Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Recent advances in the reconstruction of metabolic models and integration of omics data

Rajib Saha[1], Anupam Chowdhury[1] and Costas D Maranas

With the ever-accelerating pace of genome sequencing and annotation information generation, the development of computational pipelines for the rapid reconstruction of high-quality metabolic networks has received significant attention. Herein, we review the available biological databases and automated/semi-automated reconstruction tools. In addition, we describe available methodologies for the integration of high-throughput omics data to increase metabolic phenotype prediction accuracy. Data heterogeneity and lack of better integration of metabolic reconstruction pipelines with omics data generation protocols have hampered rapid progress thus far.

Addresses
Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA

Corresponding author: Maranas, Costas D (costas@engr.psu.edu, costas@psu.edu)
[1] Joint first authors.

Current Opinion in Biotechnology 2014, **29**:39–45

This review comes from a themed issue on **Cell and pathway engineering**

Edited by **Tina Lütke-Eversloh** and **Keith EJ Tyo**

Available online XXX

0958-1669/$ – see front matter,

http://dx.doi.org/10.1016/j.copbio.2014.02.011

Introduction

A metabolic network captures the inter-conversion of metabolites through chemical transformations catalyzed by enzymes. To this end, a metabolic model describes reaction stoichiometry and directionality, gene to protein to reaction associations (GPRs), organelle-specific reaction localization, transporter/exchange reaction information, transcriptional/translational regulation and biomass composition [1]. By defining the metabolic space, a metabolic model can assess allowable cellular phenotypes under specific environmental and/or genetic conditions [2,3]. The number of metabolic models developed in the past several years is a testament to their increasing usefulness and penetration in many areas of biotechnology and biomedicine [4,5,6••]. Initially, metabolic models have been used to characterize biological systems and develop non-intuitive strategies to reengineer them for enhanced production of valuable bioproducts [7]. More recently, models have been developed and applied for a variety of goals ranging from metabolic disease drug, target identification, study of microbial pathogenicity and parasitism (as highlighted in [5]).

The validation of high-quality [8] models is critical for not only recapitulating known physiological properties but also improving their prediction accuracy. Towards this end, strategies have been developed to incorporate other cellular processes such as gene/protein expression to better understand the emergence of complex cellular phenotypes [9,10]. For example, genome-scale metabolic models of pathogens have been reconstructed to develop novel drugs for combating infections and also minimize side effects in the host [11]. An integrated model [12] of *E. coli* has been developed by combining **M**etabolism with gene **E**xpression (i.e. ME model) to increase the scope and accuracy of model-computable phenotypes corresponding to the optimal growth condition. In addition, by combining all of the molecular components as well as their interactions, a whole-cell model [13••] has been developed for *Mycoplasma genitalium*, a human pathogen, to study previously unexplored cellular behaviors including protein-DNA association and correlation between DNA replication initiation and replication itself. Tissue specific models have also been developed for eukaryotic organisms, such as *Homo sapiens* [14] and *Zea mays* [2], to scope out novel therapeutic targets and characterize metabolic capabilities, respectively. Moving beyond the single cell/tissue level, multi-cell/multi-tissue type metabolic models have been reconstructed for higher organisms. For example, *Homo sapiens* [14,15] models have been employed for biomedicine applications and a *Hordeum vulgare* [16] model has been deployed for studying crop improvement and yield stability.

With rapid improvements in sequencing (and annotating) tools and techniques, the number of complete genomes (and annotations) is increasing at an exponential pace [17]. Metabolic models can greatly facilitate the assessment of the potential metabolic phenotypes attainable by these organisms. Therefore, rapid development of high-quality metabolic models and algorithms for analyzing their content are of critical importance. The recent genome-scale metabolic models, their automated generation, improvements and applications have been reviewed elsewhere [4,18,19,20••] and will not be covered in detail in this review. Rather, in this mini-review we will critically evaluate the available repositories, model-building and data integration techniques and existing challenges

