

Systems biology

Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*

Markus W. Covert^{1,*}, Nan Xiao¹, Tiffany J. Chen² and Jonathan R. Karr¹¹Department of Bioengineering, Stanford University, 318 Campus Drive, Stanford, CA 94305-5444 and ²Program in Biomedical Informatics, 251 Campus Drive, Stanford, CA 94305-5479, USA

Received on April 14, 2008; revised on June 17, 2008; accepted on July 8, 2008

Advance Access publication July 10, 2008

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: The effort to build a whole-cell model requires the development of new modeling approaches, and in particular, the integration of models for different types of processes, each of which may be best described using different representation. Flux-balance analysis (FBA) has been useful for large-scale analysis of metabolic networks, and methods have been developed to incorporate transcriptional regulation (regulatory FBA, or rFBA). Of current interest is the integration of these approaches with detailed models based on ordinary differential equations (ODEs).

Results: We developed an approach to modeling the dynamic behavior of metabolic, regulatory and signaling networks by combining FBA with regulatory Boolean logic, and ordinary differential equations. We use this approach (called integrated FBA, or iFBA) to create an integrated model of *Escherichia coli* which combines a flux-balance-based, central carbon metabolic and transcriptional regulatory model with an ODE-based, detailed model of carbohydrate uptake control. We compare the predicted *Escherichia coli* wild-type and single gene perturbation phenotypes for diauxic growth on glucose/lactose and glucose/glucose-6-phosphate with that of the individual models. We find that iFBA encapsulates the dynamics of three internal metabolites and three transporters inadequately predicted by rFBA. Furthermore, we find that iFBA predicts different and more accurate phenotypes than the ODE model for 85 of 334 single gene perturbation simulations, as well for the wild-type simulations. We conclude that iFBA is a significant improvement over the individual rFBA and ODE modeling paradigms.

Availability: All MATLAB files used in this study are available at <http://www.simtk.org/home/ifba/>.

Contact: covert@stanford.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Can we build a model that accounts for all of the gene products in a cell? Certainly the effort to build a whole-cell model will depend on the development of new modeling approaches, and in particular, the integration of models for different types of processes, each of which may be best described using different representation. Moreover, such

efforts will likely identify novel and important cross-talk between different networks.

One approach that has been particularly successful in enabling large-scale modeling of carbon and energy metabolism is called flux-balance analysis (FBA). FBA has been used to model metabolism in a host of microbial species, and has been expanded for a variety of applications (reviewed in Price *et al.*, 2004). Two extensions of FBA of interest here are the use of multiple FBA steps to simulate growth dynamics (Luo *et al.*, 2006; Mahadevan *et al.*, 2002; Varma and Palsson, 1994), and the incorporation of transcriptional regulatory network models (Covert *et al.*, 2001; Shlomi *et al.*, 2005, 2007). These expansions enabled us to integrate a regulatory network including 104 regulatory proteins with an existing model of 906 gene products involved in *Escherichia coli* metabolism (Covert *et al.*, 2004). We found that this integrated model (called regulatory FBA, or rFBA), significantly increased our ability to predict knockout strain phenotypes in a variety of environmental conditions (10 800 correct predictions out of 13 750 cases total, Covert *et al.*, 2004). We also demonstrated the power of a model-driven approach to discovery, identifying over 100 putative components and interactions in the *E.coli* metabolic and regulatory networks. Several of these have recently been verified experimentally (Reed *et al.*, 2006).

A major advantage of rFBA—requiring few kinetic parameters—could be a weakness in situations where the kinetic parameters have been determined and capture information not contained in rFBA. For example, *E.coli* catabolite repression and its consequences on glycolysis have been modeled in great kinetic detail (Bettenbrock *et al.*, 2006). It, therefore, seemed useful to create a framework which combines rFBA's ability to capture not only the metabolic pathways, but also the transcriptional regulation of an entire system, with the kinetic model's greater level of detail. Other groups have integrated FBA with additional kinetic information, such as lin/log kinetics (Smallbone *et al.*, 2007) and coarse-grain time-scale information (Lee *et al.*, 2008). Furthermore, Yugi and colleagues (2005) showed that integrating metabolic flux analysis with more detailed kinetic descriptions reduces the amount of training data required to add additional reactions and metabolites to dynamic models.

Here, we report the development of the integrated FBA (iFBA) framework, and the application of this framework to combining existing rFBA and kinetic models of *E.coli* central metabolism. Beyond the application to *E.coli*, our approach differs significantly

*To whom correspondence should be addressed.

