Modeling Complex Networks: Integrating Rules (BioNetGen) and Data Mining (BioPAX Ontology) into the Virtual Cell Framework

Michael L. Blinov, Ion I. Moraru, Oliver Ruebenacker, James C. Schaff, Anuradha Lakshminarayana, Fei Gao, and Leslie M. Loew

University of Connecticut Health Center
Farmington, CT, USA

Richard D. Berlin
Center for Cell Analysis and Modeling

Combinatorial complexity often arises when detailed quantitative models of signaling networks are being sought. A receptor that has 9 phosphorylation sites can exist in 512 different phosphoforms, many of which must be accounted for to simulate the time course of signaling. When details of all protein complexes are being included, this number can easily increase by a few orders of magnitude, and validation, visualization, and understanding of the model can become virtually intractable. A solution for this challenge is provided by 1) automatic extraction from pathway databases of re-usable model components, and 2) rules of interaction based on protein modularity. This way, models of large, complex networks can be assembled from separately constructed and validated components, either directly or via rules.

Problem of Combinatorial Complexity

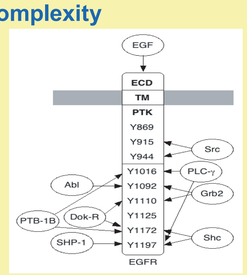
Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

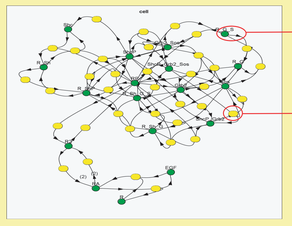
9 sites ⇒ 2^9=512 phosphorylation states

Each site has ≥ 1 binding partner
⇒ more than 3^9=19,683 total states

EGFR must form *dimers* to become active
⇒ more than 1.9×10^8 states



Problem of Manual Specification of Models



Species: One for every possible modification state of every complex.

Reactions: One for every transition among species

Manual specification of each and every species is time-consuming and error-prone...

Needs:

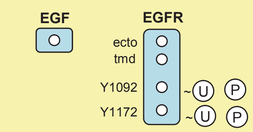
- Automate model creation to make it less error-prone:
  - Create models by specifying only some features of the system and let computer do the rest
  - Create models of signaling networks by retrieving information from public databases
- Enable better visualization of the model
- Enable collaborative modeling efforts.

Solution: modular approach

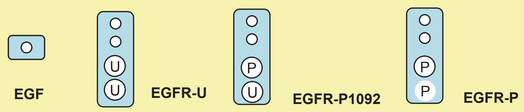
- Introducing re-usable modeling components.
- Using rules of interaction to automatically generate models (rules are based on protein modularity, e.g. when kinetics of binding is independent of other sites).
- Developing standards that allow exchange of information between different tools:
  - rule-based formats,
  - reusability through annotations

SOLUTION 1: Rule-based approach

Specification of multi-state, multi-molecular complexes



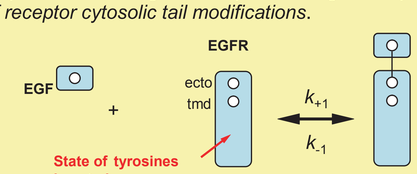
Components of proteins may have attributes, e.g. conformation or phosphorylation state.



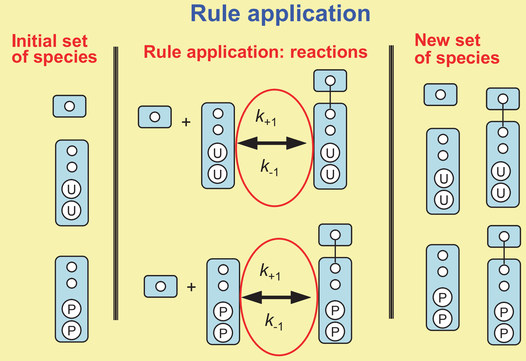
Rule-based specification of the model

Each rule usually specifies an experimentally-derived and testable feature of the system

Example: kinetics of ligand-receptor binding is independent of receptor cytosolic tail modifications.



Rule application



All reactions inherit the same rate law.

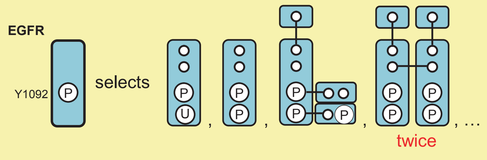
Rule-based model generation

Input: initial species **S**$_0$ | Input: reaction rules $\mathcal{R}$

Rules application 1 $\mathcal{R}(S_0) = R_0, S_1$
Rules application 2 $\mathcal{R}(S_0 \cup S_1) = R_1, S_2$
....
Rules application n $\mathcal{R}(S_n) = R_{n+1}, S_{n+1}$
Termination **Terminate if** $S_n = S_{n+1}$

Model: species S_n and reactions R_{n+1}

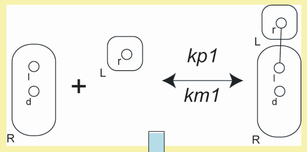
Predictions are reported as "observables", corresponding to groups of species with the same properties, i.e. obeying the same pattern



Pattern that selects EGFR phosphorylated at Y1092.

