Proceedings of the 2009 Winter Simulation Conference
M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, eds.

COMPARTMENTAL RULE-BASED MODELING OF BIOCHEMICAL SYSTEMS

Leonard A. Harris
Justin S. Hogg
James R. Faeder

Department of Computational Biology
University of Pittsburgh School of Medicine
Pittsburgh, PA 15260, USA

ABSTRACT

Rule-based modeling is an approach to modeling biochemical kinetics in which proteins and other biological components are modeled as structured objects and their interactions are governed by rules that specify the conditions under which reactions occur. BIONETGEN is an open-source platform that provides a simple yet expressive language for rule-based modeling (BNGL). In this paper we describe compartmental BNGL (cBNGL), which extends BNGL to enable explicit modeling of the compartmental organization of the cell and its effects on system dynamics. We show that by making localization a queryable attribute of both molecules and species and introducing appropriate volumetric scaling of reaction rates, the effects of compartmentalization can be naturally modeled using rules. These properties enable the construction of new rule semantics that include both universal rules, those defining interactions that can take place in any compartment in the system, and transport rules, which enable movement of molecular complexes between compartments.

1 INTRODUCTION

Proteins in cellular regulatory systems can interact in a combinatorial number of ways to generate myriad protein complexes (Hlavacek et al. 2003). These interactions, because of the multicomponent composition of proteins, can be modeled by rules that specify the classes of reactions that can occur and define reaction networks that account comprehensively for the consequences of protein-protein interactions (Hlavacek et al. 2006). The assumption underlying this modeling approach, which is consistent with the modularity of regulatory proteins (Pawson and Nash 2003), is that interactions are governed by local context, that is, properties of the components of the interacting molecules that are proximal to the site of interaction. As long as this is the case, rule-based models can achieve a compact description of a network with a large (even infinite) number of complexes (Hlavacek et al. 2006). The complex spatial topology of the cell, however, can also have profound effects on the regulation of cellular processes by controlling both the reactions that can occur and the rate at which they do so. Membranes are key players not only because they separate molecules in different compartments but also because they mediate the flow of material and information from one compartment to another (Alberts et al. 2002). Chemical reactions that occur at a cell membrane also have an enhanced rate because of the drastically smaller volume of the membrane compartment, a thin fluid layer roughly 100-1000 times smaller than the volume of the cell (Haugh and Lauffenburger 1997; Kholodenko, Hoek, and Westerhoff 2000). The spatial localization of a reacting species in a biochemical network is thus a critical property that affects its reactivity. A rule-based model of the network must therefore describe, either implicitly or explicitly, its effects on the constituent reactions. The main goal of this paper is to describe an extension to an existing rule-based modeling language for biochemical systems that explicitly represents the topology of cellular compartments, the localization of species to specific compartments, and the effects of localization on biochemical reaction rates.

BIONETGEN is an open-source software package that provides tools and a language [the BIONETGEN language (BNGL)] for rule-based modeling of biochemical systems (Faeder, Blinov, and Hlavacek 2009). The syntax and semantics of the language are formally rooted in graph theory (Blinov et al. 2006), but the language itself is simple, intuitive, and accessible to modelers with a wide range of mathematical backgrounds. Models can be developed and simulated using standalone software (Faeder, Blinov, and Hlavacek 2009), a web interface designed to facilitate collaborative modeling (Hu et al. 2009), or through a portal to the Virtual Cell modeling platform (Blinov, Ruebenacker, and Moraru 2008). The standalone software

allows both deterministic and stochastic simulations in the well-mixed limit, and separate programs have been developed to perform particle-based simulations of BNGL models (Colvin et al. 2009; Sneddon, Faeder, and Emonet 2009). A module to perform automated coarse-graining of deterministic (ODE-based) models is also available (Borisov et al. 2008). Recent examples of models that have been developed in BNGL appear in Nag et al. (2009), An and Faeder (2009), and Barua, Faeder, and Haugh (2009). Other approaches to rule-based modeling are discussed in Section 5.1.

BNGL does not currently allow explicit representation of cellular compartments and does not systematically account for the effects of spatial localization either in the selection of species that can undergo reactions or in the calculation of reaction rates. The effects of localization can be modeled in an ad hoc manner but it obscures the generality of molecular interactions, may require additional restrictions [such as the `include` and `exclude` commands (Barua, Faeder, and Haugh 2009)], and may require extensive enumeration of complexes in the rules. In this paper, we describe an extension of BNGL, which we call *compartmental* BNGL (cBNGL), that enables explicit modeling of the compartment topology of the cell and its effects on system dynamics. We show that by introducing a compartment topology, making localization a queryable attribute of both molecules and species, and introducing appropriate volumetric scaling of reaction rates the effects of compartmentalization can be naturally included in a rule-based model. The outline of the remainder of the paper is as follows. In Section 2, we motivate our introduction of cBNGL with a schematic cell regulatory model that captures many essential features of intracellular biochemistry. We then give a brief overview of BNGL in Section 3 before introducing the compartmental extension in Section 4. In Section 5, we summarize the strengths and weaknesses of cBNGL and compare it to various related approaches.

2 A COMPARTMENTAL MODEL OF THE CELL

The function of a signal transduction pathway is to detect an extracellular signal, relay this information inside the cell, and induce a change in cell function (Alberts et al. 2002). In Figure 1, we present a model of receptor-mediated signaling coupled with nuclear transport and transcriptional gene regulation that highlights the role of compartmental localization and transport.

Signaling is initiated when an extracellular ligand (L) is detected by membrane-localized receptors (R) that bind the ligand (R_1; Figure 1). Because L can bind to itself (R_2), ligand-bound receptors can dimerize (also R_2) and be brought inside the cell by endocytosis ($R_{3–5}$), a process in which a small region of the plasma membrane is pinched off, forming a small bubble called an endosome (Alberts et al. 2002). Localizing a receptor complex to an endosome has the effect of trapping the ligand and receptor molecules in a small volume, which reduces entropy and enhances the free energy of binding. Note that during receptor internalization the receptor complex moves from PM to EM, bound ligands move from EC to EN, and receptor-bound molecules in CP remain in CP.