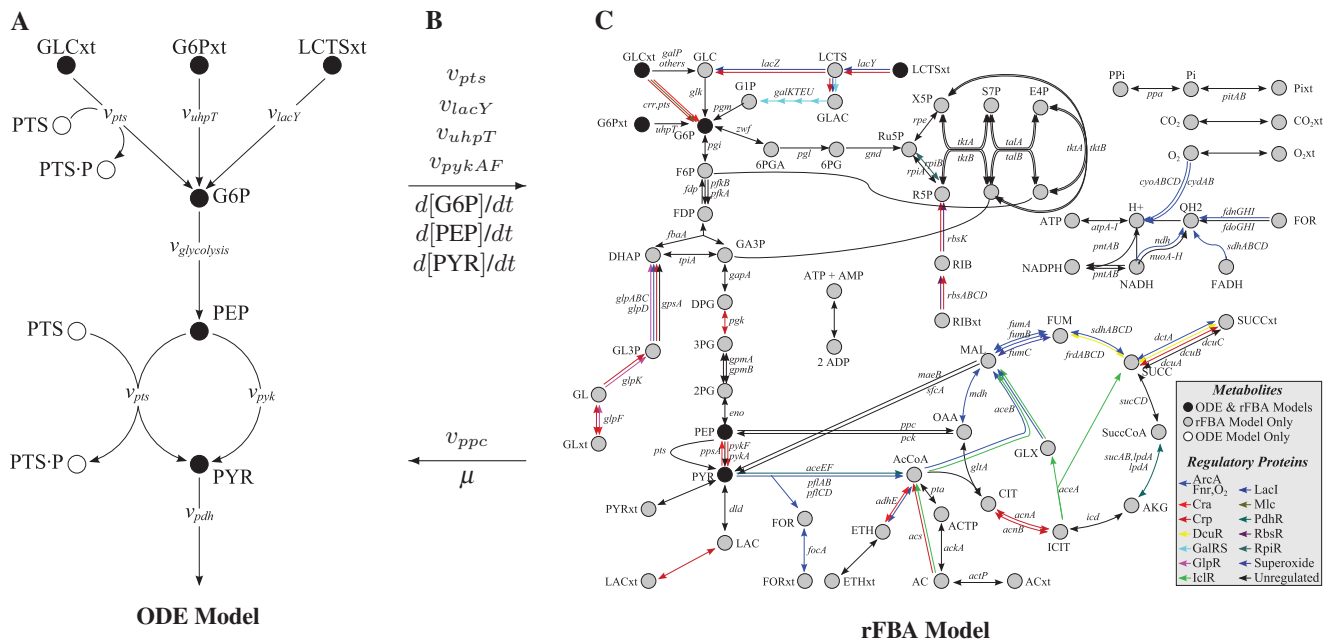


Fig. 1. Diagram of the metabolic, regulatory and signaling networks used to build the iFBA model. (A) A schematic of the complete ODE model (Kremling *et al.*, 2007). (B) A list of the variables passed between the models as part of iFBA. Arrows indicate whether a value is being passed from the ODE model to rFBA or vice versa, and metabolites common to both the rFBA and ODE models are colored black. (C) A schematic of the complete rFBA model (Covert and Palsson, 2002). Regulated fluxes are indicated by the key (lower right).

from the studies listed earlier in that we (1) integrate an FBA metabolic network with a Boolean transcriptional regulatory network as well as with a set of ODEs, and (2) incorporated two independently created models of the same system and integrate them with minimal changes to either model. We see the resulting framework as an essential stepping-stone to development of a whole-cell model, enabling the integration of a wide variety of models of cellular processes.

We compared wild-type and single gene perturbation diauxic growth (glucose/lactose and glucose/glucose-6-phosphate) time courses predicted by each of the integrated and individual models with experimental data (Bettenbrock *et al.*, 2006). We find that the integrated model is a significant improvement over the individual rFBA and ODE-based models, generating simulations which are more globally accurate and informative than the ODE-based model, and more accurate in their details than the rFBA model alone.

2 METHODS

To create the integrated model, we combined a kinetic model of *E. coli* phosphotransferase (PTS) catabolite repression developed by Kremling and colleagues (2007) (Fig. 1A), with an rFBA model of the same system (Fig. 1C, Covert and Palsson, 2002). The rFBA model was expanded from that described in Covert and Palsson (2002) by the inclusion of two fluxes to describe glucose-6-phosphate uptake by UhpT. The modified rFBA model describes the uptake and production of 11 carbohydrates, glycolysis, the pentose phosphate pathway, the TCA cycle and the production of intermediate energy stores using biomass, 77 metabolites, 87 enzymes and 16 transcription factors that regulate 46 out of 113 metabolic reactions. The ODE model describes the regulated uptake of 3 carbohydrates and their subsequent metabolism via the glycolytic pathway to produce biomass. This model includes 6 metabolites, 4 proteins (3 transporters and PTS system component

EIIA^{Crr}), as well as 16 metabolic and transport reactions. The integrated model describes biomass, 77 metabolites, 151 genes and 113 reactions.

As illustrated in Figure 1B, we integrated the rFBA and ODE models by identifying values to pass from either model to the other. First, we identified the complete set of metabolites and fluxes common to both models (common metabolites shown as black circles, all ODE reactions are in this case, although not necessarily in every case, common to both models). The variables passed from the ODE model included fluxes which were not directly subject to global effects. These included enzyme fluxes v_{pts} , v_{lacY} , v_{uhpT} and v_{pykAF} , as well as changes in metabolite concentrations, which we call ‘metabolite pooling fluxes’ $d[G6P]/dt$, $d[PEP]/dt$ and $d[PYR]/dt$. The variables passed from the rFBA model include the growth flux (μ) and the flux through phosphoenolpyruvate carboxylase (v_{ppc}). The growth flux was passed from rFBA because the ODE calculation of growth depends only on substrate uptake and neglects important global growth requirements. v_{ppc} was determined by rFBA because it was not included in the ODE model but can have an important effect on phosphoenolpyruvate concentration.

