Introduction

Genome-scale constraint-based metabolic models can be used to predict or describe cellular behaviors, such as growth rates, uptake/secretion rates, and intracellular fluxes. These models have been used for a variety of applications, involving studies on drug discovery [1], metabolic engineering [2], evolution [3], genome annotation [4–6], and multi-species interactions [7–10]. Constraint-based metabolic models are developed by integrating genomic, biochemical, and physiological information for an organism, in a process that has been recently reviewed [11]. Computational and database efforts facilitate the construction of such models by automating some of the steps in the development process; for example, mapping genes to biochemical reactions or adding/removing reactions based on physiological data [4,5,12–14].

The variables used in constraint-based models include the fluxes through transport and metabolic reactions, and model parameters include reaction stoichiometry, biomass composition, ATP requirements, and the upper and lower bounds for individual fluxes. A common misconception is that these metabolic models rely on detailed kinetic parameters; however, such kinetic parameters are not required and are generally absent from most constraint-based models. Because there are often more variables (i.e., fluxes) than equations, no unique solution exists. The large number of solutions that satisfy the model’s constraints define the model’s solution space, which can be queried using a number of approaches [15]. Most of these constraints-based approaches utilize optimization to identify a subset of solutions of interest from within the solution space that are predicted to be physiologically relevant. For example, flux balance analysis (FBA) is often used to identify flux distributions that maximize biomass yields [16].

Given the non-uniqueness of constraint-based model solutions, a growing number of methods have focused recently on incorporating additional constraints to reduce the solution space and thereby improve the precision and accuracy of model predictions. This editorial reviews recent methods that utilize additional biological information (e.g., gene or protein expression, metabolite concentrations, and kinetic parameters) to further restrict metabolic fluxes, many of which are available in a variety of software packages [17–20]. A brief description of the standard constraints used in all constraint-based models is first presented, followed by a survey of how additional constraints have been included into models that often make use of additional types of experimental data (Figure 1).

Standard Constraints

All constraint-based models use two types of fundamental constraints. Steady-state mass-balance constraints ensure that for each metabolite in the network the net production rate equals the net consumption rate. Additional inequality constraints are used to place restrictions or bounds on the values of individual fluxes based on measured rates (e.g., metabolite uptake/secretion rates) or reaction reversibility, where irreversible fluxes have a zero lower bound. Most models to date base reversibility on biochemical characterization of enzymes or consideration of network properties (e.g., no free ATP production). In standard models, none of these constraints limit metabolic fluxes based on metabolite, mRNA, or protein concentrations; however, a variety of additional constraints can be included based on thermodynamic, molecular crowding, gene expression, and regulatory and kinetic considerations.
Thermodynamic Constraints

Thermodynamic constraints are used to place restrictions on the directionality of reactions by considering metabolite concentrations and Gibbs energies of formation. From thermodynamics, the change in Gibbs free energy for a reaction ($\Delta G$) depends on the temperature, concentrations of substrates and products, and change in Gibbs free energies in a reference state ($\Delta G^0$). If a reaction is to proceed, the change in Gibbs free energy for a reaction must be negative. A few different approaches for incorporating these types of thermodynamic-based directionality constraints have been proposed. One of the first methods, network-embedded thermodynamic (NET) analysis, uses the directionality of reactions (based on pre-existing knowledge, experimental flux measurements, or constraint-based model results) to calculate $\Delta G$ or metabolite concentration ranges [21]. In NET analysis the reaction directions are determined a priori and a set of concentrations are found that are consistent with the thermodynamic constraints. However, analysis of thermodynamic constraints can also be done to identify reaction directionalities using specified metabolite concentrations (or concentration ranges). The results can then be used to limit the directionality of reactions in constraint-based models. Given the uncertainty in the Gibbs free energies of formation and metabolite concentrations, many reactions can operate in either direction and so probabilities can be used to assign uni-directional reactions [22].

