

Modeling without Borders: Creating and Annotating VCell Models Using the Web

Michael L. Blinov, Oliver Ruebenacker, James C. Schaff, and Ion I. Moraru

Center for Cell Analysis and Modeling
University of Connecticut Health Center
Farmington, CT 06030, USA
{blinov,oruebenacker,schaff,moraru}@exchange.uchc.edu
<http://vcell.org/sybil>

Abstract. Biological research is becoming increasingly complex and data-rich, with multiple public databases providing a variety of resources: hundreds of thousands of substances and interactions, hundreds of ready to use models, controlled terms for locations and reaction types, links to reference materials (data and/or publications), etc. Mathematical modeling can be used to integrate this complex data and create quantitative, testable predictions based on the current state of knowledge of a biological process. Data retrieval, visualization, flexible querying, and model annotation for future reuse, are some of the important requirements for modeling-based research in the modern age. Here we describe an approach that we implement within the popular Virtual Cell (VCell) modeling and simulation framework in order to help connect the modeling community with the web of machine-processable systems biology knowledge. A new software application, called SyBiL (Systems Biology Linker), has been designed and developed for simultaneous querying of multiple systems biology knowledge bases and data sources, such as web repositories, databases, and user files, and converting the extracted and refined data into model elements. Integration of SyBiL as a component of VCell makes these capabilities easily available to a wide modeling community.

Keywords: Biological databases, mathematical modeling, VCell, data conversion.

1 Introduction

Mathematical modeling is increasingly necessary to investigate the function of molecular pathways and networks, and a growing number of resources that facilitate computational approaches are becoming available. On the one hand, a large collection of public databases can help researchers gather existing knowledge about molecular interactions: databases such as Reactome [1], the BioCyc collection of Pathway/Genome databases [2], Pathway Interaction Database (PID) [3], BioModels repository of computable models [4], Integrating Network Objects with Hierarchies (INO) database [5], Kyoto Encyclopedia of Genes and

Genomes [6] store information about thousands of molecular species and their interactions. On the other hand, a lot of software tools can help researchers to create and simulate mathematical models of such interactions: a few examples of the more widely used platforms are VCell [7], [8], Copasi [9], CellDesigner [10]. However, formats that are typically used to store molecular pathway information (e.g. Biological Pathway Exchange standard, BioPAX, [11]) are database-centric and formats used by modeling software (e.g. Systems Biology Markup Language, SBML, [12]) are simulation-centric, with significant semantic and structural differences. This makes exchange of information between databases and modeling tools problematic. There are several tools that provide conversion from BioPAX to SBML. Each of them has distinct advantages for particular uses. BiNoM [13] is a plugin for Cytoscape [14], a powerful tool for analysis of data represented as graphs with rich data representation and visualization (primary graph layouts and scaling) capabilities. Paxtools library [15] is the primary tool for working with BioPAX format and provides excellent capabilities for analysis of BioPAX data. Some pathway databases (e.g. Reactome, INOH) now provide output in the form of SBML file. However, none of these tools is truly modeling-oriented.

In this manuscript we describe a framework, called **Systems Biology Linker, SyBiL** [16], [17], that is designed to assist a modeler in using data from multiple online sources. The unique strength of the proposed framework is a model-oriented approach, which can handle issues that can not be tackled either by current tools such as Cytoscape and PaxTools, or by straightforward export/format conversion from databases. A few examples are the capabilities to query a variety of resources from a modeling point of view, such as:

- What database entities can be used as species?
- Should a subset or superset of database entities be used as species in the model?
- Should variants of a given protein (e.g. different phosphoforms, different conformational states) be made into distinct species?
- Should multiple ligands be converted into a single or multiple species?
- Can we keep track of all such assumptions, so that when future changes are made (like dividing a single species corresponding to all ligands into several species), this can be easily done?

In Section 2 below we analyze the multiple problems faced by researchers when attempting to create mathematical models based on the knowledge from pathway databases. In Section 3 we describe the approach to integrate data sources into a modeling process. In Section 4 we describe a framework for querying multiple resources to use in model building. In Section 5 we discuss how this framework will be incorporated into the Virtual Cell (VCell) modeling and simulation platform in order to provide an expert system guiding users in building and annotating computational models. Finally, in section 6 we talk about future plans and problems that we are facing implementing our approach.

2 Challenges for Building Quantitative Models of Molecular Interaction Networks

2.1 Accessing Multiple Systems Biology Resources

Progress in systems biology crucially depends on the capacity to share machine-processable knowledge (knowledge data), which is the product of costly curation. The benefits of knowledge data manifest in many ways, from more powerful querying to new approaches in data analysis to building and annotating models. This has spawned a diverse multitude of knowledge bases on the web offering free knowledge data about substances, interactions, pathways, models, anatomical features and literature references.

