Initiatives to Build a Whole-Cell Modeling Community: 2019 Update

Jonathan R. Karr[1], Maria Lluch-Senar[2], Damjana Kastelic[2], Luis Serrano[2], and Herbert M. Sauro[3]

[1]Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA.
[2]Centre for Genomic Regulation, Barcelona, Spain.
[3]University of Washington, Seattle, WA, USA.

May 20, 2019

Abstract
Whole-cell (WC) models that predict phenotype from genotype have the potential to transform biology, bioengineering, and medicine. Achieving WC models will likely require collaboration among modelers, experimentalists, mathematicians, computer scientists, and engineers. In 2012, we and others began to build a WC community by organizing a central website, a primer, schools, hackathons, and challenges. This year, we have continued to build a WC community by developing a new website, expanding the primer, organizing a third school, and launching an online seminar. Here, we summarize these initiatives, their impact to date, and our plans to continue to build a WC community.

Introduction

Whole-cell (WC) models that predict phenotype from genotype are needed to advance biology, bioengineering, and medicine[1–3]. Despite recent progress, significant work remains to build WC models[4–7]. Achieving WC models will likely require intense collaboration among modelers, experimentalists, mathematicians, computer scientists, and engineers.

In 2012, we began to build a WC community by developing WholeCell.org, a central website for WC modeling[8]. Since then, we have continued to form a community by co-organizing a challenge on calibrating WC models in 2013[9], co-organizing a hackathon on encoding WC models into the Systems Biology Markup Language[10] (SBML) in 2015[11], organizing schools in 2016 and 2017[12,13], starting a primer in 2017[14], and surveying the goals and needs of the community in 2017 [7]. This year, we have continued to foster a community by developing a new version of WholeCell.org, expanding the primer, organizing a third school[15], and launching an online seminar[16].

Here, we summarize the new website, the expanded primer, the 2019 school, and the new seminar. We also describe the impact of our efforts to date and our plans to continue to build a WC community.

WC modeling website 2.0

In 2012, we launched WholeCell.org[8] to centralize information about WC modeling. Initially, the website included information about the first WC model[4], a wiki, and a forum. Through 2017, we accumulated additional information including links to relevant publications and events. To disseminate more information, we launched a new version of WholeCell.org this year. The new version describes the motivation for WC modeling, outlines a roadmap for achieving WC models, summarizes the existing models and modeling tools, links to articles, and advertises WC modeling events and research groups.

Expanded primer
In July 2017, we began to write *An Introduction to Whole-Cell Modeling* to synthesize the emerging concepts in WC modeling[14]. In particular, the primer is intended to provide new investigators with an overview of WC modeling, code samples, and exercises.

Initially, the primer focused on the mathematical representation of WC models, the fundamental simulation algorithms of cell modeling, model composition, and how to develop WC modeling software tools. This year, we expanded the primer by adding information about multi-algorithmic co-simulation of mixed-grain composite models, running WC models with Docker containers, revisioning WC datasets with Quilt, and grant writing.

2019 EMBO Whole-Cell Modeling Summer School
We organized the third WC school on April 7-13, 2019 at the Center for Genomic Regulation in Barcelona, Spain[15]. Nineteen participants, three speakers, six instructors, and four organizers from sixteen countries participated in the school. The school was primarily supported by EMBO.

The school included two introductory lectures, three research talks to excite the participants about WC modeling, ten four-hour hands-on tutorials on the principles of WC modeling, two breakout discussions about overcoming the challenges to WC modeling, a poster session, and multiple social activities. Each tutorial began with a brief presentation on one or more concepts, followed by hands-on computational exercises to reinforce the concepts.

Based on feedback from previous schools, we improved the school in several ways:

- We doubled the length of the school to provide the participants more time to work on the exercises and engage in discussions.
- To enable the participants to work together on the exercises and learn from each other, we held the school in a larger room and divided the participants into interdisciplinary groups.
- We organized a room block at a nearby hotel to further facilitate networking over breakfasts, dinners and walks to and from the school.
- We reduced the cost of participating in the school by securing funding from EMBO for accommodations and travel scholarships.

The feedback from the participants was universally positive. All of the participants reported that they were glad that they participated in the course, the average rating for the scientific content of the course was 4.5 out of 5, and the course exceeded the expectations of 80% of the participants.

Online Whole-Cell Modeling Seminar
To encourage more interaction among the emerging WC community, we recently launched a monthly online seminar[16]. To date, we have organized six seminars. The first seminars focused on the potential impact of WC modeling. Each seminar began with a 25-minute thought-provoking presentation about a topic in WC modeling, followed by a 35-minute group discussion.

On average, 45 investigators from ten countries have participated in the seminars. We are also sharing recordings of the meetings via YouTube. Nagasuma Chandra from the Indian Institute of Science is using these recordings to moderate a second discussion at a better time for investigators in Asia.

