechanisms for a given metabolic network. A leave-one-out cross-validation test was also performed to assess the flux prediction capability of the model. In addition, 35% and 77% of identified \( K_m \) and \( K_{ass} \) parameters, respectively, are consistent with the previously reported kinetic parameter values and also the predicted concentrations for 68% of metabolite concentrations are in good agreement with experimental measurements.

2. Materials and methods

2.1. Model scope and experimental data

A metabolic model composed of 138 reactions and 93 metabolites representing E. coli core metabolism (The Core E. coli model) was used to simplify the mathematical analysis. In general, however, the prediction quality deteriorates as fluxes move away from the reference state (Heijnjen, 2005; Liebermeister and Klipp, 2006). For example, approximate enzyme kinetic equations (Chakrabarti et al., 2013; Hatzimanikatis and Bailey, 1996; Hatzimanikatis et al., 1998; Hatzimanikatis et al., 1998; Heijnjen et al., 2005; Nielsen, 1997; Pozo et al., 2011; Savageau, 1972; Smallbone et al., 2013; Sorribas et al., 2007; Stanford et al., 2013; Visser and Heijnjen, 2003) or a combination of in vitro derived lumped and approximate rate equations (Drager et al., 2009; Costa et al., 2010) are frequently used to simplify the mathematical analysis. In general, however, the prediction quality deteriorates as fluxes move away from the reference state (Heijnjen, 2005; Liebermeister and Klipp, 2006).
coli Model) was constructed. A pictorial representation of the model is shown in Fig. 1.

This model spans reactions in glycolysis/gluconeogenesis, pentose phosphate (PP) pathway, TCA cycle, major pyruvate metabolism and anaplerotic reactions and a number of reactions in other parts of the metabolism, such as alternative carbon metabolism, nucleotide salvage and oxidative phosphorylation and glutamate metabolism (see Fig. 2).

Fluxomic data for wild-type and multiple mutant strains of E. coli (Ishii et al., 2007) were used to parameterize the model. In particular, metabolic flux data for 25 mutants growing under aerobic conditions with glucose as the carbon source were extracted (Ishii et al., 2007). Upon excluding mutant flux data corresponding to reactions absent in the model and aggregating results for isozymes (with the exception of pykA and pykF), eight sets of experimental measurements were assembled. These
included metabolic fluxes for the wild-type and seven different mutant strains ($\Delta$gnd, $\Delta$zwf, $\Delta$rpe, $\Delta$ppsA, $\Delta$pykA, $\Delta$pykF and $\Delta$pgi). Note that since the enzymatic activities of both pykA and pykF in wild-type E. coli are available in the literature (Karp et al., 2000; Al Zaid Siddiquee et al., 2004; Hoque et al., 2005) we were able to account for the mutant flux data of both of their mutant strains by using two separate reactions in our model. In total, we used flux measurements for approximately 25 reactions in each mutant strain, which include the major intercellular reactions in glycolysis, PP pathway and TCA cycle (Ishii et al., 2007), as shown in Fig. 1. Any flux measurement lower than one (per 100 mmol/gDW/h of glucose uptake) was assumed to be statistically insignificant and was excluded from the analysis. Standard Gibbs free energy estimates for all reactions in the model were obtained from the iAF1260 model of E. coli (Feist et al., 2007). In accordance with the EM procedure (Tran et al., 2008), all reactions in the model were decomposed into elementary steps and all kinetic parameters were scaled with respect to the metabolite and enzyme concentrations/activities for the reference (wild-type) strain. The initial flux distribution for the EM simulations was obtained as follows: The maximum biomass yield was obtained first using the flux balance analysis for the iAF1260 model (Feist et al., 2007) while imposing the experimental flux measurements (for 43 reactions) (Ishii et al., 2007) as constraints under aerobic minimal glucose conditions. Flux Variability Analysis (FVA) (Mahadevan and Schilling, 2003) was used next to identify the flux ranges for reactions without experimental measurements upon fixing the biomass flux at the value obtained in the first step. The obtained flux ranges were used to constrain the reactions without available measurements in our model. A feasible flux distribution was then obtained by imposing the experimental and FVA-driven flux ranges in our model. The intracellular reactions carrying a zero flux in the reference strain but a non-zero flux in at least one mutant are adjusted to carry a minimal amount of flux (i.e., equal to 0.05 mmol/gDW/h per 100 mmol/gDW/h of glucose uptake) in the reference strain. This ensures the participation of the reaction in the construction of the original EM ensemble. Examples of such reactions are those catalyzed by Edd and Eda which do not carry flux in the wild-type but become active in $\Delta$pgi (Ishii et al., 2007).