Receptor dimerization brings the cytoplasmic domains of receptors into close proximity, allowing transautophosphorylation of a tyrosine amino acid by the receptors' catalytic domains ($R_{6–7}$). Inactive transcription factor (TF) can bind phosphorylated receptors (R_8), leading to transautophosphorylation of TF in complexes containing receptor dimers ($R_{10–11}$). Phosphorylated TF tends to unbind from the complex (R_9) and has a high affinity for forming dimers (R_{12}). Dimerized TF forms an active transcription factor that is escorted into the nucleus, through a nuclear pore, by an importin (Im) molecule (R_{24}, R_{28}). Inside the nucleus, the TF dimer binds to a promoter on DNA activating transcription of mRNA (R_{13}), which is transported out of the nucleus to the cytosol (R_{16}) and translated into P1 (R_{18}). Cytosolic P1 is also escorted into the nucleus by Im (R_{28}), where it binds a second promoter to activate transcription (R_{15}, R_{17}, R_{19}) and expression (R_{30}) of a second protein (P2).

3 OVERVIEW OF BNGL

The syntax and semantics of BNGL have been described in Faeder, Blinov, and Hlavacek (2009). Briefly, a BIONETGEN model is comprised of five basic elements defined in separate blocks of a BNGL input file: `parameters`, `molecule types`, `seed species`, `reaction rules` and `observables`. A sixth block, `actions`, specifies the operations that are to be carried out on the model, such as generating a network or performing a simulation. BIONETGEN currently supports four simulation methods: `ODE`, which solves the ordinary differential equations arising from a fully enumerated network; `SSA`, which uses Gillespie's algorithm (Gillespie 1976) to stochastically simulate a fully or partially enumerated network; `PS`, which calls the particle-based simulator DYNSTOC (Colvin et al. 2009), which implements a generalized version of the STOCHSIM algorithm (Morton-Firth and Bray 1998); and `NF`, which calls a different particle-based simulator NFSIM (Sneddon, Faeder, and Emonet 2009), which generalizes the rule-based kinetic Monte Carlo algorithm of Yang et al. (2008).

Molecules, the basic building blocks of a BIONETGEN model, are declared in the `molecule types` block. Molecules may contain components, which represent the functional elements of molecules and may bind other components, either in

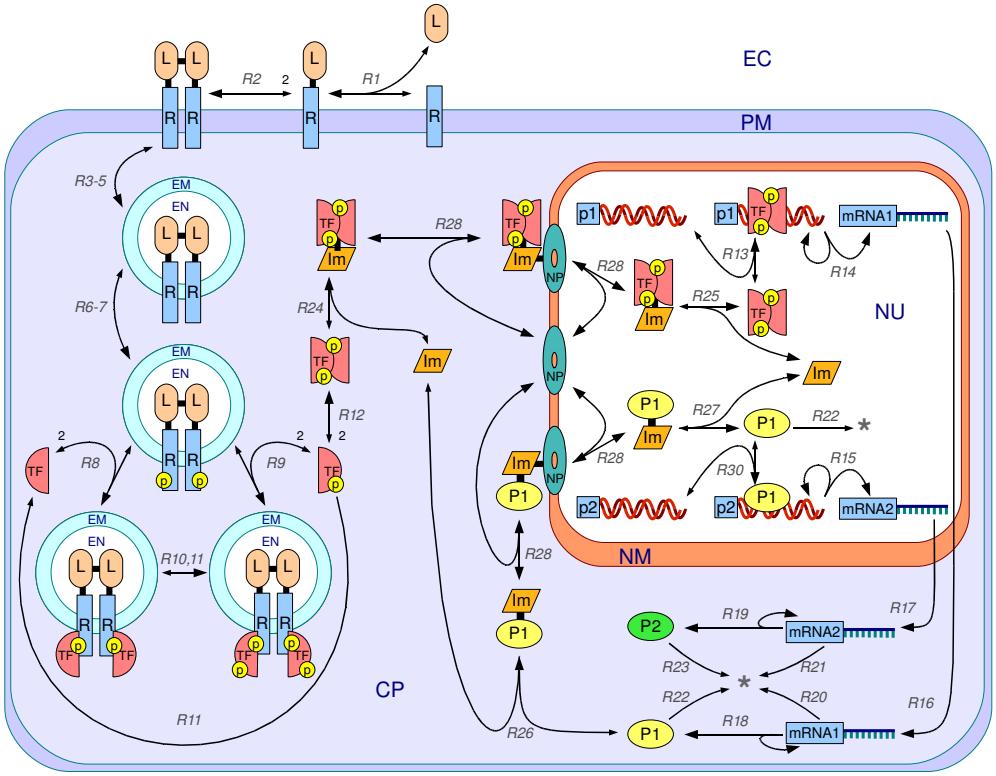


Figure 1: A compartmental model of the cell. The model couples simplified processes of signal tranduction, nuclear transport and transcriptional regulation in a single eukaryotic cell. The system consists of four volume compartments: extracellular (EC), cytosol (CP), endosomal (EN) and nuclear (NU). These are separated by three membrane surfaces: plasma (PM), endosomal (EM) and nuclear (NM). The model is presented as a pathway that proceeds from ligand (L) binding to expression of protein P2. The underlying rule-based model defines a set of 354 reactions between 78 species. Bonds between molecules are shown as black lines. Black arrows between species represent reactions. Gray arrow labels correspond to the rule number that describes the reaction (see model files at www.bionetgen.org/wsc09). Black integer-valued arrow labels represent reaction stoichiometry (if not equal to unity). DNA promoters are pictured as a double helix icon.

the same molecule or another. Components may be associated with state variables with a finite set of possible values, each representing a conformational or chemical state of a component, such as phosphorylation status. The name of the molecule type is given first followed by a comma-separated list of its components in parenthesis. The allowed values of state variables are indicated by '~' followed by a value, as in `L(r,d,loc~EC~EN)`, which declares a ligand molecule `L` that contains a receptor binding component `r`, a dimerization component `d`, and a location component `loc` that takes on values EC or EN.