Molecular Crowding Constraints

Recent efforts have used spatial constraints to place upper limits on a sum of fluxes, rather than individual fluxes. Molecular crowding constraints were first proposed by Beg et al. to restrict the total amount of enzyme that could be packed into a cell [24]. An upper limit on total enzyme volume was used and the volume of enzyme needed to sustain a given flux value was based on each enzyme’s properties (e.g., kinetics and size). This molecular crowding constraint results in a restriction on the weighted sum of the fluxes, where the weights ($w_j$) depend on an enzyme’s volume and activity (less active, larger protein will have higher weights). Molecular crowding constraints have been used to predict cellular growth rates and acetate production in Escherichia coli [24,25], to predict enzyme activities and metabolite concentrations in yeast [26], and to explain the Warburg effect of inefficient glucose catabolism in cancer cells [27]. Zhuang et al. recently extended this concept to impose limitations on the approach, thermodynamic metabolic flux analysis (TMFA), directly imbeds the thermodynamic constraints into the models. TMFA uses integer variables to identify flux distributions that are consistent with thermodynamic constraints. In TMFA, fluxes and metabolite concentrations are variables in the models and constraints ensure that non-zero fluxes and $\Delta G$ values have opposite signs [23].
amount of enzymes that could reside in the cell membrane [28],
thus placing restrictions on the weighted sum of fluxes through
reactions that take place at the cell membrane. The authors
investigated how this crowding constraint imposes a trade-off
between glucose transport and respiratory pathways and showed
that it was able to explain acetate production by *E. coli* under
glucose aerobic conditions.

**Gene Expression Constraints**

Gene expression is one of the most widely accessible measure-
ments that can provide a global snapshot of a cell’s metabolic state.
A number of studies have compared constraint-based model flux
predictions to expression data, to find consistencies and inconsis-
tencies (e.g., [29–31]). For example, genes associated with
reactions predicted to be essential for growth were found to have
higher expression than those associated with reactions predicted to
be inactive in *E. coli* [31]. On the other hand, fluxes predicted to
be inactive in *Shewanella oneidensis* but whose genes were expressed
identified pathways that were reducing biomass yields [29]. In
these cases, the expression data are not used to help predict flux
values, but instead are compared against flux predictions. As an
alternative, a number of computational tools have been developed
to integrate expression data into constraint-based models and
restrict metabolic fluxes directly (Table 1).

Most current methods for incorporating gene expression data
into the models compare gene expression levels in a single
condition and disfavor fluxes through reactions that are associated
with lowly expressed genes. The E-flux method uses gene
expression values to set upper limits on metabolic fluxes, where
reactions associated with more highly expressed genes will be
allowed to take on higher flux values [32]. While E-flux places
hard constraints on fluxes based on expression data, other methods
instead use soft constraints that can be violated. GIMME tries to
minimize the total inconsistency between fluxes and gene
expression, where inconsistency depends on the flux value and the
difference between a gene’s expression value and a chosen
threshold [33]. In this case, GIMME will try and reduce fluxes
through reactions whose associated gene’s expression falls below
the threshold. Another method, developed by Shlomi et al. [34],
tries to encourage flux through reactions whose associated genes
are highly expressed and discourage flux through reactions whose
associated genes are lowly expressed. With this method, high and
low expression thresholds are chosen and used to assign reactions
to high, low, or moderate groups. Using optimization, fluxes are
then favored through reactions belonging in the high group and
disfavored through reactions belonging to the low group.

All of these previous methods typically use expression data from
a single condition to constrain fluxes. A more recent approach
(MADE) uses expression data from multiple conditions (or a time-
series) to identify patterns of increased/decreased expression based
on significant changes in expression across conditions [35]. With
MADE, the measured patterns of expression increases and
decreases are used to find gene on/off patterns in the model
across all conditions, where more significant expression changes
are weighted more heavily. In another study, Moxley et al. used
expression changes between two conditions to predict flux changes
[35]. Using two global parameters they were able to accurately
predict flux changes from gene expression changes using non-
linear functions that account for metabolite-enzyme interaction
densities.

**Transcriptional Regulatory Constraints**

The methods described above for using gene expression-based
constraints require expression data under the condition(s) of
interest. In other words, to predict flux in a particular condition
the methods would need gene expression data from that condition.
Other methods can instead use models of transcriptional regula-
tory networks to predict the effects of transcriptional
regulation on metabolic fluxes. In this case, integrated models
of metabolism and regulation can predict metabolic fluxes under
conditions (e.g., gene knockout mutants) for which gene expression
data are not available. Transcriptional regulatory networks can be
reconstructed from high-throughput data, such as gene expression,
ChIP-chip, and genome sequencing datasets using a variety of
approaches [reviewed in [36–38]]. To date, two different types of
approaches have been used to incorporate transcriptional regula-
tory constraints into genome-scale metabolic models. The first set
of approaches used a Boolean (on/off) representation of
transcriptional regulation, where Boolean rules are used to
determine the state of transcription factors (active or inactive)
and metabolic genes (expressed or not expressed). Based on the
expression states of metabolic genes, the reactions in the metabolic
network can (if necessary genes are expressed) or cannot (if
necessary genes are not expressed) carry flux [39]. Analysis of these
Boolean types of models can be done by solving the regulatory and
metabolic models separately in an iterative fashion (rFBA) or
simultaneously (SR-FBA) by introducing integer variables to
represent the transcription factor/gene expression/reaction on/
off states [40–42]. Not all regulation can be captured using a
Boolean approach; for example, essential genes must always be on
even though their expression may be regulated. To overcome this
limitation, another type of approach has recently been used to
formulate regulatory constraints based on a probabilistic regula-
tory model, where a continuous rather than a Boolean flux
constraint is used. Here, the regulatory model predicts the
probability that a given gene is expressed and this probability is

