SPIKE User's Manual

Version 1.0

Table of Contents

Introduction

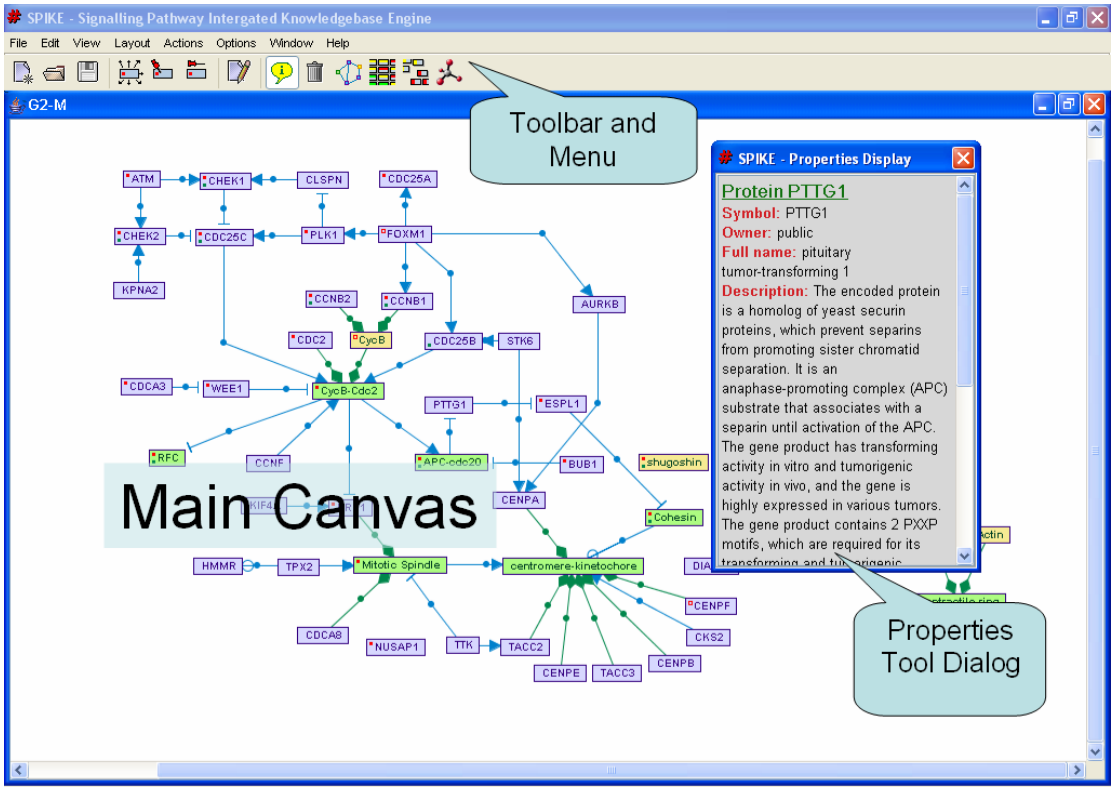
The goals of SPIKE are to help researchers integrate, visualize and interpret information related to biological networks, and to provide the research community with up-to-date, comprehensive, curated data on signal transduction pathways. SPIKE is also designed to enhance functional analysis of data generated by high-throughput functional genomics technologies.

This manual describes how to work with the SPIKE version 1.0 package. It includes instructions on how to view, analyze and define new pathways. We hope you enjoy the system. Please do not hesitate to contact us at spike@post.tau.ac.il if you encounter problems or have suggestions for improvement.

1. Overview

The SPIKE main window, shown in the next figure, is built from the following windows:

1. **Main canvas** – this region holds the graphic representation of the network, as well as any functional analysis data.
2. **Toolbar and menu** – used to control the behavior of the system, and to show/hide the tool dialogs.
3. **Tool dialogs** – used to present information on selected items, and/or to control some of the analysis characteristics.



2. Exploring pathways

To start the application, go to the SPIKE website at http://www.cs.tau.ac.il/~SPIKE, select your location (academic institution), and press "Launch!".