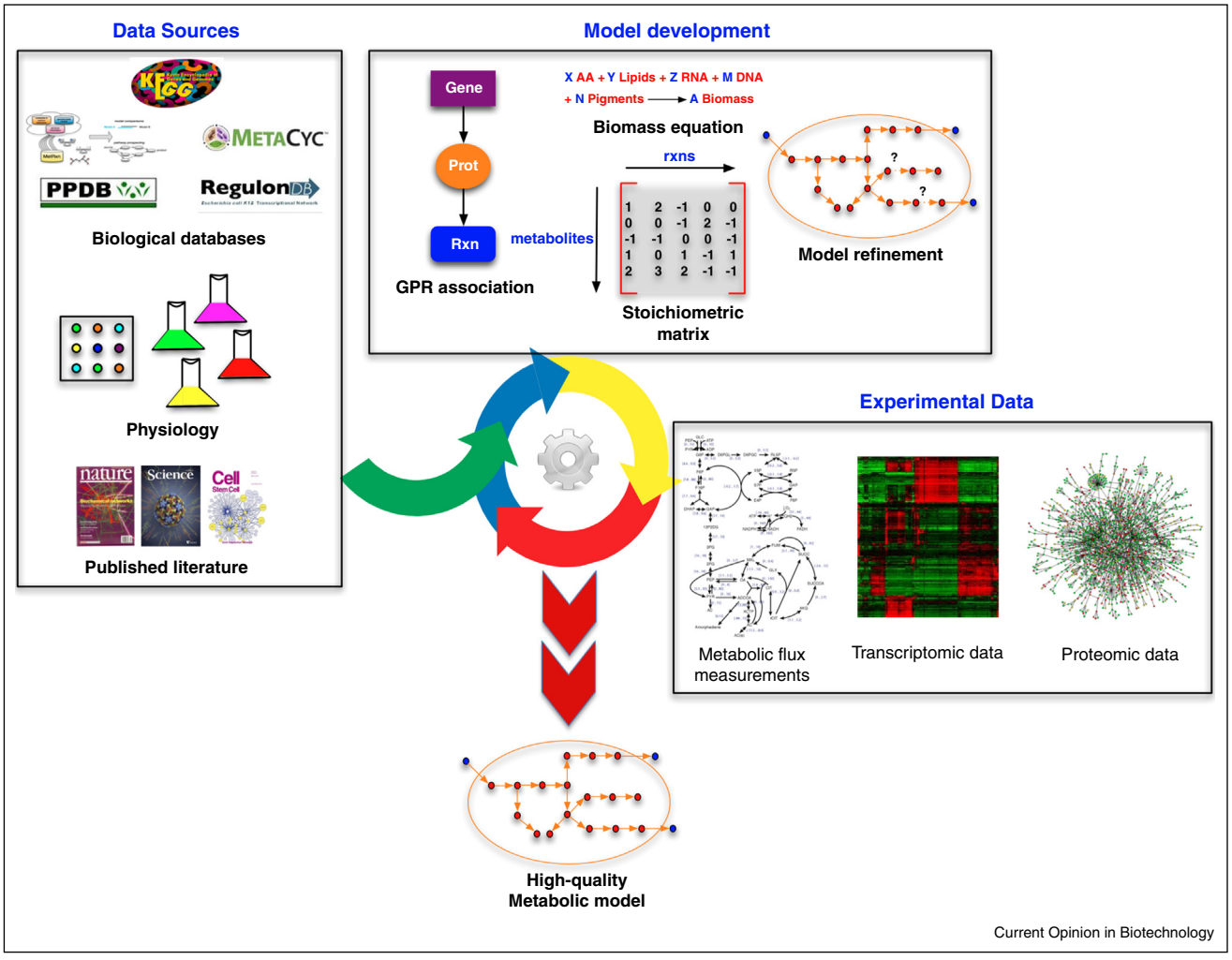
related to rapid reconstruction of high-quality metabolic models.

Metabolic model reconstruction approaches

Metabolic network/model reconstruction process follows three major steps (as highlighted in Figure 1). Initially, upon sequencing and annotating a genome of interest, literature sources and/or homology searches are used to assign function to all the **O**pen **R**eading **F**rames (i.e. ORFs). For every function with a metabolic fingerprint a specific chemical transformation is assigned. Therefore, by iteratively marching along the entire genome, a compilation of reactions encompassing the entire chemistry repertoire of the organism can be achieved. It must be noted that these models are not necessarily predictive but instead have a scoping nature by allowing us to assess what is metabolically feasible. Regulatory constraints on reaction fluxes are incorporated based on the thermodynamic (i.e. reaction reversibility) and omics (i.e. transcriptomic/proteomic) data that can further sharpen predictions.