Impact of our community building efforts
Now that several years have passed since we began to build a WC community, we can start to evaluate the long-term impact of our efforts. At least five publications and one meeting have stemmed from our efforts. The inability to encode the *Mycoplasma genitalium* model[4] into SBML at the 2015 hackathon encouraged several investigators to analyze the need for new standards for

WC modeling and develop these standards. This has resulted in two perspectives and one research report. Dagmar Waltemath from the University of Rostock and Jonathan Karr led a community perspective on the standards needed for WC modeling[11]. Kyle Medley from the University of Washington published a perspective on how to conduct WC modeling reproducibly[17]. Chris Meyers from the University of Utah and Matthias König from Humboldt University of Berlin published a scheme for encoding dynamic Flux Balance Analysis models in SBML[18].

The difficulty of encoding the *M. genitalium* model into SBML also encouraged Paulo Burke from University of São Paulo and his colleagues to develop a simplified network representation of the *M. genitalium* model[19]. The hackathon helped Burke recognize the need for simpler models and identify collaborators. Their project will produce the first WC network.

Currently, Joshua Rees, Oliver Chalkley, Lucia Marucci and Claire Grierson from the University of Bristol are developing algorithms for using WC models to design genomes[20] stemming from conversations with Jonathan Karr, Maria Lluch-Senar, and others at the 2017 school. At the school, Rees and Chalkey recognized the potential utility of WC models for genome design and formed collaborations with Karr and Senar. Their project aims to produce the first algorithms for using WC models to design genomes. The success of this project has also encouraged Marucci and Grierson to organize the first meeting on designing genomes with WC models which will take place in July 2019 at the University of Bristol.

Plans to continue to build a WC community

Despite our efforts, more work remains to build a WC community with shared goals, projects, and practices. Here, we outline our plans to continue to build a WC community.

Launch community modeling projects

Ultimately, we would like to launch community projects to model key cells such as *Mycoplasmas*, *Escherichia coli*, and human stem cells. We anticipate that such projects would encourage deep collaboration. Unfortunately, we think that such community projects are premature at this point without a firm technological foundation for collaboration. Thus, we plan to focus our short-term efforts on developing the technological foundation needed for collaboration.

Develop the technological foundation needed for collaborative modeling

As described above, we plan to focus our efforts on developing the technologies needed for collaborative WC modeling. This includes tools for integrating the data needed for WC modeling, a language for describing WC models, and a casually-correct multi-algorithmic simulator.

Assemble introductory WC materials

We plan to continue to integrate information into WholeCell.org and the primer as new models and tools are published. Ultimately, we would like the primer to be a comprehensive introduction to WC modeling. As WC modeling matures, we would also like to create an online course. This would reduce the need for schools, which would enable more workshops and hackathons.

Organize additional in-person schools

As the core concepts of WC modeling continue to evolve, we plan to continue to organize schools. Based on feedback on the schools, we plan to make several improvements to the schools:

- We plan to provide additional instruction on molecular biology by adding a tutorial on core biochemical processes such as metabolism. The tutorial will outline the role of each process, the types and granularity of data about each process, the molecules that participate in each process, the reactions that constitute each process, and their kinetics.
- We will make the tutorials more consistent by replacing the tutorial on data-driven modeling of the structure of chromosomes with a shorter lecture on the same topic.

- We will also make the tutorials more consistent by focusing the examples on a single organism.
- We will make the code samples easier to follow by converting them to Jupyter notebooks and adding additional documentation.
- To generate more excitement about WC modeling, we are considering pairing the schools with hackathons in which participants use the concepts learned in the school to build submodels.
- To help the participants focus, we will reduce the length of each day by one hour.
- To reduce the cost of participating in the school, we will help the participants find roommates by sharing their contact information before the school.

Organize monthly online meetings
We plan to continue to organize monthly online meetings to facilitate sustained discussion among the emering community. We plan to make several improvements to facilitate discussion better:

- **Support sustained discussion by focusing the content of the seminar.** For the next six months, we plan to focus the seminar on developing a shared conceptual basis for WC modeling by organizing meetings on emerging technologies for WC modeling. For Fall 2019 and Winter 2020, we are considering arranging several seminars on model verification. This will include the need to verify models, foundational concepts in unit testing and formal verification, and emerging frameworks for testing models.
- **Shift the seminar to working group meetings.** Ultimately, we plan to shift the majority of the meetings to discussions about community projects. This would include brief progress reports followed by community feedback and discussion on next steps.
- **Increase participation outside the US.** We plan to hold the meetings earlier in the day to make them more accessible to investigators in Europe.

Organize in-person workshops and hackathons
We also plan to support workshops on WC modeling. This year, we plan to contribute to two workshops on verifying WC models and using WC models to design genomes. The goals of the verification workshop will be to outline the properties of models that should be verified; identify the tools needed to verify these properties; and distribute the workload among the community.

