Supplementary Note 1:
Symbols Used SBGN Process Diagram Level 1

The SBGN Process Diagram specification defines a comprehensive set of symbols with precise semantics, together with detailed syntactic rules defining their use. It also describes how such graphical information is to be interpreted. The essence of a process diagram is *change*. It shows how different entities in a system transition from one form to another. Glyphs (symbols) are the graphical units that represent concepts in SBGN. Each one is uniquely identified by a term from the Systems Biology Ontology (SBO). There are two types of glyphs in SBGN: nodes (which are subdivided into entity pool nodes, container nodes, process nodes, and logical operator nodes), and arcs (edges) that characterize the quantitative effect of a substance on a process or vice versa.

Entity pool nodes *(EPNs)* represent ensembles of entities, such as molecules, that are considered indistinguishable from each other in the context of a given graph. Level 1 of the SBGN Process Diagram defines six distinct glyphs for the following concepts: *unspecified entity*, *simple chemical*, *macromolecule*, *nucleic acid feature*, *perturbing agent* (such as light, temperature, etc.) and *source and sink*. The *EPNs* associated with molecular entities can be duplicated and stacked to represent multimers of identical elements. An additional construct, the *complex*, can be also used as an *EPN*. The semantics of *EPNs* can be modified by *auxiliary units*, which represent a particular state, the fact that the *EPN* has been cloned in the maps, or some additional information that may be encoded using controlled vocabularies. Finally, *tags* can be used to identify an *EPN* used in two or more physically different maps, thereby allowing the modular decomposition of diagrams.

Process nodes *(PNs)* describe the way in which *EPNs* are transformed into other *EPN*s. SBGN Process Diagram Level 1 defines five *PNs*: *process* (used to represent most of the transformations between *EPNs*), *omitted process* (when several transitions are known to exist but not represented), *uncertain process* (when the transition may or may not exist), *association*, and *dissociation* (representing the

Nature Biotechnology: doi:10.1038/nbt.1558

whereabouts of non-covalent complexes). In addition, a particular type of PN is the *phenotype*, which can be modulated but does not consume or produce anything. More types of transition may be defined in the future in higher levels of the SBGN Process Diagram notation.

Connecting arcs link *EPNs* and *PNs*, and indicate how entities influence processes. In addition to *consumption* and *production* arcs, which indicate the effect on the flux of matter through *PNs*, the SBGN Process Diagram specification also provides arcs for representing different possible modifications of a process, such as *modulation*, *stimulation*, *catalysis*, *inhibition* and *trigger* (or absolute activation). Finally, some arcs link nodes of the same type, such as the *equivalence* arc, linking *EPN* and *tag*, and the *logic arc*, linking two *logical operators* or an *EPN* to a *logical operator.*

Logical operators provide the means of indicating boolean combinations of influences from *EPNs* onto *PNs*. The three possibilities are conjunction (*and*), disjunction (*or*), and negation (*not*).

Compartments and ***submaps*** are containers that permit to gather together EPNs and PNs, either by spatial proximity or as map "modules". *Submaps* are "folded" in the main map, represented by a symbol. The unfolded *submaps* can be retrieved, for instance, in other windows of a software-based system, or on other pages if the diagrams are printed on paper.

Nature Biotechnology: doi:10.1038/nbt.1558

Rules for building an SBGN process diagram

The SBGN Process Diagram specification prescribes a number of rules in an effort to help eliminate ambiguity in an SBGN diagram. These rules must be complied with in order for the diagram to be a valid Level 1 SBGN Process Diagram. It is important to realize that SBGN does not dictate how to represent something, but rather how to interpret the representation. There is generally more than one way to represent a concept (for instance, hemoglobin can be represented as a *macromolecule*, a *complex* of four macromolecules, or a *multimer* with cardinality 4). However, everyone should interpret an SBGN Process Diagram the same way. Here we list some of the most important rules. Note that these are not the only rules defined by the SBGN Process Diagram Level 1 specification—users should consult the official specification documents at http://sbgn.org/ for a complete list of rules.

- An *entity pool node* belongs to only one *compartment*. If no compartment is drawn, it is assumed to belong to a “default” compartment.

- *Compartments* cannot be nested, and they represent disjoint spatial containers. *Compartments* may overlap visually, but such overlap does not imply any kind of physical containment; i.e., a *compartment* is never “part” of another.

- The layout or organization of a *compartment* does not imply anything about its topology.

- A *complex* may contain subunits that belong to different *compartments* (however, the complex itself belongs to only one).

- The layout or organization of the *EPNs* in a *complex* does not imply any information about topology.

- *Complexes* can be nested, making a given *complex*’s topology explicit.

Nature Biotechnology: doi:10.1038/nbt.1558

- A *complex* should consist of different *EPNs*. If two or more elements of the *complex* are identical then they should be replaced by a *multimer*.

- All substrates of a *Process node* should be different. If several copies of the same *EPN* are involved in the process, the cardinality label of the *consumption* arc should be used. The same rule applies to all products.

- Once the cardinality label is added to one arc, all other arcs connected to a *PN* must display a cardinality label.

- A *PN* should correspond to only one process or series of connected processes. If the same set of *EPNs* are consumed and produced by alternative processes, they should be connected by different *PNs*.

- A *PN* with no modulations has an underlying “basal rate” which describes the rate at which it converts inputs to outputs.

- When *modulations*, *stimulations* and *inhibitions* connect to the same *PN*, their effect on the basal rate of the process is combined. If their effects are independent of alternatives, different *PN* or *logical operators* must be used.

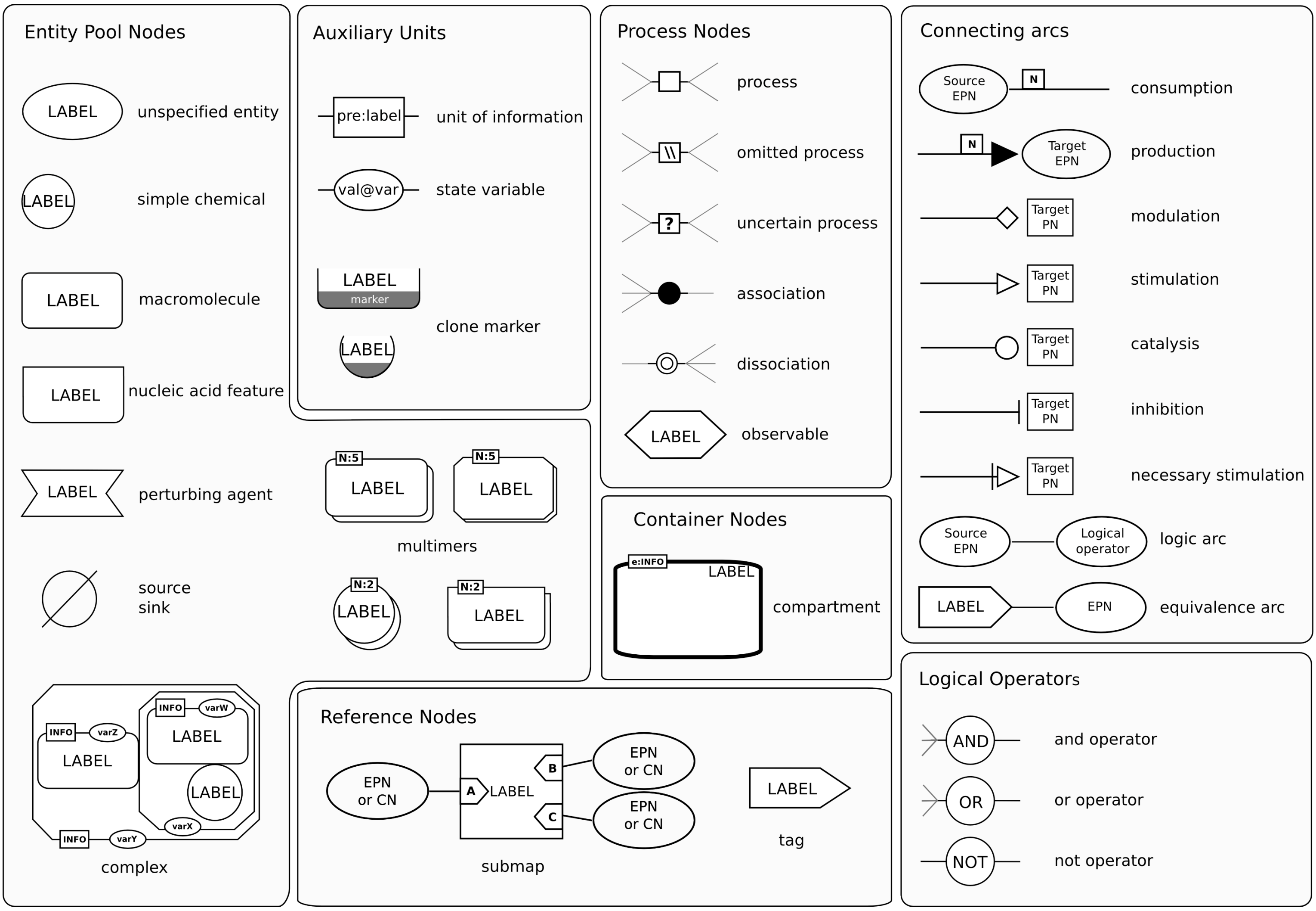
- Modulators that do not interact with each other in the manner above should be drawn as modulating different process nodes. Their effect is therefore additive.