2.1 iFBA simulation algorithm

The following paragraphs describe each step of the iFBA simulation algorithm illustrated in Figure 2. Briefly, starting from initial conditions or those calculated in the previous time step, we first numerically integrated the ODE model and computed the regulatory constraints using the Boolean regulatory model. Next, we constrained the primal of the FBA linear programming problem using the ODE and regulatory models, updated the right-hand-side of the FBA linear programming problem according to the pooling fluxes calculated by the ODE model, and solved for the FBA fluxes. Finally, we updated the biomass and external metabolite concentrations for use in subsequent time steps. The length of each time step was chosen to be large enough that the FBA assumption that the concentrations of internal metabolites are time-invariant holds, and yet small enough for the ODE model to calculate the system dynamics without accumulating numerical

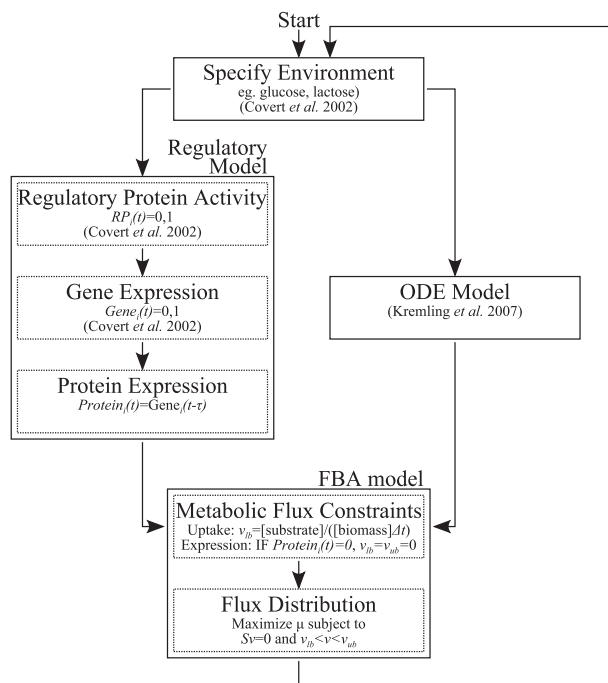


Fig. 2. Schematic of the simulation algorithm.

error. Although we used a time step of 3 min, we empirically found at any time step in the range 30 s to 5 min gave the same results.

Specify initial environment—Initial conditions for the biomass, enzymes and metabolites in the ODE model and the corresponding rFBA biomass and metabolites, where applicable, were obtained from Kremling *et al.* (2007), and are listed in Table S1. Initial conditions for the 16 regulatory proteins were determined by the regulatory model under the additional assumption that the bacteria were in steady state with the external environment prior to the start of the simulation.

Calculate regulatory protein activity, gene and protein expression—Transcriptional regulation imposes time-dependent constraints on the metabolic network. The activity of each regulatory protein, as well as expression of regulated genes and proteins was described using the Covert *et al.* (2001) Boolean regulatory model with time delay, except that in cases where activity and expression were encapsulated by the ODE model—Crp, galEKMPT, lacYZ, pgk, ptsG, and pykF—the ODE model-determined values superseded the Boolean regulatory values.

Solve ODEs—At each time step we used the MATLAB ode15s function to numerically integrate the ODE model using the growth rate and *ppc* flux computed by the FBA model at the previous time step. Next, we calculated the ODE rates at the end of the time step to later constrain the FBA linear programming problem.

Determine metabolic flux constraints and metabolite pooling fluxes—There are several types of metabolic flux constraints in iFBA: (1) irreversibility constraints, where the lower bound of the reaction is set to zero for reactions which can only proceed in the forward direction; (2) environmental constraints, where the maximum flux through an exchange reaction is limited by the amount of substrate in the culture medium; (3) transport constraints, which are represented as a maximum substrate uptake or by-product secretion rate; (4) regulatory constraints, where the flux through an enzyme is restricted by the expression of the corresponding protein(s); and (5) ODE matching constraints, where fluxes passed by the ODE model are completely specified by the ODE model. Irreversibility constraints are determined from the literature (Covert and Palsson, 2002). Environmental constraints on the exchange fluxes, v_{ex} , were computed

according to the scheme described by Varma and Palsson (1994). These constraints are then compared to the transport constraints (listed in Table S1), and the more restrictive constraints were used to bound the exchange reaction for the given time step. Regulatory constraints were derived from the expression profile of regulated proteins in the metabolic network over time. If the Boolean rule indicated at some time t that protein i is expressed, then the corresponding reaction was not constrained and the metabolic flux distribution was allowed to calculate any value for that reaction, given the other non-regulatory constraints that also control the system. However, if the rule indicated that protein i is not expressed at time t , then the corresponding reaction flux was constrained to zero,

$$v_i(t) = 0. \quad (1)$$

Finally, ODE matching constraints included any flux represented in both the ODE and FBA models— v_{pts} , v_{lacY} , v_{uhpT} and v_{pykAF} , and were implemented by setting the upper and lower bounds of the FBA fluxes equal to the corresponding rate calculated by the ODE model.

