Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Chemical Engineering

Modeling metabolic systems: the need for dynamics

Hyun-Seob Song, Frank DeVilbiss and Doraiswami Ramkrishna

Living organisms exhibit dynamic shifts of metabolic pathways to cope with various perturbations. Such temporal change of modus operandi is one of the key mechanisms characterizing metabolic behavior. In this review, we highlight the importance and potential of dynamic modeling in understanding and harnessing metabolic systems, in particular, for bioprocess optimization and metabolic engineering. In a field that harbors a variety of approaches, their relative assessment calls for appropriate quantitative measures. Toward this, we present rational criteria for the evaluation of metabolic models, and recent advances in dynamic modeling that include accounting for dynamic regulation at the whole network level, development of frameworks able to handle large-scale networks on the basis of pathway analysis, and whole-cell modeling of a simple microbe.

Addresses
School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA

Corresponding author: Ramkrishna, Doraiswami (ramkrish@ecn.purdue.edu)

Current Opinion in Chemical Engineering 2013, **2**:373–382

This review comes from a themed issue on **Biotechnology and bioprocess engineering**

Edited by **Wei-Shou Hu** and **James C Liao**

For a complete overview see the Issue and the Editorial

Available online 19th September 2013

http://dx.doi.org/10.1016/j.coche.2013.08.004

Introduction

An essential aspect of quantitatively assessing the behavior of any system is through mathematical modeling which is a process of capturing the salient features of the system by compromising its complexity to an extent dictated by the level at which we seek understanding. Such a model is an approximation to the system with a quality that may have varying attributes suitable for different goals. However, a basic quality of a model is its ability to be predictive under circumstances beyond those that were used for its identification. Both the extent of this outreach and the speed with which computation can be accomplished could serve to characterize the quality of the model.

Frequently, different model frameworks for a system become available to a system because of varying premises, necessitating a discrimination strategy to assess their relative effectiveness. Rational measures of comparison would depend on the modeling goals to which reference was made earlier. We will review such measures and demonstrate their use in modeling metabolic systems with some examples.

The focus of this paper is on dynamic modeling which aims to capture the temporal evolution of the system. The basic phenomenological input to such modeling would consequently be the *rates* of the various process components appearing in a dynamic formulation that will necessarily involve differential equations with temporal derivatives. Most process systems feature *nonlinearities* that result in complex dynamic and steady state behaviors. Avenues exist for their mathematical treatment that produces experimental scenarios otherwise unavailable for testing major hypotheses in the model.

In relating the foregoing general comments about modeling to metabolic systems, we are confronted with several peculiar features. First, there is the enormity of the number of reactions, second the elaborate gene regulatory structure in which the syntheses of a multitude of enzymes are controlled by signal transduction processes, and third the allosteric control of the activities of enzymes. Comprehensive accounting for regulatory phenomena is a forbidding task as it pervades through all of metabolism.

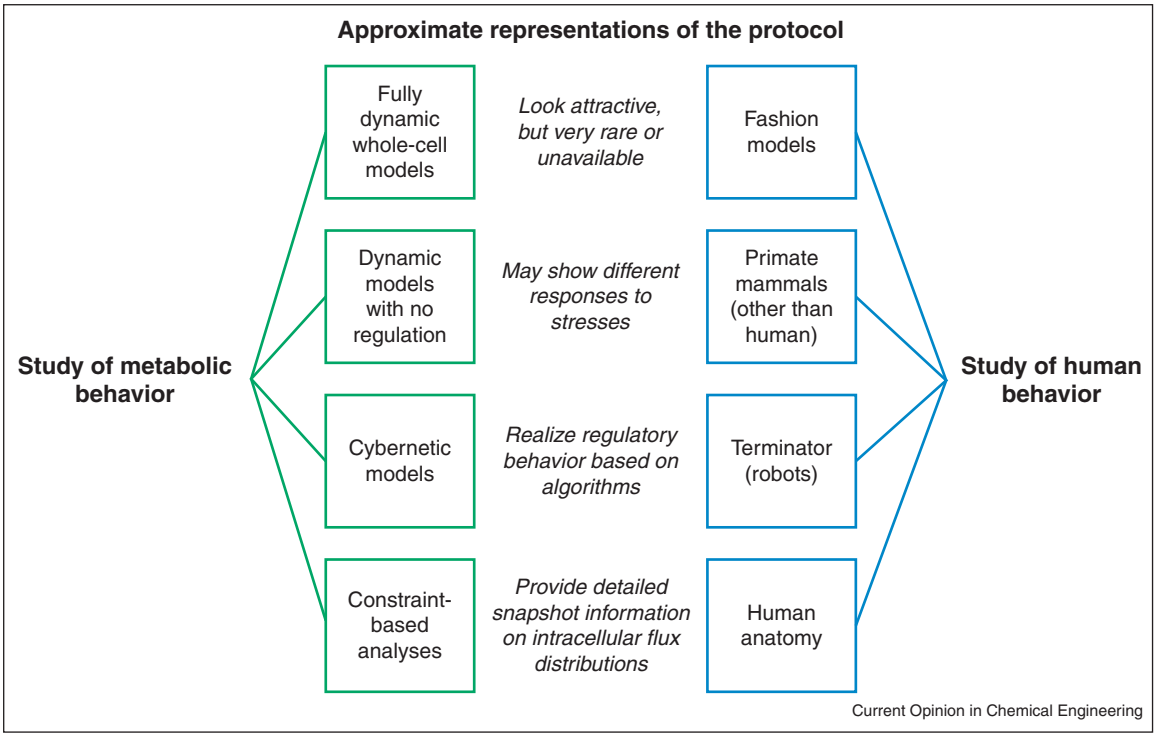
The need for dynamic modeling arises from having to predict performance rates. Thus the *productivity* of a metabolic product is the overriding issue in metabolic engineering as a design purely based on yields is often foiled by impaired growth of the organism thus not reflecting an increase in productivity.

Our objective in this paper is to examine the metabolic modeling landscape with respect to how the modeling goals are realized by the various approaches that are currently available in the literature.