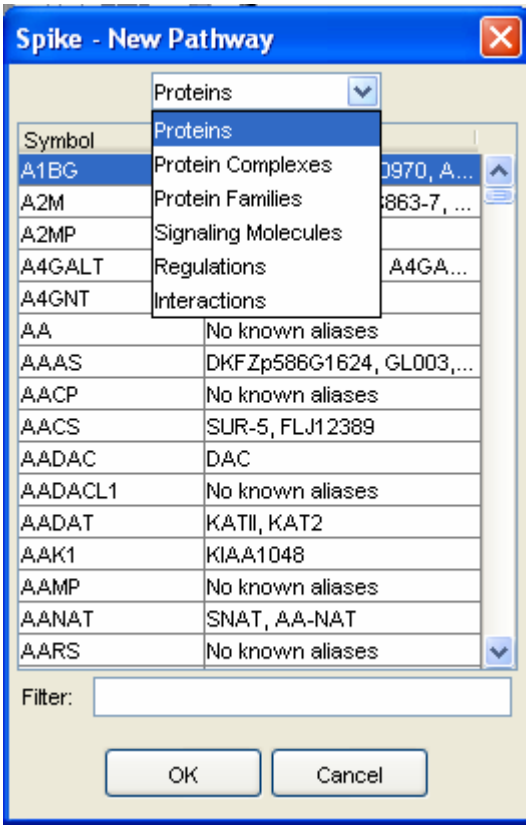
After the SPIKE splash screen shows, the SPIKE main window appears, with the login dialog at its center. If the software does not launch, contact the SPIKE system administrator at your location.

The login dialog allows different users to log in to the system, each with their set of permissions, such as the ability to modify the database, to administrate the system behavior or just to review existing pathways. The default user, "spike_guest", has no password, but can only review existing pathways. The default set of users and their passwords, is described in Appendix B – Spike default system users. For other features you will need to be a registered user. If you are not a registered user, contact your SPIKE system administrator.

Selecting your SPIKE username and password and clicking the OK button logs you into the system and starts the working session.

Creating a new pathway

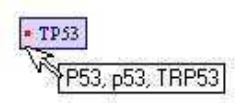
- Select the "New" button in the toolbar ( ). The following dialog appears, allowing you to select the first biological entity (protein, protein complex, protein family, signaling molecule, regulation or interaction) in your pathway.



- Locating the desired entity from the list can be done by typing the first letters of the name in the filter field. Press "OK" or double-click when the desired entity is selected. For example, select the ATM protein.
- The SPIKE canvas now opens a new pathway window with the selected entity.

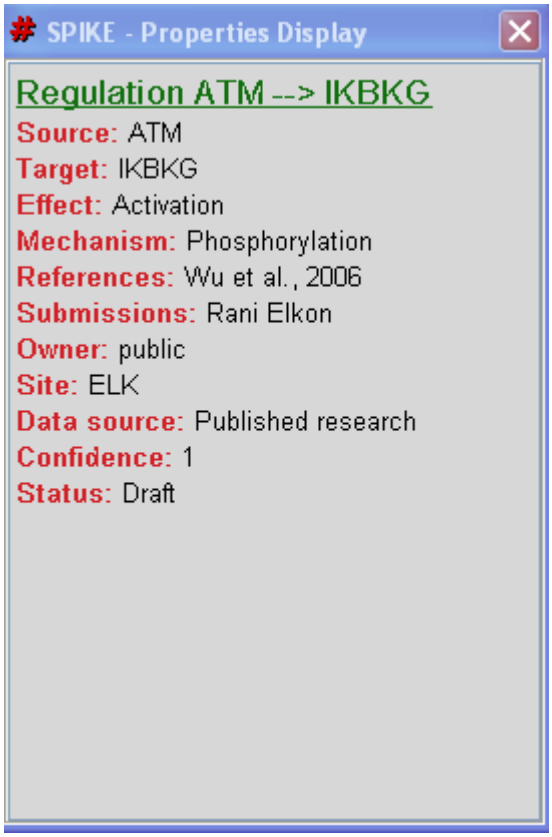
Handling nodes

SPIKE uses the HUGO official gene symbols for node labeling. Placing the mouse pointer on top of a gene's node shows a tooltip tag displayed in the next figure, which indicates its "aliases" (other familiar symbols it is known by).

