Interrogating topology of signaling networks with large-scale data within a Boolean framework

J. Saez-Rodriguez[1,2,3], L. G. Alexopoulos[1,2,3], R. Samaga[4], S. Klamt[4] , D. Lauffenburger [2,3], P. Sorger[1,2,3]

***Short Abstract* — We present a methodology and the corresponding software to create executable models specific to a particular experimental data set. Models are based on a Boolean formalism, which allows fast simulation of large networks. As a proof of principle, we generate models specific to cancerous and normal hepatocytes which allow to uncover mechanistic differences between both cell types.**

***Keywords* — signaling networks, Boolean models, hepatocytes**

I. Background

Novel high-throughput methods allow the generation of large data sets, which represent the ideal substrate for a system-wide computational analysis. Sophisticated insight can be obtained by setting up and subsequently analyzing detailed, mechanistic models. However, due to the large number of unknown parameter values, modeling very large networks is an arduous task. As a means to model complex phenomena, intermediate approaches based on Boolean (logical) descriptions are very useful, as the prior knowledge required is normally available [1]. These models encapsulate the topology and causality of the network without dealing with kinetic parameters.

II. Results

Here, we use this framework to test the consistency between the literature-mined network topology and experimental data across different cell types, conditions, and time scales. Measurements of protein activation under different conditions can be automatically compared with the predictions of a model based on a certain topology. Furthermore, we have developed a methodology to identify the network topology that optimally describes a data set based on a priori knowledge of the network connectivity.

These methods are embedded in CellNetOptimizer (CNO), an open source MATLAB toolbox that uses CellNetAnalyzer [2] as simulation engine. CNI works in concert with DataRail [3] a complementary toolbox for managing, transforming, visualizing, and modeling data, in particular the varied high-throughput data encountered in Systems Biology.

The approach is applied to unravel the differences in the signaling networks between primary and cancerous hepatocytes. Towards this end, a data compendium of 6,560 experimental points was generated in primary hepatocytes and the hepatocellular carcinoma cell line HepG2. Both cell types were stimulated with 10 different pro-growth and inflammatory cytokines, in orthogonal combination with 7 small-molecule inhibitors targeting key signaling mediators. For all these conditions, the phosphorylation levels of 17 essential readouts were measured at 30 minutes and 3 hours. Based on this data and a priori knowledge retrieved from curated databases, we were able to delineate mechanistic models specific for both primary and cancer hepatocytes. By comparing the models we could uncover significant differences in the signaling networks, in particular at the level of IGF-mediated signaling. Furthermore, data that could not be reconciled with the a priori knowledge suggested gaps in our current knowledge that pointed at potential new connections involving crosstalk between inflammatory signals and survival pathways.

Acknowledgements: This work was funded by NIH grant P50-GM68762 and by a grant from Pfizer Inc. to P.K.S. and D.A.L.

[1]Harvard Medical School - Department of Systems Biology
[2] Bioengineering Department, M. I. T.
[3]Cell Decision Process Center
email: julio@hms.harvard.edu

References

[1] Klamt et al., *BMC Bioinf*, 7:56, 2006.
[2] Klamt et al. *BMC Syst. Biol.* 1:2, 2007.
[3] Saez-Rdriguez et al., *Bioinformatics*, 24(6): 840-7, 2008. http://code.google.com/p/sbpipeline/