Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

The principles of whole-cell modeling

Jonathan R Karr[1], Koichi Takahashi[2,3] and Akira Funahashi[4]

Whole-cell models which comprehensively predict how phenotypes emerge from genotype promise to enable rational bioengineering and precision medicine. Here, we outline the key principles of whole-cell modeling which have emerged from our work developing bacterial whole-cell models: single-cellularity; functional, genetic, molecular, and temporal completeness; biophysical realism including temporal dynamics and stochastic variation; species-specificity; and model integration and reproducibility. We also outline the whole-cell model construction process, highlighting existing resources. Numerous challenges remain to achieving fully complete models including developing new experimental tools to more completely characterize cells and developing a strong theoretical understanding of hybrid mathematics. Solving these challenges requires collaboration among computational and experimental biologists, biophysicists, biochemists, applied mathematicians, computer scientists, and software engineers.

Addresses
[1] Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2] RIKEN Quantitative Biology Center, RIKEN, Osaka 565-0874, Japan
[3] Institute for Advanced Biosciences, Keio University, Fujisawa 252-8520, Japan
[4] Department of Biosciences and Informatics, Keio University, Yokohama 223-8522, Japan

Corresponding author: Karr, Jonathan R (karr@mssm.edu)

Current Opinion in Microbiology 2015, **27**:18–24

This review comes from a themed issue on **Microbial systems biology**

Edited by **Eric Brown** and **Nassos Typas**

http://dx.doi.org/10.1016/j.mib.2015.06.004

Introduction

Whole-cell models are computational models which describe how phenotype arises from genotype [1,2,3•]. The primary goal of whole-cell modeling is to enable rational bioengineering and precision medicine. Combined with genome synthesis [4] and transplantation [5], whole-cell models could enable bioengineers to maximize objectives such as biofuel production by optimizing genomes [6,7]. Such models could also enable clinicians to individualize therapy [8–10]. Furthermore, whole-cell models could be powerful scientific tools.

Shuler *et al.* introduced the first coarse-grained ordinary differential equation whole-cell model in 1979 [11,12•]. Twenty years later, when sequencing provided the first biological parts list, Tomita *et al.* [13•] developed the first large-scale fine-grained dynamical model. Researchers have continued to develop increasingly sophisticated dynamical models [14–16]. In parallel, Varma and Palsson used flux balance analysis (FBA) to create the first static genome-scale metabolic models [17]. The latest FBA models represent over 1000 genes [18]. Researchers have since expanded FBA to represent transcriptional regulation [19], transcription and translation [20•], and signaling [21]. Logical methods have also been used [22]. Recently, we and others used a hybrid methodology to construct the first dynamical model which represents every known molecular species and gene function [23••,24,25]. Simultaneously, Roberts *et al.* developed the first cell-scale structural model [26•].

Here, we describe the core principles of whole-cell modeling. We also outline our model construction process, highlighting existing tools and the challenges to achieving complete models.

The principles of whole-cell modeling

Building on Roberts' discussion [27], we outline 11 fundamental and practical principles of whole-cell modeling to illuminate a path toward complete models (Figure 1).

Single-cellularity

First, whole-cell models should represent individual cells. Single-cell models can account for how temporal dynamics and stochastic variation affect behavior. Single cells are also tractable because they are independent and directly result from molecular biochemistry. Furthermore, single-cell models can take advantage of the growing wealth of single-cell data.