Sharing knowledge data depends on standardized formats which typically result from sustained community efforts around specific interests such as BioPAX for pathways and SBML for models. While the domains of these still evolving standards overlap (e.g. both BioPAX and SBML represent substances and their interactions), conversion usually results in loss of information.

Access protocols differ naturally between knowledge bases. While downloading a data set in a popular format may be a similar experience across knowledge bases, the interface of the query and selection capabilities differs widely. Knowledge data from various sources has been aggregated by Pathway Commons [18] and Bio2RDF [19], but such aggregation necessarily lags behind the original. Also, both projects change the original modifiers and require considerable technical skills to be used efficiently, which prevent most users from using them. Moreover, neither makes use of SBML models.

More problematically, when delivery and end use of data are divorced, the mode of retrieval may be poorly adjusted to the requirements of the use. In the worst case, a user will need three different tools for retrieval, evaluation and final use of the data. If however, a single tool is responsible for retrieval and final use, it can help the user decide what data is needed and where it can be obtained, and it can filter retrieved data.

If the entire online knowledge data was stored in a single base, in a single format and with a single access protocol, a simple query could yield all available data. In reality, we have different knowledge bases, formats and protocols to serve different priorities, and the best combination of base, format and protocol depends on the given interest. Some interests are best served by a specific combination of source, format and protocol: for example, to find a model, we can query BioModels database and obtain it in SBML, or to find a pathway, we can query Pathway Commons and obtain it in BioPAX, or to find a protein, we can enter its name in UniProt database [20] and get a listing.

Other interests are best served by querying multiple sources, which requires integrating multiple formats and protocols. For example, if the interest is to find substances with which EGFR interacts, candidates are found in BioModels, Pathway Commons and UniProt. Interfaces differ and results are available in SBML, BioPAX and site-specific format respectively. Moreover, the interest does not require entire models or pathways, but only fractions thereof. Checking out

multiple sources separately and submitting a query to each, using the necessary formats and protocols, is an enormous task, especially since sources, formats and protocols evolve over time.

At the same time, efficient queries exploit the fact that resources are extensively interlinked. For example, to investigate EGFR, a user typically would start with the name EGFR, locate a UniProt identifier (for the desired organism) and then use the identifier to find models and pathways. This can be easily automated. What is needed therefore is a tool that accepts, for example, a protein name and an organism, and generates a list of possible reaction partners. This involves looking up identifiers, submitting queries simultaneously to multiple sources, supporting multiple formats and multiple protocols, and gathering the results into a single list. It requires an integration of multiple formats and multiple protocols.

2.2 Model-Driven Use of Pathway Data Requires Human Decision Making

In a mathematical model, whether two compounds constitute the same or different substances usually depends on whether they behave differently or not in that particular model scenario. Knowledge present in databases, however, aims to make statements about entities independent of any particular scenario. Due to these differences, many elements do not map one-to-one between models and pathway data.

- A species in a model can be a subset of a database entity. For example, Epidermal Growth Factor receptor protein is typically a single entity in a database, but some mathematical models have species that correspond to multiple phosphoforms of it [21].
- A species in a model can represent two or more database entities. For example, multiple isoforms of a protein represented by distinct database entities can be lumped into a single species [21].

As we see, sometimes the question of whether a species will be converted to one or more physical entities, and vice versa, is not trivial. In [16] we discuss how this question can be answered automatically by extensive analysis of extensions and annotations in the source file or import from other sources. While most common cases can be automated, we pointed out that a few cases remain where user intervention might be necessary [22], such as determining the topology of locations, substituting a substance with another one that is equivalent in a particular model, or introducing certain assumptions into a model.

2.3 Recording Modeling Assumptions

As we have seen above, a conversion between database information and models requires certain human decisions. These decisions have to be stored in some format, so it can be reused in future data manipulations (reverse conversions,

merging with other data, etc.). We need a systematic approach for storing and reusing these relationships and thus making conversions between knowledge and model reproducible and reversible. We introduced a bridging ontology, SBPAX (Systems Biology Pathway Exchange), described in [22] and summarized in section 3. The SBPAX bridging ontology allows the conversion between multiple formats (modeling and database) to be performed as two consecutive one-to-one mappings, with an intermediate refinement step that is performed on, and recorded in, the SBPAX data.

