On Data and Visualization Models for Signaling Pathways

Nattakarn Ratprasartporn, Ali Cakmak, Gultekin Ozsoyoglu
Department of Electrical Engineering and Computer Science
Case Western Reserve University
(nxr27, cakmak, tekin) @case.edu

Abstract

Signaling pathways are chains of interacting proteins, through which the cell converts a (usually) extracellular signal into a biological response. The number of known signaling pathways in the biological literature and on the web has been increasing at a very high rate, thus demanding a need for efficient ways of storing, visualizing, querying, and mining signaling pathways. In this paper, first we briefly compare the data modeling and visualization capabilities of existing signaling pathways systems. Then, we present a signaling pathway data model and its visualization that subsumes the existing models.

Our model visualizes a signaling pathway (a) as a nested graph, (b) with explicit location information (e.g., cell, tissue, organelle, nucleus, etc.), and (c) in four abstraction levels, namely, the levels of molecule-to-molecule signaling steps, collapsed sub-pathways, molecule-to-pathway connections, and pathway-to-pathway connections. We model (1) the effects of specific signaling steps, (2) state changes of signaling molecules, (3) various (extensible) structural/physical changes of signaling molecules such as complex formation, dissociation, assembly, oligomerization, di-/trimerization, cleavage and degradation, (4) condensation/hydrolysis signaling steps, and (5) exchanges and translocations as signaling steps. The visualization model gracefully models incomplete information and hierarchical levels of signaling molecules. Finally, we introduce a completely new visualization dimension for pathways, namely, Gene Ontology (GO)-based functional visualizations of pathways. We believe that functional visualizations of pathways provides new opportunities in understanding, defining and comparing existing pathways, and in helping discover new ones.

1. Introduction

Knowledge about the principal mechanisms of signal transduction and regulation mechanisms of individual macromolecules in signaling pathways has been growing at a fast rate, thus demanding a need for (a) efficient ways of storing, organizing, and querying signaling pathway data, and (b) functional visualizations of signaling pathways.

PathCase, a biochemical pathways storage, visualization, and querying tool available on the web [18], presently only deals with metabolic pathways. Here we present PathCase signaling pathways design, namely, (a) a new signaling pathways data model that subsumes the existing data models, and its accompanying visualization model, and (b) the functional template-based visualization of pathways; implementations of both parts are underway.

A signaling pathway is visualized as a nested graph consisting of multiple signaling steps and subcellular compartments, specialized into organisms. We present the visualization of a signaling step based on its effects, such as phosphorylation, ubiquitination, condensation, translocation, etc.

To control visual complexity, we employ a graph model with four abstraction levels. In addition to the traditional molecule-to-molecule signaling pathway visualization, we use (a) *manual subgraph folding*, (b) *folding of all state changes*, and (c) *collapsing of steps into one virtual step*. Moreover, signaling pathways often contain incomplete information or ambiguous information involving hierarchical levels/ states of molecules. A subgraph inside the pathway graph represents such information, and a connection between the subgraph and a node represents incomplete or ambiguous information.

Finally, we introduce a new visualization dimension for pathways, namely, Gene Ontology (GO)-based *functional visualization of pathways*. Intuitively, pathways inherit GO annotations of their gene products, and GO-based visualizations of pathways may provide new opportunities in understanding, defining and comparing existing pathways, and in helping discover new ones.

This paper (full version at [19]) is organized as follows. Sections 2 and 3 present a brief background, overview and comparison of the leading signaling pathway systems. Sections 4 and 5 describe PathCase signaling pathways data and visualization models. Section 6 presents the functional template visualization of PathCase pathways.

2. Background

Cells often communicate by means of *extracellular signaling molecules* which are molecules that carry signals. *Receptors* in cells recognize and react to specific signaling molecules, and *signaling cells* synthesize signaling molecules,

Proceedings of the 18th International Conference on Scientific and Statistical Database Management (SSDBM'06)
0-7695-2590-3/06 $20.00 © 2006 IEEE

which in turn produce a specific response only in the *target cells* that have receptors for the signaling molecules. *Signal transduction* is the process of converting (mostly) extracellular signals into cellular responses. There are also intracellular signaling molecules that mediate the signaling inside the cell, e.g., second messengers. Communication by extracellular signaling usually involves five steps: i) *synthesis*, release of signaling molecule by signaling cell, ii) *transport* of signal to a target cell, iii) *detection* of signal by a receptor protein, iv) a *change* in cellular metabolism/function or development triggered by receptor-signal complex, v) *removal of signal* to terminate further cellular response [1].

Signaling pathways also interact with each other and metabolic pathways. The interaction produces properties that are not seen in isolated pathways, e.g., self-sustaining feedback loops.

3. Brief Comparative Feature Evaluation

This section briefly compares the data and visualization models of four existing signaling pathway systems and one proposal, namely, CSNDB [2, 3], TRANSPATH [4-6] , aMAZE [7-9], PATIKA [10-12], and Fuguda-Takagi proposal [13].

Cell Signaling Networks Database (CSNDB): CSNDB [2, 3] is a database for signaling pathways of human cells, and has a web-based user interface allowing web-based queries. Pathways between or around specified molecules are visualized using an "Expert System", which infers, through hierarchical relationships, existing signaling steps.