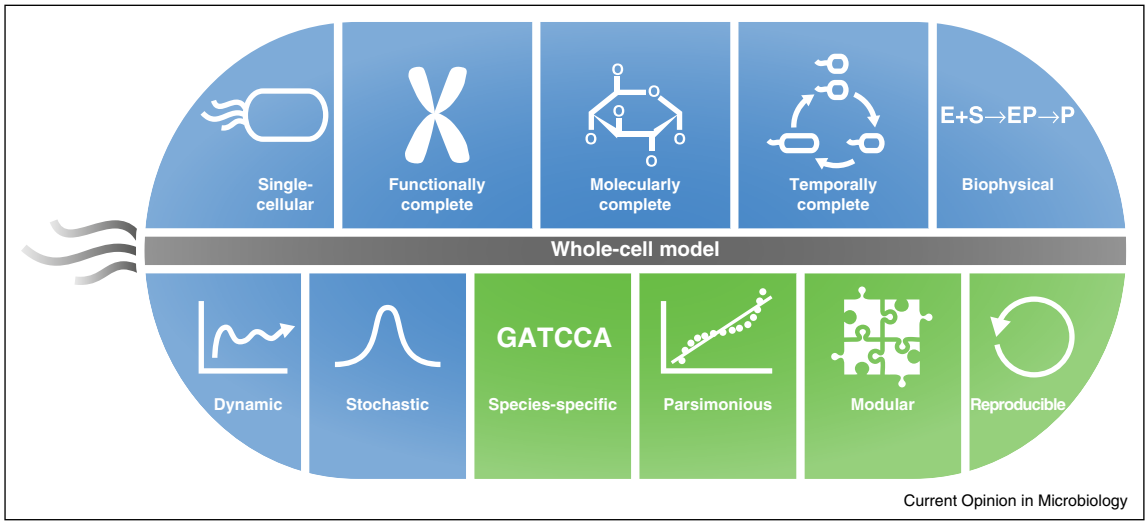
Functional closure

Behavior is determined by interacting pathways and genes. Consequently, whole-cell models should represent every known cellular and gene function. Models which represent every known function are powerful tools. For example, genome-scale metabolic models which represent every known metabolic reaction and enzyme have been used to identify missing reactions and enzymes [28].

Molecular closure

Whole-cell models should represent the cell and its environment as a closed system. Models should explicitly account for exchanges among pathways and the environment and not have arbitrary sources and sinks. This ensures

Figure 1



Fundamental (blue) and practical (green) principles of whole-cell modeling. No existing model satisfies every principle. The most advanced functional models are incomplete and do not fully represent molecular biophysics. The most advanced structural models do not represent cellular-scale behavior. Further work is needed to merge functional and structural modeling and expand their scope.

that models recognize important and often ignored connections such as the common energy carrier ATP. In turn, this enables models to capture pathway interactions that are often missed by studying pathways in isolation, such as how the energy charge affects phosphorylation and signaling.

Temporal closure

Whole-cell models should also represent the entire cell cycle. This ensures that models account for how cells regulate pathways in time to coordinate their life cycle. For example, models should account for how the dynamics of DNA replication affect dNTP concentrations and metabolism. Temporally complete models can also leverage cell theory, the fact that cells come from other cells, to constrain their dynamics. Assuming constant external conditions and absent evolution, cell theory implies that cellular populations are stable across generations. This provides a periodic constraint which enables dynamical modeling with minimal dynamical data.

Biophysics

In addition, whole-cell models should represent cellular biochemistry and biophysics, including mass conservation, thermodynamics, and spatial organization. This provides a recipe for bottom-up model construction, reducing the space of possible models. Takahashi *et al.* have reviewed several mathematical frameworks which are capable of representing cellular biophysics [29].

Dynamics

In particular, whole-cell models should be constructed from differential descriptions of molecular biochemistry and predict the emergence of cellular-scale dynamics. Emergent dynamics are valuable opportunities for experimental validation and discovery.

Stochasticity

Furthermore, whole-cell models should be discrete and stochastic. Stochastic models naturally predict the emergence of cellular variation. For example, stochastic models can account for how stochastic transcription initiation creates variation in gene expression and growth. This variation is another valuable opportunity for experimental validation.

Species-specificity

Whole-cell models must be evaluated by comparison to experimental data. Consequently, whole-cell models should represent specific genomes. This constrains the space of training data.

Parsimony

Despite the explosion in experimental data, limited data is available. For example, there is little data about non-coding RNA. Consequently, models should be parsimonious. This minimizes the need to identify unmeasured parameters.

Modularity

Absent an *ab initio* theory of biochemistry, whole-cell models must be based on many experimental descriptions of molecular biology. Consequently, like other large engineered systems, whole-cell models are best developed by combining multiple pathway submodels. This enables submodels to be developed and tested independently by different investigators using different representations.

Reproducibility

Finally, whole-cell models should be transparent, well-annotated, and reproducible. Researchers should be able to reproduce models from their primary sources, as well as reproduce simulations using multiple simulators. Models should also be described using transparent languages like SBML [30]. This is essential for collaborative modeling.