The `seed species` block defines the species initially present in the system. For example, the line

```
L(r,d,loc~EC)  Lig0
```

specifies that the initial amount of free ligand monomers in EC is `Lig0`, a parameter defined in the `parameters` block. Molecular complexes may also be specified, as in `L(r,d!1,loc~EC).L(r,d!1,loc~EC)`, where a bond linking the `d` components of each `L` molecule is indicated by a shared bond label, `!1`.

The `reaction rules` block contains rules that define how molecules interact. A rule is comprised of a set of reactant patterns, a transformation arrow, a set of product patterns, and a rate law. A pattern is a set of molecules that select species through a mapping operation (Blinov et al. 2006). The match of a molecule in a pattern to a molecule in a species depends only on the components specified in the pattern (including wildcards), so that one pattern may select many different species. The '+' operator separates two reactant patterns that must map to distinct species (i.e., they may not reside in the same complex). The '.' operator separates molecules that are part of the same species. The transformation arrow may be either unidirectional (`->`) or bidirectional (`<->`). Five basic types of operations are carried out by the rules in the example system

of Section 2: binding and unbinding of two molecules through a specified pair of components, transformation of a state variable, synthesis, and degradation. An example of a binding rule is

```
R8: R(tf~pY) + TF(d~Y,r,loc~CP) <-> \
      R(tf~pY!1).TF(d~Y,r!1,loc~CP)  kp_R_TF/vol_CP, km_R_TF
```

where the underlined portions identify the reaction center, which is the set of components modified by the rule. (`R8` indicates that this is rule #8 in the BNGL model file.) This rule specifies that *any* R molecule containing an unbound, phosphorylated `tf` component may bind to a `r` component of an unphosphorylated TF in CP. Here, the `tf` component of `R` is bound to the `r` component of `TF` by the addition of an edge labeled `1`, indicated by the two bond labels (`!1`) in the products.

`kp_R_TF/vol_CP` and `km_R_TF` specify the rate constants in the forward and reverse directions. (Assuming that the bimolecular rate constant is given in standard units of M^{-1} s^{-1}, `kp_R_TF` is that value divided by Avogadro's number.) In this case, the rate constant for the forward direction is a formula comprised of a bimolecular rate constant divided by the volume of CP. Parameters in BNGL have no explicitly defined units and bimolecular rate constants are generally given on a *per molecule per cell* basis. BNGL supports elementary (`Ele`) rate laws as well as two non-elementary types, Michaelis-Menten (`MM`) and saturation (`Sat`) (see Faeder, Blinov, and Hlavacek 2009).

The `observables` block contains definitions of model outputs, which are sums over the populations of species matched by patterns. Several examples are provided in the model files at www.bionetgen.org/wsc09

4 cBNGL

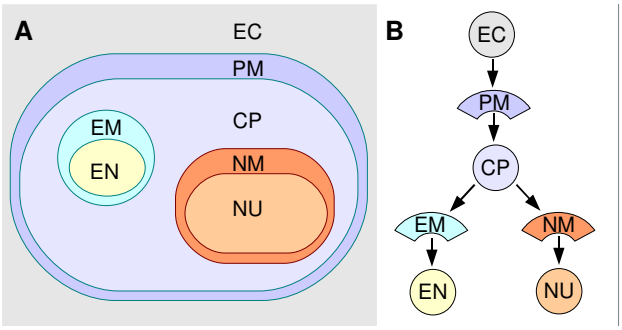
As discussed in Section 1, compartmental models such as the one in Figure 1 can, in principle, be modeled using standard BNGL. For example, we have seen in Section 3 that in order to specify location in BNGL a component (`loc`) can be added to the component lists for molecules. Rules can then be written with the appropriate location states as context for the interacting molecules. In the common case where identical interactions can take place in different compartments, this means that multiple versions of the same interaction rule must be enumerated. Each will differ by only the values of the location labels and, in the case of bimolecular interactions, by the volume-dependent rate constant. Besides being tedious, this approach obscures the generality of interactions specified by a rule. In more complex cases, for example when there are many different ways that a particular molecule can be tethered to a membrane, the required enumeration will be prone to error; it is exactly such enumeration that the rule-based approach was developed to avoid. A similar situation can arise in transport reactions that depend only on the presence of a particular molecule or component state; a different rule is required for each possible stoichiometry of the transported complex. In cBNGL, this enumeration is avoided by introducing localization as a property of both molecules and species, with the localization of a species being a derived property of the localization of its constituent molecules. Species localization also permits restricting the scope of rule application to reactants in the same or adjacent compartments and the determination of volume-dependent rate constants for bimolecular and higher-order reactions. The local nature of molecular bonds also imposes natural topological constraints on complexes. If a molecule is tethered to the plasma membrane, then it cannot be transported to the nucleus without first breaking the tether. Rule application is restricted to adhere to these constraints. In the following, we describe version 1.0 of cBNGL in detail.

4.1 Units

Species counts are assumed to be in population units or moles, not concentrations, in keeping with standard BNGL (Faeder, Blinov, and Hlavacek 2009). Rate constants for bimolecular reactions, however, are assumed to be given in units of $volume/time$ and Michaelis constants for `MM` and `Sat` rules in units of $volume^{-1}$ (this assumes that the appropriate factor of Avogadro's number, N_A, is included in the value, e.g., k_{bi}/N_A and $K_M \times N_A$). This allows for the specification of universal reaction rules that apply across compartments and whose rates are automatically scaled by the appropriate compartment volume (see Section 4.5.1).

4.2 Compartment Topology

Cellular topology, as depicted in Figure 2A, implies that the compartment graph, in which nodes represent compartments and directed edges represent containership, is a tree (Figure 2B). The structure is essentially the same as the compartment structure used in the Systems Biology Markup Language (SBML) (Hucka et al. 2008). Rules for defining compartment topologies in cBNGL are as follows:



C

```
begin compartments
  EC   3  vol_EC                  # extracellular fluid
  PM   2  sa_PM*eff_width  EC     # plasma membrane
  CP   3  vol_CP           PM     # cytoplasm
  NM   2  sa_NM*eff_width  CP     # nuclear membrane
  NU   3  vol_NU           NM     # nuclear space
  EM   2  sa_EM*eff_width  CP     # endosomal membrane
  EN   3  vol_EN           EM     # endosomal space
end compartments
```

Figure 2: (A) Illustration of a cell model with extracellular (EC), cytoplasmic (CP), nuclear (NU) and endosomal (EN) volumes and plasma (PM), endosomal (EM) and nuclear (NM) membrane surfaces. (B) Representation of the compartment topology by a directed graph. Volumes are represented by circles and membranes by arcs. A directed edge points from C1 to C2 if compartment C1 is immediately outside of compartment C2. (C) cBNGL specification of the topology shown in B.