Why is dynamics important in modeling metabolism?

A model is an approximate representation of the protocol. Various frameworks currently available for modeling metabolism include kinetic models, cybernetic models [1••], and constraint-based approaches [2] (Figure 1). Among those, metabolic networks have been studied most intensively by steady state analyses such as constraint-based approaches based on genome-scale stoichiometric models. These frameworks provide detailed images of metabolism at high resolutions, and have served

Figure 1



Different forms of mathematical models used for the study of metabolism. Fully dynamic whole-cell models account for dynamics of individual components, and their interactions and regulation without introducing simplifying approximations. Although attractive, dynamic models at this level are not available yet. Thus, alternatively, we may use other dynamic models based on the kinetic description only, or the cybernetic approach. For the detailed steady-state analysis of metabolic networks, constraint-based approaches such as flux balance analysis are most popular. To facilitate the understanding, models that may be used for the study of human behavior are presented as counterparts of metabolic models.

as a useful tool for the study of metabolism in broad areas, including metabolic engineering [3] and biomedical research [4].

Then, why are we interested in dynamic modeling frameworks? The primary reason for this is, simply and correctly, that metabolism is an intrinsically non-stationary process, and thus, its underlying characteristics are best captured by dynamic descriptions, rather than by still ones. This means that the study of dynamic features is essential for improved understanding and utilization of metabolic systems.

System understanding

Distinct features of a metabolic system are revealed from its dynamic behavior. In the field of system identification, the most informative data on a system are generated by dynamic perturbations determined by the optimal design of experiments [5]. Dynamic data generated as such not only help to characterize a system, but also serve as a test bed for the choice of a right model among alternative candidates [6••].

Understanding of dynamic features is particularly important in the study of cyclic and oscillatory phenomena observed in many living organisms. Dynamic oscillations such as those observed in yeast [7,8] require dynamic models for their elucidation. Similarly, the metabolic cycles in photosynthetic organisms in light and dark periods [9], and in pathogens adapting each developmental stage to their host environment [10] would call for dynamic modeling approaches. These dynamic behaviors and other complex phenomena such as steady state multiplicity [11••,12] are examples of emergent properties that cannot be understood or explained by focusing individual entities apart, but arise when components collectively interact.

Productivity

The value of dynamic tools would be fully realized in the application to industrial bioengineering for the production of biofuels, medicines, and other consumable products. To ensure economic viability of these bioprocesses, it is of crucial importance to maximize the productivity, that is, the *rate* at which the target metabolites are produced.

Enhancement of productivity can be achieved by two complementary approaches [13]: process optimization and metabolic engineering. In regard to the former, dynamic metabolic models have been typically employed for optimal design, configuration, and control of bioreactors [14,15]. Dynamic models are essential in metabolic engineering as well, while they have been infrequently used in that area. Metabolic control analysis (MCA) is a theoretical tool most commonly used in metabolic engineering. MCA identifies controlling fluxes from control coefficients that quantitate the extent of the system response to the changes or perturbations of network parameters, such as the substrate availability or enzyme activity [16,17]. Dynamic experimental data, or a dynamic model providing enzyme kinetics, is an essential input for the implementation of MCA. In silico tools that are currently most popular for metabolic engineering are constraint-based methods based on stoichiometric models. Due to their intrinsic adherence to the steady state assumption, however, the scope of these methods has been limited to *yield*, rather than *productivity*.

Yield and *productivity* are totally different performance criteria, while they have been used in the literature without clear distinction by somewhat confusing terms such as *efficiency* or *effectiveness*. Yield is the ratio of the formed amount of a target product to the consumed amount of a substrate, and thus *dimensionless*. On the other hand, productivity is the substrate consumption rate multiplied by yield, and thus has *the unit of rate*. These two quantities are not necessarily proportional, and often show conflicts [18].

More precisely, two concepts of productivity are available, that is, specific and volumetric [19]. They differ in the definition of substrate consumption rate. That is, specific and volumetric productivities define the substrate consumption rate per unit cell mass and per unit culture volume, and have the units of [g/(gDW h)] and [g/(l h)], respectively. By multiplying cell concentration [gDW/l], the former is converted to the latter. We may view specific productivity as a measure of performance at the cell level, and volumetric productivity at the reactor level. It is important to note that, in addition to yield, growth and uptake rates are major factors determining productivity, and should be accounted for in designing new strains. It is often observed that while the product yield is increased, the production rate is reciprocally low due to reduced growth rate [20–22]. One mistake to avoid is separating the goals, that is, aiming at yield maximization through metabolic engineering, and productivity enhancement by process optimization. The highest productivity comes from the synergistic and consistent efforts in both directions.

Metabolic regulation

Dynamic behaviors of metabolic systems are determined primarily by their regulatory mechanisms. Although complete understanding of metabolic regulation is not available yet, we address two main features.

First, it may be viewed that metabolic regulation takes place at the whole-cell level to achieve a certain goal. In other words, all cellular components directly or indirectly involved in metabolic reactions, such as genes, RNAs, proteins, and metabolites, are elegantly orchestrated toward the global objective, that is, survival. Consequently, metabolic systems can be studied most effectively by a systems biology approach treating a cell as a system regulating all components in the same context. Systems biology explores network behavior of biological systems at the whole system level, in particular their dynamic nature [23], and ultimately aims to uncover the design and operation principles of metabolic networks [24].

Another crucial feature is that cellular regulation is severely constrained by metabolic burden due to the limited internal resources, such as ribosomes, RNA polymerase, and ATP [25]. Consideration of metabolic burden is particularly important in designing new strains, because the addition of synthetic circuits is potentially perturbing the availability of these resources, leading to negative effects on the host cell metabolism. Metabolic models should provide a reasonable description on these regulatory features by which the dynamics of metabolism is determined.

How complex should a model be?

Various classes of dynamic metabolic models are currently available, ranging from simple unstructured to complex whole-cell models. Trade-off exists between complexity (i.e. the number of variables and parameters) and realism (i.e. accuracy and prediction range). For the choice of a most suitable one, therefore, candidate models should be evaluated and compared based on rational criteria such as those provided below.