TRANSPATH: TRANSPATH [4-6] is a signal transduction pathways database system with a web-based user interface that focuses on pathways involved in the regulation of transcription factors. The visualization tool, called *PathwayBuilder*, provides automated graph drawings at many detail levels. Hand-drawn maps for known pathways are also provided. The PathoSign [12] module in TRANSPATH collects information about defective cell signaling molecules causing human diseases.

aMAZE : aMAZE [7-9] provides the representation, management, annotation, and analysis of information on gene expression, catalyzed chemical reaction, regulatory interaction, protein assembly, metabolic and signaling pathways. Presently, only simple browsing is available through aMAZE LightBench.

PATIKA: PATIKA [10-12] aims to define ontology for representation of signaling pathways, and develops software tools for modeling cellular processes. Currently, PATIKA provides a thin-client web interface for read-only access to the PATIKA database and a stand-alone application for querying, visualizing, editing, and analyzing pathway graphs.

Fuguda-Takagi proposal : Fuguda and Takagi [13] propose a signaling pathway visualization design using compound graphs that are designed to manage incomplete and/or complicated pathway data.

Data models and visualization features of existing systems and *PathCase* are compared in tables 1 and 2, respectively.

Modeling Feature	**C**	**T**	**P**	**A**	**PC**
Hierarchical Relationship					
Superfamily/Subfamily molecule	y	y	n	n	y
Complex/Component molecule	y	y	y	y	y
Parent/Child Tissue	y	n	n	n	y
Location Information					
Organisms	y	y	n	y	y
Tissue	y	y	n	y	y
Cell type	n	y	n	y	y
Subcellular compartment	y	y	y	y	y
Positive/Negative location	n	y	n	n	n
Location for each molecule in a reaction	n	n	y	y	y
Molecule and reaction Attributes					
Molecule types (basic, protein, gene, etc.)	n	y	y	y	y
Molecular states	n	y	y	y	y
Molecule roles in signaling steps (hormone, transcription factor, etc.)	n	n	n	n	y
Molecule roles that each molecule can have	y	y	n	n	y
Effects of signaling steps	y	y	y	y	y
Relationship Entities for reactions, transitions, or interactions	y	y	y	y	y

Table 1. *Data Model Comparisons (C: CNSDB, T: TRANSPATH, P: PATIKA, A: aMAZE, and PC: PathCase)*

Visualization Feature	C	T	P	F	PC
Different shape/color nodes for different types of molecules	n	n	y	n	y
Different shape/color edges for different types of reactions	n	y	y	y	y
Nested graph visualization (a node can contain many nodes)	n	n	y	y	y
Number of detail levels for pathways	n	n	y	y	y
Subnetwork collapsing/expanding	n	n	y	-	y
Feedback regulation	y	y	y	y	y
Organelle visualization	n	n*	y	n	y
Tissue/Cell visualization	y	n	n	n	y
Organism visualization	n	n	n	n	y
Crosstalks between specified signaling pathways	n	y	n	y	y
molecule-molecule connection	y	y	y	y	y
molecule-pathway connection	n	n	y	n	y
pathway-pathway connection	n	n	y	y	y
Old/new state visualization	n	y	y	y	y
Visualization of incomplete information	y	y	y	y	y
Graph can be modified (moving node or edge, layout, etc.)	n	n	y	-	y

** Hand-drawn map*

Table 2. *Visualization Feature Comparisons (C: CSNDB, T: TRANSPATH, P: PATIKA, F: Fukuda & Takagi, and PC: PathCase)*

4. PathCase Signaling Pathway Data Model

This section briefly describes PathCase signaling pathway data model, which contains four main entities: *Molecular Entities*, *Signaling Steps*, *Signaling Pathways*, and *Locations*. *Molecular Entities* represent all molecules in the database. Captured relationships between molecules are hierarchical, and include state, motif, complex/component, and superfamily/subfamily relationships. For example, a molecule can be at a state such as the *unphosphorylated* or *phosphorylated* form of a molecule. Molecules are distinguished by their types with many levels of *Is_a* relationships among *basic molecules*, *gene products (proteins and RNAs)*, and *gene*.

Signaling steps are distinguished by their effects, with each signaling step possibly having many effects. The effect can be of type: *activation, inhibition, assembly, indirect, acetylation, binding, dissociation, condensation, conjugation, cleavage, degradation, dephosphorylation, exchange, expression, farnesylation, glycosylation, hydrolysis, hydroxylation, methylation, oligomerization, phosphorylation, processing, transactivation, translocation, transcription, translation, transregulation,* and *ubiquitination.*

For the molecules involving in a signaling step, PathCase model captures various roles, including: *signaling molecule, signaled molecule, cofactor, enzyme, and inhibitor*. Enzymes can be further specified with roles *hormone*, *cytokine*, *neurotransmitter*, *receptor*, *ion channel*, *effector*, *messenger*, *transcription factor*, and *adapter protein.*

Locations information of the PathCase model captures hierarchical relationships among each type of location, i.e., *organism group*, *organism*, *cell, tissue*, and *organelle*. This information is used (a) in modeling incomplete information, (b) to handle multiple levels of abstractions, and (c) to visualize signaling pathways in a given location, e.g., humans.