To capture the internal metabolite concentrations in iFBA, we incorporated metabolite pooling fluxes. The normal FBA mass balance equations assume that the concentrations of internal metabolites are time-invariant. However, the ODE model calculates time-variant metabolite concentration profiles, in our case for glucose-6-phosphate, phosphoenolpyruvate and pyruvate. Metabolite pooling fluxes were implemented by setting the corresponding entries in the right-hand side of the FBA linear programming problem equal to the rates of change of their concentrations calculated by the ODE model— $d[G6P]/dt$, $d[PEP]/dt$ and $d[PYR]/dt$.

Calculate flux distribution—Fluxes were calculated by maximizing biomass production subject to the FBA mass balance equations using the open-source COIN-OR Linear Program Solver (CLP, freely available at <http://www.coin-or.org/>). The biomass mass balance equation was based on experimental data (Ingraham *et al.*, 1983).

Calculate new environment—The growth rate and fluxes computed by the FBA model were next used to update the biomass and metabolite concentrations according to the scheme described by Varma and Palsson (1994).

$$[\text{biomass}](t + \Delta t) = \beta[\text{biomass}](t)e^{\mu\Delta t} \quad (2)$$

$$[\text{met}_i](t + \Delta t) = [\text{met}_i] + \frac{v_{ex}}{\mu}[\text{biomass}](t)(1 - e^{\mu\Delta t}), \quad (3)$$

where β is a growth rate scaling factor introduced to fit the experimental data obtained by Bettenbrock *et al.* (2006) for *E.coli* diauxic growth on glucose/glucose-6-phosphate and glucose/lactose with the biomass equation experimentally determined by Ingraham *et al.* (1983) for *E.coli* B/r growth on glucose minimal medium.

At the following time step the growth rate and flux through Ppc are used to correct the ODE phosphoenolpyruvate pooling flux to account for conversion to oxaloacetate by Ppc, and to calculate ODE rates and states.

2.2 Single gene perturbations

Single gene knockouts were implemented by setting the upper and lower bounds of the corresponding FBA flux(es) to zero, setting the values of the corresponding ODE kinetic parameter(s) to zero, and setting the expression of the corresponding transcription factor(s) to zero. For regulatory proteins, we also simulated knock-in of a constitutively active transcription factor by setting the activity of the corresponding transcription factor(s) to one. The correspondences between ODE kinetic parameters and rFBA relationships are listed in Table S2.

3 RESULTS

We evaluated the integrated model by comparing the model's predictions for wild-type and single gene perturbation *E.coli* diauxic growth on glucose/lactose and glucose/glucose-6-phosphate with