Modeling goals

Appropriate selection of a particular model can be made considering the goal of modeling [14]. In general, main goals of modeling include understanding of system characteristics, prediction of system behavior in new conditions, and discovery of new strategies for system improvement [26]. Under a given goal, the model complexity is determined by the level at which we explore the system.

In case that several frameworks show similar performance in fulfilling a goal, the simplest model would be preferred if we follow the principle of Occam's razor [27]. In order to determine optimal operating conditions of fermenters, for

instance, we may not need a whole-cell metabolic model. Genome-scale networks used for this purpose may be regarded as a sledgehammer to crack a nut. A much simpler model may provide an accurate description of dynamic interaction between an organism and environment (i.e. culture conditions). Utility of models with reduced complexity is, however, constrained by their narrowed scope resulting from oversimplification or neglect of some essential features. As such, the degree of complexity is to be determined by the modeling goal we set.

Information theoretic tools

The determination of appropriate model complexity for a given biological application can be made more systematically using information theoretic and statistical tools [28•,29,30]. A formalized extension of Occam's razor, particularly, presents a uniquely potent way of approaching the model comparison problem which is referred to as the Minimum Description Length (MDL) principle [31]. MDL is used to reinterpret the model selection problem as one related to data compression [32].

For an arbitrary dataset D and a certain model M, we can define a specific length of the dataset, $L(D)$, and similarly, a length of the model, $L(M)$, that can be measured using information theory. In information theory, $L(D)$ and $L(M)$ are the amount of information it takes to communicate the data set or model encoded into binary across a channel of communication and measuring the number of bits. In a special case that the dataset is perfectly reconstructed by the model without any error of description, the model compresses the data if the inequality $L(M) < L(D)$ is satisfied. It is often, however, that models do not provide an exact description of a given dataset. To perfectly encode D, $L(D)$ is defined by $L(M)$ plus the error between model prediction and data, $L(D|M)$ [33], that is,

$$L(D) = L(D|M) + L(M) \tag{1}$$

This implies that compression is limited by an additional term, $L(D|M)$. The MDL principle states that, for a set of models, the one that compresses the data most or captures best the regularity in data is deemed most useful [34]. $L(M)$ is often represented as a function of the parameter number of a specific model. This model selection criterion provides a measure of how well a model compresses data, and Eqn. (1) applies to both mechanistic and non-mechanistic models.

For illustration, let us consider a set of n data points generated from a process with a given level of noise [35]. One can attempt to describe this dataset using polynomials of varying degrees, from a line to a Lagrange polynomial of degree $n-1$ (Figure 2). The Lagrange polynomial perfectly generates the dataset at the cost of complexity. On the other hand, the linear model is simple but fails to capture the data trend. A statistically sound metric helps to select a model that represents the best compromise between model complexity and accuracy of data fit. These concepts may be readily demonstrated via the application of Akiake Information Criterion (AIC) [36] and Bayesian Information Criterion (BIC) [37] formulated in more digestible terms with a mean squared error approximation of model fit, $\hat{\sigma}_k^2$ [38]:

$$\mathrm{AIC} = \frac{\log(\hat{\sigma}_k^2) + 2(k+1)}{n} \tag{2}$$

$$\mathrm{BIC} = \frac{\log(\hat{\sigma}_k^2) + k\log(n)}{n} \tag{3}$$

where k denotes the number of model parameters. AIC is formulated to estimate the Kullback-Liebler divergence of a model and penalizes parameters less severely than BIC which is built from a Bayesian framework. The AIC and BIC values in Figure 2d show the preference of the polynomial of degree 3, as opposed to higher degree polynomials which fit the dataset with a smaller sum of squared errors.

Robustness test

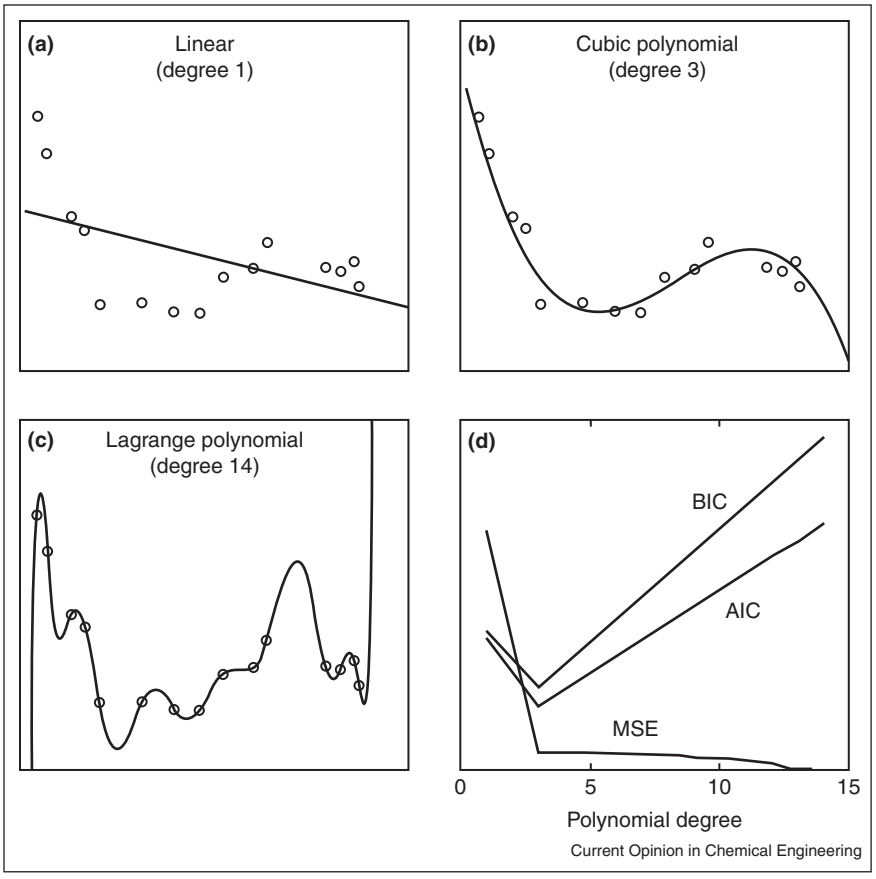
Another useful check for model reliability is robustness test. Robustness is the capability of a system to maintain a certain function in face of perturbations, a fundamental property possessed by living organisms [39,40]. For instance, cells still grow despite frequent genetic mutations, or unfavorable changes of environmental conditions. Therefore, it is important that metabolic models hold a similar robust nature. In general, models that account for regulation tend to be more robust and reliable in comparison to others that do not [41].