All molecules except molecules of type *basic molecules* have specified locations. The locations of basic molecules are left *unknown*.

5. PathCase Visualization Model

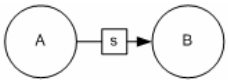
PathCase signaling pathway is visualized as a nested graph. A signaling step is visualized in two ways, (a) *full-form*, which is a multi-edge composed of input and output molecule nodes, interaction edges, and a label in the center, and (b) *simple-form*, which is usually a subview of the full-form, and captures only the signaling and signaled molecules. The signaling pathway graph consists of multiple signaling steps and subcellular compartments, specialized into organisms.

5.1. Signaling Steps

PathCase visualization of a signaling step is based on its effect, which are enumerated below.

Activation, inhibition, indirect:

For signaling steps with effects “activation”, “inhibition”, or “indirect” (and others), the main components are only an input node representing a signaling molecule, and an output node representing a signaled molecule. These steps are in the simple-form, to hide information.



Note: "A" is the signaling molecule, "B is the signaled molecule, and s is the label of the edge from A to B

Figure 1. *Signaling step "s" in the simple form*

Signaling Molecules with State Changes:

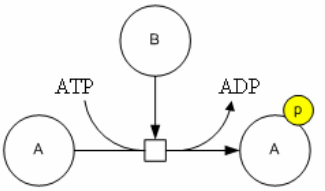


Figure 2. *Protein kinase "B" phosphorylates "A"*

Figure 2 visualizes the *phosphorylation* effect of a step, which causes a state change on the signaled molecule. The visualization issue is then how to represent both the old and the new state of the signaled molecule gracefully. For phosphorylation, the typical visualization adds the letter P to the molecule's visualization. Other effects, such as *acetylation*, *farnesylation*, *glycosylation*, *methylation*, *ubiquitination*, *conjugation*, that cause the addition of molecules can be visualized in the same way as phosphorylation. Figure 3 illustrates ubiquitination.

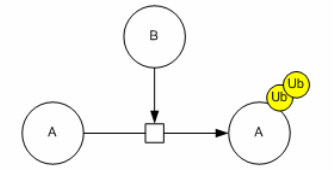


Figure 3. *Ubiquitination*

Effects that cause the removal of molecules like dephosphorylation can be visualized in the reverse direction as illustrated in Figure 4.

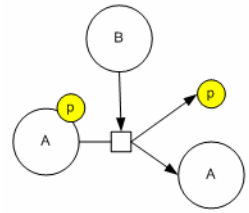


Figure 4. *Dephosphorylation of "A" by "B"*

Signaling molecules as a result of forming/breaking up of a molecule:

Two or more molecules may form a new molecule by chemical or physical bond. Figure 5, which is an example of a *complex formation*, illustrates our choice of visualization in such cases. *Dissociation*, the reverse process of the complex formation, can be visualized in the reverse order.

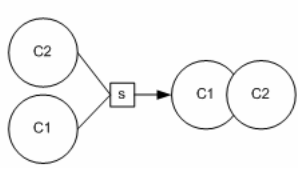


Figure 5. *Signaling step S: forming of a complex*

Assembly, which is the congregation of molecules in a certain subcellular location, can be visualized similar to figure 5, but, in such cases, the location must be specified. *Oligomerization* is another type of formation which can be visualized as in figure 3. For *di-/trimerization*, the new molecule is composed of two/three components. *Cleavage* and *degradation* represent the break-up of a molecule, and can be visualized as a step with one input molecule and one or more output molecules. Example of degradation is the break-up of a protein into amino acids.

Condensation/ hydrolysis:

Condensation forms a new chemical bond and sets water free. In contrast, *hydrolysis* is the splitting of a chemical bond with the consumption of water. We represent these effects as illustrated in figures 6 and 7.

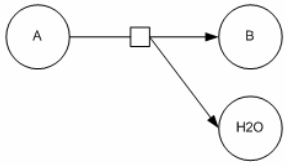
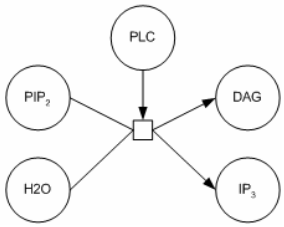


Figure 6. *Condensation*

Exchange:

Some proteins, such as Ras, are active when bound to GTP and inactive when bound to GDP. These types of signaling steps, referred to as *exchange*, can be visualized as shown in figure 8.



Note: *PLC catalyzes the hydrolysis of PIP2 producing two distinct second messengers: DAG and IP3*

Figure 7. *Hydrolysis of PIP2*

Translocation:

For *translocation*, the input and the output of the signaling step stay the same, but they are in different locations. Figure 9 illustrates the translocation of molecule "A". The visualization of translocation is by explicitly visualizing location boundaries.

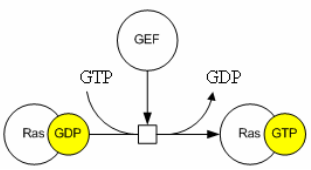


Figure 8. *Conversion to the active GTP-bound of Ras*

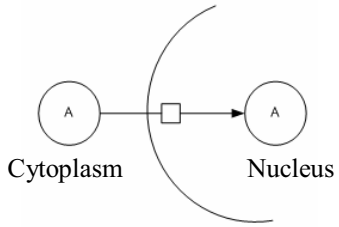


Figure 9. *Translocation of molecule "A" from cytoplasm to nucleus*

Reversible steps

Reversible signaling steps are visualized via bi-directional edges, as shown in figure 10.