Sixty regulatory interactions for reactions of the central metabolism were extracted from BREnda (Schomburg et al., 2013) and ECOYC (Karp et al., 2000) databases and were included in the model following the EM procedure (see Supplementary file S4 for a complete list). These regulatory reactions introduced new elementary reactions and dead-end enzyme complexes in the model (Cnornish-Bowden, 2012). The regulatory reactions are not active in the reference strain and carry a zero net flux (Rizk and Liao, 2009).

2.2. Identification of kinetic parameter values satisfying experimental data for the wild-type and mutant strains

The original EM procedure relies on the construction of an initial ensemble of kinetic models by sampling reaction reversibilities and enzyme fractions followed by successive screening steps using the flux data for the perturbed strains to identify minimal sets of kinetic parameters satisfying all sets of flux data (Tran et al., 2008). A limitation associated with this type of screening is that it may result in an empty ensemble, i.e., all models in the current ensemble may be rejected after one or more rounds of model screening even if only a single reaction deviates from the experimental measurements. To avoid empty ensembles, we developed a new procedure combining the original sampling method of EM with a parameter identification step that minimizes the deviation of the model predictions from the available flux measurements for all mutant strains (see Fig. 3).

Formal model parameterization involves solving an optimization problem minimizing the deviation of model predictions (e.g., for reaction fluxes) from experimental measurements (Costa et al., 2010; Tohsato et al., 2013; Balsa-Canto et al., 2010). However, identification of a set of kinetic parameters satisfying multiple sets of experimental data is a challenging task due to the very large feasible space of kinetic parameter values as well as the high number of kinetic parameters that need to be estimated (Moles et al., 2003) (as each reaction in the original model is decomposed into multiple elementary reaction steps). In order to reduce the search space, we decided to discretize the feasible space of the kinetic parameter values by taking advantage of the EM formalism. To this end, an initial ensemble of models is constructed by uniformly sampling the enzyme fractions and reaction reversibilities within the identified feasible ranges (see panels 2.1 and 2.2 of Fig. 3). Choosing the size of initial ensemble is a compromise between parameter space coverage and memory requirements per run. In this study, we used an ensemble size of $2^{17}=131,072$, as no improvement in the convergence speed and correlation of model predictions with experimental measurements was achieved for a larger ensemble size of $2^{18}$. An ensemble size of $2^{17}$ provides an approximate resolution of $7 \times 10^{-16}$ for each sampled parameter.

Despite this discretization, we found that solving the parameter estimation problem still remains challenging when considering multiple sets of experimental flux data simultaneously. In an effort to overcome this challenge, we developed an integrative procedure where kinetic parameter identification is carried out by successively adding one mutant data set at a time. The optimal kinetic parameter values obtained upon iteratively solving a two-step optimization problem are used to initialize the next iteration where a new set of flux data is included. This procedure is repeated until all mutant strain data are considered (see Fig. 4).
Model parameters include both discrete (EM decomposed reaction parameters) and continuous (elementary reactions for regulatory interactions and metabolic reactions carrying a zero net flux in the reference strain). We first optimize over the set of discrete parameter values followed by minimization of the model’s predicted deviation with respect to the continuous parameters.

2.3 Identification of the kinetic rate constants for metabolic reactions carrying a nonzero flux in the reference and mutant strains

In this optimization problem, the optimal combination of the sampled parameters in the ensemble is identified by minimizing the average relative deviation between experimentally measured and predicted values. Ultimately, a set of kinetic model that is tested and validated along different fluxomics is identified.