Robustness test requires the specification of a performance index. In the analysis of metabolic networks using the frameworks such as flux balance analysis [42] and metabolic pathway analysis [43,44], steady state-based measures (e.g. maximal biomass yield or the number of pathways producing biomass) have been used as this performance function. A more realistic test is, however, to use dynamic performances such as growth rate or uptake rate, in place of yield-based criteria.

Frameworks for dynamic metabolic modeling

In this final section, we provide a brief overview of currently available dynamic modeling frameworks, along with their advantages and limitations, and the connectivity among different approaches. This discussion covers unstructured and structured models, accounting for dynamic regulation at a network level, the usefulness of quasi steady-state (QSS) approximation, and whole-cell modeling.

Figure 2



Evaluation of the model performance with different complexity. For tutorial purpose, data points (15 in total) are generated from randomly selected polynomial with noise added. **(a)** Linear regression, **(b)** cubic polynomial approximation of data, **(c)** fit of Lagrange polynomial of degree 14, **(d)** values of information theoretic measures (AIC and BIC) and mean squared error (MSE) of polynomials of varying degree from 1 to 14.

Unstructured versus structured models

Unstructured models are based on a lumped description of metabolic networks. They assume only a limited number of reaction steps (often, a single step) from substrates to products. These simple models have been used for simulating various growth patterns of a single or multiple strains. In an environment of multiple substrates, cells exhibit a sequential or simultaneous consumption pattern depending on a pair of substrates. These regulatory behaviors have been successfully modeled using the cybernetic control laws [1,45]. This framework is termed Lumped Cybernetic Modeling (LCM). Despite such simplicity in their model structure, LCM has provided accurate simulation of complex dynamic oscillations observed in the aerobic growth of yeast [46]. The same could be modeled using purely kinetic representation with an increased number of parameters. Kinetic models of this kind have been used for the optimal design and control of bioprocesses [47–49].

The scope of model application can be expanded by accounting for detailed structure of metabolic networks. Those structured models are useful in a multitude of applications, including design of new strains for industrial use, investigation of fundamental principles underlying cellular functions, such as robustness, optimality and adaptability [50], as well as bioprocess optimization [51]. Frameworks of interest include metabolic ensemble modeling [52,53], and the approach by Smallbone *et al.* [54]. These methods are able to circumvent the difficulty arising from rigorous parameter estimation by using steady state information in place of time course data. Ensemble modeling builds up a large set of candidate models that achieve a certain steady state flux distribution in a given condition. The subsequent reduction of candidate models to a smaller set is enabled by acquiring additional sets of data from perturbation experiments such as flux shift, for example, in response to enzyme overexpression. Smallbone *et al.* proposed a framework to construct a large-scale kinetic model based on linlog

kinetics. As essential inputs to the model, it is prerequisite to have fluxes and metabolite concentrations at a reference condition in the whole network.

Accounting for dynamic regulation at a network level

For reliable and robust prediction, metabolic models must necessarily incorporate regulation. Accounting for dynamic regulation at a network level poses a serious challenge due to incomplete knowledge on the details of regulatory mechanisms. Although so-called genetically structured models (e.g. [55,56]) are available for a tiny network, they are not considered a whole network-based model.

The cybernetic control principle offers a promising alternative in this regard. The basic hypothesis of cybernetic modeling is that cells allocate limited internal resources in an optimal way so that a given metabolic objective is maximized. Young and Ramkrishna [57] establish a theoretical foundation on the cybernetic control laws based on optimal control theory in a general form, providing a rational way to develop a cybernetic model for a large size metabolic system.

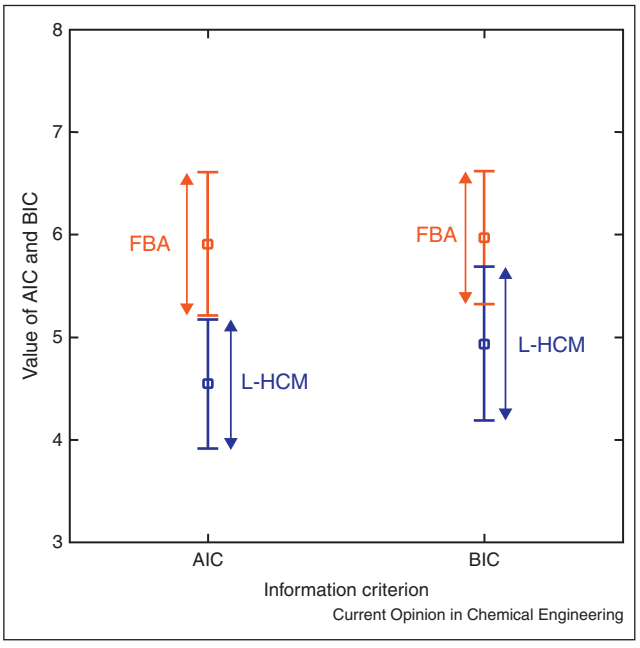
Quasi steady-state (QSS) approximation and QSS models

Developing fully dynamic models for a large scale network is hampered, not only by a heavy computational burden in simulation, but also by a proportionally increased number of parameters. Accurate parameter determination is practically impossible as many of them are insensitive to experimental data. For instance, in a large-scale kinetic single-cell model for mammalian cell cultures, only 37 parameters of 357 in total were found to be estimable within 25% of their nominal values [58].