BioNetGen software

BNGL: A textual language for graphical rules



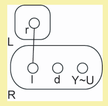
reactant patterns **product pattern** **rate law(s)**

```
L(r)  +  R(l,d)  <>  L(r!1).R(l!1,d)  kp1, km1
```

molecule | components | a chemical bond

Bonds and complexes

Bonds are indicated by edges in the species graph. Bonds are indicated by an `!<number>`, where `<number>` is the index of the bond. It is used only to identify bond endpoints. All bonds are otherwise equivalent.



bond

`L(r!1).R(l!1,d,Y~U)`

'.' used to join molecules

BioNetGen@VCell

Specify rules in BNGL (BioNetGen Language)

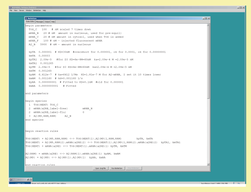
Generate SBML model in BioNetGen

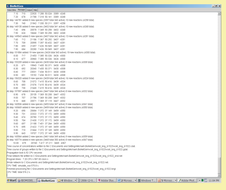
Virtual Cell BioModel

1. Launch BioNetGen from the Virtual Cell.

2. Write a model in BNGL language

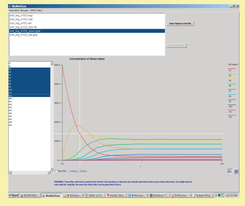
3. Run BioNetGen to generate a reaction network (and, optionally, solve it).

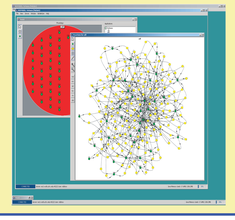




4. Check steps of reaction network generation and timecourses simulation.

5. Look at and save timecourses.





6. Select .xml file and create a Virtual Cell BioModel.

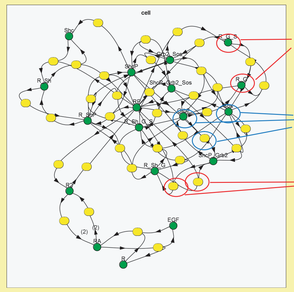
7. BioModel is launched in the Virtual Cell window.

Available at http://www.vcell.org/bionetgen
E-mail to blinov@uchc.edu

Collaborators: BioNetGen software was developed at Los Alamos National Laboratory by Michael L. Blinov, James R. Faeder and William S. Hlavacek.

SOLUTION 2: Enable automated modeling and better visualization through Data Mining

Need: structural and modular representation



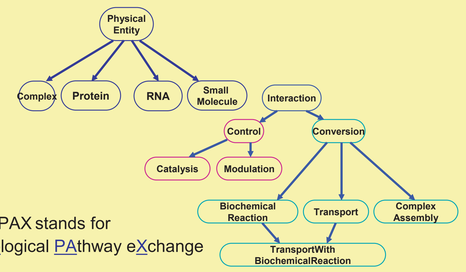
Standard groupings - types of species: proteins, small molecules, complexes, etc.

Arbitrary user-specified groupings: ligand-binding module, Grb2 interactions, etc.

Standard groupings - types of reactions: biochemical reaction, complex assembly, etc.

Using pathway data stored in BioPAX ontology

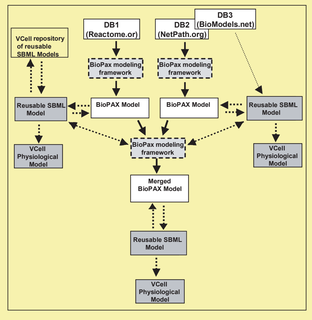
http://reactome.org http://pid.nci.nih.gov/PID



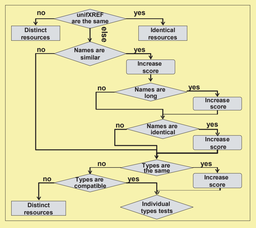
BioPAX stands for Biological PAthway eXchange

BioPAX@VCell modeling framework

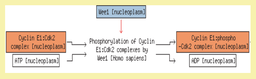
From public data in BioPAX format to a reusable model

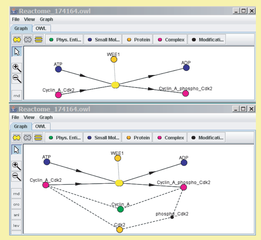


Using several data sources: merging BioPAX files



Reactome data to the Virtual Cell model





Screenshot of BioPAX modeling framework.
(a) Phosphorylation of Cyclin E1:Cdk2 complexes by Wee1 as displayed by Reactome database.
(b) A compact representation: only species and reactions are shown in the form compliant with the VCell GUI .
(c) An extended view, where a user gains an understanding that both complexes contain the same components Cyclin_A and Cdk2; Cdk2 is a protein, which enters one of these complexes in phosphorylated form.