- At most one *necessary activation* can be connected to a process node. If several *necessary activations* affect a process, their combination must be explicitly expressed using *logical operators*.

- At most one *catalysis* arc can connect to a process node. If several enzymes catalyze the “same” biochemical reaction, each catalytic process should be represented by a different *process node*.

Nature Biotechnology: doi:10.1038/nbt.1558

Figure S1: List of all glyphs specified by SBGN Process Diagram Level 1

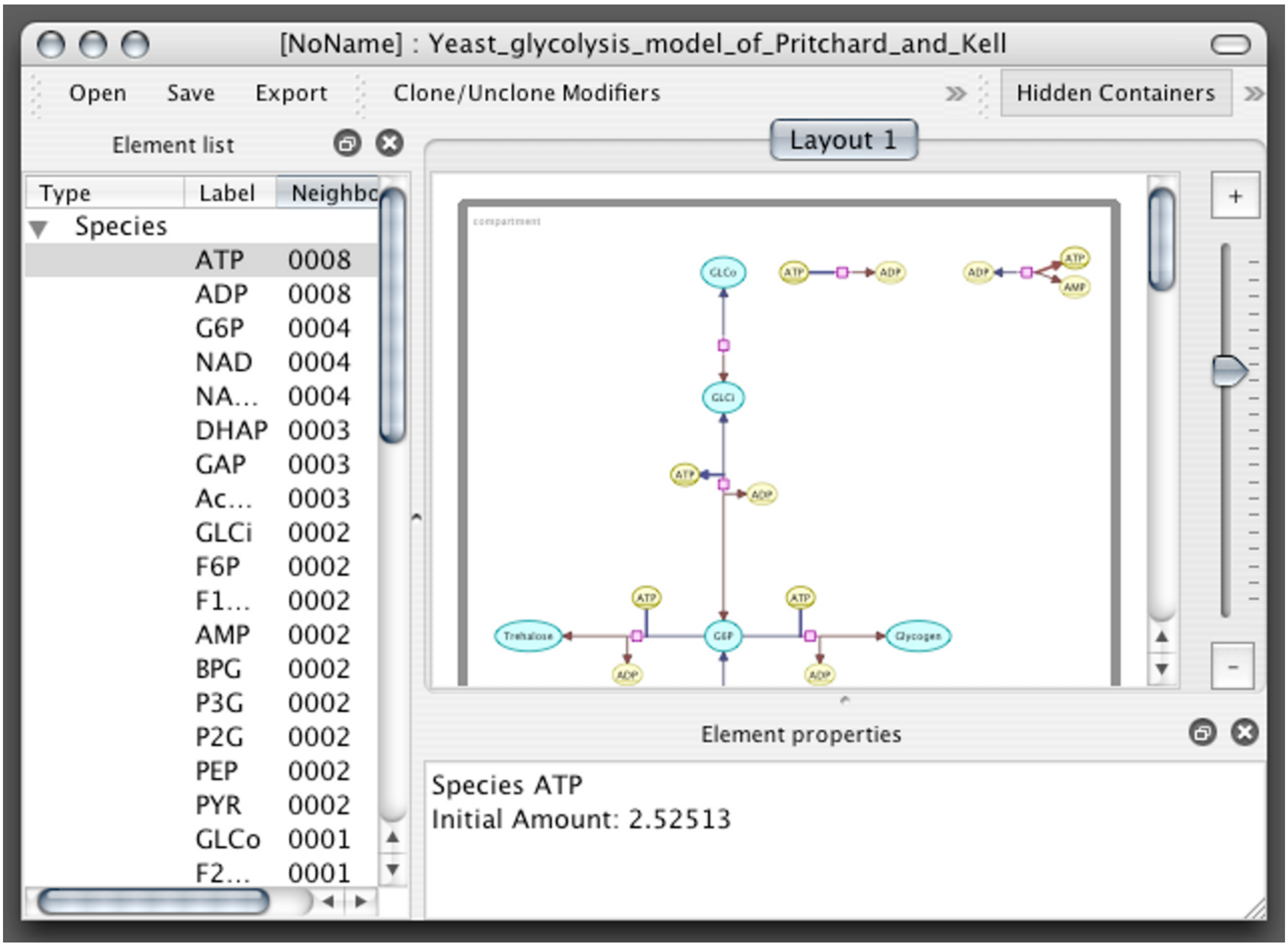


Nature Biotechnology: doi:10.1038/nbt.1558

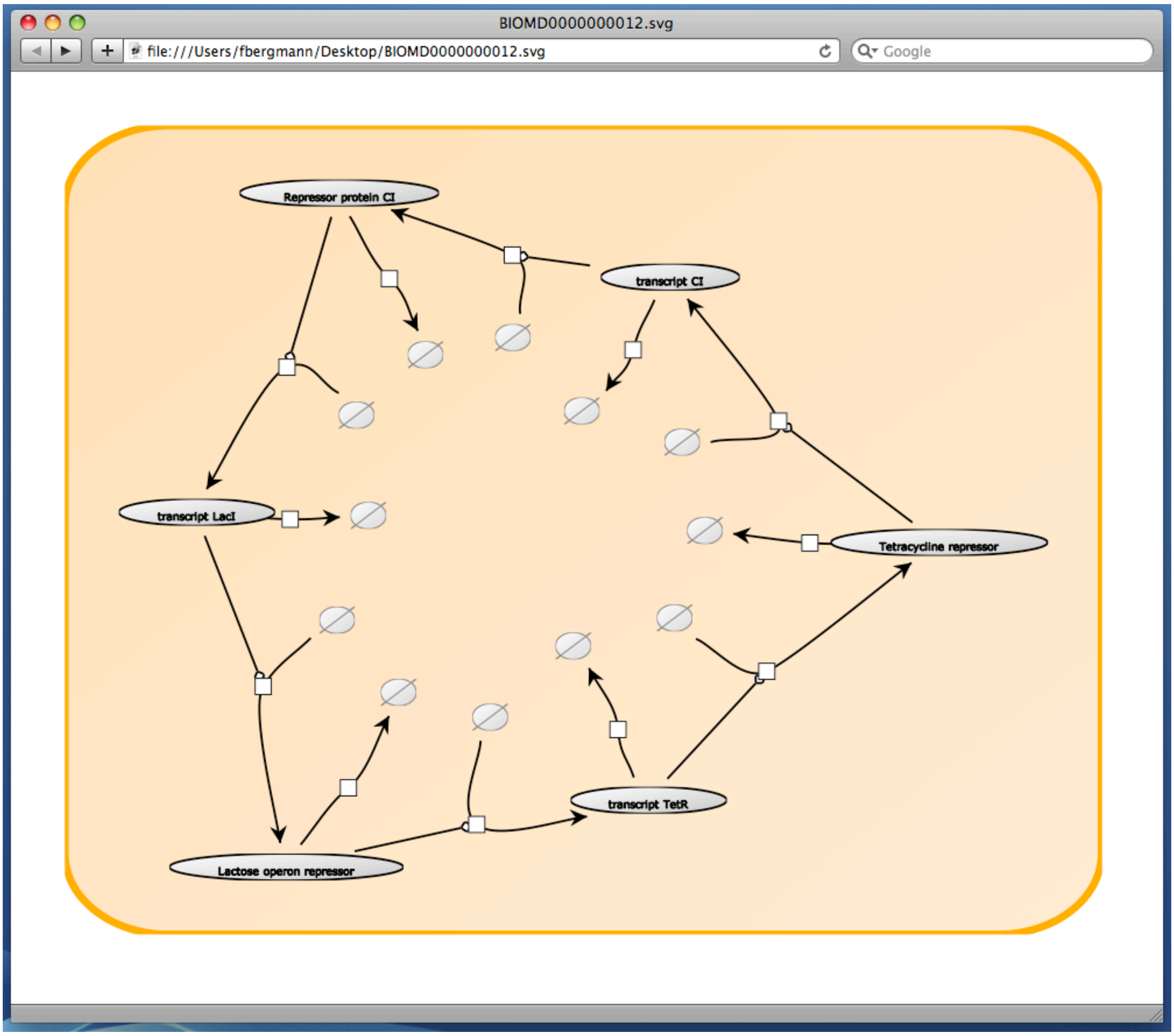
Supplementary Note 2:

Software Support for SBGN Process Diagram Level 1

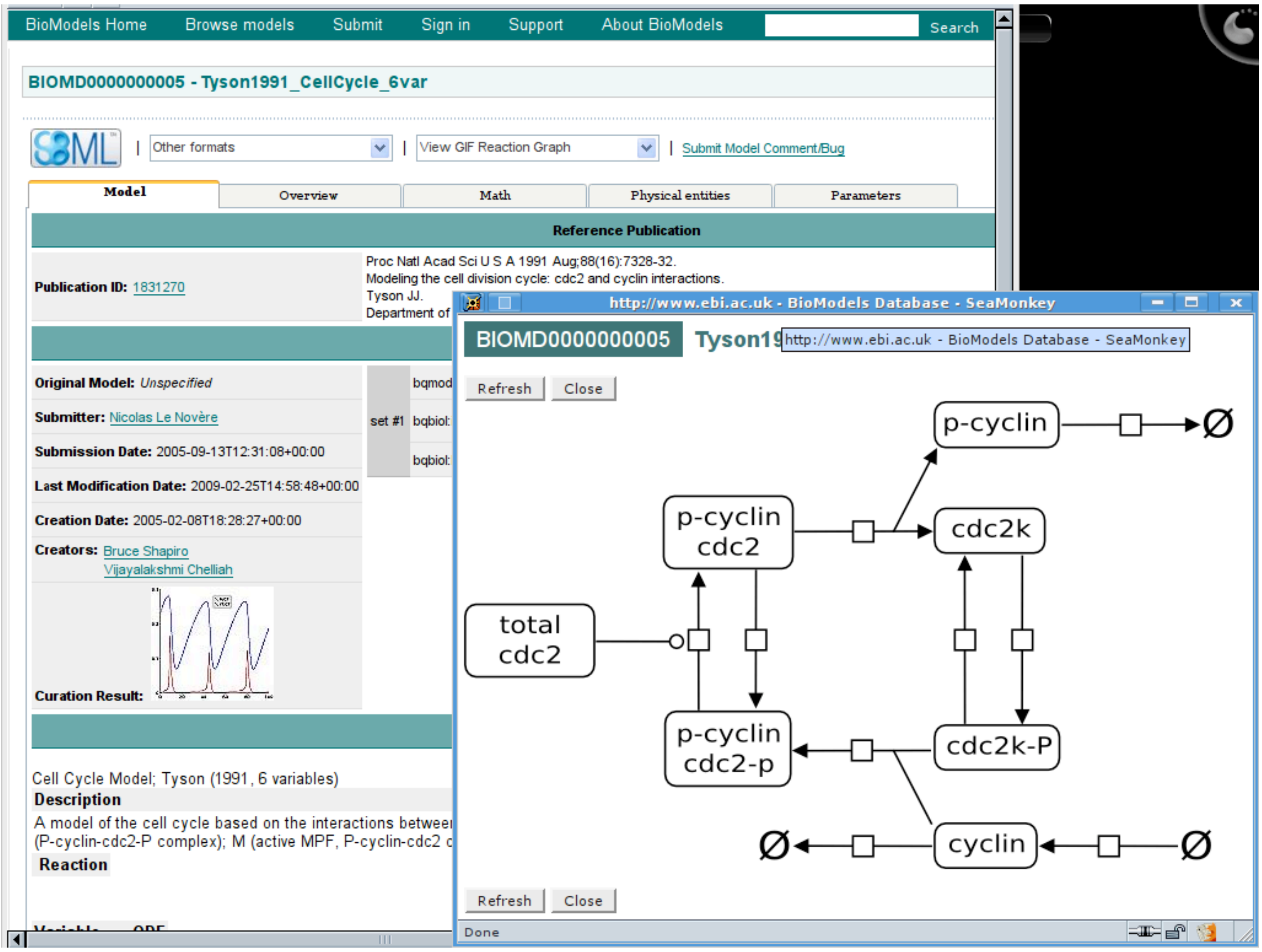
July 2009



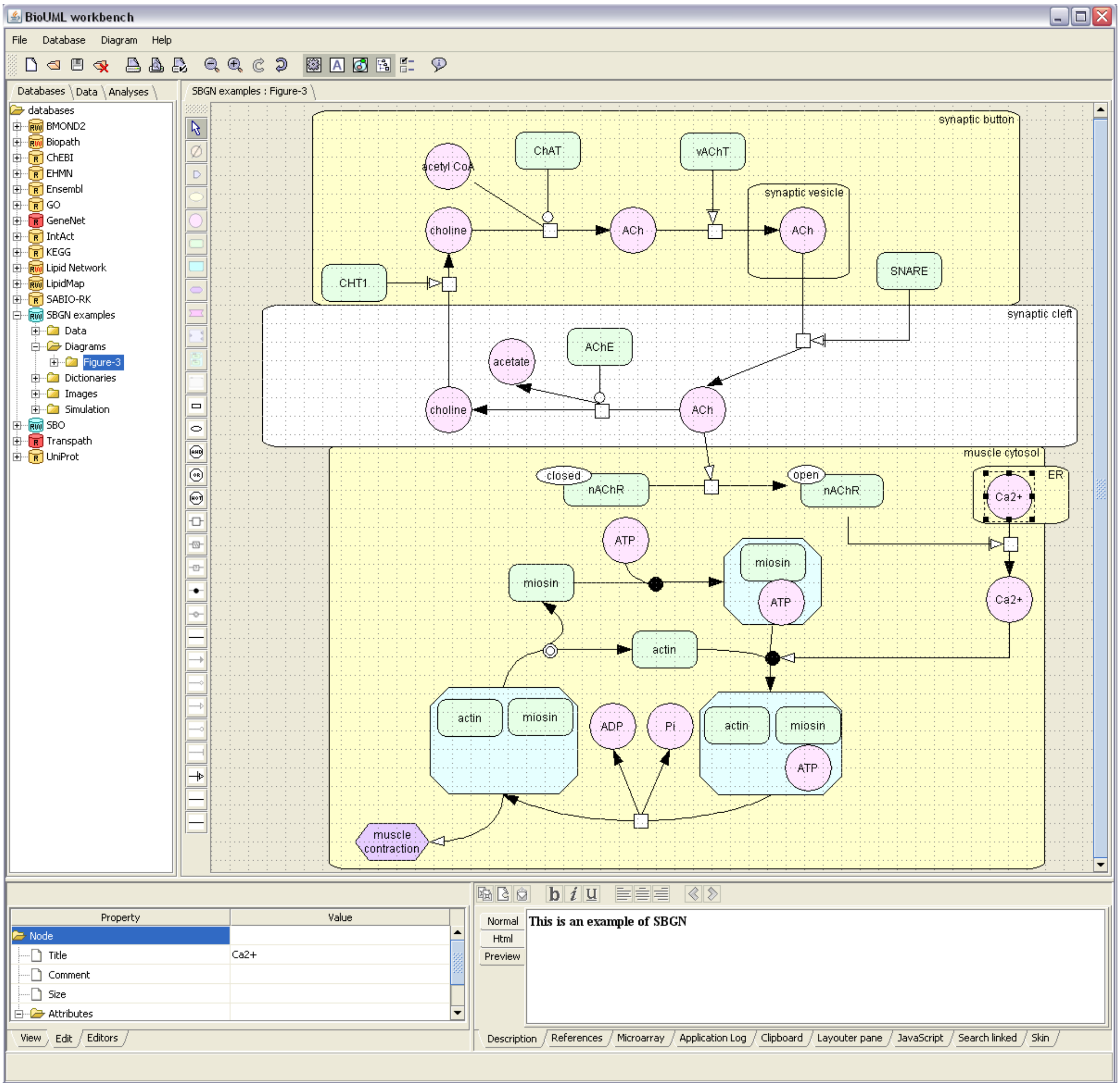
Model of glycolysis represented in the software package Arcadia, version 1.0a1 (http://arcadiapathways.sourceforge.net/). The graph is drawn from an SBML file.The layout is semi-automatic, and the glyphs are chosen based on annotations contained inside the SBML file.