To avoid this problem, we can consider the dynamics of only the slowly changing variables (such as extracellular metabolites) by assuming fast variables (such as intracellular variables) to be in QSS [59]. With the appropriate identification of the time domain of variables, this treatment significantly reduces the order of dynamic models. The apparent structure of QSS models is as simple as unstructured models due to a small number of parameters, but the range of their application is overlapped with that of fully dynamic models. That is, QSS models predict intracellular fluxes that are immediately redistributed in accordance with the change of extracellular variables, and can be applied to metabolic engineering.

QSS models have two classes: One is constraint-based approach such as dynamic flux balance analysis (DFBA), and the other is metabolic pathway-based frameworks, including macroscopic bioreaction models (MBMs) [60], and hybrid cybernetic models (HCMs) [61,62] and lumped HCMs (L-HCMs) [63,64••].

Figure 3

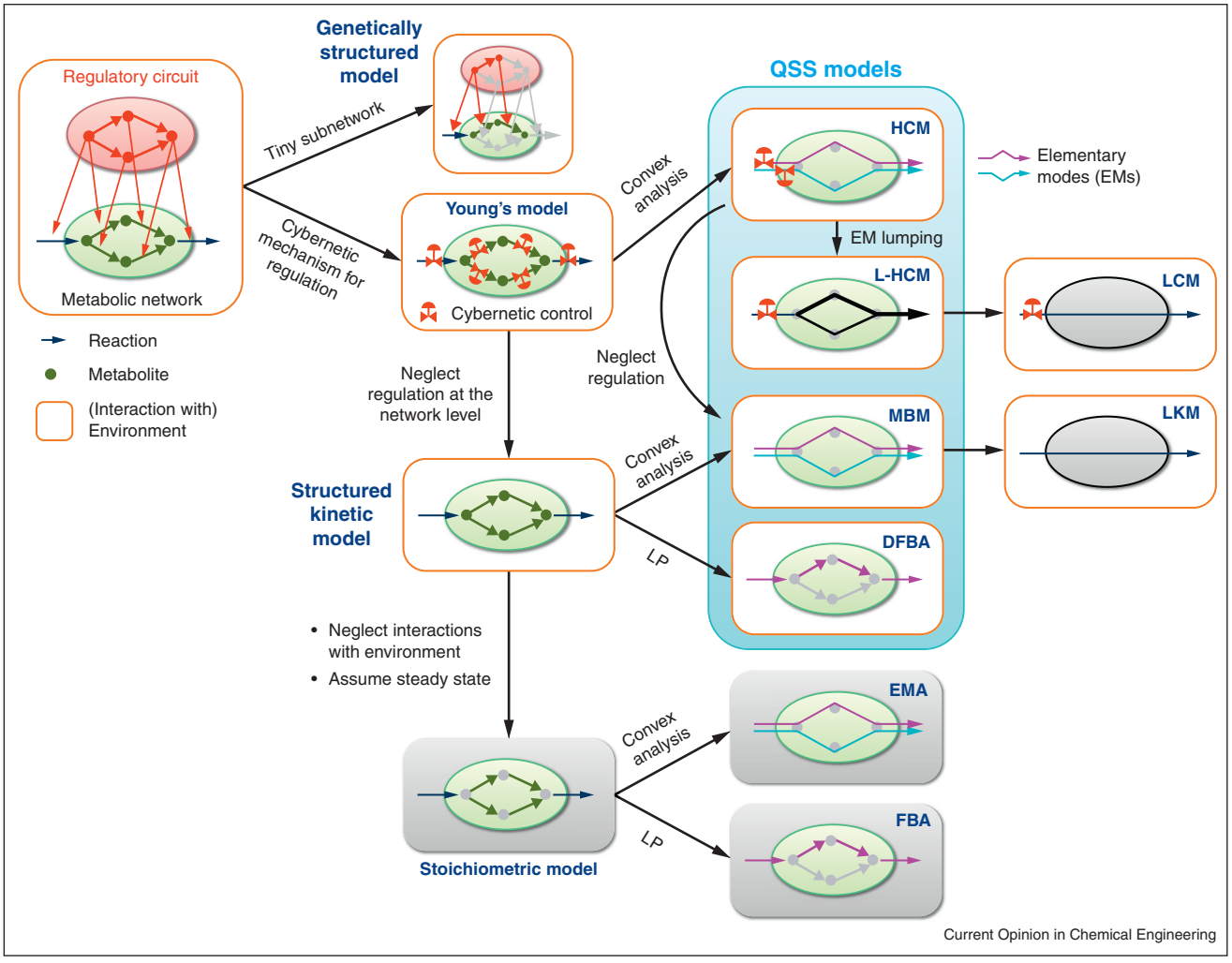


Comparison of FBA and L-HCM with respect to their prediction of intracellular flux distribution of *S. oneidensis*. Metabolic models and flux data are taken from Song *et al.* [69] and Tang *et al.* [73]. The range of predictions is due to variability in interpretation as to how a parameter is defined. In the case of L-HCM, only 7 parameters are trained upon data as the remaining parameters are generally insensitive ones. For FBA, several combinations of substrate uptake levels were used leading to another range of values.

DFBA predicts flux distribution from uptake fluxes (kinetically modeled) using linear programming (LP) such that a chosen objective (e.g. biomass yield) is maximized [65]. The use of LP allows for genome scale network-based dynamic simulation. As the flux vector obtained from LP represents a particular solution among many alternative optimal pathways connecting substrate and products, additional treatments should be considered for reliable simulations [66].