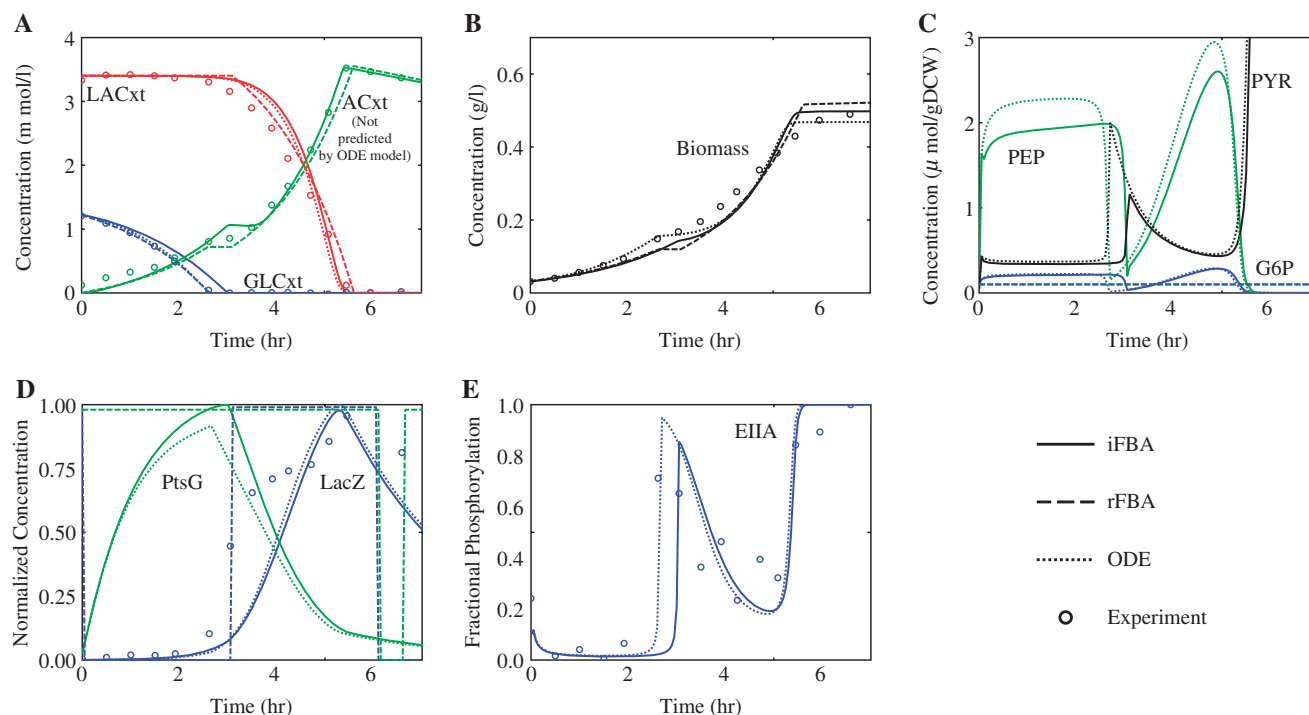


Fig. 3. Growth of the iFBA (solid lines), ODE (dotted) and rFBA (dashed) wild-type models in an aerobic environment with glucose and lactose as carbon sources, together with experimental data (Kremling *et al.*, 2007) where available (circles). Dynamic time profiles of external (A) acetate, glucose, lactose and (B) biomass concentrations; (C) internal pyruvate (PYR), phosphoenolpyruvate (PEP) and glucose-6-phosphate (G6P) concentrations; (D) key protein concentrations; and (E) degree of phosphorylation of regulatory protein EIIA^{Crr}.

that of the individual rFBA and ODE models, and where available, experimental data (Bettenbrock *et al.*, 2006, 2007).

3.1 Diauxic growth on glucose/lactose

Growth on glucose and lactose as carbon sources involves catabolite repression, leading to the preferential uptake of glucose and subsequent lactose uptake. Figure 3 shows the iFBA, rFBA and ODE wild-type simulations together with experimental data. We found that although all three types of simulations were equally able to predict carbon source uptake and biomass production (Fig. 3A, B), they were significantly different in other aspects. For example, acetate secretion was observed by Bettenbrock *et al.* (2006) under these environmental conditions. iFBA and rFBA were both able to account for acetate secretion under the given environmental conditions, but not the ODE model.

Additionally, the internal concentrations of phosphoenolpyruvate and pyruvate differed significantly between all iFBA and the individual models. Because the rFBA model assumes that the concentrations of internal metabolites are time-invariant, it did not encapsulate any of the dynamics seen in the ODE and iFBA simulations (Fig. 3C). The differences between the iFBA and ODE models in predicting internal concentrations reflected two fluxes which are included in the iFBA model but not considered in the ODE model. First, the iFBA growth flux is a drain on several key metabolites, including glucose-6-phosphate, glyceraldehyde-3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate and pyruvate in glycolysis. Loss of these metabolites to biomass results in a small but significant reduction in phosphoenolpyruvate conversion to pyruvate. Second, the Kremling model assumes

that pyruvate kinase and Pts are the dominant enzymes which utilize phosphoenolpyruvate, and that the phosphoenolpyruvate carboxylase flux is negligible (Kremling *et al.*, 2007). However, the iFBA model predicts that the Ppc flux is 5–15% of the total flux utilizing phosphoenolpyruvate (Figs 4, S2).

Modeling the transporter UhpT and PtsG concentrations led to similar results for the ODE and iFBA model, while the rFBA simulations exhibited step-like dynamics, due to the underlying Boolean rules (Fig. 3C). Similarly, rFBA did not encapsulate the complex behavior of EIIA^{Crr} because its behavior has been shown to be very complex and vary across carbon substrates (Bettenbrock *et al.*, 2007). Consequently, its behavior is not fully described by the logic rules of the rFBA Boolean regulatory model (Fig. 3D).

The iFBA model includes over 100 additional genes and corresponding regulation or reactions, beyond what is included in the ODE model. The iFBA simulation therefore includes a large amount of additional data, such as changes in global gene expression and flux distributions. For example, the metabolic flux distributions at 1 h and 5 h are shown in Figure 4. At 1 h, when bacteria are consuming glucose there is significant flux from internal glucose-6-phosphate through the pentose phosphate pathway. However, at 5 h, when bacteria are consuming lactose, the flux from glucose-6-phosphate has shifted toward glycolysis. Additionally, at 5 h the lactose-related transcription factors GalE, GalM, GalK, and GalT are now expressed, while PtsG is suppressed and bacteria secrete ethanol in addition to acetate. Changes in gene expression were calculated using iFBA and rFBA as shown, and involve induction of the proteins required to utilize lactose as a carbon source.