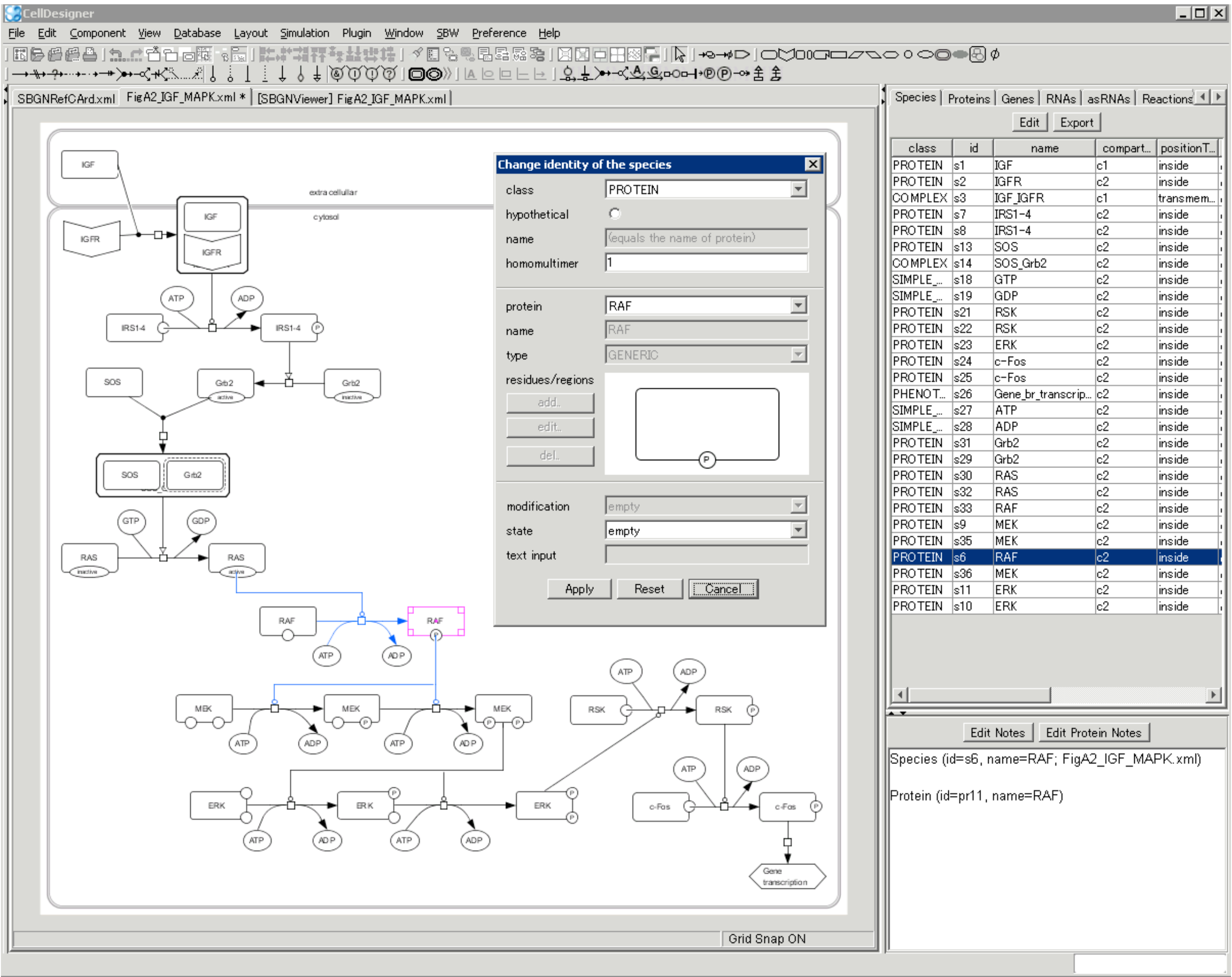
Graph of the model BIOMD0000000012 from BioModels Database (http://www.ebi.ac.uk/biomodels), generated by the software tool Athena (http://www.codeplex.com/athena), by interpreting Systems Biology Ontology (SBO; http://www.ebi.ac.uk/sbo) annotations contained within the SBML model file.



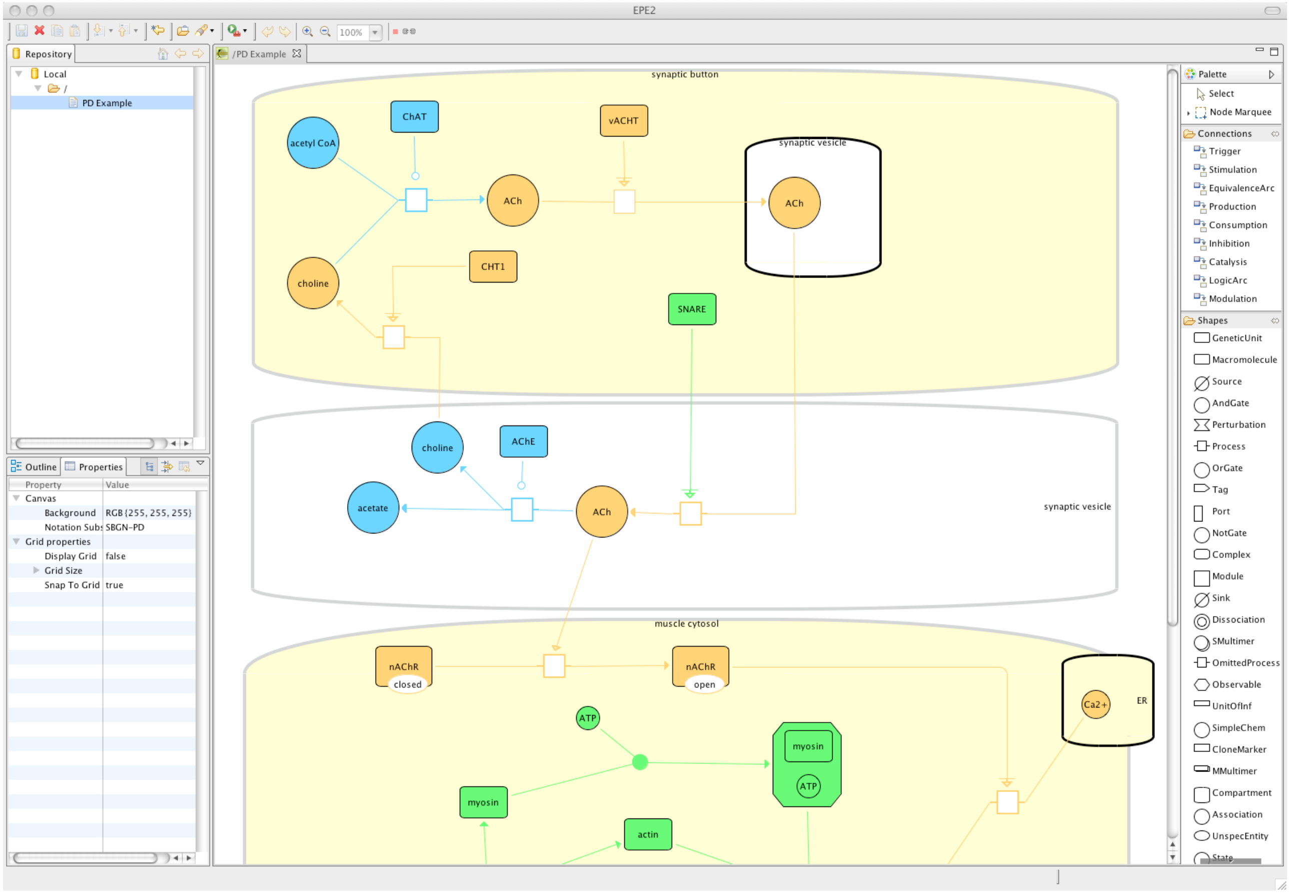
A model of the cell cycle, as represented in BioModels Database (http://www.ebi.ac.uk/biomodels).



