PATIKA: An informatics infrastructure for cellular networks

C. Aksay, A. Ayaz, O. Babur §, C. Bilgin, A. Cetintas, A. Civril, R. Colak, G. Cozen, E. Demir §
U. Dogrusoz §, †, Z. Erson, O. Gerdaneri, E. Giral, G. Gulesir, G. Nisanci, O. Sakarya, H.Yildirim

Center for Bioinformatics and Computer Engineering Department, Bilkent University, Ankara 06800, Turkey
§ Presenting authors † Corresponding author

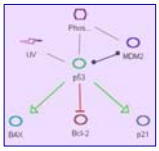
THE PATIKA PROJECT aims to develop methods and software tools for effective analysis of complex biological data at a functional level, consisting of following work-packages:

- Define an **ontology** for a comprehensive representation of cellular pathways.
- Develop **software tools** and construct an associated database using this ontology and provide an effective environment for pathway data integration, storage, access, visualization and analysis.
- Design methods for **automatic population and annotation** of the pathway database.
- Design methods for effective, **advanced querying** of the pathway database.
- Design methods for **inferring pathway activity** using temporal data such as gene expression data.
- Develop techniques for **effective visualization** of pathway and gene expression data.

Ontology

We define an intuitive, comprehensive, uncomplicated representation of cellular networks.

Basics

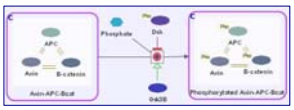


Bioentities: actors of the cellular events; genetic (e.g., DNAs, proteins), chemical (e.g., ions), or physical (e.g., heat).

Bioentity Interactions: high level, imprecise relations: protein-protein interaction, transcriptional regulation or generic.

States: different forms of Bioentities via chemical modification (acetylated protein), localization (cytoplasmic ion), aberration (mutant gene), homomerization (dimers), etc.

Transitions: changes that states undergo.

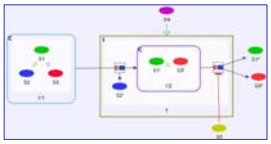


Interactions: relations of states with transitions such as substrate, product, activator and inhibitor.

Molecular Complexes: Non-covalently bound clusters of molecules behaving as a single state.

Cellular compartments: part of the model.