Fig. 3. A schematic representation of the kinetic model construction procedure. (1) First, a steady state flux distribution is obtained by imposing the available fluxomic data and refining the flux ranges (see Materials and methods section). (2.1) EM procedure is used to decompose each reaction into elementary mechanistic reaction steps ($\tilde{v}_j^{\text{f}}$ and $\tilde{v}_j^{\text{r}}$ represent the forward and reverse flux of the elementary reaction $j$ in elementary step $l$, respectively). Thermodynamic constraints are employed to confine the ranges of reactions reversibilities. In the absence of experimental measurements for metabolite concentration, they are normalized to their steady-state values (see Supplementary text S1). Conservation of mass is written for the total available enzyme in the system for each enzyme. Elementary kinetic parameters are then expressed as a function of reaction reversibilities and enzyme fractions. (2.2) An ensemble of elementary kinetic parameters of size $P$ is constructed by uniformly sampling reaction reversibilities and enzyme fractions. (2.3) For a given set of kinetic parameters the system of ODEs representing the conservation of mass for each metabolite and enzyme fraction is integrated until reaching a steady-state. (3) In order to improve model fitness, the optimizer provides a new set of model parameters (see Section 2 Materials and methods section). Ultimately, a set of kinetic model that is tested and validated along different fluxomics is identified. (4) The model predictions are validated by a comparison between the available metabolomics, kinetic constants and cross-validations.

2.3. Identification of the kinetic rate constants for metabolic reactions carrying a nonzero flux in the reference and mutant strains

In this optimization problem, the optimal combination of the sampled parameters in the ensemble is identified by minimizing the average relative deviation between experimentally measured and predicted values.
determined fluxes and those predicted by the model. The terms in the objective function are scaled by using the coefficient of variation (Becskei and Serrano, 2000; Isaacs et al., 2003) to capture the uncertainty in the experimental flux measurements. As a result, the reactions with tighter confidence interval have a larger contribution in the objective function. This procedure maintains convergence of all new models, generated by combining kinetic parameter values, to the steady-state of the reference strain. The rules that need to be followed in combining parameter values are described in Supplementary text S1. In essence, these rules require that for a given reaction, elementary kinetic parameters from the same model in the ensemble need to be chosen simultaneously (see Supplementary text S1 for more detail).

2.4. Identification of the kinetic rate constants for metabolic reactions carrying a zero flux in the reference and those representing regulatory interactions

Reactions that do not carry any flux for the wild-type may carry a nonzero flux in a perturbed strain. The kinetic parameters for these reactions are not constrained in the initial ensemble. The same holds true for regulatory reactions. The EM procedure assigns arbitrary values to both forward and reverse elementary reaction steps so as the net flux becomes zero. Therefore, in the second optimization problem the rate of decomposed forward (or reverse) reaction with zero net flux is treated as continuous non-negative variables whereas the model parameters (i.e., elementary kinetic parameters) for reactions with non-zero fluxes are retained at the values obtained from the first optimization problem (see Supplementary text S1 for more details).

2.5. GA implementation of the discrete and continuous optimization problems

Both of the discrete and continuous optimization problems described above are solved using a genetic algorithm (GA) implementation in MATLAB (MathWorks Inc.). The GA representation of each one of these two problems is shown pictorially in Fig. 5 and is detailed as follows:

Genotype and phenotype of individuals: Each kinetic model serves as a chromosome in the GA representation. Each gene in the chromosome for the discrete optimization problem corresponds to overall (un-decomposed) reaction and can take an integer value between 1 and $P$, where $P$ is the total number of models in the initial ensemble. Alternatively, every gene in the chromosome in the continuous optimization problem corresponds to an elementary reaction step and can take any value between zero and an upper bound on the flux value of the forward (or reverse) reaction (see Fig. 5 and Supplementary text S1 for more detail). The elementary flux values are then used to obtain the corresponding elementary kinetic parameters (Tran et al., 2008).

Population size: In all simulations the population size was set to three to five times the number of variables (i.e., genes) in the chromosomes.

Initialization of population: Only one chromosome was initialized with the solution obtained in the previous iteration and the rest were initialized randomly (using the built-in GA function in MATLAB).

Crossover operator: We examined algorithm performance with different crossover fractions to find an optimal one. We found that a crossover fraction between 0.7–0.8 for early generations provides and a crossover fraction between 0.8–0.9 for later generations provides the best performance.

Termination criterion: The GA procedure is terminated when the fitness of the elite chromosome is considered sufficiently high (i.e., as good as the elite chromosome before addition of new flux dataset) or if no improvement is observed in the elite chromosome after 20 generations. We also set a maximum number of generations (i.e., 100) for each optimization problem after which the procedure was terminated.