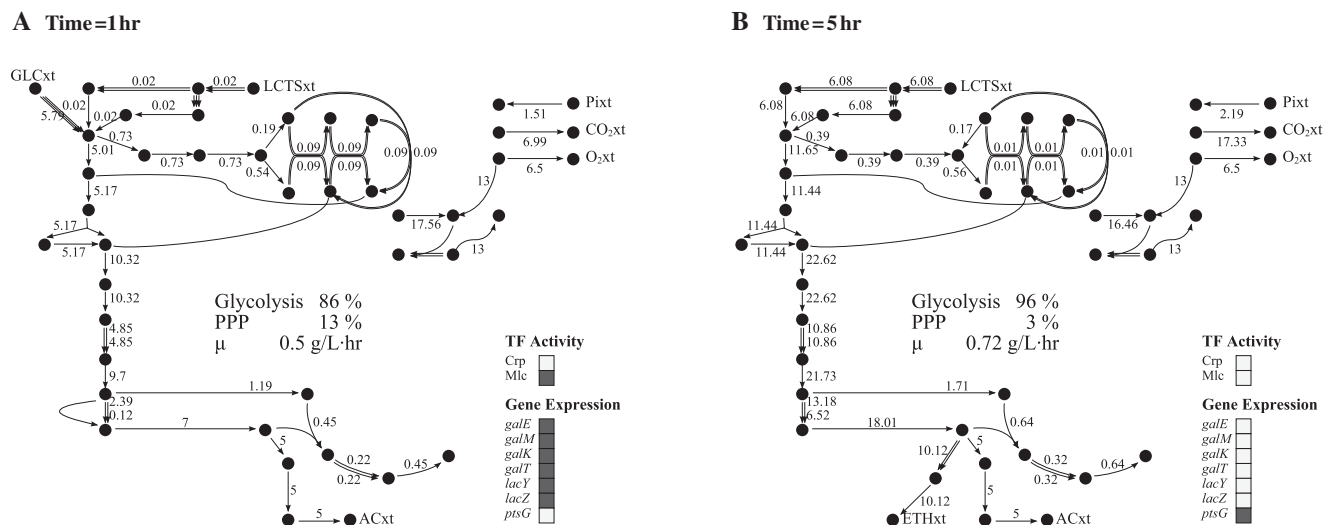


Fig. 4. Flux distributions for iFBA simulation of glucose/lactose wild-type diauxic growth, at (A) 1 h and (B) 5 h. Detailed labels for the network are shown in Figure 1, and all values are in mmol/gDCW/hr. Selected qualitative gene expression values calculated using the rFBA module of iFBA are also shown, where light gray denotes expression and dark gray denotes repression.

3.2 Diauxic growth on glucose/glucose-6-phosphate

Escherichia coli uptake of glucose and glucose-6-phosphate is concurrent, with some repression of the glucose transporter. Figure S1 shows the iFBA, rFBA and ODE wild-type simulations together with experimental data. As with glucose/lactose diauxie, we observed that all three models describe the wild-type experimental external glucose and glucose-6-phosphate, and biomass data equally well (Fig. S1A, B). Again the iFBA and rFBA models predicted similar rates of acetate secretion, and again the predicted concentrations of the internal metabolites glucose-6-phosphate, phosphoenolpyruvate and pyruvate, differed between the iFBA and ODE models (Fig. S1C) due to consideration of phosphoenolpyruvate conversion to oxaloacetate by Ppc and more detailed consideration of metabolite conversion to biomass. The concentration and activity profiles for transporters UhpT and PtsG, as well as EIIA were also similarly represented by the iFBA and ODE models, but not by rFBA (Fig. S1D, E). Finally, at the network level, in contrast to glucose/lactose diauxie, we saw large shunting of the flux away from glucose-6-phosphate from glycolysis to the pentose phosphate pathway—65% of the input flux was shifted from glycolysis to the pentose phosphate pathway between 1 h and 6 h (Fig. S2).

3.3 Single gene perturbation analysis

To further compare the iFBA, rFBA and ODE models we simulated diauxic growth of 167 *E.coli* single gene perturbations on glucose/glucose-6-phosphate and glucose/lactose—151 knockouts of 135 enzymes and 16 transcription factors, and 16 cases where we forced each transcription factor to be constitutively active. We found that the iFBA model predicted different phenotypes than the ODE model for 41 and 44 of the mutants on glucose/glucose-6-phosphate and glucose/lactose, respectively. As illustrated in Fig. 5, the genes corresponding to these mutants can be grouped into five classes—6 TCA cycle genes, 24 intermediate energy storage

genes, 3 carbohydrate transport genes, 7 glycolysis genes and 5 transcriptional regulatory genes.

For most of the 85 cases, the iFBA and rFBA models predicted the correct outcome of gene perturbation observed in various reports (reviewed in Covert and Palsson, 2002), while the ODE model failed to predict the correct outcome in many of these cases. We investigated these differences in more detail. For the TCA cycle and intermediate energy storage genes, the ODE model was unable to predict the effects of gene deletion, because it does not consider metabolic pathways beyond glycolysis and therefore does not include these genes. For transport and glycolysis, the ODE model includes the corresponding genes, but incorrectly predicts that these deletions will be non-lethal. This failure is because the ODE model's equation for biomass is based only on transport of the extracellular metabolites and not on the ability to produce biomass components. The *pgi* deletion, represents an unusual case where the ODE model predicts a more negative impact on growth. We found that this difference is because the rFBA and iFBA models include the pentose phosphate pathway which is used as an alternate route from fructose-6-phosphate to the TCA cycle when *pgi* is deleted. Forced constitutive activation of four transcription factors also led to repression of key genes whose absence had a negative impact on growth in the iFBA and rFBA models but were not included in the ODE model.