In MBMs and HCMs, on the other hand, a flux distribution is represented as a combination of multiple metabolic pathways called elementary modes (EMs) [67]. HCMs view EMs as metabolic options and describe the uptake flux to be optimally distributed among EMs such that a metabolic objective (such as growth rate or total uptake flux) is maximized. Dynamic regulation of EM fluxes is not considered in MBMs. In both models, however, the large number of EMs adds considerably to the kinetic parameters associated with substrate uptake rates leading to an overparameterization problem as measurements available for identification are restricted to the limited number of extracellular variables. In this connection, the work of Song and Ramkrishna [68] shows

Figure 4



Connectivity of metabolic models. Green and red networks imply metabolic and genetic circuits. Arrows and circles indicate fluxes and metabolites, respectively. Solid green circles in the network denote account for dynamics of intracellular metabolites, while gray circles assume steady state. The valves on the pathways indicate flux regulation by the cybernetic control laws. Models with open orange box consider interactions with environment, while gray box does not. *Acronyms*: DFBA, dynamic flux balance analysis; EMA, elementary flux analysis; FBA, flux balance analysis; HCM, hybrid cybernetic model; LCM, lumped cybernetic model; L-HCM, lumped hybrid cybernetic model; LKM, lumped kinetic model; LP, linear programming; MBM, macroscopic bioreaction model; QSS, quasi steady state.

how substantial reduction can be effected on the number of EMs from an inspection of experimental data on the yield vector space for extracellular products.

Recently, Song and Ramkrishna [63,64••] developed the L-HCM which integrates the merits of LCM and HCM. That is, the L-HCMs describe distribution of cellular resources in the network in terms of lumped EMs which are weighted averages of EMs in families classified according to similarity of metabolic function. As a result, L-HCMs can accurately predict dynamic cellular responses to environmental changes [69] and genetic perturbations [70••], from a limited set of measurements. Extension of the L-HCM framework to genome scales is in progress using optimization-based algorithms (e.g. [71•,72]) that enable the sequential identification of EMs from a large size network.

L-HCM and DFBA are two interesting frameworks to compare due to similarities in the data requirement for parameter identification, and the size of manageable metabolic networks. Information theoretic measures introduced above are used to compare FBA and L-HCM for flux predictions of *Shewanella oneidensis* [69]. Starting from the flux data [73] as a means for testing the accuracy of model predictions, AIC and BIC values were

calculated for each respective model (Figure 3). It is shown that L-HCM delivers, relative to FBA, accuracy at a low cost of model complexity.

Whole-cell modeling

In a recent work by Karr *et al.* [74,75••], the life cycle of the human pathogen *Mycoplasma genitalium* was modeled at the entire cell level. This whole-cell model accounts for all essential biological processes such as DNA replication, RNA transcription, protein synthesis, metabolism, and even cell division, showing how a whole-cell model can be constructed by combining all genome-level information of individual components and their interactions. This may not be considered, however, a *fully* dynamic model due to the difficulty in identifying a large set of parameters, and uncertainties on the details of dynamic regulatory mechanisms. Instead, simulation was performed by integrating probabilistic and constraint-based approaches.

Model connectivity

Figure 4 displays the level of simplification introduced to metabolic models and their mutual connectivity. Although no existing models may be satisfactory in all respects, their reliability is determined by the capability of appropriately accounting for interactions among genotype, phenotype and environment.

Conclusions and future prospects

Dynamic shifts of metabolic states are an inherent strategy of living organisms to ensure the survival in face of various irregular perturbations. Such defense mechanisms determine the key metabolic characteristics of a given organism, which manifest themselves most distinctively during the occurrence of dynamic changes. Therefore, dynamic models provide the most appropriate framework to understand and harness metabolic systems. In particular, they serve as an indispensable tool for productivity enhancement, which is beyond the scope of steady state-based models. Despite the great potential of dynamic models in a wide range of applications, their use in metabolic engineering has been less frequent than steady state network analyses. The main reason for this is the difficulty in extending dynamic models to a large scale network. To resolve this issue, recent frameworks integrate network analysis with a small number of dynamic (extracellular) variables based on QSS approximation.

Due to the ability of QSS models to incorporate large-scale networks, they are expected to serve as a useful alternative to fully dynamic models currently unavailable. Success of QSS models may be enhanced if regulation is accounted for, as by using cybernetic control laws. In parallel with QSS modeling, we anticipate significant progress in whole-cell modeling. Even so, the value of QSS models lies in the trade-off between complexity and accuracy.

Acknowledgment

The authors are grateful for the support by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

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. •• Ramkrishna D, Song HS: **Dynamic models of metabolism: review of the cybernetic approach**. *AIChE J* 2012, **58**:986-997.
This article overviews the development of the cybernetic modeling framework over the past three decades, highlighting the capability of accounting for dynamic regulation at the whole network level and the prospect of incorporating genome-scale networks.

2. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245-248.

3. Xu C, Liu L, Zhang Z, Jin D, Qiu J, Chen M: **Genome-scale metabolic model in guiding metabolic engineering of microbial improvement**. *Appl Microbiol Biotechnol* 2013, **97**:519-539.

4. Kim HU, Sohn SB, Lee SY: **Metabolic network modeling and simulation for drug targeting and discovery**. *Biotechnol J* 2012, **7**:330-342.

5. Asprey S, Macchietto S: **Designing robust optimal dynamic experiments**. *J Process Control* 2002, **12**:545-556.

6. •• Flassig RJ, Sundmacher K: **Optimal design of stimulus experiments for robust discrimination of biochemical reaction networks**. *Bioinformatics* 2012, **28**:3089-3096.
The authors present an optimal design of experiments for robust discrimination. The suggested sigma-point approach is useful for evaluating dynamic biological models with widely distributed parameters due to intrinsic variability as well as experimental variances. Thanks to several numerical such as derivative free estimation of expectation and covariance, the algorithm is able to handle non-smooth functions such as the one included in cybernetic models.

7. Finn RK, Wilson RE: **Population dynamics of a continuous propagator for microorganisms**. *J Agric Food Chem* 1954, **2**:66-69.

8. Tu BP, Kudlicki A, Rowicka M, McKnight SL: **Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes**. *Science* 2005, **310**:1152-1158.

9. Johnson CH, Egli M, Stewart PL: **Structural insights into a circadian oscillator**. *Science* 2008, **322**:697-701.

10. Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z: **Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *Plasmodium falciparum***. *Mol Cell Proteomics: MCP* 2011, **10** M110 006411.