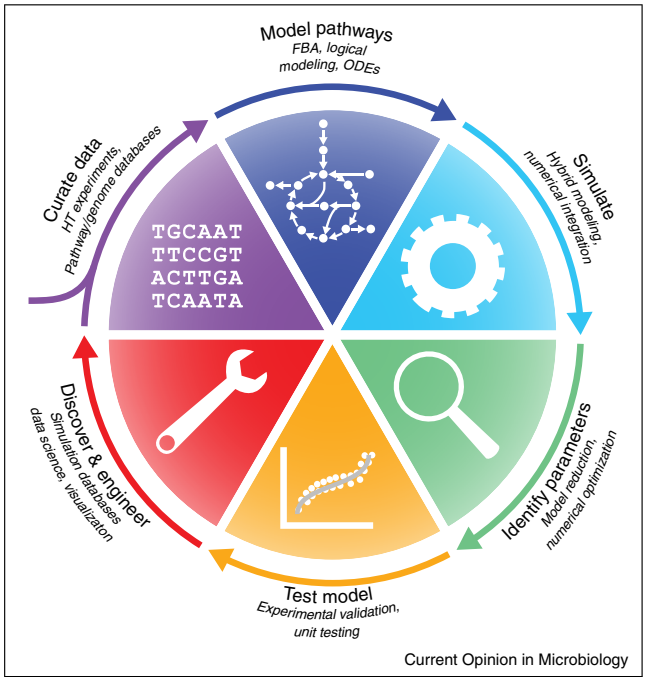
Model construction

Achieving these principles requires new approaches and tools. We briefly outline our approach to constructing whole-cell models (Figure 2), highlighting important areas for further research.

Experimental curation

The first step to constructing a model is to choose an organism and assess the feasibility of modeling it by assembling the available experimental knowledge. We have manually assembled training data from public databases and journal articles. Tables S1 and S2 list the most informative technologies and databases. Higher annotation standards are needed to enable modelers to take more advantage of published data [31]. We have organized our training data using model organism database tools such as Pathway Tools [32], WholeCellKB [33], BioMart [34], and Intermine [35].

Figure 2



Whole-cell modeling process. Experimental data is organized into a database, pathway submodels are constructed, submodels are combined, parameters are identified, the model is simulated and tested, and the model is used to guide discovery and bioengineering. The process is iterated using additional data to refine the model until an accurate model is achieved.

New experimental methods which fully characterize cells are needed to enable more comprehensive and accurate models. Improved metabolomic methods which are capable of quantitating the concentration of every metabolite are needed to train metabolic models. New proteomic methods are needed to characterize macromolecular complexes including their rates of formation and subunit composition dynamics. Improved interaction screens are needed to resolve the function of each individual interaction including every individual chaperone-substrate, protease-substrate, and miRNA–mRNA interaction. New high-throughput methods are needed to comprehensively quantitate reaction kinetics. Additional tools are also needed to comprehensively characterize single-cell temporal and population variance including of metabolite and protein concentrations, as well as of systems properties such as the growth rate, cell cycle phase durations, cell size, and reaction fluxes.

This manual reconstruction approach has been feasible for small bacteria. More scalable approaches will be needed for more complex bacteria and eukaryotes. One potential solution is to automatically reconstruct knowledge using cognitive computing [36] or other machine learning techniques [37–39]. A second potential solution is to engage a large community of scientists. Both of these will require additional molecular databases such as BRENDA [40] and UniProt [41] which are either manually assembled by single researchers, community assembled by self-curation during publication [42], or automatically assembled using natural language processing [43].

Mathematical formulation

Second, a mathematical description of how cells evolve over time must be constructed. We have described cells as thoroughly as possible given our current knowledge, desire to predict cellular behavior, and limited time and resources. This strategy takes full advantage of our existing knowledge, avoids unknown parameters and expensive computations of processes such as diffusion which minimally affect behavior, and enables one model to be used for many scientific questions. In practice, until whole-cell models are complete, modelers will need to focus on the pathways most relevant to their research.

In our experience, the easiest way to construct a whole-cell model, like any other large engineered system, is to assemble multiple pathway submodels. This approach is scalable because it enables pathways to be modeled and tested independently by different investigators using different mathematical formalisms.

Individual submodels must be implemented and/or constructed from experimental data. BioModels [44] and the CellML model repository [45] contain many existing pathway models. However, most pathways have not been

modeled, and most existing models must be modified for integration with other models. The primary obstacle to modeling pathways is the lack of quantitative data. New experimental technologies are needed to characterize more pathways.