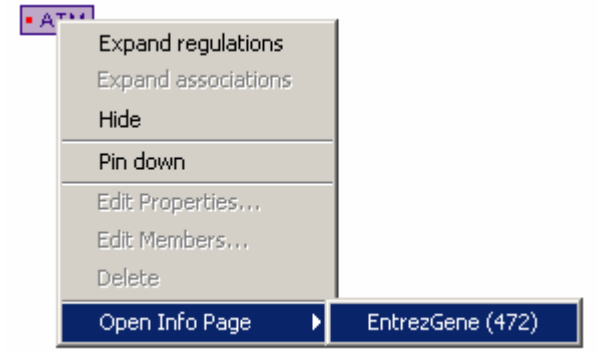


The mouse pointer shows additional information when it is placed elsewhere on gene nodes or on nodes of other entities described next.

In order to see more details on the current entity, you can open the properties tool dialog by selecting the "Properties" button in the toolbar ( ). This tool dialog, like all the other tool dialogs, can be resized and moved according to your preference. After opening the tool dialog, place the mouse pointer on the node of your choice, and its details will be displayed in the dialog.



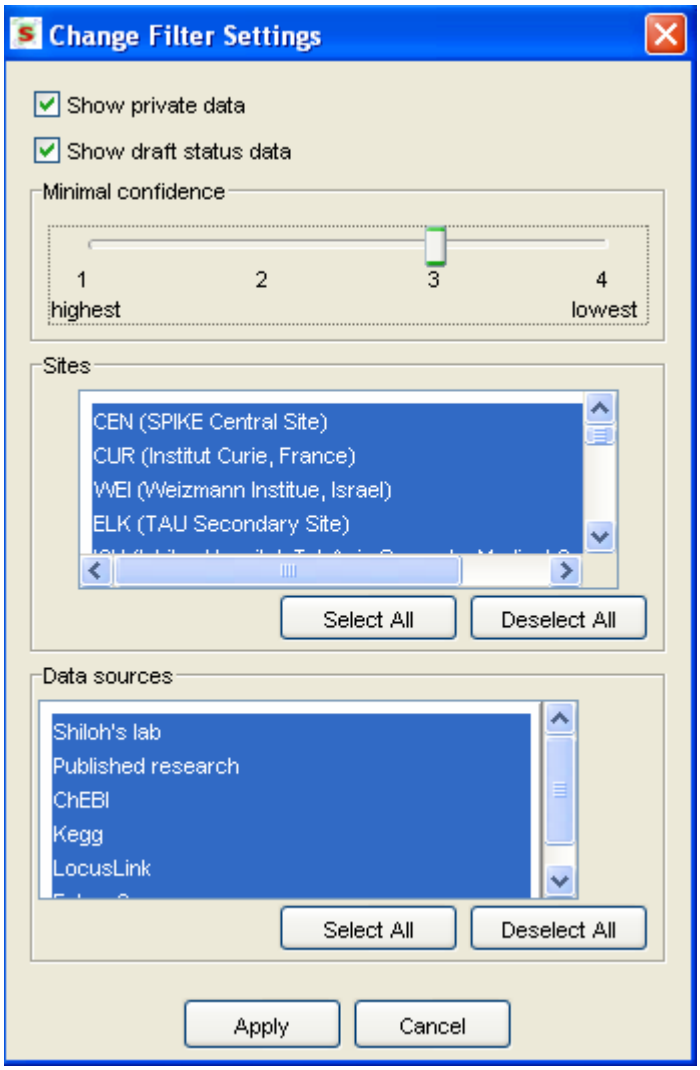
For genes, additional information can be viewed. Place the mouse pointer on top of a gene node and right-click it. This opens a menu, with some basic operations that can be performed on the node.



One of these is "Open Info Page". When selected, it opens up your internet browser with the Entrez Gene link record of the selected gene.

Expanding Pathways

The red dot on the ATM node indicates that it has additional regulations that are currently not shown. Double-clicking on the node shows all the regulations in the database involving this node. **Note that by default only regulations of the highest quality level (i.e., quality level 1) are shown.** To include also data of lower quality, select from the tool menu: "View -> Filter Configuration". In the Filter Setting panel that pops up select the desired minimal quality as shown in the figure below.