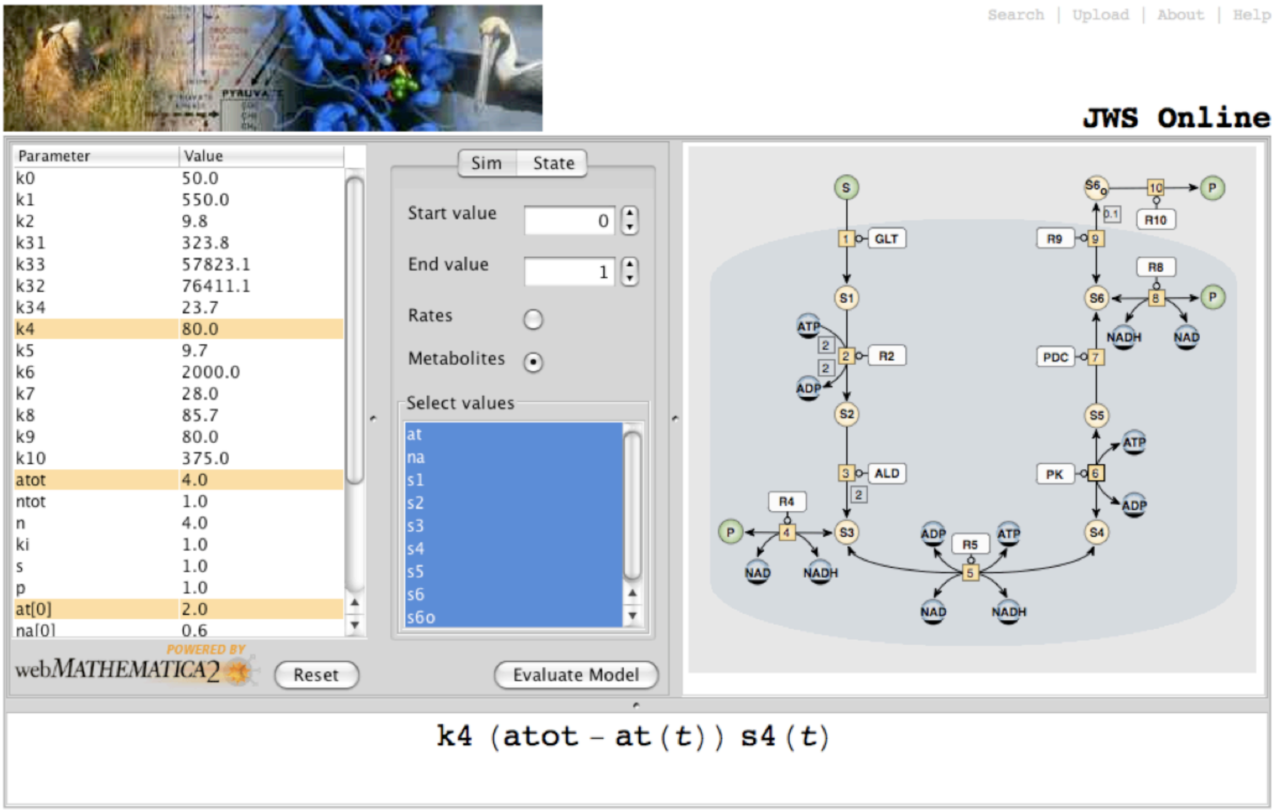
One of the examples present in the SBGN Process Diagram Level 1 specification is drawn with BioUML (http://www.biouml.org/).



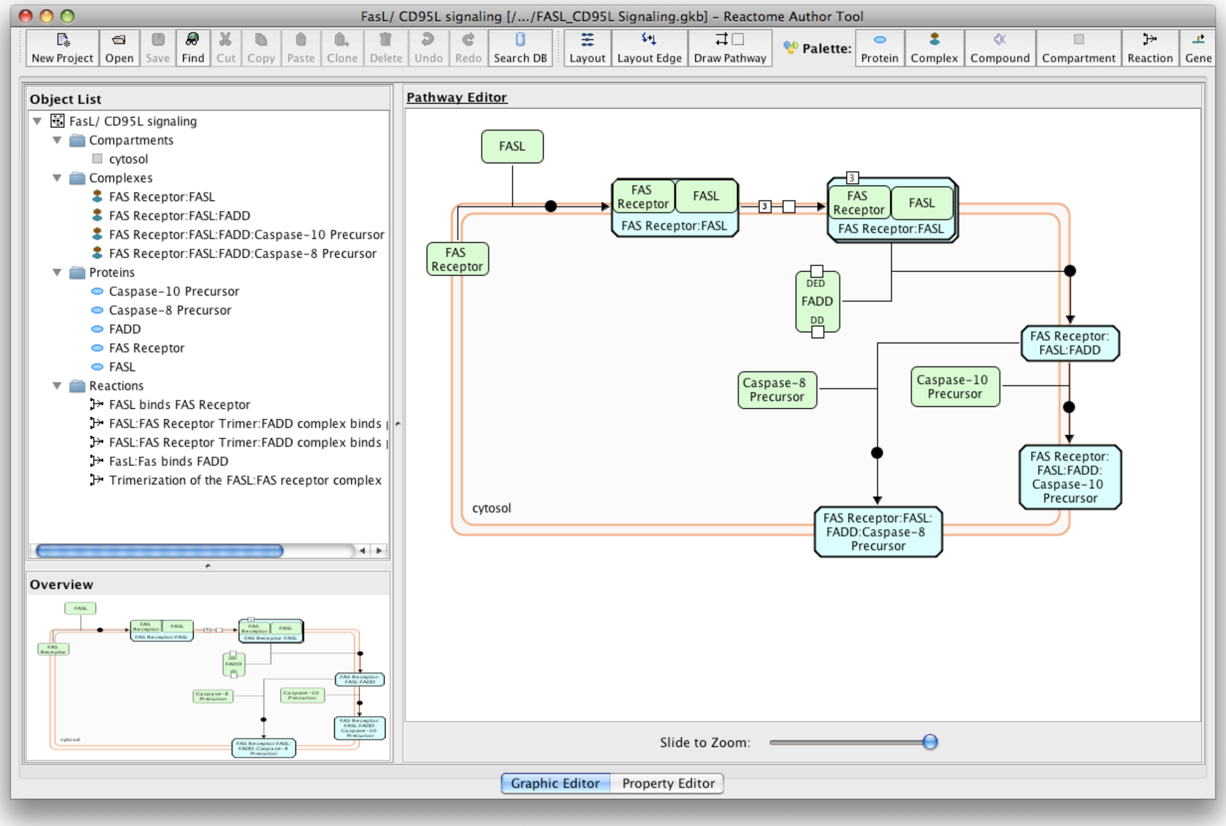
One of the examples present in the SBGN Process Diagram Level 1 specification is drawn anew with CellDesigner version 4.0.1 (http://www.celldesigner.org/). The diagram is stored in the SBML file as annotations. Most of the graphical information, such as location, shapes, colors, modifications of the species and reactions are stored in a CellDesigner-specific annotation format. CellDesigner 4.0.1 provides a function called "SBGN viewer", in which the CellDesigner-specific graphical notation can be converted into pure SBGN notation.