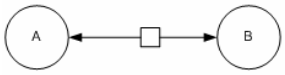


Figure 10. *Bi-directional step*

5.2 Information Hiding

Visualization of pathway graphs can quickly become very complicated; therefore, it is beneficial to visualize pathway graphs at arbitrary depths and multiple abstraction levels in order to reduce the visualization complexity. There are many types of abstractions for a pathway graph. One is *manual subgraph folding*. The user chooses some parts of the pathway graph to be folded into a single node. Another abstraction is the *folding of all state changes* of a given molecule into a single node.

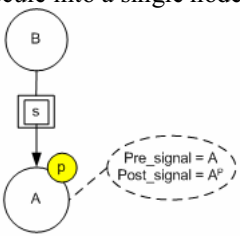


Figure 11. *Abstraction of All phosphorylated states from figure 2.*

In Figure 11, all phosphorylated states of "A" from figure 2 are folded into a single node "A". A dashed line shows a tooltip, which explains all states of "A", when a mouse is positioned over the node "A". Signaling step "s" is of type *activation* and *phosphorylation*. The step label "s" is visualized as a double rectangle to indicate that the real mechanisms of this step are hidden, and the user can view all the details by right-clicking at this label and choosing to unfold the node from the pop-up menu.

Some components in signaling steps with "binding" effects, such as complex formation, can be hidden also. For example, one component of a complex formation of figure 5 can be hidden, as illustrated in figure 12.

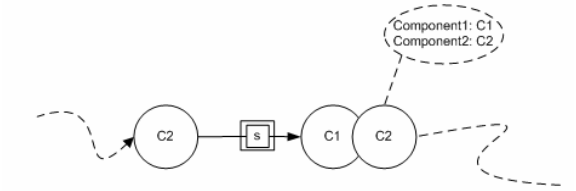


Figure 12. *Abstraction of the component "C1"*

Some common molecules may be hidden. For example, in condensation or hydrolysis, water can be hidden, as shown in figure 13.

The exchanging step of G-protein can also be hidden, as shown in figure 14.

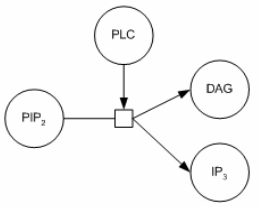
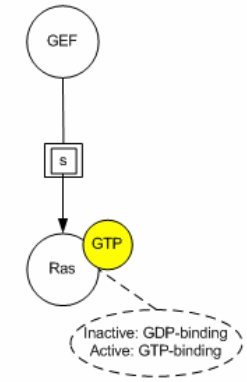


Figure 13. *Hydrolysis from figure 7: H_2O is hidden*



Note: Inactive GDP-binding of Ras in figure 8 is hidden. The step "s" can be expanded to show all details.

Figure 14. *Abstraction of Inactive GDP-binding of Ras*

The last abstraction is the collapsing of multiple steps into one "virtual step". Visualizing many steps between two nodes may not always be needed by biologists. For example, one may want to know only a gene and its product, so the transcription and translation steps can be collapsed into a single step. This step has the effect *indirect* to indicate that there are several steps involved in it. Of course, such a visualization is expandable, in which case the visualization returns to display all the details.

5.3. Incomplete information and hierarchical levels of a molecule

Consider a pathway with incomplete information:

1. $A_{phosphorylated}$ + Protein phosphatase B -> $A_{nonphosphorylated}$ + Phosphate
2. C+D -> CD (A is an inhibitor of this step)

In the above example, in step 2, we know only that A is an inhibitor, but we do not know whether the phosphorylated state or nonphosphorylated state of A is the inhibitor for step 2. This can be resolved by drawing a box covering all of the states of "A" and representing this box as molecule "A", as illustrated in figure 15.

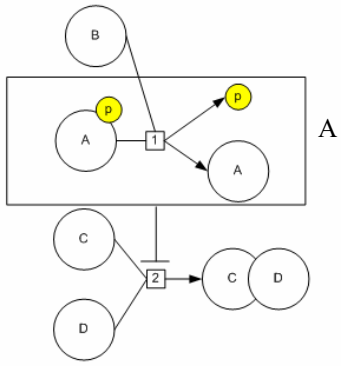


Figure 15. *An abstraction of states for molecule A*

We capture different types of hierarchical relationships among molecules, which include (a) "has-component" for complex/component relationships, (b) "has-state" for multiple-state relationships, (c) "has-subfamily" for family-subfamily relationships, and (d) "has-motif" which represents a molecule and its motif.

One way to visualize multiple molecule hierarchies is to draw them as a tree with multiple types of edges, where different edge types are drawn in different colors. TRANSPATH uses such trees. Figure 16 illustrates such a tree representing family/subfamily relationships. The family "MEKKs" is a parent of subfamilies "MEKK1", "MEKK2", "MEKK3" and "MEKK4". Step 1 is a reaction from "Rac1" to subfamily "MEKK1", and step 2 is a reaction from family "MEKKs" to "MKK4".