2.4 Merging Models

Even the well-annotated models stored in BioModels repository that are compliant with MIRIAM (Minimum information requested in the annotation of biochemical models) standard present challenges [23], [24]. For example, currently there is no standard way in SBML to distinguish between different states of the same molecules. Thus, all phosphoforms of the same receptor will be annotated with the same reference identifier (GO term from Gene Ontology [25], UniProt key, etc.). This means that there still will be often the case that it is impossible to automatically tell whether species X of model A and species Y of model B, which have the same reference identifier, are indeed identical and should be mapped into the same species in a merged model. Linking each species to the (multiple) sources of knowledge and recording assumptions on how these sources were used can greatly facilitate automatic merging of models.

2.5 Using Models to Appreciate the Complexity of Underlying Biology

Each model is usually based on a selection of a small part of a larger network of interactions. However, this point is often missing when a model is presented [26]. Linking each model element to biological knowledge will provide an opportunity to create model-based knowledge bases. In such a case, when logging in to a public model, for example, clicking on a component of interest brings up a battery of potential modifications, interactions, and activities, and the likelihoods and potential consequences of each under a variety of typical sets of conditions, or specific conditions set by the user.

3 Infrastructure for Data-Driven Modeling

3.1 SBPAX Bridging Ontology

To support the integration of datasets related to molecular networks and pathways from different sources in different formats, we need to accomplish three tasks: converting data from one format to another, gluing corresponding data sets in different formats, and merging multiple datasets into one. The core of the underlying technology we developed to accomplish these tasks is an RDF/OWL-based ontology called SBPAX, short for Systems Biology Pathway Exchange [22].

It is designed to integrate multiple formats and based on a Web Ontology Language OWL [27]. Advantage of using OWL include: (1) other ontology-based formats can be seamlessly integrated by providing relationships between their classes and properties and those of SBPAX; (2) other structured formats not based on OWL can be mapped one-to-one to OWL and linked after the mapping. This includes XML, relational databases or even comma-separated values.

SBPAX is designed to be able to express any molecular reaction network that can be expressed in SBML, BioPAX and formats similar to these. To accomplish this, SBPAX covers substances, processes and interactions, locations and stoichiometric coefficients, and hierarchies existing among these. In the case of an RDF/OWL-based format like BioPAX, data can be directly linked to SBPAX data. With other formats like SBML, data can be mirrored one-to-one to RDF/OWL-data in SBPAX that can be linked with other SBPAX terms. SBPAX data created by import from multiple sources can be exported to SBML, and every basic SBML term (e.g. species, reaction, species type or compartment) will contain, as annotation, the URI of its equivalent SBPAX object, which will be linked by relationships to all related terms in the imported files.

To give a simple example, a substance in SBPAX is defined as any group of molecules or other compounds. SBPAX provides properties to define a substance as a superset or subset of another substance, or as the union or intersection of two other substances (useful for substances defined by constraints, e.g. on their phosphorylation state). This way, we can create a substance hierarchy and identify substances that can cover more or less than one database entity.

3.2 Prototype of a Data-Driven Modeling Interface: Systems Biology Linker (SyBiL)

We have prototyped Systems Biology Linker (SyBiL) modeling framework [16], [17]. The initial version is geared towards work with two common standards BioPAX and SBML. It is designed to obtain, store and merge data in BioPAX format, and to facilitate generating of kinetic models expressed in SBML (<http://vcell.org/sybil>). The tool is providing a modeling access to complete range of BioPAX data, which is not essential for simulations, but valuable for understanding and reusing the models (such as organisms, different names, linking species to a variety of databases, etc.). It takes an advantage of easy visualization of BioPAX, where different BioPAX object classes (proteins, small molecules, complexes etc.) are represented by nodes differing in shape or color, and each object is linked to biological information from public databases. The SyBiL converts BioPAX data into a computable kinetic model in SBML. Remarkably, the model is first generated in SBPAX format and stores all the modeling decisions regarding conversion of BioPAX data into SBML. After conversion to SBML BioPAX data is used to provide each SBML species and interaction with a unique identifier.

To facilitate the organization of the data and to make selections, SyBiL provides a graphical interface that can handle any OWL data. OWL (Web Ontology Language) is designed for use by applications that need to process the content

of information instead of just presenting information to humans. OWL provides a framework for controlled vocabulary along with a formal semantics. Specific support is provided for the BioPAX ontology. Ontologies provide a great deal of flexibility in data representation, analysis and visualization, by organizing data in a way that allow intelligent selections, for example allowing the user to select which resources are visible and which are hidden, similar to the Cytoscape visualization framework. Furthermore, arbitrary sets of objects can be visually collapsed and expanded. The user can decide which kinds of property relationships are represented by graph edges. As the framework is based on the existing VCell software, it displays a graph consisting of interactions among BioPAX physical entities in the VCell style as a bipartite graph in its fully flattened form (with nodes for both physical entities and interactions). For each class of physical entities and interactions the framework provides a separate symbol. Each physical interaction is connected by an edge with the physical entities participating in it. Components of complexes can be shown. All other objects are hidden, until they are properties of an object which becomes selected.