| Method | Thresholds | Description of Solutions |
|--------|------------|--------------------------|
| E-Flux | None       | Finds solutions with fluxes whose upper limits are proportional to relative expression values |
| GIMME  | One        | Finds solutions with low flux through reactions associated with lowly expressed genes |
| Shlomi 2008 | Two | Finds solutions with non-zero flux through reactions associated with highly expressed genes and zero flux through reactions associated with lowly expressed genes |
| MADE   | None       | Finds solutions whose gene’s on/off states most closely match significant changes in gene expression across multiple conditions |
| Moxley 2009 | None | Finds changes in flux values based on changes in gene expression values |

doi:10.1371/journal.pcbi.1002662.t001
Kinetic Constraints

A variety of approaches have been developed to capture kinetic limitations in the models. These approaches involve constraining either the uptake/secretion rates using empirical rate laws that depend on extracellular concentrations or constraining intracellular fluxes using enzymatic rate laws that depend on intracellular and extracellular concentrations. Incorporating constraints on the uptake or secretion rates of metabolites often requires material balance equations for the bioreactor environment, in addition to the standard metabolic constraints for the cells. Empirical rate laws are found by fitting metabolite uptake/secretion rates to measured reactor concentrations. These rate laws are then used as additional constraints in the models. The resulting dynamic FBA (dFBA) models can then use bioreactor concentrations to constrain metabolic fluxes, which in turn affect the bioreactor concentrations. Feng et al. recently included rate laws for the uptake and secretion of organic acids into a genome-scale model for Shewanella, and Cyano bacteria species, as well as computational tools for metabolic engineering, model comparison, model refinement, and experimental design.

Traditional kinetic models already take into account the kinetic relationships between metabolic fluxes, metabolite, and protein concentrations. However, such detailed models are often available for only a few pathways in well-characterized organisms, such as E. coli and Salmonella enterica, since the kinetic properties of their enzymes have been biochemically characterized. Databases, such as BRENDA and SABIO-RK, contain an extensive collection of kinetic parameters assembled from the biochemical literature, and these in vitro estimates can be used to formulate kinetic constraints. While kinetic models exist for central metabolism and other isolated pathways, expanding these models to a genome scale is an active area of research. Vizhak et al. recently developed an approach called IOMA, which uses kinetic expressions for a subset of enzymes to constrain metabolic fluxes. By incorporating multi-omics datasets using kinetic constraints for 11 reactions into an E. coli model, the authors were able to improve flux predictions in 23 gene deletion strains.

Conclusions

As we continue to be able to measure intracellular levels of biological components with greater accuracy and precision, the need for computational approaches to integrate and analyze such large-scale datasets grows. As reviewed above, a variety of constraint-based approaches are available that use these types of datasets to reduce the solution space and improve model predictions of metabolic phenotypes. Over the coming years, more computational approaches for integrating individual and multiple types of experimental measurements will likely appear, as new biological measurement approaches are developed and more data becomes available. For example, we are likely to see integration of datasets into models of microbial communities, as multi-species models and datasets become available. With recent advances in the ability to rapidly build genome-scale models, there will also be a need to design experiments whose results would best reduce the metabolic solution space. One of the future challenges is then to prioritize what types of data are important to measure and for which components. Other related questions need to be answered as well. How important is it to have absolute versus relative concentration measurements? What experimental precision is needed for different types of data? The authors answers to all of these questions will depend on both the biological hypotheses that are being investigated and the desired precision for predicted fluxes, which specifies how much the solution space needs to shrink.

Acknowledgments

The author would like to thank Josh Hamilton and Joonhoon Kim for their helpful comments on the manuscript.
11. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5: 93–121.

12. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nucleic Acids Res 28: 27–30.

13. Kanchia M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 130–132.

14. Kummel A, Panke S, Heinemann M (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol Syst Biol 2: 295–301.

15. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10: 291–305.

16. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28: 245–248.

17. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6: 1290–1307.