However, multiple tree-structured molecule hierarchies may confuse the viewer as some parent/child edges may be misunderstood as signals--even when color-coding is used to differentiate them. We choose an alternative which represent hierarchical relationships of molecules as nested subgraphs (similar to Fukuda-Takagi model [13]), and different subgraph types are represented by different border colors.

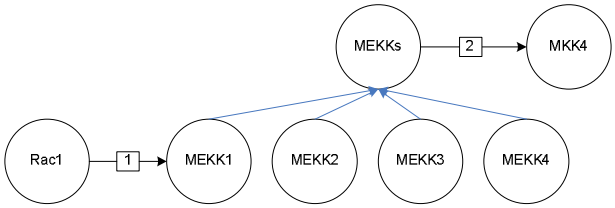


Figure 16. *Representing family and subfamilies relationship as a tree*

Figure 17 illustrates a subgraph representing family/subfamily relationships of "MEKKs" of figure 16. All subfamilies are inside the family node (subgraph). In general, the graph may contain nested subgraphs, one inside the other; for example, MEKK1 may also be a parent of other molecules. This nested graph view may complicate the drawing, in which case the information hiding technique of section 5.2 is used to simplify the visualization.

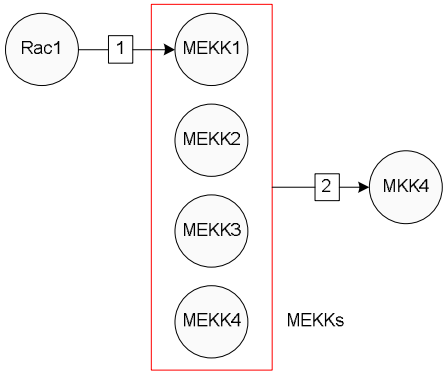


Figure 17. *Representing family and subfamilies relationship as a subgraph*

5.4. Signaling Pathways

The signaling pathways graph contains multiple signaling steps connected together and subcellular location associated with each molecule in each signaling step.

Next we illustrate the visualization enhancements to the TGF-Beta/Smad signaling pathway, using our extensions above:

1. Initiation of the signal occurs when TGF-beta binds to serine/threonine kinase, Type2 receptor.

   TGF-beta + Type2 receptor → TGF-beta/Type2 receptor complex

2. Ligand-bound TGF-beta type2 receptor forms a complex and phosphorylates TGF-beta type1 receptor.

TGF-beta/Type2 receptor complex + Type1 receptor → TGF-beta/Type2 receptor/Phosphorylated Type1 receptor complex

3. Activated TGF-beta type1 receptor then phosphorylates Smad3.

   (Smad3 + Phosphorylated Type1 receptor → Phosphorylated Smad3)

4. Activated Smad3 associates with Smad4.

   (Phosphorylated Smad3 + Smad4 → Smad3/Smad4 complex)

5. The Smad complex translocates into the nucleus. It then associates with DNA binding cofactors and activators or repressors to acts as a transcription factor of target genes.

   (Smad3/Smad4 complex (cytoplasm) → Smad3/Smad4 complex (nucleus))

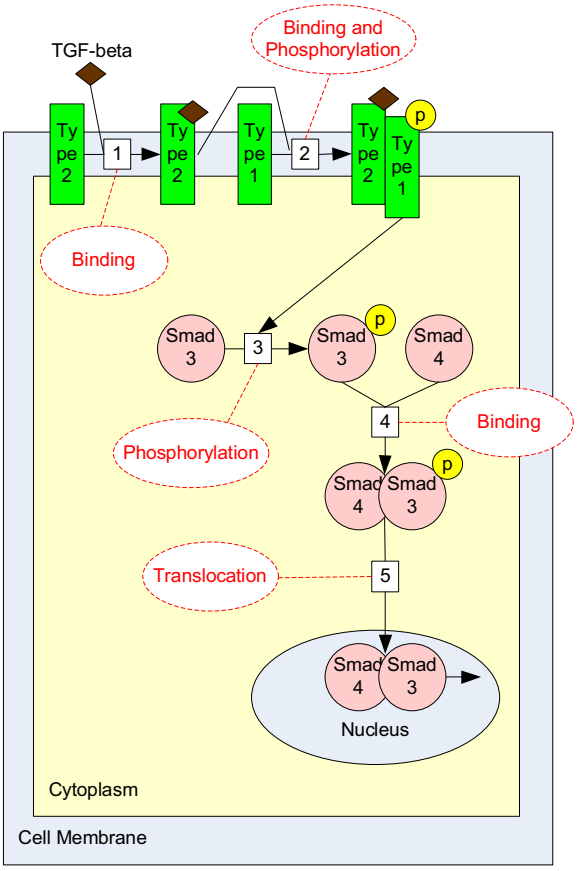


Figure 18. *TGF-beta/Smad Pathway*

Figure 18 illustrates the visualization of the TGF-Beta/Smad pathway using our extensions.

In figure 18, the types of signaling steps 1-5 are "binding", "binding and phosphorylation", "phosphorylation", "binding", and "translocation", respectively. The Smad complex in nucleus will lead to the transcription of a target gene.