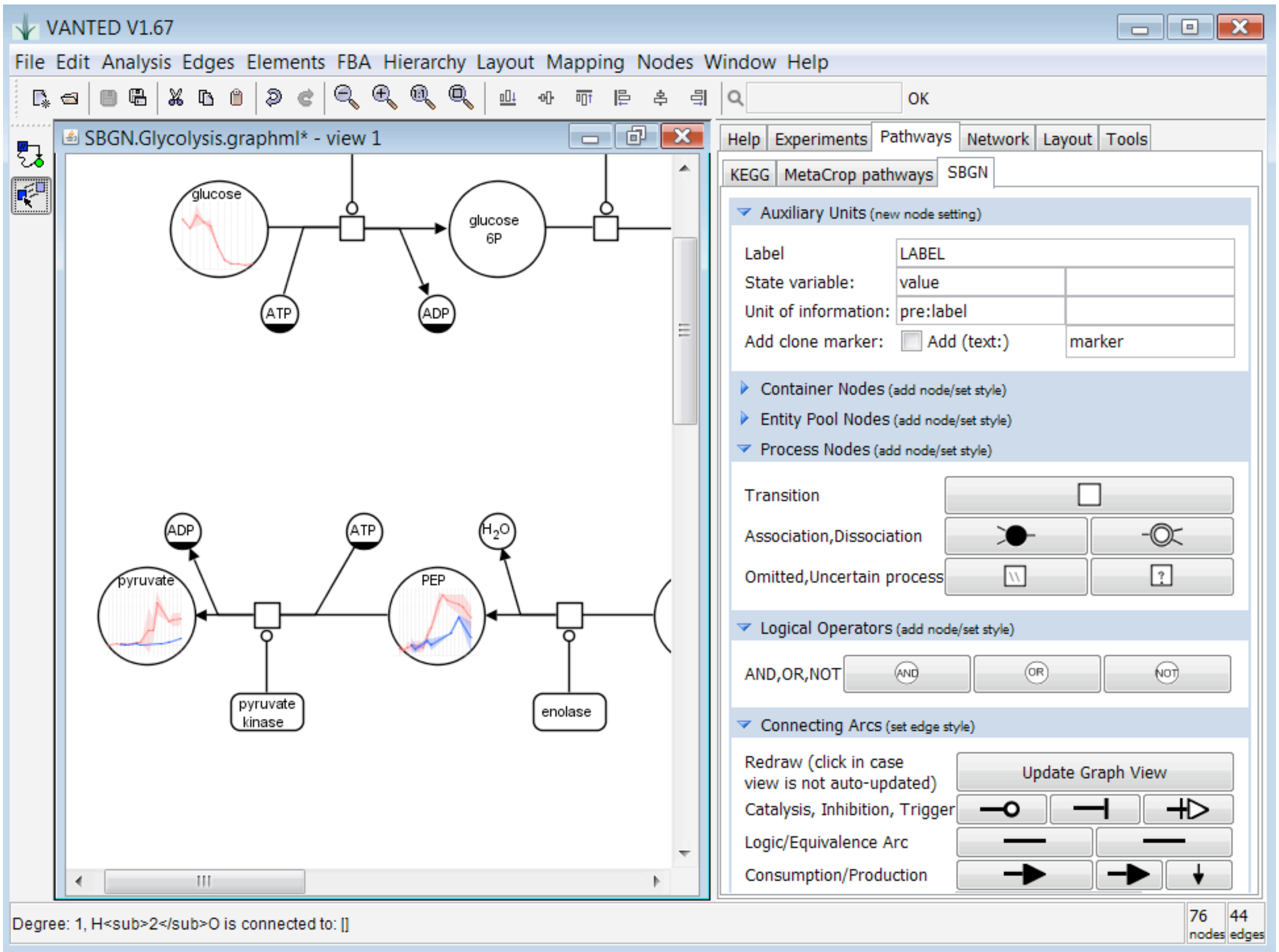
One of the examples present in the SBGN Process Diagram Level 1 specification is drawn with the forthcoming version 3 of the Edinburgh Pathway Editor (http://www.pathwayeditor.org/).



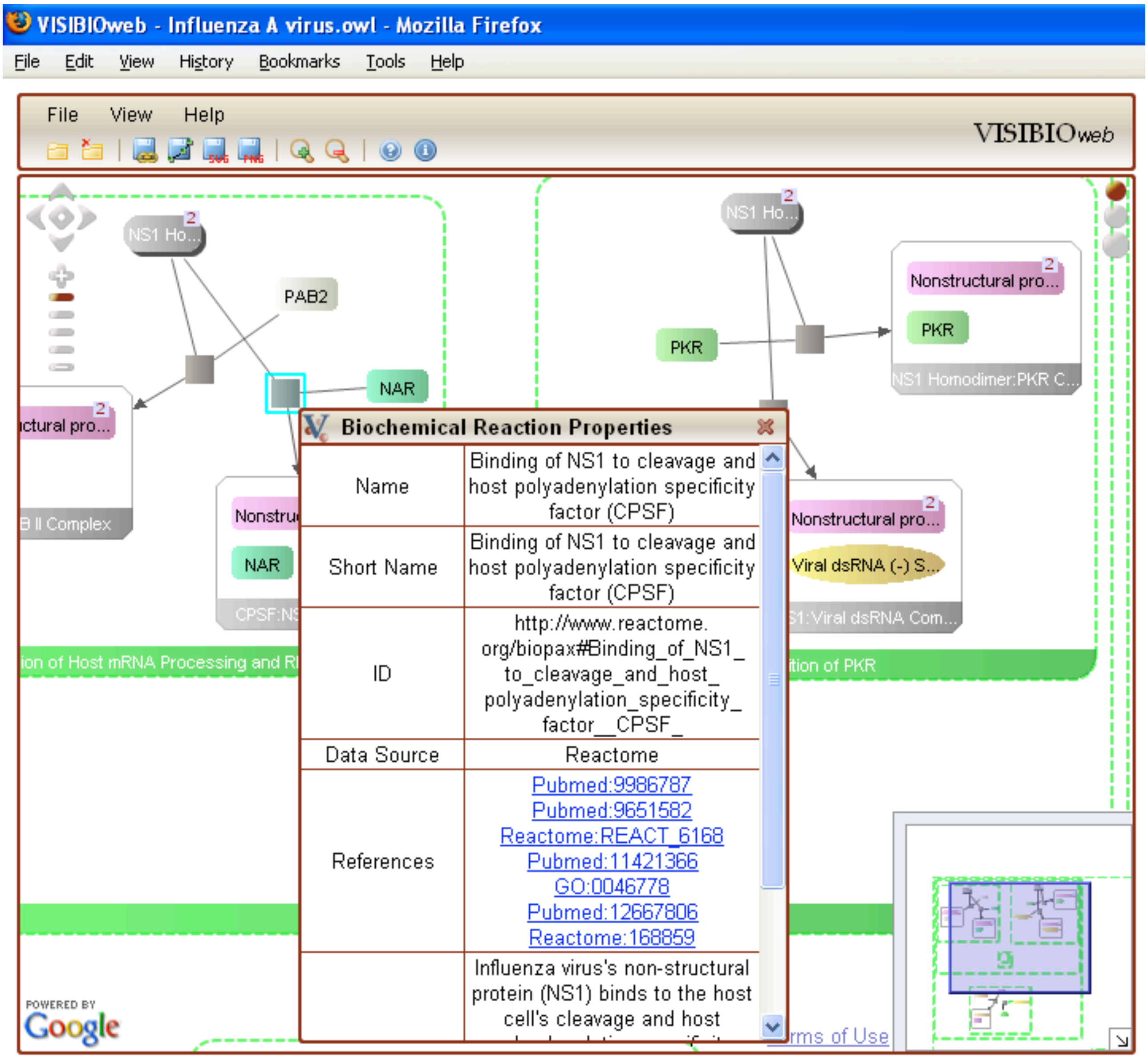
The forthcoming version of JWS Online (http://jjj.biochem.sun.ac.za/) will support SBGN-compliant model rendering. The image shown here was drawn in OmniGraffle Pro 5.1.1 using an SBGN template (available upon request from the developers of JWS Online). The schema is linked to the JWS Online applet via AppleScript code in the graph. The graph information is derived from an SBML model description translated to the JWS Online input format.



A diagram of the FAS pathway designed with the Reactome curation tool (http://brie8.cshl.org/download/tools/curatorTool/install.htm).



One of the examples present in the SBGN Process Diagram Level 1 specification is drawn with Vanted version 1.67 (http://vanted.ipk-gatersleben.de/). The layout of the diagram was made by hand, and experimental data mapped onto some of the nodes using Vanted's facilities.



An example from VISIBIOweb (http://www.bilkent.edu.tr/~bcbi/pvs.html).

Supplementary Note 3:

Symbols Used in SBGN Entity Relationship Diagram Level 1

The SBGN Entity Relationship Diagram specification defines a set of symbols with precise semantics, together with syntactic rules that control their assembly. It also describes how such graphical information is to be interpreted. The essence of an entity relationship diagram is to depict the influences of entities upon the behavior of others. The entities are things that exist, either on their own or when statements become true. For instance, an entity can exist, different entities can interact, or a value can be assigned to an entity's property. “Influences” in this context therefore can be understood as logical consequences of this existence. Contrary to the Process Diagram notation, where the different processes affect each other, the relationships here are independent. One can imagine that each of the relationships represents a specific conclusion of a scientific experiment or article. Their addition to a drawing represents the knowledge gained about the effects of the entities upon each other. The independence of relationships is the key to avoiding the potential for combinatorial explosion inherent in the Process Diagram notation.

Entity nodes *(ENs)* represent elements of truth, i.e., things that exist. In ontology parlance, they are “continuants”. Entity nodes are the sources of influences. The SBGN Entity Relationship Diagram Level 1 specification provides three different types of *ENs*: the *interactors* (*entity* and *outcome*), the *logical operators* (*and*, *or*, *not* and *delay*) and the *perturbing agent.*

The semantics of SBGN entity relationship diagrams is carried by *Relationships*. Relationships are rules that decide the existence of entity nodes, based on the existence of others. In ontology parlance, they are “occurants”. Level 1 of the SBGN Entity Relationship notation provides two types of relationships, the *statements* and the *influences*.