Finally, we found that the iFBA model predicted different phenotypes than the rFBA model for two mutants on glucose/glucose-6-phosphate—*galP* and *glk* and one mutant on glucose/lactose—*pykA*. These predictions highlight the advantage of the iFBA model over the rFBA model to include the subtle effects of the dynamics of the internal metabolites glucose-6-phosphate, phosphoenolpyruvate and pyruvate.

4 DISCUSSION

The iFBA model described earlier has several strong advantages over both the rFBA and ODE models. First, the kinetic description in the

iFBA model contains a much greater level of detail than the rFBA approach to modeling regulatory activities and events. This would be critical in cases where genes have multiple stable expression states, as has been observed with lac operon regulation (Setty *et al.*, 2003), as well as with the action of Crp and EIIA^{Crp} described here. Similarly, because of FBA's quasi steady-state assumption described earlier, the concentration of internal metabolites is not calculable without the kinetic model. This is also critical where regulatory protein activities depend on internal metabolite concentrations, which previously have been approximated by either external metabolite concentrations or combinations of metabolic fluxes (Covert and Palsson, 2002). Including the set of ODEs makes such crude approximations unnecessary. The importance of incorporating detailed kinetic information was also underscored by the *pykA*, *galP* and *glk* knockout simulations, where the iFBA model made significantly different predictions than the rFBA model due the effects of internal metabolite concentrations on the system.

A second advantage of iFBA over rFBA is that certain enzymes are expressed and active, which would never be part of a strictly optimal growth scenario. This is because they are utilized not for their metabolic contribution to growth, but for other important functions such as signal transduction. As an example, the adenylate cyclase enzyme catalyzes the conversion of ATP to cyclic AMP (cAMP), a key mediator of catabolite repression. cAMP is not required as part of the growth objective function in the FBA model, and therefore the adenylate cyclase flux wastes ATP and would never be used. In fact, adenylate cyclase is one of many reactions in reconstructed metabolic networks listed as 'dead ends' in FBA because they lead to production of metabolites that would never be used as part of a growth-optimal solution (Reed *et al.*, 2003). However, the phosphorylation of EIIA^{Crp} demonstrated by iFBA would result in the utilization of this flux to generate cAMP (Bettenbrock *et al.*, 2007). This raises the possibility that many of the FBA-determined 'dead ends' are in fact 'gateways' to other important cellular networks such as cell signaling.

We also found that the iFBA model has certain advantages over the ODE model. In particular, we observed that the iFBA model enabled us to see the global effects of dynamic changes in the Kremling model, because of its ability to calculate a flux distribution for an entire network with only few additional parameters. An important illustration of this is the experimentally determined secretion of acetate under glucose/lactose diauxie, which is captured in the iFBA model but not the ODE model. Another example is the predicted shift of metabolic flux from the glycolysis to the pentose phosphate pathway as glucose-6-phosphate is depleted in glucose/glucose-6-phosphate diauxie, which could not be determined using the ODE model. In this case, we also saw that the flux through phosphoenolpyruvate carboxylase, which was assumed to be negligible in the ODE model, was a significant percentage of the flux from phosphoenolpyruvate, resulting in a substantially lower predicted internal phosphoenolpyruvate concentration. This prediction correlates with the experimental observation that *ppc* knockout strains are unable to grow with glucose as the sole carbon source (Courtright and Henning, 1970).

iFBA is also able to determine systemic properties such as the growth rate from integrated network behavior rather than from empirical correlation with substrate uptake or other parameters, and this growth rate has a significant impact on the behavior of the ODE model. This aspect of iFBA was highlighted most dramatically

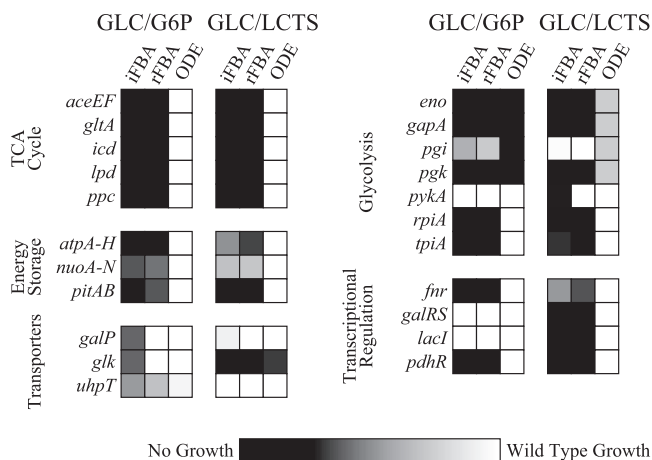


Fig. 5. Gene perturbation analysis. The ratio of mutant to wild-type biomass concentration at 8 h is shown for all mutants where differences were observed between the iFBA, rFBA and ODE-based simulations. All perturbations are knockdowns except for the catabolite repression genes where we forced the corresponding transcription factor to be constitutively expressed. All simulation results are found in Table S3.

in the gene perturbation study where we found 85 cases in which the ODE model incorrectly and the iFBA model correctly predicted the experimentally observed result of gene perturbation (Covert and Palsson, 2002). We found that these cases fell into three categories: (1) ODE model predicts lethality because it is missing an alternate pathway, (2) ODE model predicts viability because it does not account for global demands on biomass production and (3) ODE model fails to predict the correct phenotype because the function of the gene is not included in the model. The iFBA modeling framework therefore adds to the predictive power of ODE-based models, both in terms of scope and accuracy.