The pathway in figure 18 can also be visualized in a more abstract way by hiding all locations and using the signaling steps:

"TGF-beta → TGF-beta receptors",
"TGF-beta receptors → Smad3",
"Smad3 + Smad4 → Smad complex", and
"Smad complex → gene response" (see figure 19).

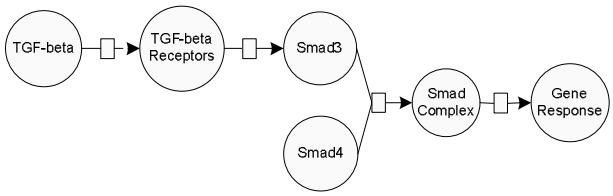
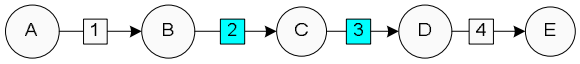


Figure 19. *TGF-beta/Smad pathway in a more abstract level*

The same pathway may occur in multiple organisms and may lead to different effects in different organisms. Therefore, pathway graph visualization should be able to represent the pathway for each organism, and allow for easy visual comparisons. One approach is to visualize the pathway for all organisms in one graph drawing, where signaling steps for each organism are distinguished by using various cues, such as color, shading, or shapes. See figure 20 for an example.



Note: *only highlighted steps 2 and 3 occur in a specified organism.*

Figure 20. *Highlighted step*

A signaling step can involve multiple cells/tissues. For such cases, in the visualization of a pathway, we employ a function that highlights signaling steps that appear in the given tissue/cell.

Two or more pathways can be visualized together to find *crosstalks* between them. However, such visualizations quickly become complicated when multiple pathways are visualized in the same cellular structure.

5.5. Connection Capabilities

In addition to molecule-to-molecule connections, which visualize the connections between signaling molecules and signaled molecules, other types of connections are available in PathCase.

First of all, PathCase provides molecule-to-pathway connections to find pathways related to a given molecule. "Related" means the given molecule

is in another pathway by being a signaling or signaled molecule or playing other roles, such as inhibitor, activator, etc. For example, IFN-gamma signaling pathway induces the expression of Smad7, an antagonist Smad, which prevents the interaction of Smad3 with the TGF-beta receptor in TGF-beta signaling pathway [14].

Another connection type is pathway-to-pathway connection. One pathway connects to another pathway through shared signaling or signaled molecules between the two pathways; or activated molecules in one pathway plays as agonists or antagonists in another pathway.

6. Functional Visualization of Pathways

In this section, we introduce a new visualization dimension for biochemical (metabolic or signaling) pathways, namely, Gene Ontology (GO)-based *functional visualization of pathways*. First, we briefly discuss Gene Ontology, which is a *controlled term vocabulary* describing the central attributes of genomic entities, *i.e.*, genes and gene products [15]. GO contains about 20,000 terms/concepts which are organized in a hierarchical manner through *is-a* and *part-of* relations between the concepts (To query and visualize GO, we have built the Ontology Viewer [16]). In the context of Gene Ontology, as central figures in almost every biological process, genes and gene products have always been the units of annotation. This has proved to be beneficial in that major characteristics of genes/gene products are defined within a common, standardized language. Nevertheless, what makes gene products interesting and, hence, worthwhile to explore for researchers, is the fact that they often take roles as part of a network of indirectly interacting gene products, i.e., biological pathways which define the steps of activities that are essential for organisms. Therefore, it seems intuitive for pathways to inherit GO annotations of its building blocks, gene products. In this regard, visualization of pathways within the context of Gene Ontology categories may provide new opportunities in understanding, defining and comparing existing pathways, and in helping discover new ones. In the remainder of this section, we consider only the molecular function GO subontology; generalizations to other subontologies are future research.

We now describe how we model functional visualizations of pathways. Once all the gene products taking roles in a pathway are annotated with the corresponding GO categories, the pathway may be considered as a sequence of GO category blocks (at different levels of granularity) as illustrated in figure 21 below. Thus, we model a pathway as a network of GO categories, each of which is associated with a set of steps of the pathway. We then construct a generic template, called *visual functionality template* for a class of pathways, which defines the functional characterization of pathways in the given class. Functional visualization models may help biologists in pathway construction to predict components of pathways. Obviously, the extent to which a functionality template would be beneficial in guiding a researcher to a focused set of enzymes totally depends on the depth of the template GO categories within the GO hierarchy. To illustrate the modeling and visualization process, we give an example.

Example 1. In KEGG pathways database [17], there are several pathways listed under the category of "Lipid Metabolism". One can annotate each of the listed pathways under this title, and generate a new GO-category-based version of each pathway by replacing the reactions with GO categories (at different levels of the GO hierarchy). We refer to the resulting graph a *GO-Annotated pathway*. Such a visualization maybe used in several ways.

i) New Pathway Discovery: One may attempt to gain new insights by viewing such graphs in a pathway category. For example, we may merge the constructed GO-annotated pathway templates into a single network of GO categories that represents all the pathways in Lipid Metabolism; call this a *visual functionality template representing the Lipid Metabolism*, which is displayed in figure 22. Once such a template is constructed, it can be used as a blueprint to guide researchers in the process of building a new pathway that is known to be listed under the title of lipid metabolism.