One of the most critical steps of metabolic model building is to establish GPR information of a specific organism from biological databases and/or literature sources. To this end, biological databases (as highlighted in [1]) such as KEGG, SEED, Metacyc, BKM-react, Brenda, Uniprot, Expasy, PubChem, ChEBI and ChemSpider provide information about reactions/metabolites and associated enzymes and genes. However, as illustrated by Kumar

Figure 1



Outline for the development of a high-quality metabolic model: the first step involves retrieving data from different biological databases, physiology and biochemistry of the organism as well as published literature. In the next step, GPR associations are established, the biomass equation is described based on experimental measurements and the model is represented in the form of a stoichiometric matrix. Furthermore, gaps in the model are identified and reconciled based on established gap filling techniques. Finally, in the third step high-throughput experimental measurements, such as transcriptomic, proteomic and fluxomic, data, are utilized to improve the model accuracy.

et al. [1], incompatibilities in data representation such as metabolites with multiple names/chemical formulae across databases, stoichiometric errors (i.e. elemental or charge imbalances) and incomplete atomistic details (e.g. absence of stereo-specificity, and presence of R-group(s)) are key bottlenecks for the rapid reconstruction of new high-quality metabolic models by combining information from these databases. Recently, databases such as MetRxn [1] have been developed to address these issues by integrating information (of metabolites and reactions) from eight such databases and 90 published metabolic models. Overall, MetRxn (as of December 2013) contains over 44 000 unique metabolites and 35 000 unique reactions that are charge and elementally balanced.

In addition to the GPR information, subcellular localization of metabolic enzymes/reactions is critical to develop the metabolic model of any eukaryotic organism. In this regard, there exist protein localization databases such as PPDB [21] and SUBA [22] for plant species (e.g. *Arabiopsis thaliana* and *Zea mays*). There are also computational algorithms [23] to predict enzyme/reaction localization (when a limited amount of localization data is available) by utilizing the embedded metabolic network and parsimony principal to minimize the number of transporters. However, none of these databases or algorithms is complete or error-proof, which necessitates manual scrutiny before making any final reaction assignments to one or multiple intracellular organelle(s). In addition to databases of metabolic functionalities, there exists a number of knowledgebases of regulation (e.g. RegulonDB [24] (for *E. coli*) and Grassius [25] (for grasses)) and kinetic parameters (e.g. Sabio-RK [26], Ecocyc [27] and Brenda [28] for *E. coli*). Nevertheless, such information is largely incomplete and unavailable for all but a few model organisms, which emphasizes the need to thoroughly refer to published literature sources.

By making use of data from different biological databases, draft metabolic models can be reconstructed in an automated [18,29–33] or semi-automated [34••,35–37] fashion. Automated methods are fast and require minimal user input while semi-automated methods are slower and require user feedback and inspection. Automated methods such as SEED [29], BioNetBuilder [30] and ReMatch [31] can integrate data from several databases. However, the user is responsible for assessing the accuracy of the network gap filling step, removing thermodynamically infeasible cycles (e.g. using the loopless FBA method [38]) and customizing the biomass composition to the organism of interest. Semi-automated methods (e.g. RAVEN [34••], MicrobeFlux [35] and other works, as highlighted in [2,3,36,37]), make use of not only available databases but also published models of closely related species. These methods allow for user-driven gap filling and growth-discrepancy reconciliation measures and use biomass compositions based on experimental measurements whenever available. However, the existing semi-automated algorithms often create thermodynamically infeasible cycles while reconciling any network gaps [39] or fixing growth inconsistencies [40,41]. Overall, the automated methods are very useful for developing initial draft models, whereas semi-automated methods can refine these draft models to bring them to a required completion level.

Estimates for free energy changes of reactions are frequently used to impose thermodynamic constraints on reaction fluxes, metabolite concentrations and kinetic parameters as highlighted elsewhere [42,43]. The group contribution method [44] or recently improved group contribution method [45•] can be utilized to estimate the reaction Gibbs free energy and ultimately predict the reaction direction. Furthermore, Hamilton *et al.* have developed Thermodynamics-based Metabolic Flux Analysis (TMFA) [46] to quantify metabolite concentrations and reaction free energy ranges and examine the effect of thermodynamic constraints on the allowable flux space (of the *i*JR904 *E. coli* model) that improve model performance such as gene essentiality prediction. Although TMFA provides some idea about directionality, the thermodynamic constraints can be too wide. Therefore, as shown in the *i*AF1260 [47] *E. coli* model, literature survey still remains the best source for assigning reaction directionality.