Rule-based modeling is a powerful and scalable approach for assembling genome-scale models [46,47]. Several rule-based and conventional tools can be used to construct and modify pathway models including BioNetGen [46], BioUML [48], CellDesigner [49], CobraPy [50], COPASI [51], E-Cell [2], iBioSim [52], and JDesigner [53]. Table S3 lists several additional tools. Further work is needed to scale up these tools for larger models.

Submodel integration

Next, individual submodels must be combined. Mathematically homogeneous submodels can be merged analytically. Heterogeneous submodels must be combined by dividing the state variables into independent subvariables dedicated to each submodel; integrating the individual submodels based on these subvariables; and merging the subvariables to update the global variables. The integration time step should be set faster than the fastest inter-pathway dynamics. Slower time steps will introduce communication delays. We and others have developed hybrid simulators which are capable of integrating heterogeneous submodels [2,3,23••,54–56]. Further work is needed to develop a deeper theoretical understanding of multi-algorithm modeling.

Parameter estimation

Once the model's structure has been implemented, the model's parameters must be identified by matching the model's predictions to experimental data. Identifying whole-cell models is challenging because they are high-dimensional, stochastic, and computationally expensive. We have followed a three-step approach to parameter identification. First, we have only created parameters whose values can be estimated from one or a few experimental observations. Second, we have used public data to estimate each parameter. Third, we have refined parameter values by numerically minimizing the prediction error of a manually constructed reduced model which approximates the full model.

Unfortunately, this approach is not scalable. Building increasingly comprehensive models requires increasingly comprehensive experimental data. Manually constructing reduced models is also not scalable.

There are many other promising local and global parameter identification strategies. Several researchers have reviewed these approaches and their application to smaller models [57–59]. Several innovations are needed to apply these methods to larger, hybrid models. Automated model reduction [60–62] is needed to construct models which are tractable to numerical optimization. Researchers should pursue both statistical and physics-based approaches. Automatic differentiation should be applied to improve the efficiency of gradient-based optimization [63]. Faster, parallelized simulation engines and distributed optimization procedures should be applied to explore parameters more quickly [64,65].

Model refinement and validation

The last step to constructing a whole-cell model is to iteratively evaluate the model's predictions and refine the model. We have focused on evaluating the predicted phenotypes of genetic perturbations. Additional data representing single-cell variation and cell cycle dynamics are needed for more rigorous validation. Experimental design based on predicting the most likely informative experiments [58], robotic and microfluidic experimentation [66], and computational gap filling [67] should be applied to automate model refinement.

Visualization and analysis

The final steps in whole-cell modeling are to simulate the model, analyze simulation results to construct new hypotheses, and conduct experiments to test those hypotheses. We have developed WholeCellSimDB to organize simulation results and facilitate large-scale analyses [68]. We have also developed WholeCellViz [69] and the E-Cell session monitor [70] to visualize simulation results. We have used these tools to gain new insights into cellular energy usage [23••], learn kinetic parameters [71], and analyze the metabolic demands of synthetic gene networks [72]. Numerous other visualization software are available including Cytoscape, Gephi, VANTED, and VisANT [73].

Conclusions

Whole-cell models promise to predict how genotype determines phenotype. Combined with genome synthesis and transplantation, whole-cell models could enable bioengineers to construct cellular factories. Whole-cell models could also enable clinicians to individualize therapy. Furthermore, whole-cell models would be unprecedented scientific tools.

Whole-cell models have several advantages over focused models. They are constructed once, but can drive many scientific, engineering, and medical questions. They can also predict non-intuitive effects by chaining together many individually intuitive interactions. Furthermore, they can systematize biological discovery and unify disparate research.

Whole-cell modeling is a new and exciting field with numerous challenges that require collaboration among computational and experimental biologists, bioinformaticists, and computer scientists. Table S4 lists several efforts to build a whole-cell modeling community.

We have proposed 11 principles to guide what whole-cell models represent and how they presently must be constructed. Our own work has followed most of these principles. However, as detailed by Macklin *et al.*, further work is required to achieve all of these principles even for the simplest bacteria [74]. A high-level declarative language is needed to describe models more transparently, theoretical studies are needed to better understand multi-algorithm models, and more efficient simulators are needed to simulate models more quickly.