4 Specific Querying of Multiple Resources for Model Building

We are developing SyBiL into a framework that will allow the modeler to get simultaneous access to most systems biology knowledge bases from within a modeling-oriented interface. By obtaining references through simple text queries or more advanced queries, and then getting related data from references, the user will get information sufficient to create and annotate model elements in one step.

Previously, SyBiL was able to create models from locally stored BioPAX files [16], [17], [22]. Recently, we added a new capability for generating models from data retrieved from Pathway Commons and UniProt web resources.

The interface forwards a keyword query to Pathway Commons. After entering a search string (EGFR) and clicking the search button, the matches are displayed as a tree (Fig. 1). Each match contains brief information (name, synonyms, excerpts, and organism) as well as a list of cross references and a list of pathways that include the matched entity. Since Pathway Commons returns multiple UniProt cross references for human EGFR without distinction, SyBiL then sends these cross references to UniProt, which lists P00533 as the one in use and the other ids as obsolete.

After a pathway is selected, it can be retrieved in the BioPAX format from Pathway Common by clicking on the “Get pathway” button. The entities contained in the pathway are presented as a list (Fig. 2), grouped into processes and various kinds of substances (DNA, RNA, proteins, metabolites and complexes). For example, selecting the protein EGFR_HUMAN imports from this pathway human EGFR as well as the complexes which it is part of and the interactions in this pathway in which EGFR or any of its complexes participate. The reaction network can be converted into a VCell model and then further edited and simulated.

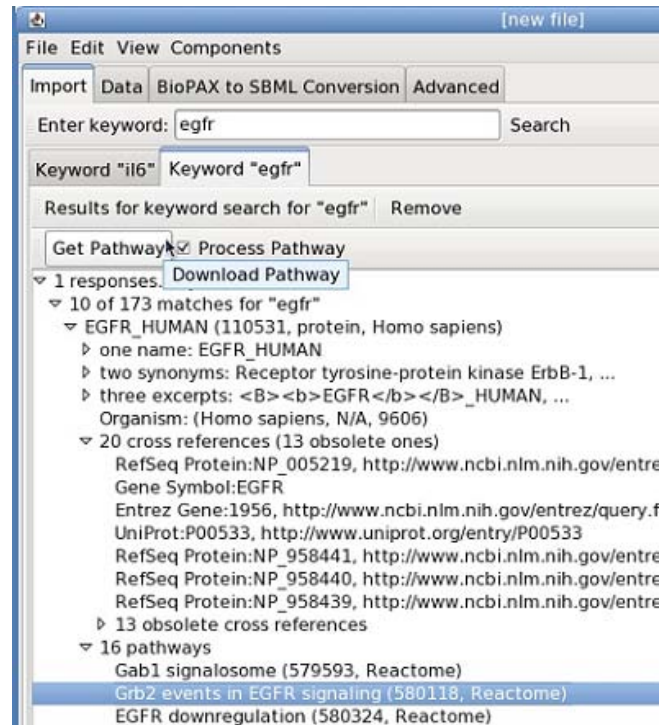


Fig. 1. Querying Pathway Commons for EGFR brings up all available information about entities that contain this term, presented as a tree. Under each protein entry there is a lot of extra oinformation, including a list of pathways that use this protein.

We plan to expand retrieval to data in other formats and from other resources, such as Bio2RDF, Uniprot Database, BioModels database, etc (Fig. 3) A user intending to use knowledge data to create, extend or annotate a mathematical model will be able to use the proposed interface as an entry point for selection of entities (e.g. substances and interactions) and their relationships to build his or her model.

The simplest starting point would be to enter a search term (such as a name of a protein) or a qualified identifier (e.g. UniProt P00533) to get knowledge data that can be used to create or annotate a species, as well as information about pathways, interactions and reaction partners for a name of interest. This information will be translated into an SBML model for use in any SBML-compliant software, or can be used by VCell software for modeling. A user will be able to send queries that rely on knowledge data, such as looking for components of a given complex, or participants in a given interaction, or find all entities at a given location, etc. A more advanced example of querying knowledge data is: given two substances, find the shortest chain (a list of substances such that each substance is a reaction partner for the substance next on the list) and connect them, under certain constraints. To eliminate trivial solutions, some constraint is needed. A