Integration of omics data in metabolic models

In this section we review recent developments in integrating high-throughput omics data with metabolic models and critically analyze their contribution towards improving the genotype-phenotype prediction and metabolic network properties. Due to the underdetermined nature of genome-scale metabolic models, a lot of effort has been expended at improving the accuracy of estimation for the reaction fluxes. Metabolic Flux Analysis (MFA) [48] is a unique resource for quantifying internal metabolic fluxes by using relative enrichment of substrate labels from isotope labeling experiments (ILE) as additional information [49–51]. Detailed atom-transition information for each reaction involved in the MFA network is collected either from literature (for well-studied pathways), databases [52,53], or from motif-searching optimization algorithms [54,55•]. Subsequently, the fluxes and their confidence intervals in the network are estimated by minimizing the sum-squared error between experimental and simulated mass isotopomer distribution (MID) data using different optimization frameworks. Recent advances in the systematic identification of the input substrate labels [56] and the design of labeling experiments (e.g. [57•]) has improved the accuracy and scope of flux estimation. The inferred flux data could then be integrated into metabolic models for further sharpening the allowable flux ranges of the remaining reactions in

the model (using Flux Variability Analysis). A key impediment of MFA is that it is generally applied to core models of metabolism [58] spanning less than 5% [59] of a genome-scale metabolic model. As a result, flux information is generally available for the central carbon metabolism of the organism with limited information on flux redirections in other parts of metabolic network in response to genetic or environmental perturbations. In addition, the results are sensitive to the selection of the metabolic network used to fit the labeling data [59]. Even though recent attempts have been made at constructing large-scale MFA networks using the flux coupling method [60] and the elementary carbon modes approach [61], flux analysis at a genome-scale level has not been attempted yet.

Metabolic model prediction accuracy can further be improved by incorporating transcriptomic/proteomic data as regulatory constraints (see Figure 1). Thus, condition- and tissue-specific metabolic models can be developed to simulate specific phenotypes [62]. The main approaches for integrating omics data to abstract regulation can be broadly classified into two categories [62]: (a) the switch approach (e.g. GIMME and iMAT): on/off reaction fluxes based on threshold expression levels, and, (b) the valve approach (e.g. E-Flux and PROM): regulate reaction fluxes based on relative gene/protein expressions. To circumvent the problem of using arbitrary cutoffs for gene expression, recent approaches [63,64] use absolute gene expression levels as a penalty metric such that the sum of squared error between the gene expressions and their encoded reaction fluxes is minimized. Overall, all of these approaches make the underlying assumption that transcription of genes is linearly correlated with the flux of the reactions they encoded, which is not necessarily accurate [65]. However, faced with a lack of detailed mechanistic information between transcription and enzyme activity, these frameworks provide a 'first-guess' type estimate for correlating genotype with phenotype.

Regulatory signaling and transcription networks have been integrated as separate modules with metabolic networks [66,67]. Generally, these are simulated as boolean networks where information from signaling molecules and transcription factors is carried as on-off signals to target proteins. Similar frameworks have also been constructed for translation and post-translational regulation [68,69]. Besides providing a mechanistic basis for correlating the genotype with the observed phenotype, these integrated frameworks have the added advantage of being dynamic in nature. Each module is assumed to be in an independent quasi-steady state during a specific time interval, and is updated for the next interval by solving a system of ordinary differential equations (ODEs) of the variables (e.g. enzyme and metabolite concentrations, transcription factors, and mRNA abundances) interacting at the interface of two modules. More recently, this framework has been extended to construct the first whole-cell model of *M. genitalium* [13••], where 28 cellular functions designed as distinct modular networks have been integrated into a whole-cell dynamic framework interacting at the edges with ODEs of eight types of common variables. The whole-cell network couples metabolic and non-metabolic functions, as well as temporal information of protein localization and cell replication. However, it requires a detailed mechanistic approach to accurately describe transcription, translation and regulation of the enzyme activities, which is seldom available. In addition, the assumption that each module is at a quasi-steady state within the same time interval may not be universally applicable. Nevertheless, the whole-cell model framework is a major landmark in the reconstruction of integrated metabolic networks.