1. A surface compartment must either be an outermost compartment (i.e., no containing compartment) or be contained by a single volume compartment.
2. A volume compartment must either be an outermost compartment or be contained by a single surface compartment.
3. A surface compartment may contain only a single volume compartment.
4. A volume compartment may contain *multiple* surface compartments.

Compartments in cBNGL are declared in a `compartments` block analogous to other blocks used in BNGL (see Section 3). The syntax of each line in the block is

```
[index] compartment_name dimension volume [containing_compartment]
```

where the square brackets denote optional arguments. `compartment_name` is a standard BNGL name (see Faeder, Blinov, and Hlavacek 2009), `dimension` is either 2 or 3 depending on whether the compartment is a surface (e.g., plasma membrane) or a volume (e.g., cytoplasm) and `volume` is the compartment volume, in units consistent with those used for bimolecular rate constants in the `parameters` block. For a surface compartment, the volume is the product of the surface area and an effective width, which accounts for the enhancement of a reaction rate relative to its value in three-dimensional space (Haugh and Lauffenburger 1997). `containing_compartment` is the name of the parent compartment, if applicable (e.g., CP is contained by PM). The full topology specification for the example system of Section 2 is shown in Figure 2C.

4.3 Molecule Location

Each instance of a molecule in cBNGL has a compartment attribute, obviating the need for the location (`loc`) components of Section 3. Ligand molecules, for example, are declared in the `molecule types` block of a cBNGL input file as `L(r,d)`. In species, molecules must be given a location, specified using what we call postfix notation, e.g., `L(r,d)@EC`, which specifies a free ligand molecule in EC. Postfix notation may also be used in patterns to specify the location of a molecule, as in `L(r)@EC`. This will match ligand molecules in EC with a free `r` site.

4.4 Species Location

Complexes can be built in cBNGL from molecules in the same way as in standard BNGL. For example, a freely diffusing ligand dimer in EC is specified as `L(r,d!1)@EC.L(r,d!1)@EC`. Complexes can comprise molecules in adjacent compartments, as in the pattern `L(r!1)@EC.R(l!1)@PM`, but are currently not allowed to span multiple surface compartments. Complexes may then span, at most, a single surface compartment and its two adjacent volume compartments. The localization of a species is a global property of a species that is based on the location of each element comprising the species.

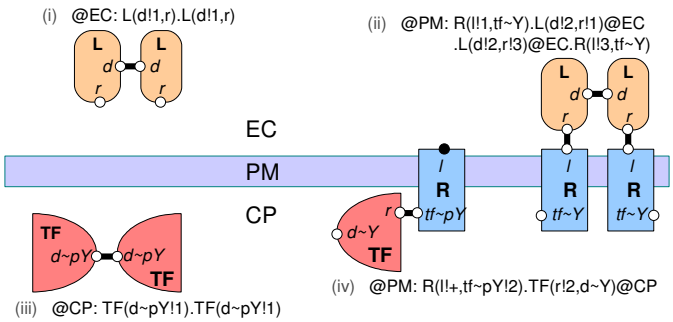


Figure 3: Species and patterns in cBNGL. (i) Species declaration for a ligand dimer localized in the extra-cellular space (EC) using compartment prefix notation. (ii) Species declaration of a receptor-dimer complex. The species is localized to the plasma membrane (PM) but two of the member molecules are in the EC. (iii) A pattern that matches all TF dimers that are phosphorylated at site `d` and localized to the cytosol (CP). This pattern is not a species since components `r` and `im` of TF are not declared. (iv) A pattern that matches complexes localized to the PM containing R and TF with additional context. Site `l` is shown as a filled black circle, indicating that R must be bound to some, unspecified molecule through site `l`.

Definition 1 (Species localization)**.**
A species ***X*** *is said to be localized to a volume compartment* ***V*** *if, and only if, all of the molecules in* ***X*** *reside in* ***V****. Conversely,* ***X*** *is said to be localized to a surface compartment* ***S*** *if any molecule in* ***X*** *resides in* ***S****. Species cannot be formed that span multiple surface compartments.*

Species can be referenced in cBNGL using a prefix notation that specifies their location in line with the above definition. For example, the ligand dimer species above can be written alternatively as `@EC:L(r,d!1).L(r,d!1)`. Prefix and postfix notations can also be used together, as in

```
@PM:R(l!3,tf~pY).L(r!3,d!1)@EC.L(r!2,d!1)@EC.R(l!2,tf~Y)
```

where both R molecules are located in PM. More examples of species and patterns in cBNGL are shown in Figure 3.

4.5 Reaction Rules

Rules in cBNGL are written using the same basic syntax as in standard BNGL (Section 3). The key difference is the possible specification of compartment localization for molecules or species in the reactant and product patterns. When compartments are omitted a rule is considered universal, acting on all sets of matching reactants that are in the same or adjacent compartments. If compartments are specified, we refer to the rule as scope-restricted. Incorporating compartments into the modeling language also requires the introduction of new types of rules for modeling compartment-to-compartment transport. Currently, transport rules are always scope-restricted because both the source and destination compartments must be specified. Lifting this restriction will require more extensive modification of the current BNGL syntax.

4.5.1 Universal Reaction Rules

The utility of universal rules is that they simplify the specification of compartmental models. As discussed above, specifying such models in standard BNGL requires use of location components and enumeration of different versions of the same rule in different compartments. This is avoided in cBNGL by using universal rules. In addition, cBNGL automatically applies the restriction that reactants be in the same or adjacent compartments and applies the correct volumetric scaling to rate constants and other reaction parameters for bimolecular and higher-order reactions. No scaling is required for first order reactions. For elementary bimolecular reactions with the rate constant given in units of *volume*/*time*, the rate constant is divided by the volume of the reactant compartment with the highest dimension (Haugh and Lauffenburger 1997; Barua, Faeder, and Haugh 2009). (The obvious generalization for higher-order reactions, dividing by the product of the $N-1$ highest-order compartment volumes, is implemented but its use is *not recommended*.) Note that volumetric scaling is performed for *all* rules not just universal ones.