18. Jensen PA, Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 27: 541–547.

19. Fleming RM, Thiele I (2011) von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. Bioinformatics 27: 142–143.

20. Zur H, Ruppin E, Shlomi T (2010) iMAT: an integrative metabolic analysis tool. Bioinformatics 26: 3140–3142.

21. Kummer A, Panke S, Heinemann M (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol Syst Biol 2: 2006.0034.

22. Fleming RM, Thiele I, Nasheuer HP (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. Biophys J 97: 45–56.

23. Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. Biophys J 92: 1792–1805.

24. Bae SK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, et al. (2007) Intracellular crowding refines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proc Natl Acad Sci U S A 104: 12663–12668.

25. Vazquez A, Bae SK, de Menezes MA, Ernst J, Bar-Joseph Z, et al. (2008) Impact of the solvent capacity constraint on E. coli metabolism. BMC Syst Biol 2: 7.

26. Vazquez A, de Menezes MA, Barabasi AL, Oliva ZN (2008) Impact of limited solvent capacity on metabolic rate, enzyme activities, and metabolite concentrations of S. cerevisiae glycolysis. PLoS Comput Biol 4: e1000195. doi:10.1371/journal.pcbi.1000195.

27. Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E (2011) Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. PLoS Comput Biol 7: e1002010. doi:10.1371/journal.pcbi.1002010.

28. Zhuang K, Vemuri GN, Mahadevan R (2011) Economics of membrane occupancy and respiro-fermentation. Mol Syst Biol 7: 300.

29. Pinchak GE, Hill EA, Geydybrukh OV, De Ingenis J, Zhang X, et al. (2010) Constraint-based model of Shewanella oneidensis MR-1: a tool for data analysis and hypothesis generation. PLoS Comput Biol 6: e1000222. doi:10.1371/journal.pcbi.1000222.

30. Raghunathan A, Reed J, Shin S, Palsson B, Daehler S (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. BMC Syst Biol 3: 38.

31. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, et al. (2010) Omic data from evolved E. coli is consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6: 390.

32. Colijn C, Brandes A, Zucker J, Lian DS, Weiner R, et al. (2009) Interpreting expression data with metabolic flux models: predicting mycobacterium tuberculosis mycolic acid production. PLoS Comput Biol 5: e1000489. doi:10.1371/journal.pcbi.1000489.

33. Becker SA, Palsson BO (2000) Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 4: 10. doi:10.1371/journal.pcbi.1000082.

34. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotechnol 26: 1093–1010.

35. Modey JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper H, et al. (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcr1p. Proc Natl Acad Sci U S A 106: 6477–6482.

36. De Smet R, Marchal K (2010) Advantages and limitations of current network inference methods. Nat Rev Microbiol 8: 717–729.

37. Rodionov DA (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem Rev 107: 3467–3497.

38. Herrgard MJ, Covert MW, Palsson BO (2004) Reconstruction of microbial transcriptional regulatory networks. Curr Opin Biotechnol 15: 70–77.

39. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213: 73–88.

40. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429: 92–96.

41. Covert MW, Palsson BO (2002) Transcriptional regulation in constraints-based metabolic models of Escherichia coli. J Biol Chem 277: 29058–29064.

42. Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol 3: 101.

43. Chaurasrakaran N, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 107: 17845–17850.

44. Feng X, Xu Y, Chen Y, Tang YJ (2012) Integrating flux balance analysis into kinetic models to decipher the dynamic metabolism of Shecavella oneidensis MR-1. PLoS Comput Biol 8: e1002376. doi:10.1371/journal.pcbi.1002376.

45. Hanly T, Henson MA (2011) Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol Bioeng 108: 376–385.

46. Schomburg I, Chang A, Elam G, Gremse M, Heldt C, et al. (2004) BRENDa, the enzyme database: updates and major new developments. Nucleic Acids Res 32: D431–D433.

47. Wittig U, Kania R, Goldehriwki M, Rey M, Shi L, et al. (2012) SABIO-RK-database for biochemical reaction kinetics. Nucleic acids res 40: D790–D796.

48. Jamshidi N, Palsson BO (2008) Formulating genome-scale kinetic models in the post-genome era. Mol Syst Biol 4: 171.

49. Fleming RM, Thiele I, Provan G, Nasheuer HP (2010) Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. J Theor Biol 264: 683–692.

50. Smallbone K, Simeoni E, Swainston N, Mendes P (2010) Towards a genome-scale kinetic model of cellular metabolism. BMC Syst Biol 4: 6.

51. Soh KC, Miskovic L, Hatzimanikatis V (2012) From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. FEMS Yeast Res 12: 129–143.

52. Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics 26: 245–250.

53. Tafti R, Anton JE, Briley J, Kay Z, Klatt CG, et al. (2009) In silico approaches to study mass and energy flows in microbial consortia: a synthetic case study. BMC Syst Biol 3: 114.