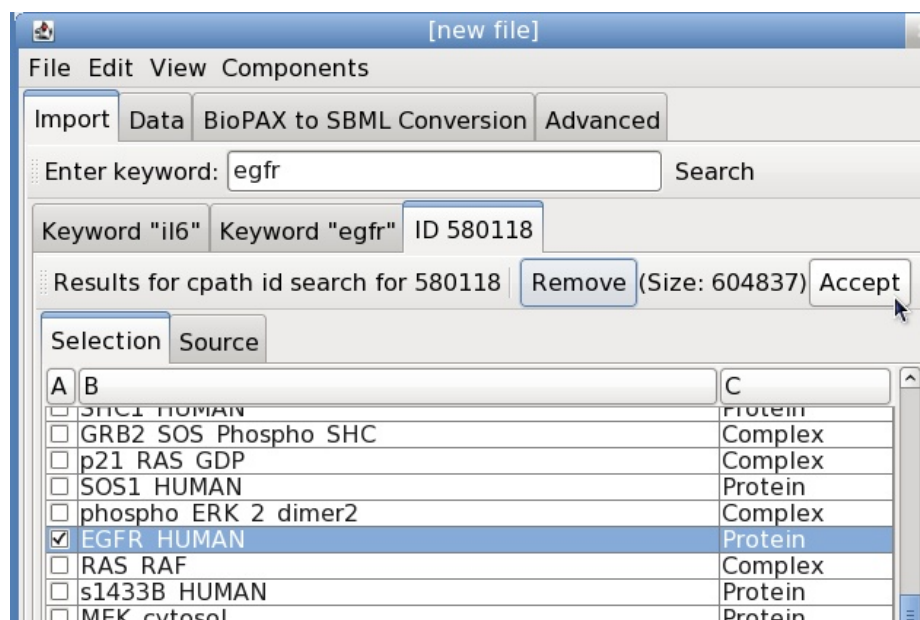


Fig. 2. Selecting a certain pathway brings up the number of entities that are used in this pathway. These entities itself can be selected to be imported into a modeling tool.

sensible constraint could be that every substance in a chain needs to be: (1) not ATP or (2) not a metabolite or (3) a protein. A reaction network will be built by aggregating related substances and interactions from multiple queries, where a substance or interaction obtained in one query can be the starting point for another query. The resulting network can be visualized with multiple levels of resolution. After the user selects substances and interactions and configures modeling assumptions, a mathematical model can be created and annotated automatically.

More specifically, the results of the generic query are presented as a raw SBPAX file (Fig. 4). It may include a variety of similar substances, such as EGFR, phosphorylated EGFR, EGFR phosphorylated at Y1174, EGFR unphosphorylated at Y1174, etc. For each substance, a modeler is provided with three basic pieces of information required to make a decision: location, chemical identity, and list of interactions. The user can set global options that determine which substances may be part of different subsets. This may depend on the system under consideration, but a typical setup would be that proteins, but not metabolites, can have different subsets. For example, this would mean that every interaction that involves ATP refers to the same substance, but different interactions may refer to different subsets of EGFR. After this global setup, the next step would be to go through the proteins and determine, for example, that EGF has no subsets, but EGFR has. BioPAX data usually contains information about the relevant post-translational modifications (as sequence features), which