Integrated frameworks for metabolic model development discussed so far do not use detailed mechanistic relations to link gene expressions with reaction fluxes. The ME model framework [12,70••] has been developed to provide a detailed mechanistic basis to quantify transcription of mRNA, translation of proteins, formation of protein complexes, catalysis of reactions and formation of macromolecules. Similar to flux balance analysis (FBA), simulations using ME models minimize the cellular machinery required to sustain an experimentally observed growth rate, where protein dilution is coupled with the growth of the organism. This framework can predict gene and protein expressions with reasonable accuracy along with an improved prediction for reaction fluxes. The ME model is also able to drive discovery of protein regulation. Despite not accounting for any post-transcriptional regulation, the ME framework provides a significant step towards a systems-wide quantitative description of biological processes.

Several attempts have also been made to link enzyme activities and metabolite concentrations with the reaction fluxes of detailed mechanistic networks. Detailed kinetic models have been constructed using steady-state phenotype information for the wild-type organism and several of its mutants. For example, Cotton *et al.* [71] have constructed the kinetic model of central metabolism for *E. coli*, where the kinetic parameters are identified by minimizing the error between the experimental and model-predicted values of metabolite concentrations and enzyme activity for the wild-type and several of its single gene mutants [72]. The kinetic expressions are imported from an earlier kinetic model for *E. coli* [73]. Likewise, Vital Lopez *et al.* [74] have constructed a kinetic network for *E. coli* central metabolism spanning over 100 reactions using mass action kinetic expressions derived from transcriptomic and fluxomic data. The major restrictions in these models are either the size of the network (for the first one [71]), or the accuracy of the kinetic expressions (for the latter one [74]). Such limitations could be resolved by using the ensemble modeling approach [75] where each

reaction is decomposed into its elementary steps (with detailed regulations, available from databases (e.g. BRENDA [28]), and the ensemble of kinetic models is filtered using fluxomic data for mutants. Genome-scale kinetic models using approximate mass action [76] or lin-log kinetics [77] have also been developed. In the latter approach [77], the kinetic parameters are estimated from metabolomic information and FBA. While these methods require significantly more refinement, especially in the construction of the kinetic expressions, they delineate a strategy for future construction of high confidence, integrated metabolic networks linking the genome to the observable phenotypes.

Concluding remarks

Metabolic network models play an important role in quantitatively assessing the allowable metabolic phenotype of an organism and thereby can be deployed to guide metabolic engineering, synthetic biology and/or drug targeting interventions. Through the coordinated use of biological databases, model building strategies and high-throughput omics-data integration techniques both the quality and scope of metabolic models is increasing. However, significant knowledge gaps and a lack of best-practice methodologies require additional scrutiny. For example, delineating the effect of different levels of regulation (transcriptional, translational and/or post-translational) on metabolic flux would help establish the connectivity and directionality of regulation in metabolic models. In addition, the design of labeling protocols that will enable the elucidation of metabolic fluxes beyond core metabolism in a high-throughput manner for a number of genetic and/or environmental perturbations will provide the basis for the parameterization and construction of more predictive metabolic models. Finally, the adoption of common standards in metabolite and reaction description will speed up sharing of information across database resources. By integrating 'best-practice' lessons learned from model organisms the development of systematic workflows will facilitate the construction of high-quality metabolic models for less studied organisms.

Acknowledgements

This study was supported by two grants from the United States Department of Energy, Biological and Environmental Research (DOE-BER) grant DESC0006870 and grant DE-FG02-05ER25684.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kumar A, Suthers PF, Maranas CD: **MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases**. *BMC Bioinform* 2012, **13**:6.

2. Saha R, Suthers PF, Maranas CD: ***Zea mays* iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism**. *PLoS ONE* 2011, **6**:e21784.

3. Saha R, Verseput AT, Berla BM, Mueller TJ, Pakrasi HB, Maranas CD: **Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp, ATCC 51142 and *Synechocystis* sp. PCC 6803**. *PLoS ONE* 2012, **7**:e48285.

4. Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY: **Recent advances in reconstruction and applications of genome-scale metabolic models**. *Curr Opin Biotechnol* 2012, **23**:617-623.

5. Pitkanen E, Rousu J, Ukkonen E: **Computational methods for metabolic reconstruction**. *Curr Opin Biotechnol* 2010, **21**:70-77.

6. •• Esvelt KM, Wang HH: **Genome-scale engineering for systems and synthetic biology**. *Mol Syst Biol* 2013, **9**:641.
Critical review on current technologies for genome-scale engineering that discusses prospects for extending efficient genome modification to new hosts, and, thereby, exploring the continued advances toward designing programmable chasses.

