Comprehending Biochemical Network Dynamics through Automatic Inference of State Transition Diagrams

Debprakash Patnaik, Vandana Sreedharan, Naren Ramakrishnan, and Yang Cao
Department of Computer Science and Genetics, Bioinformatics, and Computational Biology Program
Virginia Tech, Blacksburg, VA 24061

Abstract

The development process for biochemical network models follows the traditional pipeline:

Biochemical Network → ODE → Simulation → Time Series → System Dynamics

Understanding how small changes in the original network propagate to cause qualitative differences in dynamics are key issues in model comprehension, iterative model improvement, and validation. We present an automatic approach to summarize time series data from simulations into state transition diagrams capturing the system dynamics. Finally, we show how key behavioral features inferred from the state transition diagrams can be connected back to the network topology in a way that cannot be directly inferred from the ODE model or the raw time series dataset. Our approach for network comprehension thus opens up an important algorithmic approach to the toolkit of the systems biologist. We demonstrate applications to studying the yeast cell cycle progression both in wild-type cells and in mutants which cause cell cycle arrest at different stages. Our algorithm identifies key cell cycle states, transitions between them, and deviations in these transitions among mutants in a completely unsupervised manner.

Background

The cell cycle is a regulatory system of fundamental biological significance, governed (in Eukaryotes) by a universal mechanism that has been characterized in great detail both genetically and biochemically [Murray and Hunt, 1993]. Realistic and accurate models are available [Chen et al., 2004], which make specific predictions that can be tested experimentally. However, cell cycle modeling has now reached the limit of what can be hand-crafted, and the next level of sophistication will require powerful tools to comprehend regulatory networks and the underlying state transitions they model.

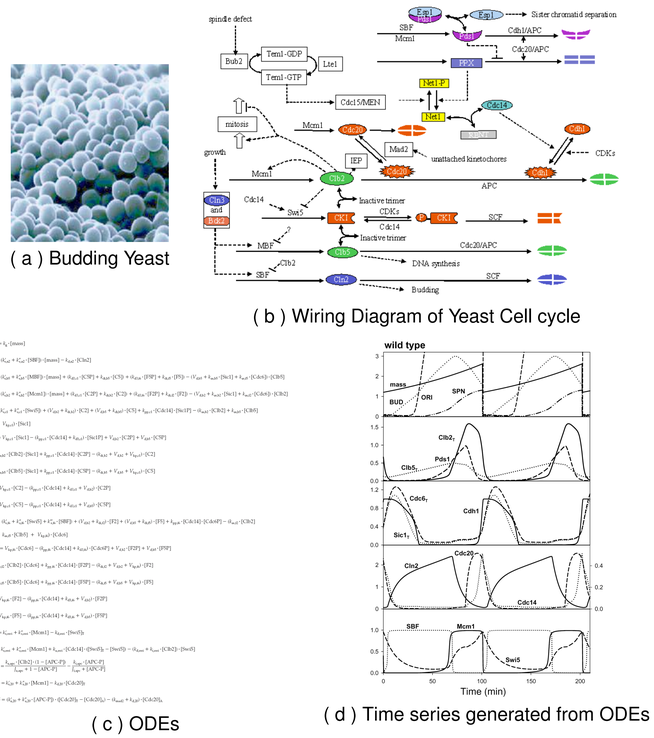


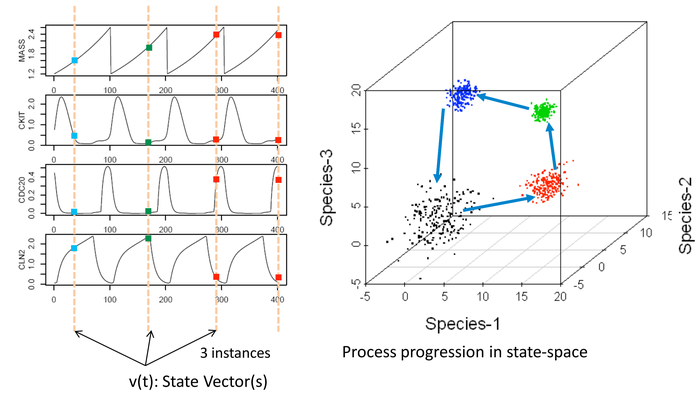
Figure 1: *Modeling the cell cycle*

Molecular biologists have painstakingly dissected and characterized individual components and their interactions to derive a consensus picture of the regulatory network. The responses of the living cell to internal and external stimulus are controlled by complex interacting protein networks. These networks are nearly impossible to comprehended by intuitive reasoning alone. Mathematical modeling, based on biochemical rate equations, provides a rigorous tool for modeling the complexities of molecular regulatory networks. Even thought such models capture the dynamics of the system well, comprehending these models is still difficult.

Methods for Temporal Redescription of Data

We outline two methods for temporal redescription: one based on clustering species concentration vectors and another based on clustering time points. The former has been applied to redescribing data from simulations and the latter has been applied to redescribing data from gene expression measurements.

Temporal information is initially stripped out from simulation results and multivariate species concentration vectors are clustered to identify dense regions of spate space. The original time series data is redescribed in terms of these clusters thus putting back the temporal relationships. Hence, by using clusters of species concentrations to define the "states" and transitions between these states to define the system trajectories, we show how we can reconstruct key dynamical features such as linear state progressions and even higher level features such as oscillations.