The goals of the genome design workshop will be to bring together investigators from WC modeling, synthetic biology, and minimal genomes to outline the challenges to using WC models to design genomes and brainstorm solutions. Tentatively, the workshop will be held in July 2019 at the University of Bristol and organized by Lucia Marucci, Claire Grierson, and Paul Race.

Conclusion
Realizing the potential of WC models will likely require a large community of expertise in cell biology, dynamical modeling, and numerical simulation. Currently, we are developing the technological foundation needed for this collaboration. Over the past several years, we have also initiated several efforts to begin to assemble a WC community. We plan to continue to grow this community by expanding our primer, improving our schools, organizing in-person workshops, and pivoting our online seminar to an online working group. Ultimately, we aim to launch community efforts to model key cells such as *Mycoplasmas*, *E. coli*, and human stem cells.

Acknowledgements
We thank Saahith Pochiraju for help building the website; Arthur Goldberg for help writing the primer; Gabriel Gonzalez Cano, Arthur Goldberg, Verónica Llorens-Ríco, Samuel Miravet-Verde, John Sekar, Anna Sole Amat, Marie Trussart, and Marc Weber for help teaching the school; and Yosef Roth and Veronica Porubsky for help organizing the seminar.

These efforts were supported by National Science Foundation awards 1548123 and 1649014 to

JRK, National Institutes of Health award R35 GM119771 to JRK, and National Institutes of Health award P41 EB023912 to Herbert Sauro. The 2019 school was also supported by an EMBO Practical Course award to JRK, MLS, and DK.

References

1. Tomita, M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol* **19,** 205–210 (2001).
2. Karr, J. R., Takahashi, K. & Funahashi, A. The principles of whole-cell modeling. *Curr Opin Microbiol* **27,** 18–24 (2015).
3. Carrera, J. & Covert, M. W. Why build whole-cell models? *Trends Cell Biol* **25,** 719–722 (2015).
4. Karr, J. R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* **150,** 389–401 (2012).
5. Macklin, D. N., Ruggero, N. A. & Covert, M. W. The future of whole-cell modeling. *Curr Opin Biotechnol* **28,** 111–115 (2014).
6. Goldberg, A. P. *et al.* Emerging whole-cell modeling principles and methods. *Curr Opin Biotechnol* **51,** 97–102 (2018).
7. Szigeti, B. *et al.* A blueprint for human whole-cell modeling. *Curr Opin Syst Biol* **7,** 8–15 (2018).
8. Karr, J. R. & Pochiraju, S. *Whole-cell modeling* `https://www.wholecell.org` (2019).
9. Karr, J. R. *et al.* Summary of the DREAM8 parameter estimation challenge: toward parameter identification for whole-cell models. *PLoS Comput Biol* **11,** e1004096 (2015).
10. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19,** 524–531 (2003).
11. Waltemath, D. *et al.* Toward community standards and software for whole-cell modeling. *IEEE Trans Biomed Eng* **63,** 2007–2014 (2016).
12. Karr, J. R., Lluch-Senar, M, Serrano, L & Carrera, J. *The 2016 Whole-Cell Modeling Summer School* `https://doi.org/10.5281/zenodo.1004027` (2016).
13. Karr, J. R., Lluch-Senar, M, Serrano, L & Kastelic, D. *The 2017 Whole-Cell Modeling Summer School* `https://doi.org/10.5281/zenodo.1004135` (2017).
14. Karr, J. R. & Goldberg, A. P. *An introduction to whole-cell modeling* `https://intro-to-wc-modeling.readthedocs.io` (2017).
15. Karr, J. R., Lluch-Senar, M, Kastelic, D & Serrano, L. *2019 EMBO whole-cell modeling summer school* `http://meetings.embo.org/event/19-whole-cell-modelling` (2019).
16. Roth, Y. D., Porubsky, V., Sauro, H. M. & Karr, J. R. *Online whole-Cell modeling seminar* `https://reproduciblebiomodels.org/seminar` (2019).
17. Medley, J. K., Goldberg, A. P. & Karr, J. R. Guidelines for reproducibly building and simulating systems biology models. *IEEE Trans Biomed Eng* **63,** 2015–2020 (2016).
18. Watanabe, L., König, M. & Myers, C. Dynamic flux balance analysis models in SBML. *BioRXiv,* 245076 (2018).
19. Burke, P. E., de Campos, C. B. & Quiles, M. G. Whole-cell representation and analysis of Mycoplasma genitalium organism using complex networks. *ComplexNet.* doi: 10.13140/2.1.3736.6084.
20. Rees, J. *et al.* Designing minimal genomes using whole-cell models. *bioRxiv,* 344564 (2019).