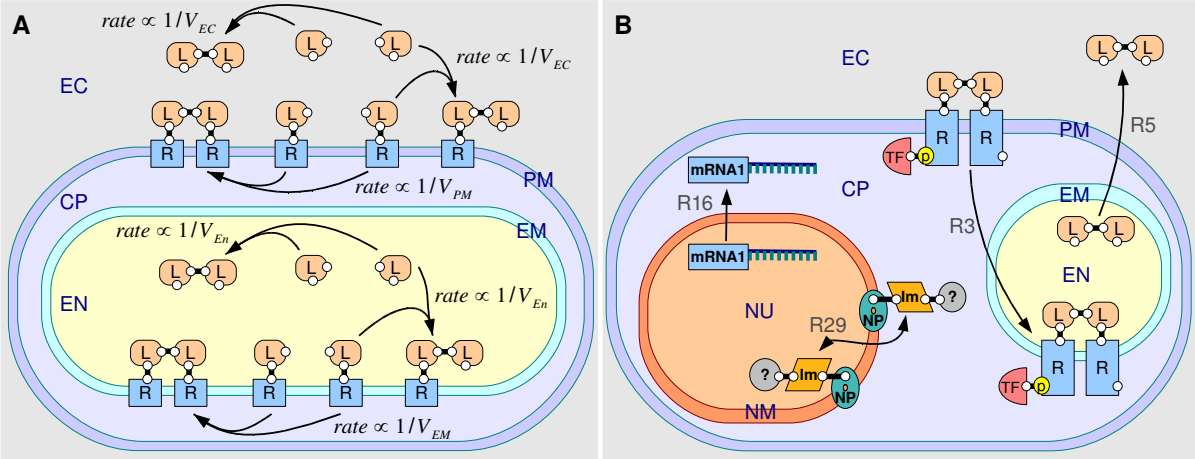


Figure 4: (A) Six instances of the universal rule for ligand dimerization. The single rule describes: (i) free ligand dimerization in the extracellular space (EC) and (ii) in the endosomal space (EN); (iii) free ligands binding to receptor-bound ligands at the plasma membrane (PM) and (iv) at the endosomal membrane (EM); and (v) receptor-bound ligand dimerization at PM and (vi) at EM. The rates for the individual reaction instances are automatically scaled by the volume of the reaction compartment (see Section 4.5.1). (B) Four types of transport reactions. (i) `R5: @EN:L -> @EC:L`. Volume-to-volume species transport. Free ligand dimers are recycled from the endosome to the extracellular space. (ii) `R3: @PM:R.R -> @EM:R.R`. Surface-to-surface species transport. Receptor-dimer complexes are internalized by endocytosis. Molecules in the dimer complex are transported along with the dimer. Molecules in PM are transported to EM, molecules in the adjacent EC are transported to EN and molecules in CP remain in CP. (iii) `R16: mRNA@NU -> mRNA@CP`. Single molecule transport. (iv) `R29: Im@CP.NP <-> Im@NU.NP`. Molecule transport with cargo. Importin bound to a nuclear pore is transported into the nucleus along with any bound cargo (specified by `MoveConnected` keyword; not shown).

An example of a universal rule from the cBNGL specification of the compartmental model in Section 2 is ligand dimerization:

```
R2: L(d) + L(d) <-> L(d!1).L(d!1)   kp_LL, km_LL
```

Figure 4A illustrates the six instances of this rule along with the appropriate scaling factors. Universal rules can also describe the synthesis of molecules, as in

```
R14: DNA(p1!+) -> DNA(p1!+) + mRNA1()   k_transcribe
```

The created molecule is placed is in the same volume as the reactant or, for bimolecular reactions, in the same volume used for scaling the rate constant (higher-order synthesis reactions are not currently supported). In the above rule, mRNA1 is placed in the same compartment as the DNA, which is always NU. Specifying a different localization for mRNA1 would override this behavior.

4.5.2 Scope-Restricted Rules

Scope restriction limits application of a rule to species in a particular compartment through the specification of localization in reactant patterns. An example of a scope-restricted reaction rule in the compartmental model of Section 2 is

```
R11:  TF(d~pY)@CP -> TF(d~Y)@CP   k_TF_dephos
```

which specifies that dephosphorylation of transcription factors can only take place in the cytoplasm, which might be the case if an implicitly modeled phosphatase were localized there.

In scope-restricted rules it is possible for a reactant pattern to match a species that upon application of the rule would create a product that does not match the product pattern. For example, consider a scope-restricted version of the ligand

dissociation rule:

```
@PM:L(d!1).L(d!1) -> @PM:L(d) + @PM:L(d)
```

The pattern on the left-hand side matches receptor-ligand complexes that contain either one or two membrane-bound receptor molecules. Only those containing two receptor molecules, however, will result in two product species localized to PM after the bond between ligands is broken. BIONETGEN checks that products of rule application match the product patterns, aborting the application if they do not.

4.5.3 Transport Rules

There are currently four classes of transport rule supported by cBNGL (see Figure 4B): (i) molecule transport, (ii) cargo-carrying (trafficking) transport, (iii) volume-to-volume species transport, and (iv) surface-to-surface species transport.