2.6. Comparison of model predictions against experimental data

Since confidence intervals were provided only for the wild-type flux and metabolite concentration measurements in the data set (Ishii et al., 2007), we constructed a confidence range for each case, separately. For the reported fluxes, the same confidence intervals reported for each reaction in the wild-type strain were used to construct a flux range around the reported values (i.e., one standard deviation confidence interval) in the mutant strains. For metabolite concentrations, the reported confidence intervals for metabolite concentrations in the wild-type strain were too wide (coefficient of variation is more than 20% for more than 85% of the reported metabolite concentrations) that would have encompassed all predicted values by the model. Therefore, we created a more conservative range by choosing 10% deviation from the
nominal (mean) values to represent experimental ranges. For the lumped kinetic parameters ($K_m$ and $k_{cat}$), given that multiple values are reported in the literature and databases such as BRENDA (Schomburg et al., 2013), we used one standard deviation from the mean of the reported values as experimental ranges (see Supplementary file S4 for more detail).

3. Results

3.1. Elementary kinetic parameter estimation

An initial ensemble of kinetic models all satisfying the reference (wild-type) flux data was first created. Starting with the Δgnd mutant data, we identified the best combination of kinetic parameter values in the initial ensemble (see Section 2) satisfying experimental flux data for this mutant. This procedure is repeated successively by adding one more mutant at a time (i.e., ΔppsA, ΔpykA, Δzwf, Δpgi, ΔpykF and Δrpe, respectively). The kinetic parameter values obtained in each step are used as an initial point for solving the optimization problems in the next step. By using this procedure we were able to identify kinetic parameter values simultaneously satisfying the experimental flux measurements for the reference (wild-type) and the seven mutant strains mentioned above. The model is available for download at http://maranas.che.psu.edu/models.htm as a mat-file (MathWorks Inc.).

Fig. 6 shows the deviation of the predicted fluxes of selected reactions by the final model from the measured fluxes. As shown in this figure the model parameterization procedure performs well in fitting the predicted fluxes for 78% of the reactions (with measured fluxes) as they fall within the reported experimental ranges (see Section 2.6 of Methods and materials). For the rest of the reactions, the model prediction is almost always within three standard deviations of the experimental ranges.

3.2. Model cross-validation

In order to avoid over-fitting the data to the model and to evaluate the accuracy of the model in predicting the cellular phenotype in response to new perturbations, we conducted a leave-one-out cross-validation test (Jesse Russell, 2012). In each validation, the flux data for one mutant strain was excluded from the training set and the resulting model was used to predict the fluxes for the excluded mutant strain (see Fig. 7). This analysis showed that excluding one at a time the flux data for Δgnd, ΔppsA, ΔpykA, Δzwf, ΔpykF and Δrpe in the cross-validation tests yields an average reduction in prediction quality by approximately 15% across all tests. We note that these six gene deletion mutants are located in the vicinity of the pay-off phase of glycolysis (i.e., $gnd$, $zwf$, and $rpe$) or in the PP pathway (i.e., $gnd$, $zwf$ and $rpe$). This means that in the absence of one mutant flux data in a cross-validation test, the model can still be trained robustly using flux data for its adjacent mutants. This is not, however, the case for Δpgi whose absence from the training set cannot be compensated by flux data from other mutants as alluded to by the much higher deviation from experimental data.
Fig. 6. Deviation of the predicted steady-state flux of some selected reactions by the constructed model from available experimental measurements in (a) Δgnd, (b) ΔppsA, (c) ΔpykA, (d) ΔpykF, (e) Δpgi, (f) Δzwf and (g) Δrpe strains. The blue circles represent the experimental measurements and the green diamonds the predicted flux distributions. The error bars denote one standard deviation confidence interval for the corresponding reaction in the wild-type strain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
(i.e., 50% reduction in prediction quality) upon exclusion. In particular, the flux of reactions in glycolysis and Entner–Doudoroff pathway are poorly predicted in the absence of flux data for the Δpgi strain. For example, the reaction catalyzed by Eda (6-PG→G3P+PYR) does not carry flux in any of the mutant strains except for Δpgi as it provides G3P (Glyceraldehyde 3-phosphate) and PYR intermediates which cannot be produced through glycolysis in the absence of pgi. Therefore, upon omitting flux data for Δpgi, the model cannot anticipate the activity of the Entner–Doudoroff pathway. Another example involves the reaction catalyzed by Fba (FDP→DHAP+G3P) whose flux is reduced significantly (by 40% on average compared to wild-type and other mutant strains) in the Δpgi strain. This significant reduction in flux of this reaction cannot be captured when excluding Δpgi from the training flux data set. Nonetheless, even though these discrepancies propagate to some extent in other part of the network, the model prediction is still acceptable for the remaining reactions (i.e., on average 12% deviation from the experimental ranges).