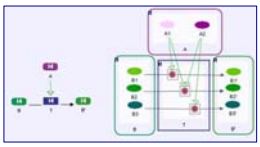
Incomplete Information

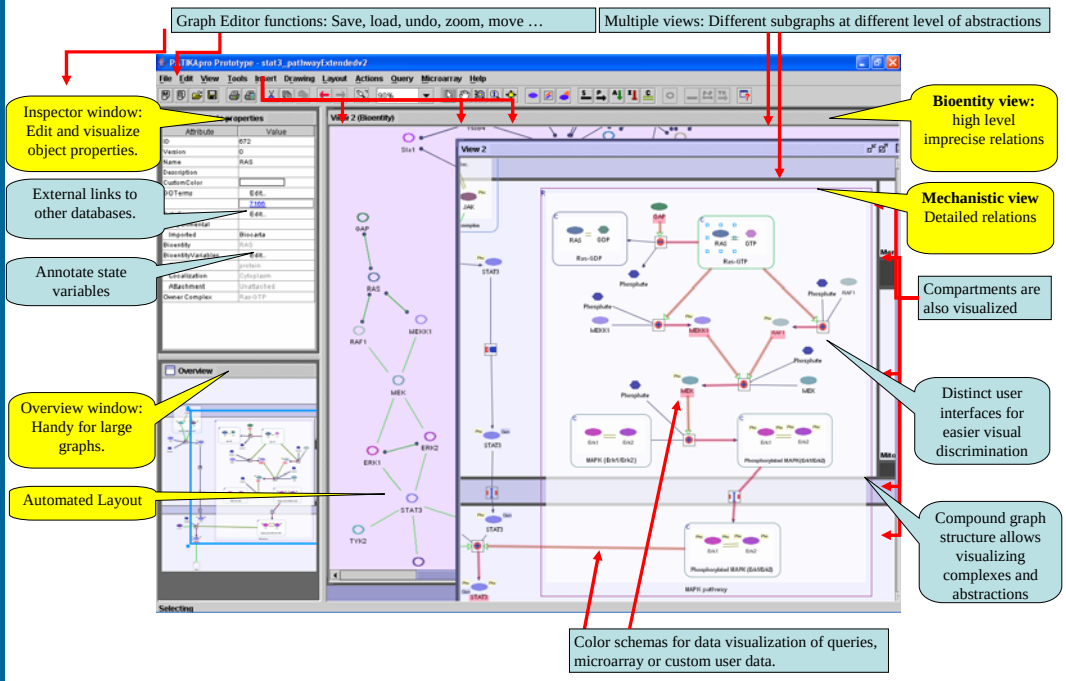


Since the data on cellular processes is incomplete, different levels of information may be available for certain events. On the left, it is unknown whether *S4* activates either of two transitions.

Homologies

B is transformed into B' by activation of A. In the actual case there are two A homologs, three B homologs and three B' homologs.



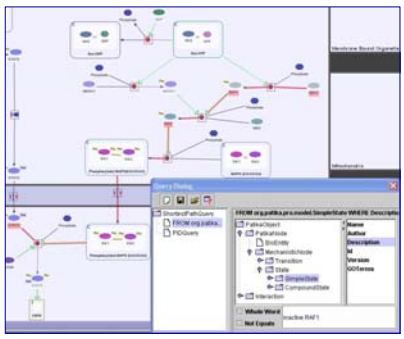


Software

A client/server architecture to provide access to PATIKA database through a state-of-the-art visual pathway editor has been implemented in pure Java™.

Querying

Advanced, graph theoretic queries may be performed through specialized GUIs, including queries by value or ID, neighborhood and shortest path querying.



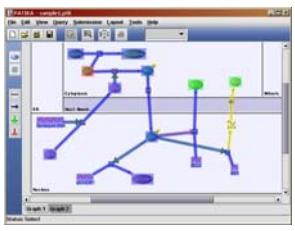
Shortest path from inactive RAF1 *to* singly phosphorylated STAT3.

Multi-User Environment

Collaborative construction and concurrent modification issues are also addressed. While a user is working on a pathway locally, others might change its topology or properties in the database.

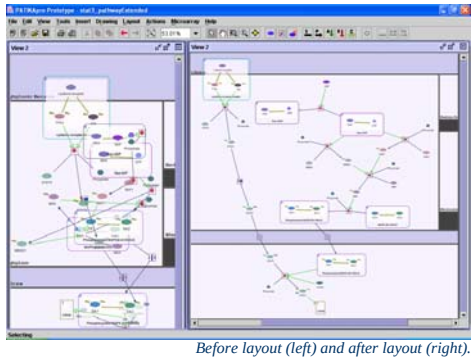
Checks for up-to-date status of graph objects result in each graph object being color-coded with respect to its status:

Blue: *Up-to-date*
Red: *Out-of-date*
Yellow: *Local*
Green: *Locally Modified*



Automated Layout

Specialized algorithms for layout of cellular pathways produce aesthetically pleasing drawings.

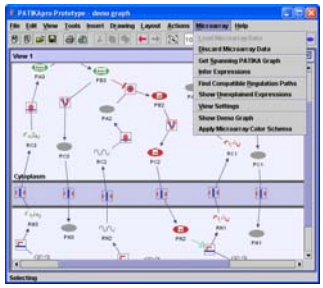


Before layout (left) and after layout (right).

Gene Expression Analysis

Please visit **POSTER F-67** for details of PATIKA's Microarray Data Analysis Facilities.

Support for analysis of gene expression data including a pathway activity inference method using gene expression data has been implemented.



Previous Contributors
Many people have previously worked on the project, to whom we'd like to thank, including A. Gursoy, R. Cetin-Atalay, M. Ozturk, S. F. Akgul, B. Caskurlu, E.D. Ozkan, C. Gerede, A. Kocatas, E. Karakoc, O. Kurt, Z. Madak, E. Sahin, E. Senel, S. Onay, B. Ozmen