

# ProcessDB: A Cellular Process Database Supporting Large-Scale Iterative Kinetic Modeling in Cell Biology

Ann K. Chasson  
Integrative Bioinformatics  
1903 Adams Ct.  
Mountain View, CA 94040

achasson@integrativebioinformatics.com

Robert D. Phair  
Integrative Bioinformatics  
12114 Gatewater Dr.  
Rockville, MD 20854

rphair@integrativebioinformatics.com

## ABSTRACT

Cellular modeling is essential. Systems with thousands of interacting parts are simply beyond the capacity of human information processing (1), and computational approaches to cellular modeling are widely regarded as one of the grand challenges of 21<sup>st</sup> century biology (2).

This challenge is being approached from the perspective of genomics and proteomics in a number of laboratories (3, 4). Other investigators have approached the problem from the perspective of experimental cell biology (5). Our contribution is also built from the experimental cell biology viewpoint and provides a flexible and dynamic database tool for managing multiple complex models of cellular and molecular systems.

The importance of databases for managing the enormous experimental data sets produced by large-scale biology is widely recognized. In our experience a database for the corresponding elements of biological theory is equally important.

ProcessDB is a database-driven software application, which manages the translation of cell biological diagrams or hypotheses into the corresponding system of nonlinear ordinary differential equations. It can track model modifications in a natural way and thus supports an iterative model development process that is potentially fully computational. Underlying ProcessDB is a relational database, designed and implemented using Oracle. ProcessDB is scalable and will support modeling and data analysis in both individual laboratories and geographically dispersed multi-lab collaborations.

We developed ProcessDB to support the methods of kinetic modeling (6, 7) in the context of very large biological systems. We refer to this combination of mechanistic kinetic modeling and database technology as *integrative bioinformatics*.

The entities in ProcessDB cover the common language of molecular cell biology and biochemical kinetics. They include molecules, complexes, places, states, processes, rate laws, models, and model realizations. Molecules and complexes have their usual definitions. Places are cellular or physiological locations such as the lumen of the endoplasmic reticulum or blood plasma. States, sometimes called compartments or state variables, are defined by a molecule or complex in a place. A process is a binding, transformation, or translocation event that produces a flux of molecules from start states to end states, mediated and controlled by other states. A rate law, often developed from chemical kinetics, is an algebraic expression that quantifies flux (molecules/sec) through a process. A model is a collection of

processes. A model realization is a model whose processes each have an assigned rate law.

Key features of ProcessDB are:

- Support for arbitrarily complex mechanistic kinetic models or hypotheses from individual laboratories or consortia
- A cell and molecular biology-based relational data model
- Explicit decoupling of model/hypothesis storage from experimental data storage
- Automatic support for model comparison and versioning
- Provision for input to the model design process from expert biologists, public databases, and cluster analysis (8)
- Provision for feedback to model formulation from parameter optimization
- A seamless integration with modeling software packages that support plain text input

We have tested the initial implementation of ProcessDB in the context of a small project (14 states) in the cell biology of the secretory pathway (9) that involved five competing alternative hypotheses, and found that its capabilities simplified and accelerated the labor intensive tasks of model modification and testing.

## REFERENCES

1. Miller, G. A. (1956) *Psychological Review* **63**, 81-97.
2. Tomita, M. (2001) *Trends Biotechnol.* **19**, 205-210.
3. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C. *et al.* (1999) *Bioinformatics.* **15**, 72-84.
4. Ogata, H., Goto, S., Fujibuchi, W. & Kanehisa, M. (1998) *Biosystems* **47**, 119-128.
5. Schaff, J. C., Slepchenko, B. M. & Loew, L. M. (2000) *Methods Enzymol.* **321**, 1-23.
6. Phair, R. D. (1997) *Metabolism* **46**, 1489-1495.
7. Jacquez, J. A. (1996) *Compartmental analysis in biology and medicine* Ann Arbor: Bio Medware, 1996).
8. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. U. S. A* **95**, 14863-14868.
9. Presley, J. F., Phair, R. D. & Lippincott-Schwartz, J. (Submitted 2001) *Nature*.