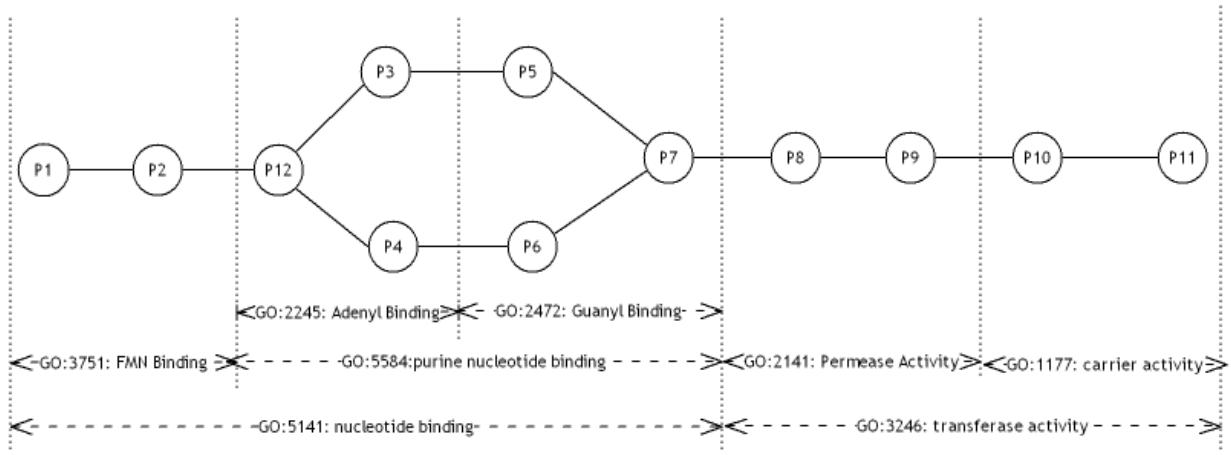


Figure 21. *GO annotation of a hypothetical pathway involving gene products*

ii) Missing Step or Gene Detection: A template such as the one in Figure 22 may help a wet-lab researcher working on a particular pathway by providing guidance on (a) the steps required within a specific pathway, or (b) the kind of genes/gene products needed to be focused in a specific organism (as candidates for a role based on the functional characterization of the pathway represented by the constructed functionality template). We give an example.

Example 2. According to the template in figure 22, assume that a biologist has been able to verify the genes/gene products which perform ion transporter activity and hydrolise activity for a specific lipid metabolism pathway. Then, by looking at the lipid metabolism template, the biologist infers that it is likely that there is a step or a set of steps that involve(s) receptor binding between ion transporter and hydrolise steps. To this end, the biologist may focus on locating smaller set of genes/gene products that are known to have a role in receptor binding.

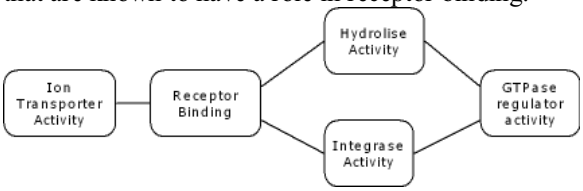


Figure 22. *Hypothetical GO Functionality Template for Lipid Metabolism Pathways*

iii) Automated Pathway Categorization: Automated pathway categorization involves template creation not only for pathway categories, but also for a particular pathway to be categorized. Functionality templates for a pathway can be directly created from GO annotations of a pathway. For instance, figure 23 displays three possible templates for the hypothetical pathway of figure 21. For a given pathway, it is possible to create a set of templates by using the *is-a* hierarchies of the GO hierarchy. Being able to organize pathways into a particular hierarchy helps researchers to browse and get a better grasp of pathways based on their vicinities with other pathways. Moreover, there may be more than one way of categorizing pathways according to different considerations or points of interest. We give an example.

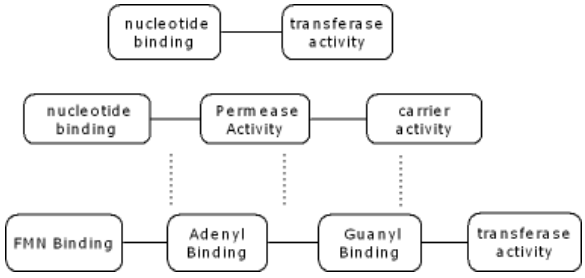


Figure 23. *Sample templates for the hypothetical pathway of figure 1*

Example 3. Consider figure 24. Assume that, for three (hypothetical) categories, the functionality templates are created as shown on the left side of figure 24. Then, the templates for each category are compared to the templates of the given pathway (displayed on the right of figure 24). Each matching can be scored by utilizing the GO hierarchies, and the pathway is classified under the best matching title.

7. Conclusion

We have presented the signaling pathway and Go-based visualization models of PathCase Database System [18]. The features presented in this paper are presently being implemented to PathCase.