Molecule transport is the simplest of these: the compartment designations of explicitly specified molecules are changed upon application of the rule. cBNGL allows molecules to transport volume-to-volume across a surface, volume to adjacent surface, and surface to adjacent volume. Surface-to-surface molecule transport is not permitted. Application of a molecule transport rule is rejected if the transport results in a bond that spans non-adjacent compartments. The rule in the compartmental model of Section 2 that transports mRNA from the nucleus to the cytoplasm is

```
R16: mRNA1@NU -> mRNA1@CP  k_mRNA_to_CP
```

In cargo-carrying transport, movement of a specified molecule causes simultaneous transport of connected molecules that are in the same compartment. Many cellular systems contain molecules that can bind and transport cargo between compartments, and these may bind a variety of molecules and complexes. In nuclear transport, for example, importin proteins bind to different protein molecules and escort them through the nuclear pore (Alberts et al. 2002). This type of molecule transport behavior is designated by adding the `MoveConnected` keyword to a molecule transport rule, as in the example

```
R29: Im(fg!1)@CP.NP(fg!1) <-> Im(fg!1)@NU.NP(fg!1) \
      k_Im_cross_NP, k_Im_cross_NP  MoveConnected
```

Here, an importin molecule in CP, bound through its `fg` site to a nuclear pore (NP), is transported into and out of NU along with its cargo, which can be a TF dimer or a P1. Thus, the `MoveConnected` keyword allows one to write a single rule that encompasses all possibilities without enumerating them. A precise description of the effect of a `MoveConnected` declaration requires the following definition:

Definition 2 (Compartment-connected component)**.**
For a molecule ***M*** *in compartment* ***C****, the compartment-connected component is the set of all molecules within the complex* ***X****, of which* ***M*** *is a member, that are connected to* ***M*** *by a path fully contained within* ***C****.*

Only molecules in the compartment-connected component of the explicitly transported molecule(s) in a rule are co-transported. Molecules that are in the same compartment as the trafficking molecule but are connected to it through a path that passes through another compartment are not co-transported. This situation may arise because trafficking molecules often bind to surface molecules that may already be bound to other molecules in the same compartment as the trafficking molecule.

Species transport rules change the locations of all molecules within a species rather than just a single molecule. These rules are specified using prefix compartment notation. Volume-to-volume species transport is straightforward: the compartment designations for all molecules in the volume-localized species are changed to the same destination compartment. An example is ligand recycling from EN to EC (see Figure 4B):

```
R5: @EN:L -> @EC:L  k_recycle
```

Surface-to-surface species transport is more involved because species localized to surfaces can contain molecules that reside in adjacent volume compartments. These molecules must be correctly assigned to a new compartment during transport. In cBNGL, surface-to-surface transport rules map molecules in volume compartments in a manner consistent with endo- and exocytosis (see Figure 1). During endocytosis, molecules exterior to the invaginating membrane move into the newly-formed

compartment, whereas molecules interior to the invaginating membrane remain in the same compartment. Surface-to-surface species transport is defined as follows:

Definition 3 (Surface-to-surface species transport)**.**

1. *Molecules located in the containing compartment* **`Vc1`** *of* **`S1`** *move into the contained compartment* **`Vc2`** *of* **`S2`***.*
2. *Molecules located in the originating surface* **`S1`** *move into the destination surface* **`S2`***.*
3. *Molecules located in the volume* **`Vs`** *that is adjacent to both* **`S1`** *and* **`S2`** *remain in* **`Vs`***.*

This type of transport is restricted to surface pairs that share an adjacent volume compartment. In the example of Section 2, receptor endocytosis (internalization) is modeled by the rule

```
R3: @PM:Rec.Rec -> @EM:Rec.Rec   k_R_endo
```

4.6 Comparison with BNGL

We have developed both BNGL and cBNGL versions of the compartmental model of Section 2. Each produces a network of 78 species and 354 reactions, giving identical numerical results (data not shown). The BNGL specification requires 45 rules whereas the cBNGL specification requires 30 rules, with most of the difference coming from the rules for ligand dimerization, receptor-ligand binding, endosome recycling and nuclear transport. Complete input files for this model system along with instructions for downloading a cBNGL version of BIONETGEN can be found at www.bionetgen.org/wsc09.

5 DISCUSSION

cBNGL introduces a compartment topology composed of membranes and volumes, a new molecule property, *molecule localization*, which ties molecules to a compartment, and a new global property of species, *species localization*, which allows queries to determine whether a species is tethered to a membrane or freely diffusing in solution. The new `@` syntax allows a modeler to construct species and pattern matches in the context of the compartment topology. The new concepts of species and molecule localization provide a basis for the construction of new semantics for universal and transport rules.

Universal rules, which do not query localization explicitly, describe reactions that can occur between any set of matching species that are able to interact. Such rules reflect the usual biological scenario where the co-localization of reactants is a sufficient condition for the reaction to proceed. In cases where a reaction occurs in specific locations, the modeler may scope restrict a pattern in a rule to a single compartment by adding location context. The compartment to use in scaling of bimolecular reaction rates is determined from the dimensionality of the reactant compartments: a surface compartment is used only when both reactants are localized to the membrane; otherwise, the three-dimensional compartment is used. In both cases the rate is divided by the compartment volume, which naturally yields a large rate enhancement when both reactants are localized to the membrane (Haugh and Lauffenburger 1997; Kholodenko, Hoek, and Westerhoff 2000).

Transport rules allow the expression of a wide variety of biological transport phenomena. Molecule transport enables the description of simple diffusion across a membrane, translocation mediated by a membrane transporter, and insertion of a molecule into a membrane. Species transport rules allow the transport of an entire species based on a pattern match to part of the species complex. Volume-to-volume species transport rules allow concise representation of biological scenarios where transport across a membrane is facilitated by an escort molecule. Cargo-carrying transport coupled with a binding reaction at the membrane permits chaperone mediated species transport with a saturable rate. Surface-to-surface species transport rules allow transport of membrane-bound complexes to and from a contained membrane compartment in a manner that preserves complex structure and correct topological relationships. A common example of such a process in biology is receptor-mediated endocytosis, where an active receptor complex initiates vesicle formation and internalization.

While cBNGL encompasses a wide variety of membrane and transport phenomenon, there are important limitations. The compartment topology is static, so cell division, vesicle budding and fusion are not described. Thus, surface-to-surface transport rules capture only the average rates of molecular transport and not the turnover of individual vesicles. Compartment volumes are also fixed, so dynamic changes in volume cannot be described. Furthermore, transport reactions are not universal but tied to specific, named compartments. In a biological setting, transport typically depends on the presence of specific channels, pores or transport proteins in the separating membrane rather than the physical properties of the specific compartments. As a practical matter, we also note that our implementation of cBNGL for the particle-based simulation methods referred to in Section 1 is in progress. Future work will address each of these issues.