In summary, the great advantage of flux balance models over traditional sets of ordinary differential equations is that they allow for analysis of the entire metabolic and regulatory networks. The advantage of the ODE models is that they capture intracellular concentrations and short time-scale dynamics, which are critical components of signal transduction. We find that the iFBA approach described here has the potential to incorporate the advantages of both perspectives.

There are several ways to potentially improve the iFBA framework. First, this model is based on an objective which maximizes the growth rate, and it has been shown that other objectives may be more accurate predictors of phenotype, depending on the growth conditions (Burgard and Maranas, 2003; Schuetz *et al.*, 2007; Segre *et al.*, 2002). Additionally, there are multiple flux distributions which could provide an equivalent growth rate, and only one of these has been selected for the simulation. Incorporating these equivalent distributions could also lead to a richer description of phenotype (Shlomi *et al.*, 2007), and possibly also account for the natural phenotypic variation between cells in a culture. Finally, although we decided to initially focus on central metabolism for the purposes of developing iFBA, with this proof-of-principle in hand our iFBA model could be improved by including the 755 additional genes described in our more comprehensive rFBA model of *E. coli* (Covert *et al.*, 2004).

Can we build on this approach to eventually create a whole-cell model of *E.coli*? Currently the largest flux-balance model of *E.coli* incorporates 1260 open reading frames corresponding to metabolism (Feist et al., 2007), and another includes an additional 104 genes corresponding to transcriptional regulation (Covert et al., 2004). Our current work suggests that these large-scale metabolic and regulatory network models may now be thought of as a scaffold with which any ODE-based or other model that has an interface with metabolism may be integrated. This integration would allow processes which have been characterized and modeled in isolation to be re-evaluated in the context of their global effects. As more ODE-based models are developed and integrated into frameworks like that described here, it may eventually be possible to capture a majority of the known biological processes which occur in *E.coli* or other organisms in a single computational model.

ACKNOWLEDGEMENTS

We thank Iman Famili, Ken Judd, Markus Herrgard and Bernhard Palsson for helpful discussions and advice, and Andreas Kremling for generous assistance in implementing his models of catabolite repression.

Funding: This work has been supported by a James H. Clark Faculty Scholarship in the Stanford School of Engineering to M.W.C., a Stanford Bio-X Fellowship to N.X., a Stanford Graduate Fellowship to J.R.K., and the National Library of Medicine (grant number LM 07033) for T.J.C.

Conflict of Interest: none declared.

REFERENCES

- Bettenbrock, K. et al. (2006) A quantitative approach to catabolite repression in *Escherichia coli*. *J. Biol. Chem.*, **281**, 2578–2584.
- Bettenbrock, K. et al. (2007) Correlation between growth rates, EIIA^{Crr} phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J. Bacteriol.*, **189**, 6891–6900.
- Burgard, A.P. and Maranas, C.D. (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol. Bioeng.*, **82**, 670–677.
- Courtright, J.B. and Henning, U. (1970) Malate dehydrogenase mutants in *Escherichia coli* K-12. *J. Bacteriol.*, **102**, 722–728.
- Covert, M.W. et al. (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, **213**, 73–88.
- Covert, M.W. et al. (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.*, **277**, 28058–28064.
- Covert, M.W. et al. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, **429**, 92–96.
- Feist, A.M. et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.
- Ingraham, J.L. et al. (1983) *Growth of the Bacterial Cell*. Sinauer Associates, Inc., Sunderland, MA.
- Kremling, A. et al. (2007) Analysis of global control of *Escherichia coli* carbohydrate uptake. *BMC Sys. Biol.*, **1**, 42.
- Lee, J.M. et al. (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.*, **4**, e1000086.
- Luo, R.Y. et al. (2006) Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. *Mol. Sys. Biol.*, **2**, 2006.0031.
- Mahadevan, R. et al. (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.*, **83**, 1331–1340.
- Price, N.D. et al. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.
- Reed, J.L. et al. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Gen. Biol.*, **4**, R54.
- Reed, J.L. et al. (2006) Systems approach to refining genome annotation. *Proc. Natl Acad. Sci. USA*, **103**, 17480–17484.
- Schuetz, R. et al. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Sys. Biol.*, **3**, 119.
- Segre, D. et al. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA*, **99**, 15112–15117.
- Setty, Y. et al. (2003) Detailed map of a cis-regulatory input function. *Proc. Natl Acad. Sci. USA*, **100**, 7702–7707.
- Shlomi, T. et al. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl Acad. Sci. USA*, **102**, 7695–7700.
- Shlomi, T. et al. (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Sys. Biol.*, **3**, 101.
- Smallbone, K. et al. (2007) Something from nothing – bridging the gap between constraint-based and kinetic modeling. *FEBS J.*, **274**, 5576–5585.
- Varma, A. and Palsson, B.O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, **60**, 3724–3731.
- Yugi, K. et al. (2005) Hybrid dynamic/static method for large-scale simulation of metabolism. *Theor. Biol. Med. Model.*, **2**, 42.