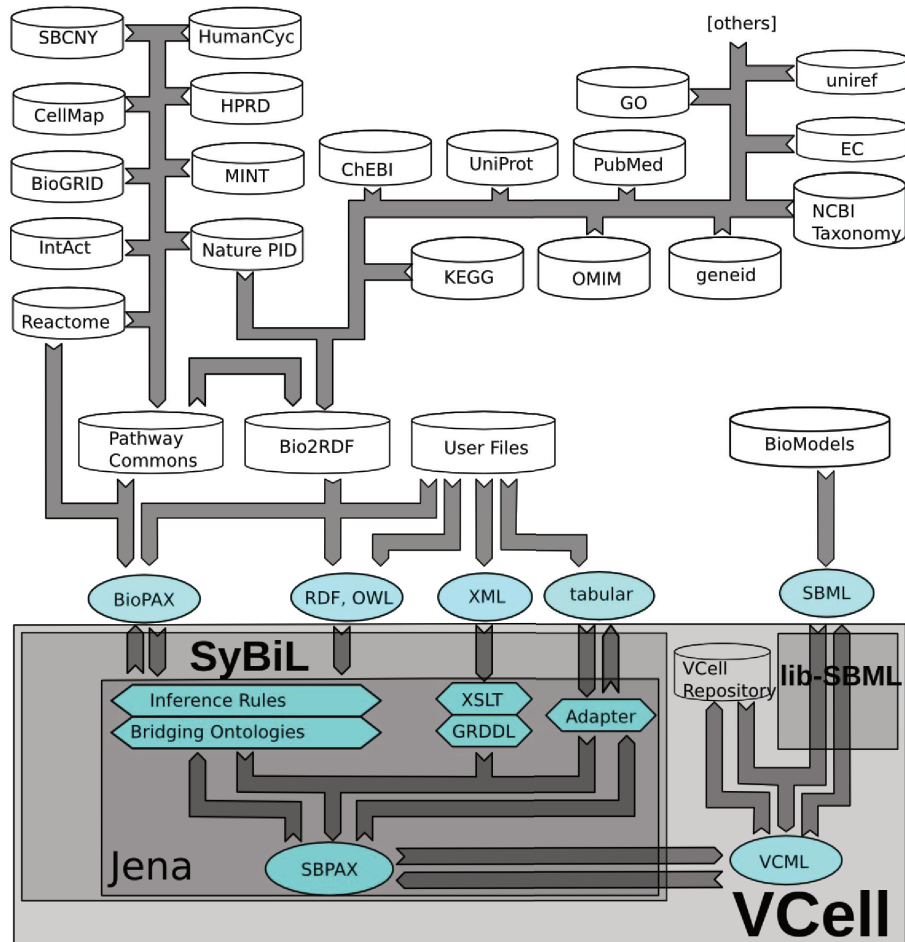


Fig. 3. Schematic description of SyBiL-VCe architecture. Arrows show the flow of knowledge from the knowledge bases (cylinders) to components of SyBiL and VCell and the formats used (ovals). Pathway Commons collects data from nine pathway bases available in BioPAX. Reactome's data in BioPAX is available from Pathway Commons, but can also be retrieved directly client. Bio2RDF collects knowledge data from a wide variety of sources, some shown here. The knowledge data collected by Bio2RDF is RDF which mirrors data that is either RDF or converted to RDF. Bio2RDF maps name spaces to internal name spaces which reveal the source of the data and can be trivially mapped back. SyBiL uses Jena to handle RDF and conversions from all above formats to SBPAX, by methods shown as hexagons. SyBiL in VCell allows conversions between SBPAX and VCML, and VCell handles conversions between VCML and SBML.

can be used to identify subsets of EGFR, which can be arranged into hierarchies in SBPAX. For example, EGFR phosphorylated at Y1174 is declared to be a subset of phosphorylated EGFR and subset of EGFR. After identifying these