Statements can be true or false. *Statements* are targets of *influences*. They are not true themselves, but can carry truth element (the *EN outcome*). SBGN Entity

Nature Biotechnology: doi:10.1038/nbt.1558

Relationships Level 1 provides four types of *statements*: *assignment*, *interaction*, *non-interaction* and *phenotype*.

Influences represent the effects of an entity on other relationships. The symbols attached to their extremities make precise their semantics. SBGN entity relationship *influences* can be viewed as logical rules linking *ENs* and other rules. The Entity Relationship Diagram Level 1 specification provides seven *influences*: *modulation*, *stimulation*, *inhibition*, *necessary stimulation*, *absolute inhibition*, *absolute stimulation*, and *logic arc*.

Nature Biotechnology: doi:10.1038/nbt.1558

Rules for building an SBGN entity relationship diagram

The SBGN Entity Relationship Diagram notation defines a number of rules to help eliminate ambiguity in a map. These rules must be followed in order for the diagram to be a valid SBGN Entity Relationship Diagram. Here we list some of the most important rules, but they are not the only ones defined by SBGN—users should consult the official specification at http://sbgn.org/ for a complete list of rules.

- Only one relationship can originate from an outcome, whether it is *influence* or *interaction*. The relationships are seen as independent rules; separate consequences of an assignment or an interaction have to originate from different outcomes, that is assertion of truth of this assignment or interaction.
- There cannot be both an *absolute stimulation* and an *absolute inhibition* targeting the same statement.
- In the case of a non-binary interaction, the “cis” or “trans” *unit of information* must be carried by the circle representing the n-ary *interaction*, and not the arc connecting this circle and a given interactor.
- If an *influence* targeting an *interaction* carries a “cis” or “trans” unit of information, at least one of the *interactors* must be the same *entity* as the origin of the *influence.*
- If more than one instance of an *entity* is involved in an *interaction* or a *non-interaction*, a *unit of information cardinality* must be associated with each entity involved in the statement.
- A *cis* or *trans unit of information* can be carried only by a relationship involving a single *entity*.

Nature Biotechnology: doi:10.1038/nbt.1558

Rules for understanding an SBGN entity relationship diagram

It is important to realize that the SBGN Entity Relationship Diagram specification does not dictate how to represent something, but rather how to interpret the representation. There is generally more than one way to represent a concept. However, everyone should interpret an SBGN entity relationship diagram the same way. SBGN entity relationships can be interpreted as logical rules describing the consequences of the existence of entities.

- An *interaction* linking the *interactors* A and B means: “A interacts with B”. An *outcome* on an *interaction* represents the cases when the statement is true, that is when the interaction effectively exists. If the interaction is a physical interaction between molecules, the *outcome* represents the complex resulting from the interaction. It is used as follows: “when (or if) A interacts with B then ...”
- An *assignment* linking a state variable value v to a *state-variable* V of an *entity* E means: “v is assigned to V of E” or “V of E takes the value v”. An *outcome* on an *assignment* represents the cases when the statement is true, that is when the variable effectively displays the value. It is used as follows: “when (or if) V of E takes the value v then ...”
- A *phenotype* P means: “P exists (can be observed)”.
- A *modulation* linking an *entity node* E and a *relationship* R means: “if E exists then R is either reinforced or weakened”.
- A *stimulation* linking an *entity node* E and a *relationship* R means: “if E exists then R is reinforced” or “if E then the probability of R is increased”.
- An *absolute stimulation* linking an *entity node* E and a *relationship* R means: “if E exists then R always takes place”.
- A *necessary stimulation* linking an *entity node* E and a *relationship* R means: “R only takes place if E exists”.
- An *inhibition* linking an *entity node* E and a *relationship* R means: “if E exists then R is weakened” or “if E then the probability of R is lowered”.
- An *absolute inhibition* linking an *entity node* E and a *relationship* R means: “if

Nature Biotechnology: doi:10.1038/nbt.1558

E exists then R never takes place”.

- An *and* linking several *logic arcs* originating from *entity nodes* E_i and an *influence* F means: “if for each i, E_i exists, then F”.
- An *or* linking several *logic arcs* originating from *entity nodes* E_i and an *influence* F means: “if for any i, E_i exists, then F”.
- A *not* linking a *logic arc* originating from an *entity node* E and an *influence* F means: “if E does not exist, then F”.
- A *delay* linking a *logic arc* originating from an *entity node* E and an *influence* F means: “if E exists then F takes place, but not immediately”.

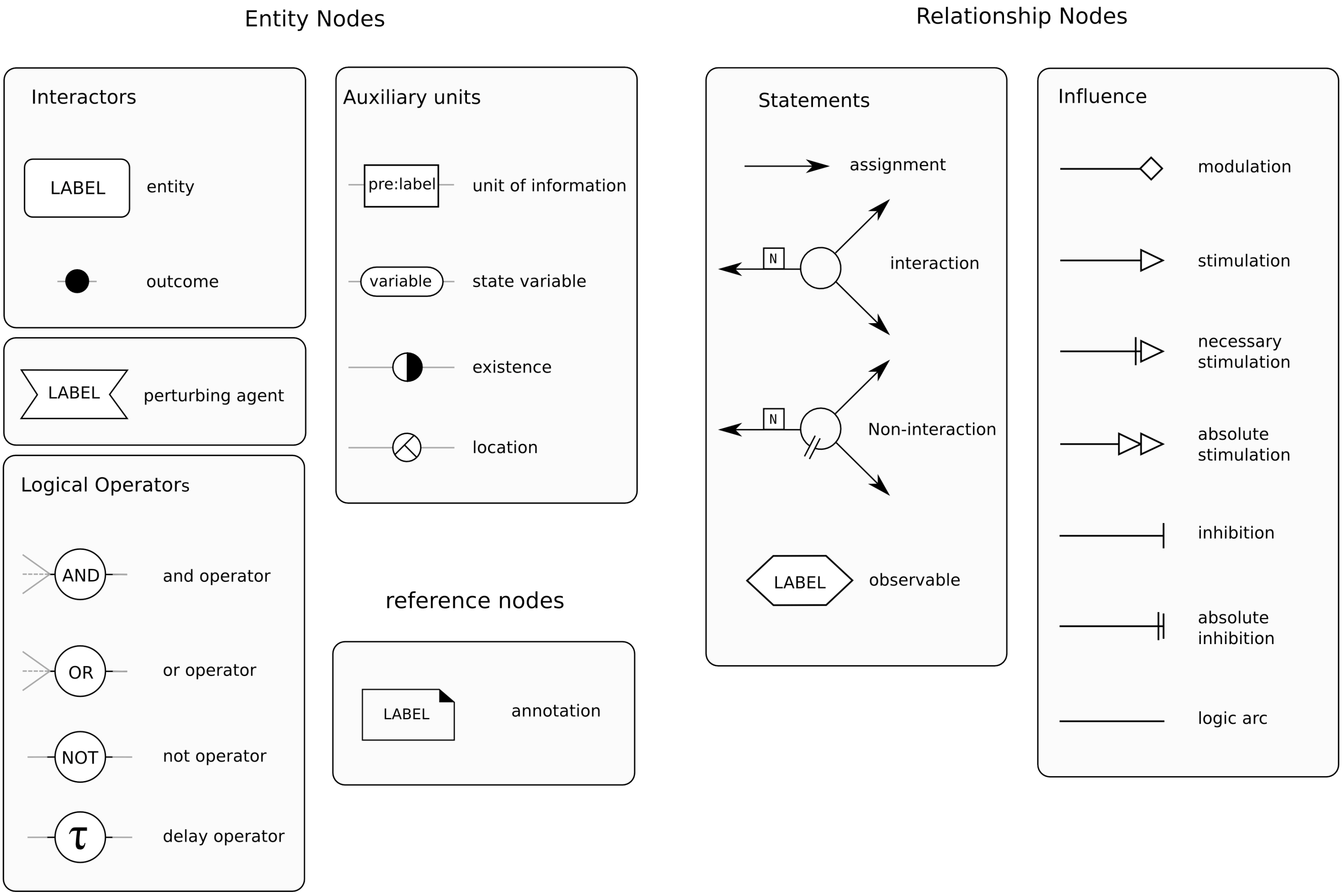
The use of “cis” and “trans” units of information on a combination of relationships brings power and versatility to entity relationship diagrams in SBGN. However, the resulting semantics may be difficult to grasp. Here are the basic rules that permit understanding the diagrams:

- The *unit of information* “cis” or “trans” carried by an *interaction* refers to the *interactors* targeted by the *interaction*.
- The *unit of information* “cis” or “trans” carried by an *influence* targeting a state variable *assignment* refers to the origin of the *influence* and to the *entity* carrying the target of the *assignment*.
- The *unit of information* “cis” or “trans” carried by an *influence* targeting another *influence* refers to the origin of the carrying *influence* and to the origin of the targeted *influence*.
- The unit of information “cis” or “trans” carried by an influence targeting an *interaction* refers to the origin of the *influence* and all the relevant *interactors* targeted by the *interaction*.