5.1 Related Work

Formal modeling in cellular systems has focused on three related, but largely disjoint, areas of biological expression: biochemistry of structured molecules, membrane-mediated biochemistry, and dynamic membrane systems. The rule-based modeling languages κ-calculus and BNGL focus on the biochemistry of structured molecules but lack a natural approach to compartments and membranes (Faeder, Blinov, and Hlavacek 2009; Danos et al. 2007). The Stochastic Simulation Compiler (SSC), another rule-based platform, allows static compartments, diffusion between compartments, and modeling of spatial and geometric effects, but lacks the notion of a membrane (Lis et al. 2009). BioCham, a modeling platform with facilities for model checking, describes biochemistry through rules over unstructured objects and includes compartments and basic transport (Calzone, Fages, and Soliman 2006).

Several platforms have been constructed with a focus on membrane-mediated reactions. Cyto-Sim, a formal language for describing reactions in the context of a membrane structure, includes a syntax and rule set that describe integral and peripheral membrane proteins, membrane recruitment reactions, and transport (Sedwards and Mazza 2007). Cyto-Sim's reactions, however, are limited to transformations of unstructured objects. Little b, a modular framework for biological modeling, includes rule-based modeling features and static membrane structures (Mallavarapu et al. 2009). Molecules can be localized to the cytosolic or extracellular face of membranes and interact with molecules in the adjacent volume. Rules are provided for basic molecular transport reactions. Simmune, a multiscale platform tying rule-based molecular interactions to macroscopic cell behavior, models cells as a plasma membrane containing a cytosolic compartment (Meier-Schellersheim et al. 2006). Membrane proteins in Simmune may have cytosolic and extracellular domains that interact with the adjoining volumes. While Simmune includes dynamic cell division and death, the topology of the cell is fixed and cellular models that required nested compartments for organelles are not handled.

Cardelli pioneered the formal description of dynamic membrane systems with the introduction of Brane Calculi (Cardelli 2005; Cardelli 2008), which formally express biological membrane processes, including division, fusion and phagocytosis, but do not include description of biochemical reactions. BioAmbients, based on the stochastic π-calculus and ambients calculus, provides a language and simulation platform for dynamic compartments with structured molecular interactions (Regev et al. 2004). It thus begins to bridge the biology of membranes and molecules, but treatment of membrane-mediated biochemistry is lacking. Similarly, Beta Workbench implements dynamic compartments through the abstraction of beta binders, an interface that wraps a collection of biological processes and controls external communication (Dematte et al. 2008). Nesting of beta binders is not permitted, so modeling cells with organelle structures is beyond its scope.

Recent efforts in the process algebra literature have made considerable progress in merging rule-based modeling with dynamical membrane systems. Laneve and Tarissan (2008) proposed bio-κ, a restricted variant of the κ-calculus combined with syntax and semantics for dynamic membrane reactions. Membranes can contain molecules and interact with adjacent volumes. While the rule-based capability of bio-κ is limited, the biological expressivity includes phagocytosis, fusion, fission and molecule translocation. Damgaard, Danos, and Krivine (2008) developed C-calculus, which introduces the concept of a channel, a type of bond that connects topologically related volumes. C-calculus is able to express complex biological phenomena that combine membrane and molecule interactions, such as clathrin-dependent vesicle formation. It can also describe chaperone-mediated transport. Though rich in their expressive capabilities, the primary shortcomings of these approaches are their current lack of freely-available end-user simulation software.

In Table 1, we compare the capabilities of cBNGL with the various modeling approaches cited above. Although none of the platforms cover the full range of capabilities, cBNGL has the advantage of being able to describe a wide variety of biological phenomena associated with compartmentalization and membranes in a language that is fully compatible with the freely-available BIONETGEN suite of modeling and simulation tools.

ACKNOWLEDGMENTS

LAH and JSH contributed equally to this work. This work was supported by the National Institutes of Health (GM076570) and the National Science Foundation (CCF-0829788). The authors thank Gary C. An, Michael L. Blinov, William S. Hlavacek, and Bin Hu for helpful comments on the manuscript and Thierry Emonet and Michael Sneddon for helpful discussions.

REFERENCES

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2002. *Molecular biology of the cell*. 4th ed. Garland Science.

Table 1: Comparisons of various biochemical modeling and simulation platforms in terms of capabilities for modeling species transport, membrane biology, dynamic creation and deletion of compartments and availability of end-user software.

Platform	Rule-based?	Species transport?	Membranes?	Dynamic compartments?	Software?	Reference
Beta Workbench	✓		limited	✓	✓	Dematte et al. (2008)
bio-κ	✓		✓	✓		Laneve and Tarissan (2008)
BioAmbients	✓			✓	✓	Regev et al. (2004)
BioCham	limited			✓†	✓	Calzone et al. (2006)
Brane calculi			✓	✓		Cardelli (2005); Cardelli (2008)
C-calculus	✓	✓	✓	✓		Damgaard et al. (2008)
cBNGL	✓	✓	✓		✓	This work
Cyto-Sim			✓		✓	Sedwards and Mazza (2007)
Little b	✓		✓		✓	Mallavarapu et al. (2009)
Simmune	✓		limited	✓	✓	Meier-Schellersheim et al. (2006)
SSC	✓				✓	Lis et al. (2009)

† Allows compartment volumes to change dynamically but cannot create nor delete compartments.

An, G. C., and J. R. Faeder. 2009. Detailed qualitative dynamic knowledge representation using a BioNetGen model of TLR-4 signaling and preconditioning. *Math. Biosci.* 217 (1): 53–63.

Barua, D., J. R. Faeder, and J. M. Haugh. 2009. A bipolar clamp mechanism for activation of Jak-family protein tyrosine kinases. *PLoS Comput. Biol.* 5 (4): e1000364.

Blinov, M. L., O. Ruebenacker, and I. I. Moraru. 2008. Complexity and modularity of intracellular networks: a systematic approach for modelling and simulation. *IET Syst. Biol.* 2 (5): 363–368.