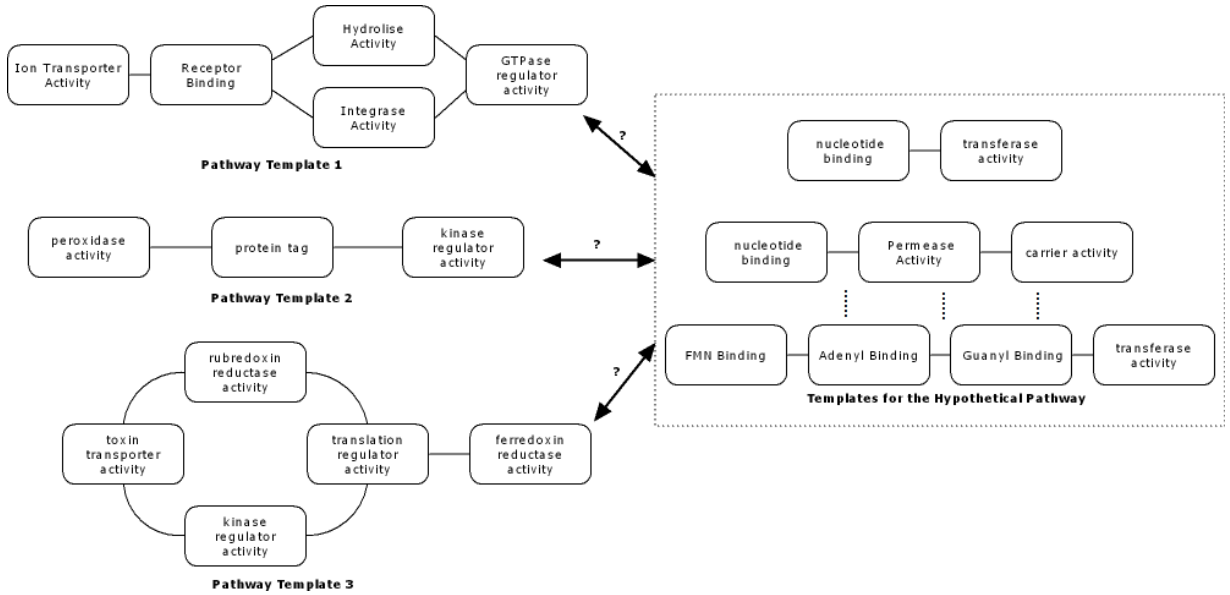


Figure 24. *Template Matching for Pathway Categorization*

8. Acknowledgment

This research is supported by the National Science Foundation grant DBI-0218061, a gift from the Charles Wang Foundation, and Microsoft Equipment donation.

9. References

[1] Lidish, H., Berk, A., Zipursky, L., Matsudaira, P., et al., Molecular Cell Biology, ,W. H. Freeman, 2000.

[2] Takai-Igarashi, T., "Guide to the Cell Signaling Networks Database", *Trends in Glycoscience and Glycotechnology*, 11(60):201-210, 1999.

[3] Takai-Igarashi, T., Kminuma, T., "A Pathway Finding System for the Cell Signaling Networks Database", *Silico Biol*., 1, 129-146, 1999.

[4] TRANSPATH, available at http://www.gene-regulation.com/pub/databases.html

[5] Krull, M., Voss, N., Choi, C., Pistor, S., et al., "TRANSPATH: An Integrated Database on Signal Transduction and a Tool for Array Analysis", *Nucleic Acids Research*, 31(1), 2003.

[6] Krull, M., Pistor, S., Voss, N., Kel, A., et al., "TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations", *Nucleic Acids Research*, Jan 1, 34 (Database issue):D546-51, 2006.

[7] aMAZE, available at http://www.amaze.ulb.ac.be

[8] van Helden, J., Naim, A., Mancuso, R., et al., "Representing and analysing molecular and cellular function using the computer", *Biol Chem* 381(9-10):921-35, 2000.

[9] Lemer, C., Antezana, E., Couche, F., Fays, F., et al., "The aMAZE LightBench: a web interface to a relational database of cellular processes", *Nucleic Acids Research,* 32, D443-D448, 2004.

[10] PATIKA, available at http://www.patika.org

[11] Demir, E., Babur, O., Dogrusoz, U., Gursoy, et al. "PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways", *Bioinformatics*, 18(7): 996-1003, 2002.

[12] Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., et al., "An Ontology for Collaborative Construction and Analysis of Cellular Pathways", *Bioinformatics*, 20(3): 349-356, 2004.

[13] Fukuda, k., Takagi, T., "Knowledge Representation of Signal Transduction Pathways", *Bioinformatics*, 17(9):829-837, 2001.

[14] Ulloa, L., Doody, J., Massague, J., "Inhibition of Transforming Growth Factor-Beta/Smad Signaling by the Interferon-Gamma/Stat Pathway", *Nature*, 397(6721): 710-3, 1999.

[15] The Gene Ontology Consortium, "The Gene Ontology (GO) database and informatics resource", *Nucleic Acids Research*, 32, D258-D261, 2004.

[16] CaseMed Ontology Viewer, available at http://nashua.case.edu/termvisualizer

[17] KEGG, available at http://www.kegg.com

[18] PathCase, available at http://nashua.case.edu/pathways.

[19] Ratprasartporn, N., Cakmak, A, Ozsoyoglu, G., "PathCase Signaling Pathway and GO-Based Functional Template Visualization", Tech. Report, CWRU, 2006.