Temporal redescription can also be viewed as a task of segmenting the time series data. Each segment is modeled as a mixture of clusters so that segment boundaries involve significant re-grouping and re-definition of clusters [Tadepalli et al., 2008]. This work has been applied to redescribing time series data from gene expression measurements.

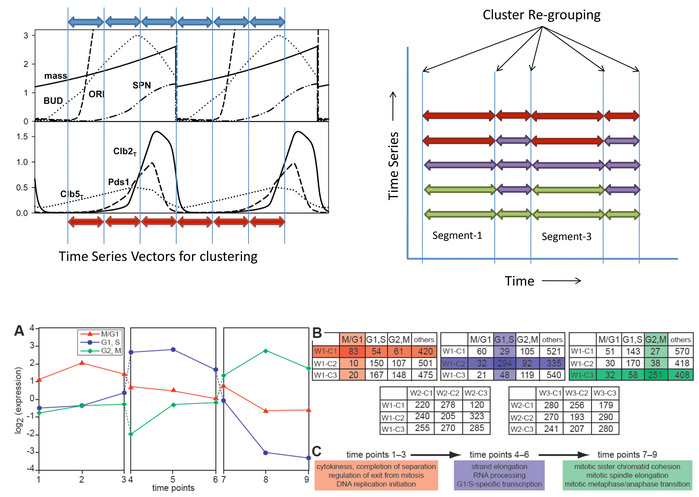


Figure 4: *Segmenting the yeast cell cycle (YCC) data. The YCC involves the staged coordination of several phases (M/G1, time points [1–3]; G1,S, time points [4–6]; and G2,M, time points [7–9]). (A) Mean expression profiles for each group of genes depict the changing emphasis across the three phases. Contingency tables capture the concerted grouping of genes within segments (B, first row) as well as the re-groupings between segments (C, first row). Observe that the contingency tables in the second row involve significant enrichments whereas the tables in the third row approximate a uniform distribution. Gantt chart views (C) depict the temporal coordination of biological processes underlying the dataset.*

Results

In automatic state discovery from cell cycle time series/trajectory data, concentration profiles and rate of change dynamics of each state map directly to known cell cycle phases.

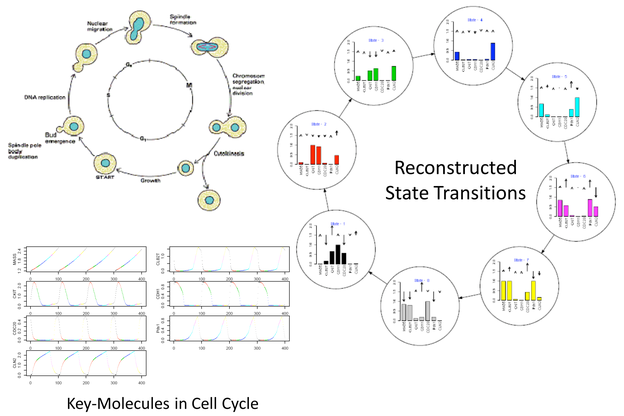


Figure 5: *Reconstruction of the state transition diagram for wild type cell cycle progression*

Mutant cell cycle models:

1. $Cdc14 - ts$: Cdc14 causes inactivation of mitotic CDK, enabling cells to exit mitosis.
2. $Cdc20 - ts$: Cdc 20 activates Anaphase Promoting Complex essential for exit from mitosis.
3. $Clb1\delta Clb2\delta$: Clb1,2 are kinases essential for entry into mitosis.
4. $Tem1\delta$: Tem1 is a GTP-binding protein active in Mitotic Exit Network pathway.

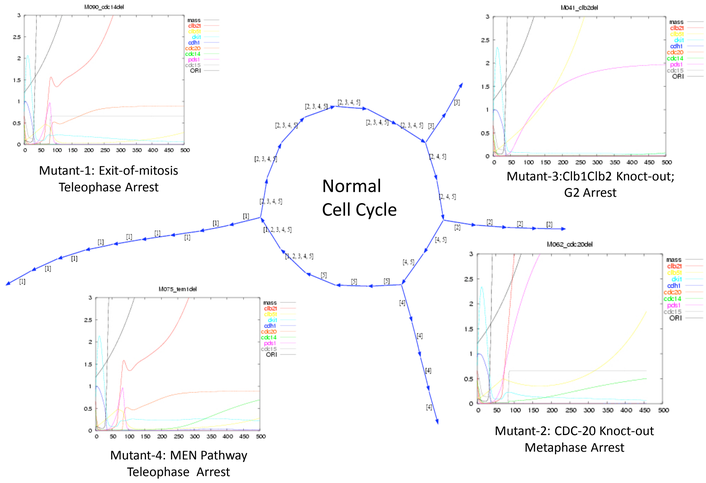


Figure 6: *Contrasting state progression of wild-type vs mutant cell cycle models.*

References

K C Chen, L Calzone, A Csikasz-Nagy, F R Cross, B Novak, and J J Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell*, 15(8):3841–3862, Aug 2004.

A. Murray and T. Hunt. *The cell cycle: An introduction*. W.H. Freeman&co., New York, 1993.

S. Tadepalli, N. Ramakrishnan, L.T. Watson, B. Mishra, and R.F. Helm. Simultaneously segmenting multiple gene expression time courses by analyzing cluster dynamics. In *APBC*, pages 297–306, 2008.