Blinov, M. L., J. Yang, J. R. Faeder, and W. S. Hlavacek. 2006. Graph theory for rule-based modeling of biochemical networks. *Lect. Notes Comput. Sci.* 4230:89–106.

Borisov, N. M., A. S. Chistopolsky, J. R. Faeder, and B. N. Kholodenko. 2008. Domain-oriented reduction of rule-based network models. *IET Syst. Biol.* 2 (5): 342–351.

Calzone, L., F. Fages, and S. Soliman. 2006. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics* 22 (14): 1805–1807.

Cardelli, L. 2005. Brane calculi – Interactions of biological membranes. *Lect. Notes Comput. Sci.* 3082:257–278.

Cardelli, L. 2008. Bitonal membrane systems. *Theor. Comput. Sci.* 404 (1-2): 5–18.

Colvin, J., M. I. Monine, J. R. Faeder, W. S. Hlavacek, D. D. Von Hoff, and R. G. Posner. 2009. Simulation of large-scale rule-based models. *Bioinformatics* 25 (7): 910–917.

Damgaard, T. C., V. Danos, and J. Krivine. 2008. A language for the cell. Technical Report TR-2008-116, IT University of Copenhagen, Rued Langgaards Vej 7, DK-2300 Copenhagen V.

Danos, V., J. Feret, W. Fontana, R. Harmer, and J. Krivine. 2007. Rule-based modelling of cellular signalling. *Lect. Notes Comput. Sci* 4703:17–41.

Dematte, L., C. Priami, and A. Romanel. 2008. The Beta Workbench: a computational tool to study the dynamics of biological systems. *Brief. Bioinform.* 9 (5): 437–449.

Faeder, J. R., M. L. Blinov, and W. S. Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. In *Methods in Molecular Biology: Systems Biology*, ed. I. V. Maly, p. 113–168. Humana Press.

Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22 (4): 403–434.

Haugh, J. M., and D. A. Lauffenburger. 1997. Physical modulation of intracellular signaling processes by locational regulation. *Biophys. J.* 72 (5): 2014–2031.

Hlavacek, W. S., J. R. Faeder, M. L. Blinov, A. S. Perelson, and B. Goldstein. 2003. The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* 84 (7): 783–794.

Hlavacek, W. S., J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka, and W. Fontana. 2006. Rules for modeling signal-transduction systems. *Science STKE* 2006:re6.

Hu, B., G. M. Fricke, J. R. Faeder, R. G. Posner, and W. S. Hlavacek. 2009. GetBonNie for building, analyzing and sharing rule-based models. *Bioinformatics* 25 (11): 1457–1460.

Hucka, M., S. Hoops, S. M. Keating, N. Le Novère, S. Sahle, and D. Wilkinson. 2008. Systems Biology Markup Language (SBML) Level 2: Structures and facilities for model definitions. *Nature Precedings*. http://dx.doi.org/10.1038/npre.2008.2715.1.

Kholodenko, B. N., J. B. Hoek, and H. V. Westerhoff. 2000. Why cytoplasmic signalling proteins should be recruited to cell membranes. *Trends Cell Biol.* 10 (5): 173–178.

Laneve, C., and F. Tarissan. 2008. A simple calculus for proteins and cells. *Theor. Comput. Sci.* 404 (1-2): 127–141.

Lis, M., M. N. Artyomov, S. Devadas, and A. K. Chakraborty. 2009. Efficient stochastic simulation of reaction-diffusion processes via direct compilation. *Bioinformatics*. in press; doi:10.1093/bioinformatics/btp387.

Mallavarapu, A., M. Thomson, B. Ullian, and J. Gunawardena. 2009. Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *J. R. Soc. Interface* 6 (32): 257–270.

Meier-Schellersheim, M., X. Xu, B. Angermann, E. J. Kunkel, T. Jin, and R. N. Germain. 2006. Key role of local regulation in chemosensing revealed by a new molecular interaction-based modeling method. *PLOS Comput. Biol.* 2 (7): 710–724.

Morton-Firth, C. J., and D. Bray. 1998. Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* 192 (1): 117.

Nag, A., M. I. Monine, J. R. Faeder, and B. Goldstein. 2009. Aggregation of membrane proteins by cytosolic cross-linkers: Theory and simulation of the LAT-Grb2-SOS1 system. *Biophys. J.* 96 (7): 2604–2623.

Pawson, T., and P. Nash. 2003. Assembly of cell regulatory systems through protein interaction domains. *Science* 300 (5618): 445–452.

Regev, A., E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. 2004. BioAmbients: an abstraction for biological compartments. *Theor. Comput. Sci.* 325 (1): 141–167.

Sedwards, S., and T. Mazza. 2007. Cyto-Sim: a formal language model and stochastic simulator of membrane-enclosed biochemical processes. *Bioinformatics* 23 (20): 2800–2802.

Sneddon, M., J. R. Faeder, and T. Emonet. 2009. In preparation.

Yang, J., M. I. Monine, J. R. Faeder, and W. S. Hlavacek. 2008. Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E* 78 (3): 031910.

AUTHOR BIOGRAPHIES

LEONARD A. HARRIS is a Postdoctoral Associate in the Department of Computational Biology at the University of Pittsburgh School of Medicine. He did his graduate studies in Chemical Engineering at Cornell University, where he developed an integrated algorithm for accelerated stochastic simulations. His email address for these proceedings is <lharris@pitt.edu>.

JUSTIN S. HOGG is a graduate student in the Carnegie Mellon–University of Pittsburgh Ph.D. Program in Computational Biology. His email address for these proceedings is <jsh32@pitt.edu>.

JAMES R. FAEDER is an Associate Professor in the Department of Computational Biology at the University of Pittsburgh School of Medicine. Prior to coming to Pittsburgh, he was a Technical Staff Member at Los Alamos National Laboratory from 2003–2007, where, in collaboration with Michael L. Blinov and William S. Hlavacek the BIONETGEN project was initiated. His research interests include simulation of large-scale biological networks, the biochemistry of signal transduction, and the role of protein-protein interactions in intracellular processes. His email address for these proceedings is <faeder@pitt.edu>.