For additional filtering options see the Filtering section below.

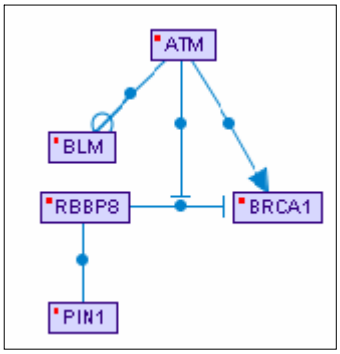
When expanding these regulations, additional nodes appear. They are placed on top of each other, and can be manually dragged to new locations. Another option for placing them is to use the auto-layout mode. Selecting the auto-layout mode by pressing the auto-layout button ( ) will place all the nodes on the canvas so that it would be clearer to view all the regulations.

Nodes can be pinned down (ATM) so that they will not change location even in auto-layout mode. To pin a node down, right-click the mouse on that node, and select "Pin down" from the pop-up menu. To pin down multiple nodes, select them all and press the "Pin selected nodes" button ( ) placed at the tool bar.

The blue edges that appear represent the various interactions and regulations between the entities. SPIKE presents four types of regulatory relations:

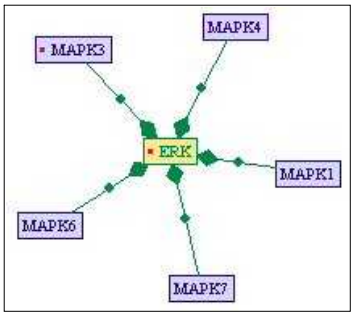
- Activation regulation
- Inhibition regulation
- Unknown regulation
- Protein-protein interaction (un-directed edge).

Activation is represented by an arrow (e.g., the activation of BRCA1 by ATM). Inhibition is represented by a "T" shape (e.g. the inhibition of BRCA1 by RBBP8). When the effect of the regulation is unknown, it is displayed with an empty circle (e.g., it was reported that ATM phosphorylates the BLM protein; however the regulatory effect of this modification is still not clear).

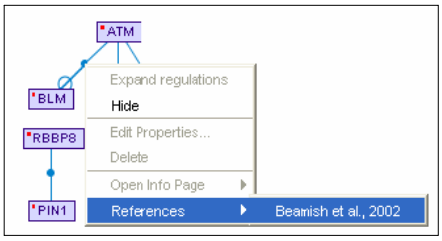


Regulations can indicate regulatory interactions between proteins and other regulations as well. For example, ATM phosphorylates RBBP8, thereby specifically disrupting its inhibition of BRCA1. This is represented by a regulation with ATM as the source and the regulation between RBBP8 and BRCA1 as the target.

Protein complexes and protein families are shown as green and yellow nodes, respectively. Association between complex or family and their members are shown by green edges. A green dot within a node indicates association relations that are currently not shown. Double-clicking on nodes with green dots will expand such associations on the canvas.



More information on regulations and associations is shown in the properties tool dialog, when placing the mouse pointer on the circle in the middle of the edge. References supporting regulations can be viewed similarly to the Entrez Gene web pages, by opening the relevant PubMed references. This will take you directly to the PubMed abstract of an article that reported the regulation.



Saving and Opening pathways

You can save the viewed pathways and their layout for instant viewing at future sessions, by selecting the "Save" button (), and choosing the desired path and filename. For saving all opened pathways, you can select "Save All" () in the menu. In order to open saved pathways, use the "Open" button (), and select the saved pathway. Several pre-built SPIKE maps are posted on our website (http://www.cs.tau.ac.il/~spike/ ; follow the Maps link). Download the spike map files to your pc and open them from the SPIKE package.

Export to an image file

You can export the reviewed pathway to an image file by selecting "File -> Export to image" in the menu, and choosing the desired path and filename. This function is very useful in preparing presentations, articles, etc.