To date, we have manually reconstructed and identified whole-cell models which represent hundreds of genes. Achieving models of more complex bacteria and eukaryotes, which represent tens of thousands of genes, demands new automated pathway reconstruction methods based on artificial intelligence techniques such as machine learning, natural language processing, and ontology engineering. These models will contain hundreds of thousands of quantitative parameters such as binding affinities and rate constants. Identifying their values will demand new high-throughput experiments which can quantitate individual molecular interactions, as well as new biophysical models which can accurately predict quantitative parameters from sequences. Ultimately, whole-organism models will require hierarchical modeling approaches which use agent-based modeling to combine multiple whole-cell models of multiple cell types. In addition, alternative incentives are needed to reward collaborative modeling. Solving these challenges will allow whole-cell modeling to fulfill its promise of enabling bioengineering and precision medicine.

Acknowledgements

We thank Markus Covert for numerous discussions on whole-cell modeling and Javier Carrera, Derek Macklin, Matthew Oberhardt, and Eytan Ruppin for valuable feedback on this manuscript. This work was supported a James S. McDonnell Foundation Postdoctoral Fellowship Award to JRK, MEXT HPCI Strategic Program Supercomputational Life Science and JSPS KAKENHI (25711012) Grants to KT, and JSPS KAKENHI Grants (23136513 and 24300112) to AF.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mib.2015.06.004.

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. Tomita M: **Whole-cell simulation: a grand challenge of the 21st century**. *Trends Biotechnol* 2001, **19**:205-210.

2. Takahashi K, Ishikawa N, Sadamoto Y, Sasamoto H, Ohta S, Shiozawa A, Miyoshi F, Naito Y, Nakayama Y, Tomita M: **E-Cell 2: multi-platform E-Cell simulation system**. *Bioinformatics* 2003, **19**:1727-1729.

3. • Takahashi K, Kaizu K, Hu B, Tomita M: **A multi-algorithm, multi-timescale method for cell simulation**. *Bioinformatics* 2004, **20**:538-546.
Describes an algorithm for integrating models composed of multiple mathematically heterogeneous submodels.

4. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA *et al.*: **Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome**. *Science* 2008, **319**:1215-1220.

5. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC: **Genome transplantation in bacteria: changing one species to another**. *Science* 2007, **317**:632-638.

6. Kim B, Kim WJ, Kim DI, Lee SY: **Applications of genome-scale metabolic network model in metabolic engineering**. *J Ind Microbiol Biotechnol* 2015, **42**:339-348.

7. Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY: **Systems biology as a foundation for genome-scale synthetic biology**. *Curr Opin Biotechnol* 2006, **17**:488-492.

8. Tian Q, Price ND, Hood L: **Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine**. *J Intern Med* 2012, **271**:111-121.

9. Loscalzo J, Barabasi AL: **Systems biology and the future of medicine**. *Wiley Interdiscip Rev Syst Biol Med* 2011, **3**:619-627.

10. Hamburg MA, Collins FS: **The path to personalized medicine**. *N Engl J Med* 2010, **363**:301-304.

11. Shuler ML, Leung S, Dick CC: **A mathematical model for the growth of a single bacterial cell**. *Ann N Y Acad Sci* 1979, **326**:35-52.

12. • Atlas JC, Nikolaev EV, Browning ST, Shuler ML: **Incorporating genome-wide DNA sequence information into a dynamic whole-cell model of *Escherichia coli*: application to DNA replication**. *IET Syst Biol* 2008, **2**:369-382.
Reports an ordinary differential equation whole-cell model of *E. coli* encompassing multiple physiological processes.

13. • Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC *et al.*: **E-CELL: software environment for whole-cell simulation**. *Bioinformatics* 1999, **15**:72-84.
Reports the first molecularly resolved whole-cell model representing over 100 gene functions. Describes a hybrid approach to whole-cell modeling combining discrete and stochastic mathematical representations.

14. Werner SL, Kearns JD, Zadorozhnaya V, Lynch C, O'Dea E, Boldin MP, Ma A, Baltimore D, Hoffmann A: **Encoding NF-kappaB temporal control in response to TNF: distinct roles for the negative regulators IkappaBalpha and A20**. *Genes Dev* 2008, **22**:2093-2101.

15. Smallbone K, Simeonidis E, Swainston N, Mendes P: **Towards a genome-scale kinetic model of cellular metabolism**. *BMC Syst Biol* 2010, **4**:6.

16. Gérard C, Tyson JJ, Coudreuse D, Novák B: **Cell cycle control by a minimal Cdk network**. *PLoS Comput Biol* 2015, **11**:e1004056.

17. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110**. *Appl Environ Microbiol* 1994, **60**:3724-3731.

18. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO: **A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism — 2011**. *Mol Syst Biol* 2011, **7**:535.

19. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO: **Matrix formalism to describe functional states of transcriptional regulatory systems**. *PLoS Comput Biol* 2006, **2**:e101.

20. • Thiele I, Jamshidi N, Fleming RM, Palsson BO: **Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical**

formulation, and its functional characterization. *PLoS Comput Biol* 2009, **5**:e1000312.
Describes a FBA-based approach to modeling multiple cellular processes including bacteria metabolism, transcription, and translation.

21. Lee JM, Gianchandani EP, Eddy JA, Papin JA: **Dynamic analysis of integrated signaling, metabolic, and regulatory networks**. *PLoS Comput Biol* 2008, **4**:e1000086.

22. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C *et al.*: **A genomic regulatory network for development**. *Science* 2002, **295**:1669-1678.

23. •• 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.
Reports the first whole-cell model which represents the specific function of every characterized gene product. This model predicts the dynamics of every molecular species over the entire cell cycle.

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

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

26. • Roberts E, Magis A, Ortiz JO, Baumeister W, Luthey-Schulten Z: **Noise contributions in an inducible genetic switch: a whole-cell simulation study**. *PLoS Comput Biol* 2011, **7**:e1002010.
Reports the first coarse-grained molecular dynamics structural whole-cell model which predicts macromolecule diffusion.

27. Roberts E: **Cellular and molecular structure as a unifying framework for whole-cell modeling**. *Curr Opin Struct Biol* 2014, **25**:86-91.

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

29. Takahashi K, Arjunan SN, Tomita M: **Space in systems biology of signaling pathways — towards intracellular molecular crowding in silico**. *FEBS Lett* 2005, **579**:1783-1788.

30. Finney A, Hucka M: **Systems biology markup language: level 2 and beyond**. *Biochem Soc Trans* 2003, **31**:1472-1473.

31. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T *et al.*: **Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project**. *Nat Biotechnol* 2008, **26**:889-896.

32. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L *et al.*: **Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology**. *Brief Bioinform* 2010, **11**:40-79.

33. 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.

34. Kasprzyk A: **BioMart: driving a paradigm change in biological data management**. *Database* 2011, **2011**:bar049.

35. Smith RN, Aleksic J, Butano D, Carr A, Contrino S, Hu F, Lyne M, Lyne R, Kalderimis A, Rutherford K *et al.*: **InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data**. *Bioinformatics* 2012, **28**:3163-3165.

36. Kelly J III, Hamm S: *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press; 2013.

37. Arakawa K, Yamada Y, Shinoda K, Nakayama Y, Tomita M: **GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes**. *BMC Bioinformatics* 2006, **7**:168.

38. Hamilton JJ, Reed JL: **Software platforms to facilitate reconstructing genome-scale metabolic networks**. *Environ Microbiol* 2014, **16**:49-59.

39. Saha R, Chowdhury A, Maranas CD: **Recent advances in the reconstruction of metabolic models and integration of omics data**. *Curr Opin Biotechnol* 2014, **29**:39-45.

40. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, Sensen CW, Schomburg D: **BRENDA in 2015: exciting developments in its 25th year of existence**. *Nucleic Acids Res* 2015, **43**:D439-D446.

41. UniProt Consortium: **UniProt: a hub for protein information**. *Nucleic Acids Res* 2015, **43**:D204-D212.

42. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S *et al.*: **Big data: the future of biocuration**. *Nature* 2008, **455**:47-50.

43. Ananiadou S, Kell DB, Tsujii J: **Text mining and its potential applications in systems biology**. *Trends Biotechnol* 2006, **24**:571-579.

44. Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating S, Knight-Schrijver V *et al.*: **BioModels: ten-year anniversary**. *Nucleic Acids Res* 2015, **43**:D542-D548.

45. Lloyd CM, Lawson JR, Hunter PJ, Nielsen PF: **The CellML model repository**. *Bioinformatics* 2008, **24**:2122-2123.

46. Faeder JR, Blinov ML, Hlavacek WS: **Rule-based modeling of biochemical systems with BioNetGen**. *Methods Mol Biol* 2009, **500**:113-167.

47. Chylek LA, Harris LA, Tung CS, Faeder JR, Lopez CF, Hlavacek WS: **Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems**. *Wiley Interdiscip Rev Syst Biol Med* 2014, **6**:13-36.