In addition, to flux comparisons, we also examined the estimated kinetic parameters in each cross-validation test. This showed that the coefficient of variation is less than 20% for more than 90% of the kinetic constants implying that the model parameterization was generally robust to the omissions of individual flux data sets with the exception of examples noted earlier.

3.3. Evaluation of the model performance in predicting metabolite concentrations

Taking advantage of the availability of experimental measurements for metabolite concentrations for both wild-type and mutant strains (Ishii et al., 2007), we also assessed that the predictive power of the constructed kinetic model for metabolite

---

**Fig. 7.** Cross-validation analysis. The gray bars represent the average scaled deviation of the predicted steady-state fluxes for different mutant strains upon their exclusion from the training set. The white bars correspond to the average scaled deviation of the predicted steady-state flux distribution from the experimental measurements while including all mutant datasets. The difference between two bars represents the reduction in flux predictive power upon excluding all flux data for the one mutant.

**Fig. 8.** (Colour online) (a) Prediction performance of the model for metabolic concentrations. (b) Comparison of the EM-based computed Michaelis–Menten constant and (c) turnover number with the values reported in literature or databases (Schomburg et al., 2013). Dashed lines denote one order of magnitude departures from the experimental values.
concentrations, which were used for model training (see Fig. 8a).
First, the normalized metabolite concentrations in the EM procedure were converted into actual concentrations for metabolites with available experimental data in the reference strain (i.e., wild type). For metabolites with no concentration data (approximately 50%) we used the ranges used in the iAF1260 model of E. coli (Feist et al., 2007). This analysis shows that 68% of the predicted metabolite concentrations are consistent with the experimental measurements implying that there is an overlap between the predicted ranges and the experimentally reported/calculated ranges (see Section 2.6 of Methods and materials).

3.4. Evaluation of the estimated kinetic parameter values

One can derive the corresponding kinetic parameters of an overall reaction, given the kinetic parameter values of its corresponding elementary reaction steps assuming a pseudo steady-state assumption (Edda Klipp et al., 2009; Cornish-Bowden, 2012) (see Supplementary text S2 for more detail). This allows comparing known kinetic parameter values for Michaelis constants and turnover numbers (i.e., $K_m$ and $k_{cat}$, respectively) reported in the literature and databases such as BRENDA (Schomburg et al., 2013) with values estimated in this study. First, the scaled $K_m$ and $k_{cat}$ values in the EM procedure were converted into actual ranges using the metabolite concentration ranges in the reference strain (i.e., wild type). A comparison between the experimentally measured $K_m$ and $k_{cat}$ values and the corresponding EM-based parameters using the constructed kinetic model is shown in Fig. 8b and c. Similar to concentration comparisons, an overlap between the predicted and experimental ranges is denoted as consistency (see Section 2.6 of Methods and materials). As shown in these figures 35% and 77% of the estimated $K_m$ and $k_{cat}$ values are consistent with experimental measurements, respectively (see Supplementary text S3 for a statistical assessment of the significance of the reported agreements). It should be noted that due to the existence of a large variation in the reported values of $k_{cat}$ in the literature (see Supplementary file S4), most $k_{cat}$ values estimated in this study fall in the reported/calculated confidence ranges, as shown in Fig. 8c. This implies that consistency of $k_{cat}$ values may not form a solid basis for testing the accuracy of the identified model parameterization.

