

Visualisierung eines Ganzzellenmodells

von Arne Hoffmann

14. Februar 2018

1. Motivation und verwandte Arbeiten

Die steigende Leistung kosteneffizient verfügbarer Rechentechnik erlaubt es immer komplexere Modelle zu simulieren. Im jungen Fachbereich der Systembiologie ermöglicht dies zusammen mit der Entwicklung neuer experimenteller Methoden den Schritt von der Simulation einzelner Zellprozesse hin zu immer komplexeren Interaktionen zwischen verschiedenen Zellprozessen und letztlich der Ganzzellensimulation[1].

Im Feld der Ganzzellenmodellierung versucht man alle Zellprozesse mit einander verbunden zu modellieren. Damit soll die Interaktion komplexer Zellvorgänge durch Simulation auf molekularer Ebene verständlich dargestellt werden. Neben der Unterstützung von Experimenten in der Molekularbiologie soll dies personalisierte Behandlung in der Medizin und Design und Simulation von Mikroorganismen in der synthetischen Biologie ermöglichen oder verbessern.

Im Zuge des Yeast Cell Model (YCM) Projekts [2] wird ein Ganzzellenmodell der Backhefe *Saccharomyces cerevisiae* entwickelt. Dieses komplexe Modell der verschiedenen Zellprozesse, welches die hoch vernetzten Interaktionen zwischen über 10000 Spezies beschreibt, benötigt nun eine Visualisierung, welche es erlaubt das Modell und seine Simulationsergebnisse zu erkunden.

Zur Visualisierung älterer Ganzzellenmodelle werden üblicherweise Konzentrationsverläufe [3][6] oder Änderungen einzelner Eigenschaften [4][5][6] über den Simulationszeitraum verwendet. Dazu werden skizzierte Netzwerkdarstellung des Zellmodells angeboten. Bei den meistens sehr kleinen Zellmodellen und der Betrachtung einer geringen Menge an Stoffen oder Eigenschaften ist diese Form der Darstellung ausreichend. Bei der hohen Komplexität des YCM Projekts, welches sich nicht nur auf eine kleine Anzahl an Stoffen konzentriert, wäre eine Darstellung aller Konzentrationsverläufe zu unübersichtlich.

Es gibt bereits mehrere Anwendungen wie Cytoscape [7] und CellDesigner [8] zur Darstellung

biochemischer Netzwerke. Diese können jedoch nur eingeschränkt Simulationsergebnisse integrieren.

Das whole-cell model Projekt [9] von Jonathan Karr *et al.* hat bereits eine umfangreiche Modellierung und Visualisierung eines Ganzzellenmodells umgesetzt, jedoch für das Bakterium *Mycoplasma genitalium*, welches für sein besonders kleines Genom aus ungefähr 500 Genen bekannt ist. Diese geringe Größe erlaubt es Jonathan Karr *et al.* ein agentenbasiertes Modell zu implementieren, d.h. dass einzelne Moleküle simuliert werden. Die vom Ganzzellenmodell des YCM Projekts simulierte Backhefe *Saccharomyces cerevisiae* ist ein Eukaryot und kein Prokaryot wie das Bakterium *Mycoplasma genitalium*, was unter anderem die Existenz eines von einer Doppelmembran umschlossenen Zellkerns bedeutet. Außerdem sorgt das bedeutend größere Genom mit ungefähr 6700 Genen dafür, dass kein agentenbasierter Ansatz gewählt werden kann, sondern gewöhnliche Differentialgleichungen (englisch kurz: ODEs) die Grundlage bilden. Allein dieser Unterschied im Modell macht WholeCellViz [10], das Projekt zur graphischen Darstellung des Ganzzellenmodells von Jonathan Karr *et al.*, inkompatibel mit dem YCM Projekt. Es wird also eine eigenständige Visualisierung für das YCM Projekt benötigt.