48. Kiselev I, Kolpakov FA: **BioUML: plugin for population-based modeling**. *Virtual Biol* 2014, **2**:7-15.

49. Matsuoka Y, Funahashi A, Ghosh S, Kitano H: **Modeling and simulation using CellDesigner**. *Methods Mol Biol* 2014, **1164**:121-145.

50. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR: **COBRApy: COnstraints-Based Reconstruction and Analysis for Python**. *BMC Syst Biol* 2013, **7**:74.

51. Mendes P, Hoops S, Sahle S, Gauges R, Dada J, Kummer U: **Computational modeling of biochemical networks using COPASI**. *Methods Mol Biol* 2009, **500**:17-59.

52. Stevens JT, Myers CJ: **Dynamic modeling of cellular populations within iBioSim**. *ACS Synth Biol* 2013, **2**:223-229.

53. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H: **Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration**. *OMICS* 2003, **7**:355-372.

54. Adalsteinsson D, McMillen D, Elston TC: **Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks**. *BMC Bioinformatics* 2004, **5**:24.

55. Salis H, Sotiropoulos V, Kaznessis YN: **Multiscale Hy3S: hybrid stochastic simulation for supercomputers**. *BMC Bioinformatics* 2006, **7**:93.

56. Pahle J: **Biochemical simulations: stochastic, approximate stochastic and hybrid approaches**. *Brief Bioinform* 2009, **10**:53-64.

57. Banga JR, Balsa-Canto E: **Parameter estimation and optimal experimental design**. *Essays Biochem* 2008, **45**:195-209.

58. Kreutz C, Timmer J: **Systems biology: experimental design**. *FEBS J* 2009, **276**:923-942.

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

60. Holzhütter HG, Drasdo D, Preusser T, Lippert J, Henney AM: **The virtual liver: a multidisciplinary, multilevel challenge for systems biology**. *Wiley Interdiscip Rev Syst Biol Med* 2012, **4**:221-235.

61. Anderson J, Chang YC, Papachristodoulou A: **Model decomposition and reduction tools for large-scale networks in systems biology**. *Automatica* 2011, **47**:1165-1174.

62. Kremling A, Saez-Rodriguez J: **Systems biology — an engineering perspective**. *J Biotechnol* 2007, **129**:329-351.

63. Rall LB: **Automatic differentiation: techniques applications**. *Lect Notes Comput Sci* 1981, **120**:154.

64. Panait L, Luke S: **Cooperative multi-agent learning: the state of the art**. *Auton Agent Multi Agent Syst* 2005, **11**:387-434.

65. Nedic A, Ozdaglar A: **Convex optimization in signal processing and communications**. In *Convex Optimization in Signal Processing and Communications.* Edited by Palomar DP, Eldar YC. Cambridge University Press; 2009:340-386.

66. King RD, Rowland J, Oliver SG, Young M, Aubrey W, Byrne E, Liakata M, Markham M, Pir P, Soldatova LN *et al.*: **The automation of science**. *Science* 2009, **324**:85-89.

67. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nat Biotechnol* 2010, **28**:977-982.

68. Karr JR, Phillips NC, Covert MW: **WholeCellSimDB: a hybrid relational/HDF database for whole-cell model predictions**. *Database* 2014, **2014**:bau095.

69. Lee R, Karr JR, Covert MW: **WholeCellViz: data visualization for whole-cell models**. *BMC Bioinformatics* 2013, **14**:253.

70. Takahashi K, Sakurada T, Kaizu K, Kitayama T, Arjunan S, Ishida T, Bereczki G, Ito D, Sugimoto M, Komori T *et al.*: **E-CELL system version 3: a software platform for integrative computational biology**. *Genome Inform* 2003, **14**:294-295.

71. Sanghvi JC, Regot S, Carrasco S, Karr JR, Gutschow MV, Bolival B Jr, Covert MW: **Accelerated discovery via a whole-cell model**. *Nat Methods* 2013, **10**:1192-1195.

72. Purcell O, Jain B, Karr JR, Covert MW, Lu TK: **Towards a whole-cell modeling approach for synthetic biology**. *Chaos* 2013, **23**:025112.

73. Secrier M, Schneider R: **Visualizing time-related data in biology, a review**. *Brief Bioinform* 2014, **15**:771-782.

74. Macklin DN, Ruggero NA, Covert MW: **The future of whole-cell modeling**. *Curr Opin Biotechnol* 2014, **28**:111-115.