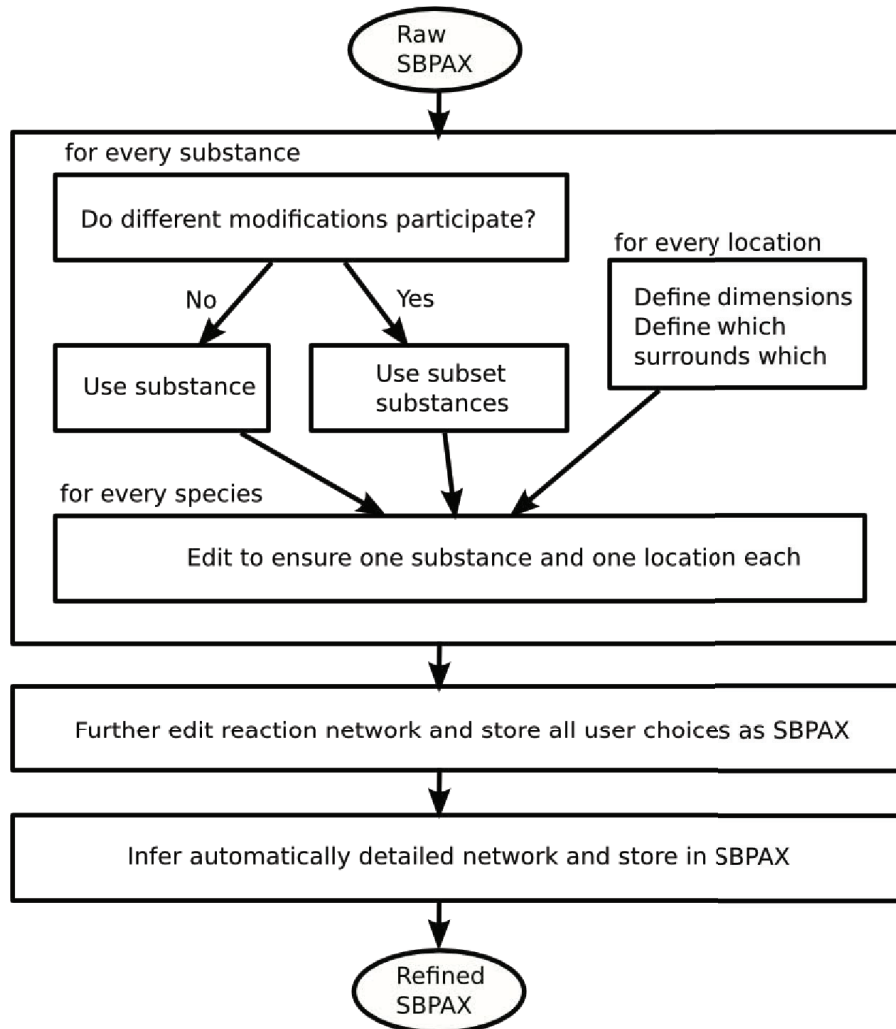


Fig. 4. Conversion of data into a model is performed as a refining step in SBPAX. Information about substances and processes extracted from databases (such as in BioPAX format) is presented as raw SBPAX which is converted into a model with user input. For example, if raw SBPAX describes an interaction of EGFR phosphorylation (multiple modifications of EGFR are involved) the way typically found in databases, we can not use database substance EGFR. Instead, we have to declare subset substances for unphosphorylated and phosphorylated EGFR respectively. The topology of the locations is usually not in the database, so it has to be added as well. Finally, a species (e.g. “EGFR in cytosol”) is defined by a substance (“EGFR”) and a location (“cytosol”). Refined SBPAX contains this inferred information.

subsets the final step would be to determine, which of these subsets participate in which interaction. After participating substances are determined, a user will be provided with an option to assign locations (if they are unavailable) and set topology of locations, such as set surroundedBy and hasDimension properties. Finally, the user will be provided with the list of processes. Some of the processes may be flagged, for example if previously the user identified reactant and product as the same species. The user will have an option to modify certain processes, which may lead to changes in substance assignment, etc.

After the selection of model elements, the next step is converting the SBPAX data into a fully annotated mathematical model, such as SBML. The SBML language specification has recently introduced features that facilitate and standardize the inclusion of additional information that is not required for the numerical interpretation of the model, but which can help describe the model and relate model and model elements to each other, both within the same file or between files from different sources.

5 VCell-Based Expert System for Building and Annotating Computational Models of Molecular Interaction Networks

The Virtual Cell (VCell; <http://vcell.org/>; [7], [8]) is a computational framework that is easily accessible to cell biologists and that permits construction of models, application of numerical solvers to perform simulations, and analysis of simulation results. Due to continuous enhancements in capabilities and to many unique features, it has achieved a fast growing user base. As of February, 2009, more than 2,000 worldwide VCell users had actually run simulations. These users are currently collectively storing over 29,000 models and the results of more than 160,000 simulations in the VCell database system, of which more than 600 models and more than 2,300 simulation results were made public by their owners to be available to the overall scientific community. Users can formulate complex models of cellular processes with a simple biology-oriented graphical user interface. VCell will automatically (i) generate the appropriate mathematical encoding for running a simulation, and (ii) generate and compile the appropriate computer code.

SyBiL was developed with VCell in mind. An expert system that enables and guides users to create, extend and annotate computational models from variety of web resources would be extremely useful for VCell framework. A typical computational model consists of compartments populated by substances interacting with each other. The expert system would support starting new models, as well as adding elements and annotations to existing models. The goal of the expert is to respond to any situation by presenting the user with a list of options that is short, but allows proceeding efficiently in any direction the user may desire.

Practically any situation can be boiled down to this scenario: the user needs to choose one specimen from a universe of particulars (e.g. a substance, an

interaction, an anatomical feature, an organism). The expert system activates a set of candidates and ranks them according to a variety of indicators of how likely it is the one the user intends and also places them into categories. Picking a candidate completes the selection process. If no single candidate can be selected, the expert may present a category of candidates, and picking a category leads to a new list of candidates.

Working in a context, an expert system will be able to make sophisticated guesses, for example when adding a new protein to a model containing human proteins, the new protein is probably human, too. Complex queries become possible, such as how do IGF1 and IL6 interact with each other?, which can only be answered by a complex series of queries.

6 Conclusions and Further Directions

We are developing the Systems Biology Linker (SyBiL) application and Virtual Cell (VCell) plugin as a tool that automatically retrieves public information while guiding the user through the process of modeling: for example, starting by choosing human as an organism and EGFR as a name, it will locate the UniProt identifier for human epidermal growth factor receptor, create a corresponding species element in a VCell model, and add standards-compliant annotations to identify human, human EGFR, related publications etc. It will offer many suggestions on how to proceed: what reactions are known to involve this protein, what are its possible subcellular locations, what are its most common binding partners, what reaction chains (signaling pathways) may lead to some desired effect (e.g. MAP kinase cascade activation), etc. All user decisions will be recorded, and at each point the user will be able to add extra information, with annotations being automatically updated. This will also ensure that modeling assumptions can easily be retrieved in the future for model adjustments, and greatly facilitate model reuse in different contexts and/or by different researchers.