2. Ziel

Ziel dieser Arbeit ist die Ermittlung der Anforderungen an die Visualisierung für das Ganzzellenmodell der Backhefe *Saccharomyces cerevisiae* des YCM Projekts. Darauf basierend soll ein Design der Visualisierung entwickelt und ein Prototyp dieses als Webanwendung implementiert werden.

Die Visualisierung soll das Modell und Simulationsdaten aus mehreren Abstraktionsstufen zeigen, wobei die gröbste Abstraktionsstufe die Zelle und ihr Wachstum und die feinste Abstraktionsstufe die einzelnen Zellprozesse mittels der Process Description Language des SBGN-Standards [11] darstellen soll.

Auch sind Informationen verschiedener Datenbanken, ChEBI [12], GO [13][14], SBO [15], SGD [16] und der Datenbank vom YCM Projekt, in die Darstellung zu inkorporieren.

3. Vorgehen

Zur Erfassung der Anforderungen an das Design wird zunächst ein Mockup der Visualisierung entworfen. Dies muss dann in regelmäßigem Austausch mit Jens Hahn und Josch K. Pauling vom YCM Projekt angepasst und erweitert werden.

Auf Basis mehrerer Interviews sind bereits folgende Eigenschaften des zugrundeliegenden Modells und Anforderungen an die Visualisierung bekannt.

Das Ganzellenmodell des YCM Projekts ist in mehrere Module unterteilt, welche jeweils einen Teilprozess der Zelle, wie Zellwachstum oder der Transport durch Membranen, modellieren. In den Modulen enthalten sind Beschreibungen von Spezies, wie zum Beispiel ATP, und ihrer Interaktionen. Die Spezies sind Compartments zugeordnet, welche subzelluläre Strukturen, wie zum Beispiel eine Membran oder Cytosol, repräsentieren. Dabei kann die selbe Spezies in mehreren Compartments enthalten sein. Jedes dieser Module wird in Form von SBML- [17] bzw. Python-Dateien vorgegeben.

Für das Design der Abstraktionsstufen ist vorgegeben, dass die feinste Abstraktionsstufe sich an der Process Description Language des SBGN-Standards orientieren soll. Das macht die feinste Ebene zu einem gerichteten Graphen, dessen Knoten einzelne Spezies und Reaktionen zwischen ihnen darstellen. Geordnet werden die einzelnen Teilgraphen durch die Compartments in denen die Einzelnen Spezies enthalten sind. Zu beachten bei der Darstellung auf dieser Stufe ist, dass es einige hoch vernetzte Spezies wie ATP gibt, welche in typischen Darstellungen biochemischer Prozesse nicht dargestellt werden. Solche Spezies und ihre Verbindungen sind dezenter darzustellen. Dies könnte zum Beispiel durch die Darstellung mit geringerer Farbsättigung erreicht werden.

Durch Anklicken eines Knotens der zu einer bestimmten Spezies gehört, sollen Informationen zu der entsprechenden Spezies von den zuvor genannten Datenbanken abgerufen und dargestellt werden.

Größere Abstraktionsstufen zeigen die verschiedenen Compartments. Da die Compartments subzelluläre Strukturen beschreiben sind mit ihnen auch Eigenschaften wie Größe verbunden, welche ebenfalls nach Möglichkeit graphisch dargestellt werden sollen.

Die größte Abstraktionsstufe stellt die Zelle als Ganzes, sowie die in ihr enthaltenen Module dar.

Die entwickelte Visualisierung sollte sich Änderungen in der Anzahl und Aufbau der Module anpassen können um eine später Modifikation des YCM Projekts zu erleichtern.

Ist die Bearbeitung des Mockups abgeschlossen, gilt es einen Prototyp der Visualisierung zu implementieren. Als Programmiersprache wird JavaScript verwendet um eine Plattform übergreifende

Webanwendung zu entwickeln. Dabei wird ECharts[18] als Bibliothek zur Darstellung der Graphen verwendet.

