



Pathway studio—the analysis and navigation of molecular networks

Alexander Nikitin, Sergei Egorov, Nikolai Daraselia and Ilya Mazo*

Ariadne Genomics, Inc., 9700 Great Seneca Hwy, Rockville, MD 20850, USA

Received on December 6, 2002; revised on April 7, 2003; accepted on April 29, 2003

ABSTRACT

Summary: PathwayAssist is a software application developed for navigation and analysis of biological pathways, gene regulation networks and protein interaction maps. It comes with the built-in natural language processing module MedScan and the comprehensive database describing more than 100 000 events of regulation, interaction and modification between proteins, cell processes and small molecules.

Availability: PathwayAssist is available for commercial licensing from Ariadne Genomics, Inc. The light version with limited functionality will be available for free for academic users at www.ariadnegenomics.com/downloads/

Contact: mazo@ariadnegenomics.com

Information about protein function and cellular pathways is central to the system-level understanding of living organism. This knowledge is scattered throughout numerous scientific publications. The need to bring the relevant information together calls for software systems to organize and study pathway data.

PathwayAssist is a Windows desktop application developed for navigation and analysis of molecular networks. It is written in C++ and runs under Windows ME, 2000 and XP. The application uses Jet engine as a back-end to store data, but can connect to other databases that support ADO or ODBC access (e.g. MySQL, Oracle). In addition, there is a second data abstraction layer implemented as COM interfaces to allow for the accommodation of different database schema.

PathwayAssist comes with a database of molecular networks automatically assembled from scientific abstracts. It contains more than 100 000 events of regulation, interaction and modification between proteins, cell processes and small molecules. The database has been compiled by the application of the text-mining tool **MedScan** to the whole PubMed.

MedScan preprocesses input text to extract relevant sentences, which are subjected to natural language processing.

The preprocessing step uses a manually curated dictionary of synonyms to recognize biological terms. Sentences that do not contain at least one matched term are filtered out. The natural language processing kernel deduces the syntactic structure of a sentence and establishes logical relationships between concepts. Finally, the results are matched against the functional ontology to produce the biological interpretation.

PathwayAssist enables researchers to create their own pathways and produce publication quality pathway diagrams (Fig. 1). For visualization purposes, pathways are represented as a graph with two types of nodes. The nodes of the first type are reserved for proteins, small molecules and cellular processes. The nodes of the second type (controls) represent events of functional regulation, chemical reactions and protein–protein interactions.

This data model aligns well with data representations of existing databases: TransPath (Schacherer *et al.*, 2001), eMaze (van Helden *et al.*, 2001) etc., and is generic enough to capture the wide range of phenomena. In particular, protein interactions can be shown as non-directed links; regulatory events are displayed as arrows, with effects being shown as ‘+’ (activation) or ‘–’ (downregulation), and mechanism of regulation (transcriptional, protein modification, etc.) being translated into color or shape of control nodes. In the case of biochemical pathways, links have different semantics and show the flow of reactants in and out of reaction.

PathwayAssist can import data from protein–protein interaction databases (BIND; <http://www.bind.ca>), (DIP; <http://dip.doe-mbi.ucla.edu/>), as well from metabolic databases (KEGG; <http://www.genome.ad.jp/kegg/>). For relationships coming from literature, mouse over the link displays the original sentence. Reference back to the PubMed abstract(s) is also supported.

To enable data analysis, the following tools are available:

Search—find and display a list of objects based on a name or a keyword.

Expand—searches the database and displays objects functionally linked to a selected node. By alternating

*To whom correspondence should be addressed.