7. Ranganathan S, Suthers PF, Maranas CD: **OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions**. *PLoS Comput Biol* 2010, **6**:e1000744.

8. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nat Protoc* 2010, **5**:93-121.

9. Blazier AS, Papin JA: **Integration of expression data in genome-scale metabolic network reconstructions**. *Front Physiol* 2012, **3**:299.

10. Reed JL: **Shrinking the metabolic solution space using experimental datasets**. *PLoS Comput Biol* 2012, **8**.

11. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, Yi KY, Rhee JH, Lee SY: **Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery**. *Mol Syst Biol* 2011, **7**:460.

12. Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimpe-Rutledge AC, Smith RD, Adkins JN, Zengler K, Palsson BO: **In silico method for modelling metabolism and gene product expression at genome scale**. *Nat Commun* 2012, **3**.

13. •• Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW: **A whole-cell computational model predicts phenotype from genotype**. *Cell* 2012, **150**:389-401.
This work introduces the first whole-cell model of the life-cycle of *M. genitalium*, where multiple cellular functions (from metabolism to DNA replication) were mathematically integrated as separate modules of a dynamic framework to accout for all annotated gene function and facilitate biological discovery.

14. Jerby L, Shlomi T, Ruppin E: **Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism**. *Mol Syst Biol* 2010, **6**.

15. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD *et al.*: **A community-driven global reconstruction of human metabolism**. *Nat Biotechnol* 2013, **31**:419-425.

16. Grafahrend-Belau E, Junker A, Eschenroder A, Muller J, Schreiber F, Junker BH: **Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale**. *Plant Physiol* 2013, **163**:637-647.

17. Pagani I, Liolios K, Jansson J, Chen IMA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC: **The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata**. *Nucleic Acids Res* 2012, **40**:D571-D579.

18. Zhou T: **Computational reconstruction of metabolic networks from KEGG**. *Methods Mol Biol* 2013, **930**:235-249.

19. Chen N, del Val IJ, Kyriakopoulos S, Polizzi KM, Kontoravdi C: **Metabolic network reconstruction: advances in in silico interpretation of analytical information**. *Curr Opin Biotechnol* 2012, **23**:77-82.

20. •• Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD: **Mathematical optimization applications in metabolic networks**. *Metab Eng* 2012, **14**:672-686.

Critical review on optimization-based frameworks developed that address challenges in the analysis and engineering of metabolic networks including exploring model predictions, correcting and improving metabolic models and redesigning metabolic networks for the targeted overproduction of a desired compound.

21. Hieno A, Naznin HA, Hyakumachi M, Sakurai T, Tokizawa M, Koyama H, Sato N, Nishiyama T, Hasebe M, Zimmer AD *et al.*: **ppdb: plant promoter database version 3.0**. *Nucleic Acids Res* 2013.

22. Tanz SK, Castleden I, Hooper CM, Vacher M, Small I, Millar HA: **SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis**. *Nucleic Acids Res* 2013, **41**:D1185-D1191.

23. Mintz-Oron S, Aharoni A, Ruppin E, Shlomi T: **Network-based prediction of metabolic enzymes' subcellular localization**. *Bioinformatics* 2009, **25**:i247-i252.

24. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, Weiss V, Solano-Lira H, Martinez-Flores I, Medina-Rivera A *et al.*: **RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more**. *Nucleic Acids Res* 2013, **41**:D203-D213.

25. Yilmaz A, Nishiyama MY Jr, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E: **GRASSIUS: a platform for comparative regulatory genomics across the grasses**. *Plant Physiol* 2009, **149**:171-180.

26. Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Algaa E, Weidemann A, Sauer-Danzwith H, Mir S *et al.*: **SABIO-RK – database for biochemical reaction kinetics**. *Nucleic Acids Res* 2012, **40**:D790-D796.

27. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M *et al.*: **EcoCyc: fusing model organism databases with systems biology**. *Nucleic Acids Res* 2013, **41**:D605-D612.

28. Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A *et al.*: **BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA**. *Nucleic Acids Res* 2013, **41**:D764-D772.

29. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C: **Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED**. *Methods Mol Biol* 2013, **985**:17-45.