The primary challenge of this project is to keep up with an evolving web. Formats are being replaced by newer versions, and sources may change their interface or may go in and out of existence. To meet this challenge, we rely as much as possible on technology that is generic rather than specific to particular versions. For example, we use a generic RDF/OWL tool (Jena, [28]) to process BioPAX, and we base queries on a generic protocol. We also rely as much as possible on configuration rather than hard-coding. In particular, processing of BioPAX is based on a bridging ontology SBPAX, which makes the code itself independent of many details of BioPAX. Support for a new version of BioPAX is primarily added by extending SBPAX, and little or no change is necessary for the main application code.

Acknowledgements. The project was supported by NIH U54 RR022232 and P41RR013186 grants.

References

1. Vastrik, I., et al.: Reactome: a knowledgebase of biological pathways and processes. *Genome Biol.* 8(3) (2007); Database is accessible, <http://www.reactome.org/>
2. Krummenacker, M., Paley, S., Mueller, L., Yan, T., Karp, P.D.: Querying and computing with BioCyc databases. *Bioinformatics* 21(16), 3454–3455 (2005); Database is accessible at, <http://biocyc.org>
3. <http://pid.nci.nih.gov/>
4. Le Novère, N., et al.: BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 34, D689–D691 (2006); Database is accessible at, <http://biomodels.net>
5. <http://www.inoh.org/>
6. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M.: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27, 29–34 (1999); Database is accessible at, <http://www.genome.jp/kegg/>
7. Slepchenko, B.M., et al.: Quantitative cell biology with the Virtual Cell. *Trends Cell Biol.* 13, 570–576 (2003); Software is accessible at, <http://vcell.org>
8. Moraru, I.I., et al.: Virtual Cell modelling and simulation software environment. *IET Systems Biology* 2(5), 352–362 (2008)
9. Hoops, S., et al.: COPASI -a COMplex PATHway Simulator. *Bioinformatics* 22(24), 3067–3074 (2006), <http://copasi.org>
10. Funahashi, A., Tanimura, N., Morohashi, M., Kitano, H.: CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 1, 159–162 (2003); Software is accessible at, <http://celldesigner.org>
11. Luciano, J.S.: PAX of mind for pathway researchers. *Drug Discov Today* 10, 937–942 (2005)
12. Hucka, M., Finney, A., et al.: The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531 (2003)
13. Zinovyev, A., Viara, E., Calzone, L., Barillot, E.: BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics* 24, 876–877 (2008)
14. Shannon, P., et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11), 2498–2504 (2003)
15. Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., Nisanci, G., Cetin-Atalay, R., Ozturk, M.: PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* 18(7), 996–1003 (2002)
16. Ruebenacker, O., Moraru, I.I., Schaff, J.C., Blinov, M.L.: Kinetic Modeling Using BioPAX Ontology. In: *Proceedings of the 2007 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 339–348 (2007)
17. Blinov, M.L., Ruebenacker, O., Moraru, I.I.: Complexity and modularity of intracellular networks: a systematic approach for modelling and simulation. *IET Systems Biology* 2(5), 363–368 (2008)
18. Pathway Commons, <http://www.pathwaycommons.org/pc/>
19. Belleau, F., Nolin, M.A., Tourigny, N., Rigault, P., Morissette, J.: Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J. Biomed Inform.* 41(5), 706–716 (2008), <http://bio2rdf.org/>
20. Universal Protein Resource, <http://uniprot.org>
21. Blinov, M.L., Faeder, J.R., Goldstein, B., Hlavacek, W.S.: A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* 83(2-3), 136–151 (2006)

22. Ruebenacker, O., Moraru, I.I., Schaff, J.C., Blinov, M.L.: Integrating BioPAX knowledge with SBML models. *IET Systems Biology* 3(5), 317–328 (2009)
23. Le Novère, N., Finney, A., Hucka, M., et al.: Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology* 23(12), 1509–1515 (2005)
24. Le Novère, N.: Model storage, exchange and integration. *BMC Neurosci.* 7 (suppl. 1), S11 (2006)
25. Ashburner, M., et al.: Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29 (2000)
26. Mayer, B.J., Blinov, M.L., Loew, L.M.: Molecular Machines or Pleiomorphic Ensembles: Signaling Complexes Revisited. *J. Biol.* 8(9), 81–95 (2009)
27. OWL Web Ontology Language, <http://www.w3.org/TR/owl-features/>
28. Jena Semantic Web Framework, <http://jena.sourceforge.net>