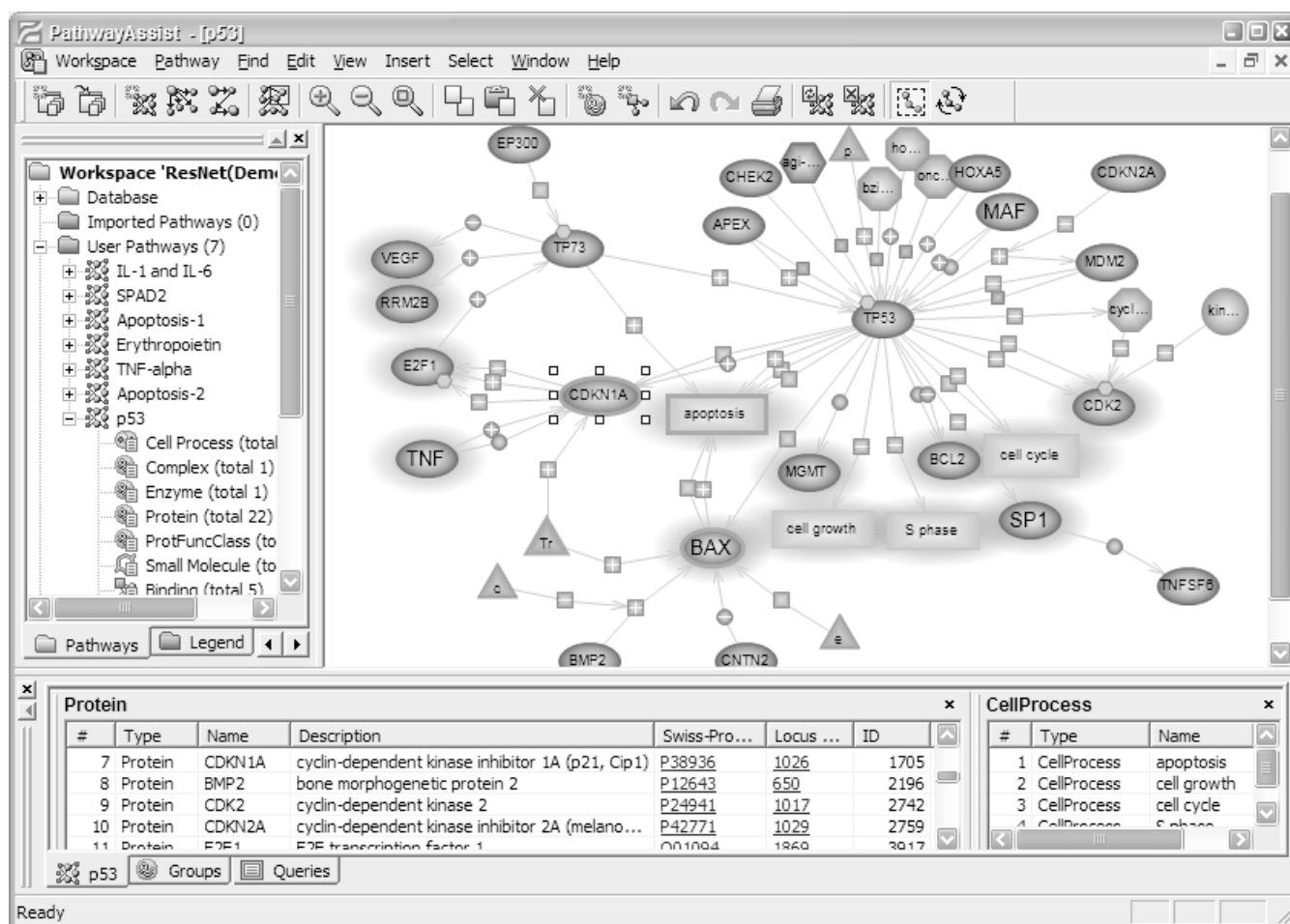


Fig. 1. PathwayAssist graphical interface.

expand and filtering options, users can browse through the database building their favorite pathways.

Build Pathway—finds set of links between two or more nodes by searching for the shortest path in the total network of all links in the database. This tool assists in finding regulatory paths between two objects.

Find common nodes—searches for common targets or regulators for the group of molecules. This tool as well as **Build Pathway** can find functional links between proteins in the lists imported from other programs (i.e. gene expression clusters).

Internet-friendly applications allow you to access data from public repositories, such as PubMed. Protein names stored in the database can be linked to HUGO, Locus Link, GenBank and SwissProt.

Scientific community has developed over the years certain informal standards for visualizing pathways and cell-signaling networks. PathwayAssist uses a proprietary graph

visualization engine to allow for the following visual features:

- Defining shape, size and color of nodes and drawing style of arrows.
- Selecting groups of nodes and highlighting them by color background.
- Anti-aliasing for smooth and 3D-like look of graphical features.
- Using cell object images as a background 'wallpaper' for pathways.

The publication quality drawing style of PathwayAssist is coupled with the set of layout options borrowed from professional graph analysis toolkits:

- Automatic layout employs the version of force-directed algorithm. It is trying to distribute graph evenly and minimize the number of intersections.

- Manual layout is done by drag-and-drop and move of nodes, selected sub-graphs and colored groups.
- Adding and removing nodes can be used to configure the map. The viewer stores the history and supports undo/redo functionality.
- Designing new pathways by copy/paste or manual editing.

To summarize, we believe that the proprietary graph engine provides for the more adequate presentation of pathways than off-the shelf graph toolkits used by other pathway software. At the same time, an advantage over pure pathway editors and systems using static images is in the formalized data storage and support of navigation tools.

The search tools are complemented by the large database of networks. Although less precise than curated sources it has

more coverage and is well poised for finding relations between molecules of interest.

Finally, the unique feature of PathwayAssist is MedScan that can translate keyword searches of PubMed into pathway diagrams, thus creating snapshot of information available in selected abstracts.

REFERENCES

- Schacherer,F., Choi,C., Gotze,U., Krull,M., Pistor,S. and Wingender,E. (2001) The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, **17**, 1053–1057.
- van Helden,J., Naim,A., Lemer,C., Mancuso,R., Eldridge,M. and Wodak,S.J. (2001) From molecular activities and processes to biological function. *Brief Bioinform.*, **2**, 81–93.