Da dieses Projekt als Studienprojekt zeitlich stark begrenzt ist, können wahrscheinlich nicht alle gewünschten Features vollständig implementiert werden, es kann sich jedoch eine Masterarbeit anschließen, welche die Entwicklung fortführt.

4. Quellenverzeichnis

- [1] Kitano, Systems Biology: A brief overview, Science 2002.Kitano, H. (2002). Systems biology: a brief overview. *Science*, 295(5560), 1662-1664.
- [2] Yeast Cell Modeling. <https://rumo.biologie.hu-berlin.de/tbp/index.php/en/research/topics2/42-research/topics/32-yeast-cell-modeling>. Accessed 4 Jan 2018
- [3] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., ... & Hutchison 3rd, C. A. (1999). E-CELL: software environment for whole-cell simulation. *Bioinformatics (Oxford, England)*, 15(1), 72-84.
- [4] Domach, M. M., Leung, S. K., Cahn, R. E., Cocks, G. G., & Shuler, M. L. (1984). Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotechnology and Bioengineering*, 26(3), 203–216.
- [5] Castellanos, M., Wilson, D. B., & Shuler, M. L. (2004). A modular minimal cell model: Purine and pyrimidine transport and metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17), 6681–6686.
- [6] Morgan, J. J., Surovtsev, I. V., & Lindahl, P. A. (2004). A framework for whole-cell mathematical modeling. *Journal of Theoretical Biology*, 231(4), 581–596.
- [7] Cytoscape. <http://www.cytoscape.org/>. Accessed 17 Jan 2018.
- [8] Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H., CellDesigner: a process diagram editor for gene-regulatory and biochemical networks, *BIOSILICO*, 1:159-162, 2003
- [9] Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., ... & Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2), 389-401.
- [10] Lee, R., Karr, J. R., & Covert, M. W. (2013). WholeCellViz: data visualization for whole-cell models. *BMC bioinformatics*, 14(1), 253.
- [11] Nicolas Le Novère, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, Sarala M Wimalaratne, Frank T Bergman, Ralph Gauges, Peter Ghazal, Hideya Kawaji, Lu Li, Yukiko Matsuoka, Alice Villéger, Sarah E Boyd, Laurence Calzone, Melanie Courtot, Ugur Dogrusoz, Tom C Freeman, Akira Funahashi, Samik Ghosh, Akiya Jouraku, Sohyoung Kim, Fedor Kolpakov, Augustin Luna, Sven Sahle, Esther Schmidt, Steven Watterson, Guanming Wu, Igor Goryanin, Douglas B Kell, Chris Sander, Herbert Sauro, Jacky L Snoep, Kurt Kohn & Hiroaki Kitano. The Systems Biology Graphical Notation. *Nature Biotechnology* 27, 735 - 741 (2009). <http://dx.doi.org/10.1038/nbt1675>

1038/nbt.1558

- [12] Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., and Steinbeck, C. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.*
- [13] Ashburner et al. Gene ontology: tool for the unification of biology (2000) *Nat Genet* 25(1):25-9. *Online at Nature Genetics.*
- [14] Gene Ontology Consortium. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, 45(D1), D331-D338.
- [15] Courtot M., Juty N., Knüpfer C., Waltemath D., Zhukova A., Dräger A., Dumontier M., Finney A., Golebiewski M., Hastings J., Hoops S., Keating S., Kell D.B., Kerrien S., Lawson J., Lister A., Lu J., Machne R., Mendes P., Pocock M., Rodriguez N., Villegier A., Wilkinson D.J., Wimalaratne S., Laibe C., Hucka M., Le Novère N. (2011), *Model storage, exchange and integration. Molecular Systems Biology*, 7:543DOI MSB
- [16] Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) *Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res.*
- [17] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A. , Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novère, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J. (2003). The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, vol. 19, no. 4, pp. 524–531.
- [18] ECharts <https://github.com/ecomfe/echarts> Accessed 14 Feb 2018