3.5. Sensitivity analysis

Parameter sensitivity is assessed by quantifying the effect of kinetic parameter value perturbations on the prediction quality of the model. Lack of sensitivity implies non-identifiability of the corresponding parameter by the available flux data (Tohsato et al., 2013). Each elementary kinetic parameter individually scaled up and down by a total of 10-fold. Any error of less than 5% from the original flux predictions by the model was assumed to be negligible. The error was computed as the average deviation of the predicted fluxes from those of the constructed model across all mutants. This analysis indicates that for 33% of the elementary

Fig. 9. Results of the sensitivity analysis. Upper and lower limits of the elementary kinetic parameters with the prediction error of less than 5% from the original flux predictions for reactions in (a) glycolysis, (b) TCA cycle and (c) PP pathway. For each reaction the most sensitive parameter is illustrated.
kinetic parameters the model predictions are sensitive to perturbations implying that it is not possible to change the kinetic parameter value and still retain a prediction error of less than 5%. In particular, 44% of the elementary kinetic parameters of glycolysis, 29% of TCA cycle and 33% of PP pathway reactions showed sensitivity to perturbations. The highest sensitivity was observed for elementary kinetic parameters of the reactions catalyzed by Gapd, Fba and Pts which were quoted among the most influential parameters in previous reports (Costa et al., 2010; Di Maggio JCDR and Diaz, 2009). Fig. 9 shows the ranges of the elementary kinetic parameters with prediction error of less than 5% from the original flux predictions, for the most sensitive parameter of a given reaction. In general, model parameters of reactions that are located at the entry point of sub-pathways (i.e., glycolysis, TCA cycle and PP pathway) were more sensitive to perturbation as detailed before (Tohsato et al., 2013). As expected, due to the global participation of cofactors in the network, the model was quite sensitive to perturbations in kinetic parameters of reactions catalyzed by transhydrogenase (THDH2 and NADTRHD), dehydrogenase (NADH16pp), adenylate kinase (ADK1) and ATP synthase (ATPS4rrp) enzymes. Given that elementary kinetic parameters were perturbed one at a time, any conclusion regarding indefinability of the remaining 67% of the parameters whose perturbations results in a prediction error of more than 5% would require accounting for potential correlations among different kinetic parameters. To gain some insight into the impact of correlations among model parameters, we performed sensitivity analysis for a number of selected pairwise combinations of parameters. This analysis revealed that model predictions are indeed sensitive to the combined perturbation of these parameters even though they were insensitive to their individual perturbations. For example, flux predictions by the model were found to be highly sensitive (i.e., error of model predictions was more than 100% from the original flux predictions by the model) to pairwise perturbations of the elementary kinetic parameters of the reaction catalyzed by MET with those of reactions catalyzed by Eda, Edd, ACONIIs and Fbp, while they were insensitive to their individual perturbations.

4. Summary and discussion

In this study, we developed a kinetic model of \( E. coli \) core metabolism by integrating the EM formalism (Tran et al., 2008) with efficient GA-based techniques and making use of the available flux data for wild-type and seven mutant strains (Ishii et al., 2007; Bennett et al., 2009). This model contains 93 metabolites, 138 reactions and 60 substrate-level regulatory interactions. The parameter estimation procedure proposed in this study was successful in fitting the flux predictions by the model to experimental measurements for 78% of the reactions while the remaining predicted reaction fluxes are within three standard deviations of measured ranges. A possible reason for not being able to fit all reaction fluxes is the presence of tight bounds on concentrations. It is the tight coupling of reaction fluxes through concentrations irrespective of kinetic parameter values that disallows matching all experimentally reported flux values. The limited scope of the current model, as it accounts for only 138 out of 2383 reactions in iAF1260, implies that the effect of interactions with absent reactions and metabolites could not be captured.

Cross-validation tests and sensitivity analysis revealed that the constructed model can be used with a high confidence to predict most of reaction fluxes and metabolite concentrations involved in central metabolism. Existing kinetic models do not typically perform well in predicting the phenotype of genetically perturbed strains as parameterization is carried out for a single flux data set. The presented model on the other hand was constructed by simultaneously accounting for multiple sets of flux data for the wild-type and several mutant strains. The presented model construction pipeline also provides a systematic approach for future improvements as more flux data are becoming available. This study paves the way for the reconstruction of a genome-scale kinetic model for \( E. coli \) and other organisms. This, however, calls for the availability of additional experimental flux measurements for mutant strains carrying mutations in diverse parts of the metabolism in order to support the robust parameterization of the larger-scale kinetic model.

Authors’ contributions

CDM, JCL conceived the study. CDM and ARZ coordinated the study, participated in its design and data analysis and helped to draft the manuscript. AK designed and performed the simulations and data analysis and drafted the manuscript. All authors have read and confirmed the manuscript.

Acknowledgment

This work was supported by the United States Department of Energy (DOE) Grant DE-ER65254.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2014.05.014.

References