Nature Biotechnology: doi:10.1038/nbt.1558

Figure S3: List of all glyphs specified by SBGN Entity Relationship Diagram

Level 1



Nature Biotechnology: doi:10.1038/nbt.1558

Supplementary Note 4:
Symbols Used in SBGN Activity Flow Diagram Level 1

The SBGN Activity Flow Diagram specification defines a set of symbols and detailed syntactic rules for their use, to allow users to create pathway diagrams—for example, diagrams of signaling pathways—resembling those often found in research publications. An activity flow diagram is designed to show how activities are propagated from one entity to another. Glyphs (symbols) are the graphical units that represent concepts in SBGN. Each one is uniquely identified by a term from the Systems Biology Ontology (SBO). There are two types of glyphs in SBGN: nodes (which are subdivided into activity nodes, unit of information node, container nodes, and logical operator nodes), and arcs (edges) that characterize the relationships between nodes.

Activity nodes *(ANs)* represent activities produced by an entity from an entity pool. Level 1 of the SBGN Activity Flow Diagram specification defines three distinct glyphs, one for each of the following concepts: *biological activity, perturbation, and phenotype*. The notation uses one glyph to represent activities from all kinds of biological entities; collectively, they are called "biological activity". The nature of the molecule that the activity comes from (e.g., simple chemical or macromolecule) can be encoded in the *units of information*. A biological activity can come from one biological entity, a part of an entity, or a combination of them; thus, biological activity is not equivalent to a biological entity per se. Each activity node can only be represented once within a given compartment.

Modulation arcs *(MAs)* describe the way in which one *AN* affects (or influences) others. Level 1 of the SBGN specification defines four *MAs*: *positive influence* (activation)*, negative influence* (inhibition)*, unknown influence,* and *necessary stimulation*.

Unit of information provides a way to illustrate the nature of the entity producing a given activity. In the SBGN activity flow notation, the *unit of information* is a

Nature Biotechnology: doi:10.1038/nbt.1558

unique type of node, in that its relationship to another node does not require an arc. Level 1 of the specification defines distinct glyphs for representing five different types of entities: *macromolecule, simple chemical, genetic, unspecified,* and *complex*.

Logical operators provide a means of indicating Boolean combinations of influences from *ANs* onto other *ANs*. The four possibilities are conjunction (*and*), disjunction (*or*), negation (*not*), and *delay*.

Compartments are containers that provide particular information about the location of ANs and MAs.

Rules for building an SBGN activity flow diagram

The SBGN Activity Flow notation defines a number of rules to help eliminate ambiguity in a diagram. These rules must be complied with in order for the diagram to be a valid Level 1 SBGN Activity Flow Diagram. It is important to realize that SBGN's activity flow notation does not impose how to represent something, but rather how to interpret the representation. There is generally more than one way to represent a concept; for instance, an EGF receptor can be represented by one *biological activity* node as EGF receptor activity, or by two *biological activity* nodes, one as EGF binding activity, the other as EGF receptor kinase activity. However, everyone should interpret an SBGN activity flow diagram the same way. Here we list some of the most important rules. Note that they are not the only rules defined by SBGN Activity Flow Level 1; users should consult the official specification at http://sbgn.org/ for a complete list of rules.

- An *activity node* can only appear once in a given *compartment*. If no compartment is drawn, the node is assumed to belong to a "default" compartment.

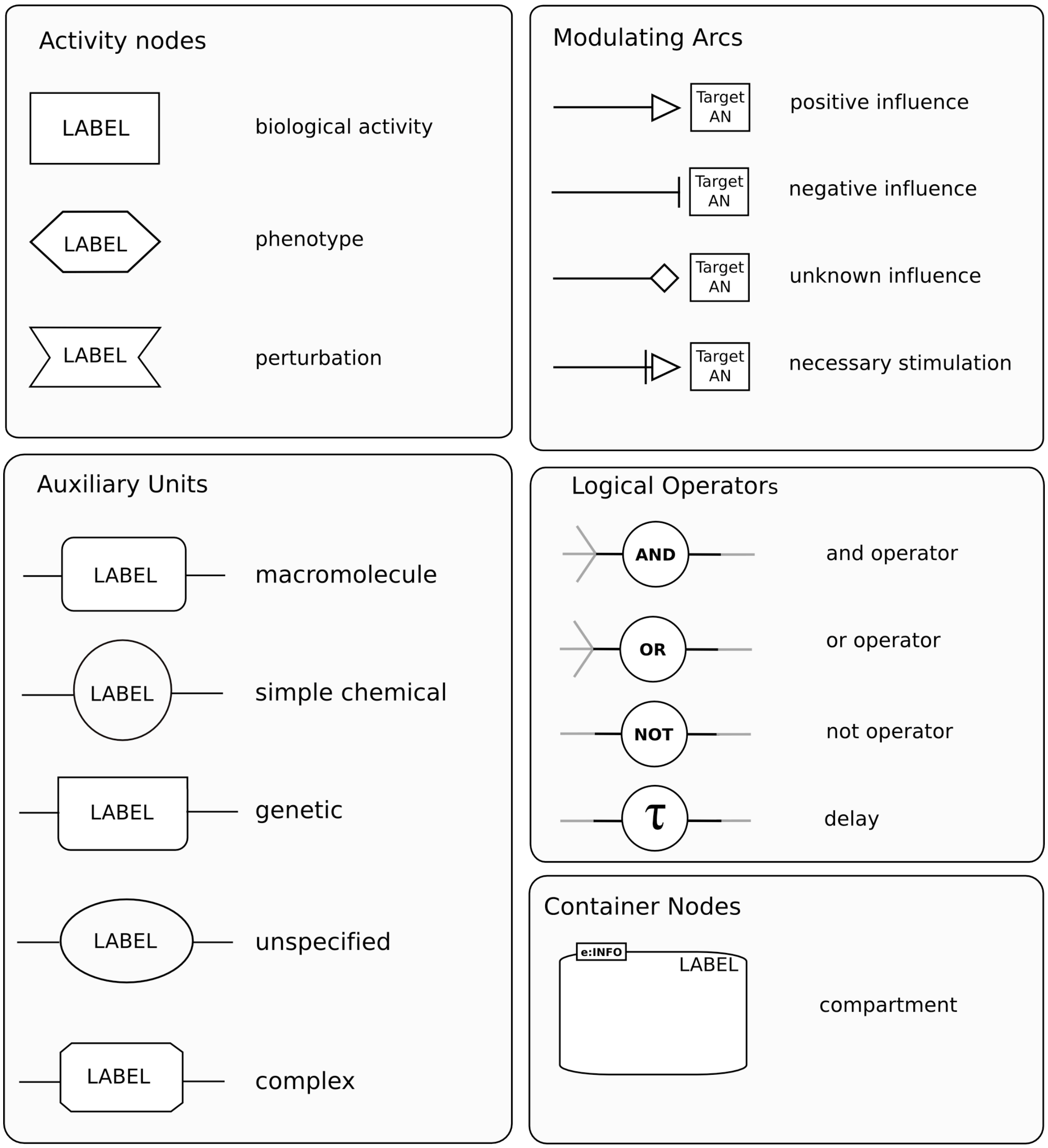
- *Compartments* cannot be nested, and they represent disjoint spatial containers. *Compartments* may overlap visually, but such overlap does not imply any kind of physical containment; i.e., a *compartment* is never "part" of another.

- The layout or organization of a *compartment* does not imply anything about its topology.

- The use of *unit of information* is not required, but should be used when it becomes necessary for proper interpretation of a diagram. In addition, *the unit of information* can be used without a label to simply indicate the nature of the entity producing the activity, or with a label to provide more information.

- *Perturbation* is used only as an origin to a modulation arc or a logic operator.

Figure S4: List of all glyphs specified by SBGN Activity Flow Diagram Level 1



Nature Biotechnology: doi:10.1038/nbt.1558