11. •• Kim JI, Song HS, Sunkara SR, Lali A, Ramkrishna D: **Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat**. *Biotechnol Progr* 2012, **28**:1160-1166.
This article provides cybernetic model-based predictions of steady state multiplicity occurring in anaerobic growth of *Escherichia coli*. The authors emphasize that the multiplicity is a consequence of the nonlinear regulation, rather than of only nonlinear kinetics.

12. Lei F, Rotboll M, Jorgensen SB: **A biochemically structured model for *Saccharomyces cerevisiae***. *J Biotechnol* 2001, **88**:205-221.

13. Song HS, Ramkrishna D: **Issues with increasing bioethanol productivity: a model directed study**. *Korean J Chem Eng* 2010, **27**:576-586.

14. Gernaey KV, Lantz AE, Tufvesson P, Woodley JM, Sin G: **Application of mechanistic models to fermentation and biocatalysis for next-generation processes**. *Trends Biotechnol* 2010, **28**:346-354.

15. Song HS, Kim SJ, Ramkrishna D: **Synergistic optimal integration of continuous and fed-batch reactors for enhanced productivity of lignocellulosic bioethanol**. *Ind Eng Chem Res* 2011, **51**:1690-1696.

16. Moreno-Sanchez R, Saavedra E, Rodriguez-Enriquez S, Olin-Sandoval V: **Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways**. *J Biomed Biotechnol* 2008, **2008**:597913.

17. Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW: **Metabolic control analysis in drug discovery and disease**. *Nat Biotechnol* 2002, **20**:243-249.

18. Song HS, Morgan JA, Ramkrishna D: **Towards increasing the productivity of lignocellulosic bioethanol: rational strategies fueled by modeling**. In *Bioethanol.* Edited by Lima MAP. InTech; 2011.

19. Ozturk SS, Hu W-S: *Cell Culture Technology for Pharmaceutical and Cell-based Therapies*. CRC Press; 2006.

20. Chu BC, Lee H: **Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation**. *Biotechnol Adv* 2007, **25**:425-441.

21. Jeppsson M, Johansson B, Hahn-Hagerdal B, Gorwa-Grauslund MF: **Reduced oxidative pentose phosphate pathway flux in recombinant xylose-utilizing *Saccharomyces cerevisiae* strains improves the ethanol yield from xylose**. *Appl Environ Microbiol* 2002, **68**:1604-1609.

22. Anesiadis N, Cluett WR, Mahadevan R: **Dynamic metabolic engineering for increasing bioprocess productivity**. *Metab Eng* 2008, **10**:255-266.

23. Cassman M: **Barriers to progress in systems biology**. *Nature* 2005, **438**:1079.

24. Alon U: *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London: Chapman & Hall; 2006.

25. Glick BR: **Metabolic load and heterologous gene-expression**. *Biotechnol Adv* 1995, **13**:247-261.

26. Bailey JE: **Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities**. *Biotechnol Progr* 1998, **14**:8-20.

27. Gauch HG: *Scientific Method in Practice*. Cambridge University Press; 2003.

28. Penny W: **Comparing dynamic causal models using AIC, BIC and free energy**. *Neuroimage* 2012, **59**:319-330.
• Author develops framework to identify models using information criteria to distinguish a generating model from class of candidate models. This embodies a practical demonstration of concepts in this review.

29. McDonald CP, Urban NR: **Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models**. *Ecol Model* 2010, **221**:428-432.

30. Symonds MR, Moussalli A: **A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion**. *Behav Ecol Sociobiol* 2011, **65**:13-21.

31. Roos T: **Short course: introduction to information-theoretic modeling**. *Fifth Brazilian Conference on Statistical Modelling in Insurance and Finance*; *Maresias, Brazil: 2011*.

32. Barron A, Rissanen J, Yu B: **The minimum description length principle in coding and modeling**. *IEEE Trans Inform Theory* 1998, **44**:2743-2760.

33. Grünwald PD: *The Minimum Description Length Principle*. MIT Press; 2007.

34. Rissanen J: *Information and Complexity in Statistical Modeling*. Incorporated: Springer Publishing Company; 2007.

35. Kirk P, Thorne T, Stumpf MP: **Model selection in systems and synthetic biology**. *Curr Opin Biotechnol* 2013, **24**:767-774.

36. Akaike H: *Information Theory and An Extension of the Maximum Likelihood Principle. International Symposium on Information Theory*. 2nd edn.. Armenian SSR: Tsahkadsor; 1973,267-281.

37. Schwarz G: **Estimating the dimension of a model**. *Annals Statistics* 1978, **6**:461-464.

38. McQuarrie AD, Tsai C-L: *Regression and Time Series Model Selection*. Singapore: World Scientific; 1998.

39. Kitano H: **Towards a theory of biological robustness**. *Mol Syst Biol* 2007, **3**:137.

40. Kitano H: **Biological robustness**. *Nat Rev Genet* 2004, **5**:826-837.

41. Young JD, Henne KL, Morgan JA, Konopka AE, Ramkrishna D: **Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control**. *Biotechnol Bioeng* 2008, **100**:542-559.

42. Edwards JS, Palsson BO: **Robustness analysis of the *Escherichia coli* metabolic network**. *Biotechnol Progr* 2000, **16**:927-939.

43. Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S: **Structural robustness of metabolic networks with respect to multiple knockouts**. *J Theor Biol* 2008, **252**:433-441.

44. Wilhelm T, Behre J, Schuster S: **Analysis of structural robustness of metabolic networks**. *Syst Biol (Stevenage)* 2004, **1**:114-120.

45. Kompala DS, Ramkrishna D, Jansen NB, Tsao GT: **Investigation of bacterial-growth on mixed substrates — experimental evaluation of cybernetic models**. *Biotechnol Bioeng* 1986, **28**:1044-1055.

46. Jones KD, Kompala DS: **Cybernetic model of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures**. *J Biotechnol* 1999, **71**:105-131.

47. Wang L, Ridgway D, Gu T, Moo-Young M: **Kinetic modeling of cell growth and product formation in submerged culture of recombinant *Aspergillus niger***. *Chem Eng Commun* 2008, **196**:481-490.