30. Avila-Campillo I, Drew K, Lin J, Reiss DJ, Bonneau R: **BioNetBuilder: automatic integration of biological networks**. *Bioinformatics* 2007, **23**:392-393.

31. Pitkanen E, Akerlund A, Rantanen A, Jouhten P, Ukkonen E: **ReMatch: a web-based tool to construct, store and share stoichiometric metabolic models with carbon maps for metabolic flux analysis**. *J Integr Bioinform* 2008, **5**.

32. Dale JM, Popescu L, Karp PD: **Machine learning methods for metabolic pathway prediction**. *BMC Bioinform* 2010, **11**.

33. Reyes R, Gamermann D, Montagud A, Fuente D, Triana J, Urchueguia JF, de Cordoba PF: **Automation on the generation of genome-scale metabolic models**. *J Comput Biol* 2012, **19**:1295-1306.

34. •• Agren R, Liu LM, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J: **The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum***. *PLoS Comput Biol* 2013, **9**.

RAVEN (Reconstruction, Analysis and Visualization of Metabolic Networks) Toolbox is a software suite for carrying out semi-automated reconstruction of genome-scale models by using published models, KEGG database, gap-filling and quality control measures. It also provides methods for performing simulations, visualizing simulation results and omics data and analyzing the results.

35. Feng X, Xu Y, Chen Y, Tang YJ: **MicrobesFlux: a web platform for drafting metabolic models from the KEGG database**. *BMC Syst Biol* 2012, **6**:94.

36. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD: **A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, iPS189**. *PLoS Comput Biol* 2009, **5**:e1000285.

37. Mueller TJ, Berla BM, Pakrasi HB, Maranas CD: **Rapid construction of metabolic models for a family of *Cyanobacteria* using a multiple source annotation workflow**. *BMC Syst Biol* 2013, **7**:142.

38. Schellenberger J, Lewis NE, Palsson BO: **Elimination of thermodynamically infeasible loops in steady-state metabolic models**. *Biophys J* 2011, **100**:544-553.

39. Satish Kumar V, Dasika MS, Maranas CD: **Optimization based automated curation of metabolic reconstructions**. *BMC Bioinform* 2007, **8**:212.

40. Zomorrodi AR, Maranas CD: **Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data**. *BMC Syst Biol* 2010, **4**.

41. Kumar VS, Maranas CD: **GrowMatch: an automated method for reconciling in silico/in vivo growth predictions**. *PLoS Comput Biol* 2009, **5**:e1000308.

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

43. Soh KC, Hatzimanikatis V: **Network thermodynamics in the post-genomic era**. *Curr Opin Microbiol* 2010, **13**:350-357.

44. Jankowski M, Henry C: **Group contribution method for thermodynamic analysis of complex metabolic networks**. *Biophys J* 2008.

45. • Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, Milo R: **An integrated open framework for thermodynamics of reactions that combines accuracy and coverage**. *Bioinformatics (Oxford, England)* 2012, **28**:2037-2044.

The study introduced Pseudoisomeric Reactant Contribution method that can systematically estimate compound formation energies by using measured '*K*' and *pKa* data. This method extends the group contribution method and can achieve a higher coverage of unmeasured reactions.

46. Hamilton JJ, Dwivedi V, Reed JL: **Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models**. *Biophys J* 2013, **105**:512-522.

47. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information**. *Mol Syst Biol* 2007, **3**.

48. Wiechert W: **^{13}C metabolic flux analysis**. *Metab Eng* 2001, **3**:195-206.

49. Antoniewicz MR, Kelleher JK, Stephanopoulos G: **Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions**. *Metab Eng* 2007, **9**:68-86.

50. Weitzel M, Noh K, Dalman T, Niedenfuhr S, Stute B, Wiechert W: **13CFLUX2 – high-performance software suite for ^{13}C-metabolic flux analysis**. *Bioinformatics* 2013, **29**:143-145.

51. Nargund S, Sriram G: **Mathematical modeling of isotope labeling experiments for metabolic flux analysis**. *Methods Mol Biol* 2014, **1083**:109-131.

52. Blum T, Kohlbacher O: **MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization**. *Bioinformatics* 2008, **24**:2108-2109.

53. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG**. *Nucleic Acids Res* 2006, **34**:D354-D357.

54. Ravikirthi P, Suthers PF, Maranas CD: **Construction of an *E. coli* genome-scale atom mapping model for MFA calculations**. *Biotechnol Bioeng* 2011, **108**:1372-1382.