48. Charalampopoulos D, Vázquez JA, Pandiella SS: **Modelling and validation of *Lactobacillus plantarum* fermentations in cereal-based media with different sugar concentrations and buffering capacities**. *Biochem Eng J* 2009, **44**:96-105.

49. Vázquez JA, Murado MA: **Unstructured mathematical model for biomass, lactic acid and bacteriocin production by lactic acid bacteria in batch fermentation**. *J Chem Technol Biotechnol* 2008, **83**:91-96.

50. Steuer R, Gross T, Selbig J, Blasius B: **Structural kinetic modeling of metabolic networks**. *Proc Natl Acad Sci U S A* 2006, **103**:11868-11873.

51. Çelik E, Çalk P, Oliver SG: **A structured kinetic model for recombinant protein production by Mut+ strain of Pichia pastoris**. *Chem Eng Sci* 2009, **64**:5028-5035.

52. Tan YK, Liao JC: **Metabolic ensemble modeling for strain engineers**. *Biotechnol J* 2012, **7**:343-353.

53. Tran LM, Rizk ML, Liao JC: **Ensemble modeling of metabolic networks**. *Biophys J* 2008, **95**:5606-5617.

54. Smallbone K, Simeonidis E, Swainston N, Mendes P: **Towards a genome-scale kinetic model of cellular metabolism**. *Bmc Systems Biol* 2010:4.

55. Lee SB, Bailey JE: **Genetically structured models for lac promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids — lac operator function**. *Biotechnol Bioeng* 1984, **26**:1372-1382.

56. Kotte O, Zaugg JB, Heinemann M: **Bacterial adaptation through distributed sensing of metabolic fluxes**. *Mol Syst Biol* 2010, **6**:355.

57. Young JD, Ramkrishna D: **On the matching and proportional laws of cybernetic models**. *Biotechnol Progr* 2007, **23**:83-99.

58. Sidoli FR, Mantalaris A, Asprey SP: **Toward global parametric estimability of a large-scale kinetic single-cell model for mammalian cell cultures**. *Ind Eng Chem Res* 2005, **44**:868-878.

59. Song HS, Ramkrishna D: **When is the quasi-steady-state approximation admissible in metabolic modeling? When**

admissible, what models are desirable?. *Ind Eng Chem Res* 2009, **48**:7976-7985.

60. Provost A, Bastin G, Agathos SN, Schneider YJ: **Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells**. *Bioprocess Biosystems Eng* 2006, **29**:349-366.

61. Kim JI, Varner JD, Ramkrishna D: **A hybrid model of anaerobic *E. coli* GJT001: combination of elementary flux modes and cybernetic variables**. *Biotechnol Progr* 2008, **24**:993-1006.

62. Song HS, Morgan JA, Ramkrishna D: **Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose**. *Biotechnol Bioeng* 2009, **103**:984-1002.

63. Song HS, Ramkrishna D: **Prediction of metabolic function from limited data: lumped hybrid cybernetic modeling (L-HCM)**. *Biotechnol Bioeng* 2010, **106**:271-284.

64. •• Song HS, Ramkrishna D: **Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function**. *Biotechnol Bioeng* 2011, **108**:127-140.

The authors present the features of lumped hybrid cybernetic model that can incorporate large scale metabolic networks but be identified with limited data. This paper reports accurate predictions of dynamic data of various different strains of *Escherichia coli* in anaerobic environments. Potentially, this framework can be extended to genome-scale networks by integrating efficient optimization algorithms developed for the computation of metabolic pathways.

65. Mahadevan R, Edwards JS, Doyle FJ: **Dynamic flux balance analysis of diauxic growth in *Escherichia coli***. *Biophys J* 2002, **83**:1331-1340.

66. Hoffner K, Harwood SM, Barton PI: **A reliable simulator for dynamic flux balance analysis**. *Biotechnol Bioeng* 2013, **110**:792-802.

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

68. Song HS, Ramkrishna D: **Reduction of a set of elementary modes using yield analysis**. *Biotechnol Bioeng* 2009, **102**:554-568.

69. Song HS, Ramkrishna D, Pinchuk GE, Beliaev AS, Konopka AE, Fredrickson JK: **Dynamic modeling of aerobic growth of *Shewanella oneidensis*. Predicting triauxic growth, flux distributions, and energy requirement for growth**. *Metab Eng* 2013, **15**:25-33.

70. •• Song HS, Ramkrishna D: **Prediction of dynamic behavior of mutant strains from limited wild-type data**. *Metab Eng* 2012, **14**:69-80.

The dynamic responses of *Escherichia coli* to various gene knock-outs are reliably predicted by the lumped hybrid cybernetic model, the parameters of which are identified from only a few wild-type data. This successful prediction offers the model as a tool for dynamic metabolic engineering.

71. • Rezola A, de Figueiredo LF, Brock M, Pey J, Podhorski A, Wittmann C *et al.*: **Exploring metabolic pathways in genome-scale networks via generating flux modes**. *Bioinformatics* 2011, **27**:534-540.

This work presents a mixed-integer linear programming which enable sequential computation of metabolic pathways from genome-scale networks.

72. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S *et al.*: **Computing the shortest elementary flux modes in genome-scale metabolic networks**. *Bioinformatics* 2009, **25**:3158-3165.

73. Tang YJ, Martin HG, Deutschbauer A, Feng XY, Huang R, Llora X *et al.*: **Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations**. *Biotechnol Progr* 2009, **25**:1254-1259.

74. Karr JR, Sanghvi JC, Macklin DN, Arora A, Covert MW: **WholeCellKB: model organism databases for comprehensive whole-cell models**. *Nucleic Acids Res* 2013, **41**:D787-D792.

75. •• Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr *et al.*: **A whole-cell computational model predicts phenotype from genotype**. *Cell* 2012, **150**:389-401.

The authors present a whole-cell model that describes molecular interactions at the entire genome. They integrate twenty eight different submodules for simulation. The resulting model not only shows close relation to existing experimental data, but also reveals new findings previously undetected in experiments.