55. • Latendresse M, Malerich JP, Travers M, Karp PD: **Accurate atom-mapping computation for biochemical reactions**. *J Chem Inf Model* 2012, **52**:2970-2982.

This work introduces a novel minimal wighted edit distance (MWED) method to accurately map the atom transition in chemical reactions using bond propensity of various functional groups as the weighted metric.

56. Antoniewicz MR: **^{13}C metabolic flux analysis: optimal design of isotopic labeling experiments**. *Curr Opin Biotechnol* 2013, **24**:1116-1121.

57. • Leighty RW, Antoniewicz MR: **COMPLETE-MFA: complementary parallel labeling experiments technique for metabolic flux analysis**. *Metab Eng* 2013, **20**:49-55.
Using singly labeled glucose tracers for each of its carbon atoms, this work underlines an novel approach of improving the accuracy and precision of measuring internal metabolic fluxes in living cells.

58. Crown SB, Antoniewicz MR: **Publishing ^{13}C metabolic flux analysis studies: a review and future perspectives**. *Metab Eng* 2013, **20**:42-48.

59. Leighty RW, Antoniewicz MR: **Parallel labeling experiments with [U-^{13}C]glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis**. *Metab Eng* 2012, **14**:533-541.

60. Suthers PF, Chang YJ, Maranas CD: **Improved computational performance of MFA using elementary metabolite units and flux coupling**. *Metab Eng* 2010, **12**:123-128.

61. Pey J, Rubio A, Theodoropoulos C, Cascante M, Planes FJ: **Integrating tracer-based metabolomics data and metabolic fluxes in a linear fashion via elementary carbon modes**. *Metab Eng* 2012, **14**:344-353.

62. Hyduke DR, Lewis NE, Palsson BO: **Analysis of omics data with genome-scale models of metabolism**. *Mol Biosyst* 2013, **9**:167-174.

63. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BO, Hyduke DR: **GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data**. *Bioinformatics* 2013, **29**:2900-2908.

64. Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, Mendes P, Swainston N: **Improving metabolic flux predictions using absolute gene expression data**. *BMC Syst Biol* 2012, **6**:73.

65. Hoppe A: **What mRNA abundances can tell us about metabolism**. *Metabolites* 2012, **2**:614-631.

66. Covert MW, Xiao N, Chen TJ, Karr JR: **Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli***. *Bioinformatics* 2008, **24**:2044-2050.

67. Berestovsky N, Zhou W, Nagrath D, Nakhleh L: **Modeling integrated cellular machinery using hybrid Petri-Boolean networks**. *PLoS Comput Biol* 2013, **9**:e1003306.

68. Wang YC, Chen BS: **Integrated cellular network of transcription regulations and protein–protein interactions**. *BMC Syst Biol* 2010, **4**:20.

69. Fisher CP, Plant NJ, Moore JB, Kierzek AM: **QSSPN: dynamic simulation of molecular interaction networks describing gene regulation, signalling and whole-cell metabolism in human cells**. *Bioinformatics* 2013, **29**:3181-3190.

70. •• O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO: **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction**. *Mol Syst Biol* 2013, **9**:693.
This work extends the metabolism-only representation of genome-scale models to incorporate detailed mechanism for gene transcription and translation of mRNA that paves the way for developing a systems-level quantivative basis for biology.

71. Cotten C, Reed JL: **Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models**. *BMC Bioinform* 2013, **14**:32.

72. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A *et al.*: **Multiple high-throughput analyses monitor the response of *E. coli* to perturbations**. *Science* 2007, **316**:593-597.

73. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M: **Dynamic modeling of the central carbon metabolism of *Escherichia coli***. *Biotechnol Bioeng* 2002, **79**:53-73.

74. Vital-Lopez FG, Wallqvist A, Reifman J: **Bridging the gap between gene expression and metabolic phenotype via kinetic models**. *BMC Syst Biol* 2013, **7**:63.

75. Zomorrodi AR, Lafontaine Rivera JG, Liao JC, Maranas CD: **Optimization-driven identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks**. *Biotechnol J* 2013, **8**:1090-1104.

76. Jamshidi N, Palsson BØ: **Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models**. *Biophys J* 2010, **98**:175-185.

77. Smallbone K, Mendes P: **Large-scale metabolic models: from reconstruction to differential equations**. *Ind Biotechnol* 2013, **9**:179-184.