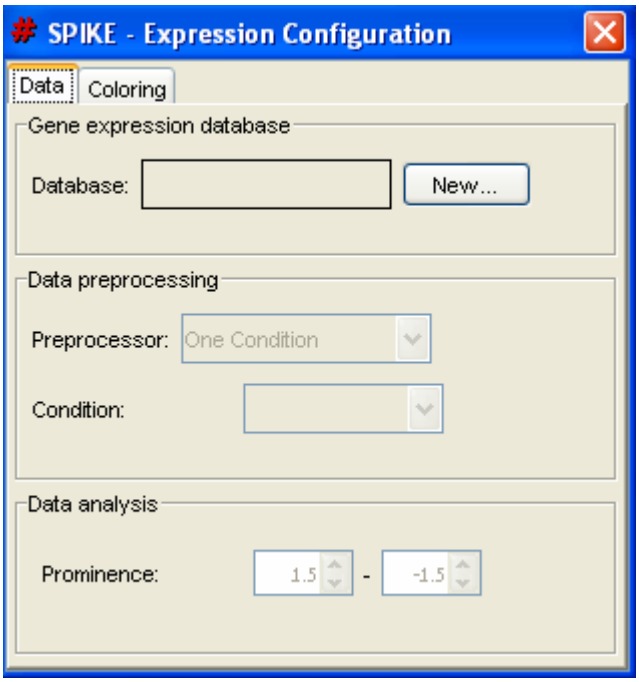
3. Analysis utilities

This section explores how SPIKE enhances the analysis of functional genomics or proteomics datasets.

Superposition of gene expression data

SPIKE enables the superposition of microarray data on top of the signaling maps. SPIKE supports both absolute (e.g., Affymetrix chips) and relative (e.g., cDNA microarrays) gene expression data. For further information regarding the expression data file format, see Appendix A.

To start the analysis, select the "Gene Expression" button on the toolbar ( ). The expression configuration tool dialog opens.

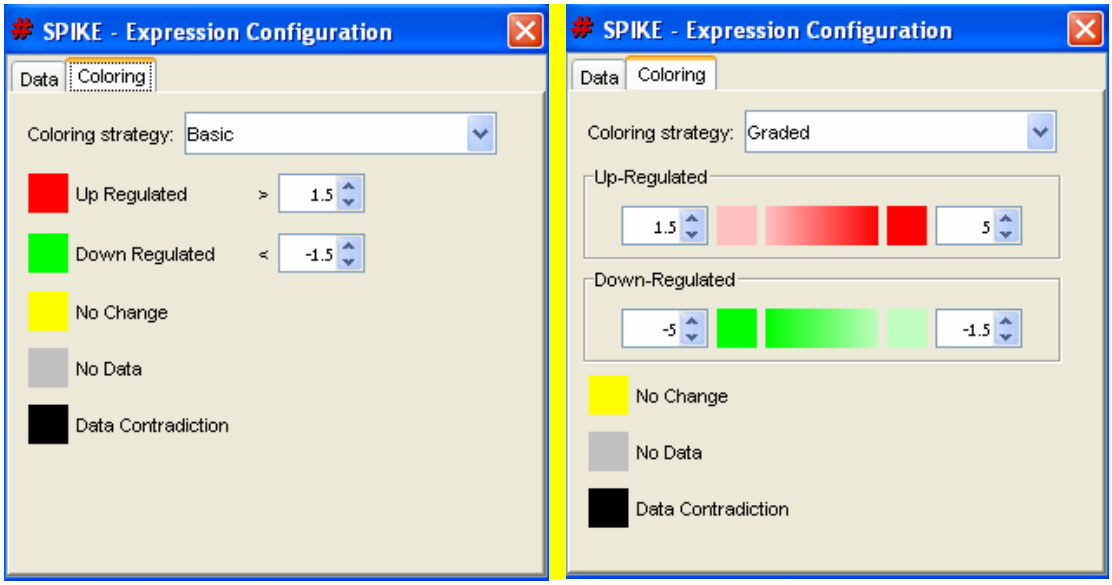


To open the gene expression data file, select "New..." on the tool dialog. Next, select the expression data file and its type (absolute or relative expression values; see Appendix A for further details).

SPIKE shows for each displayed gene the fold change in expression between two selected conditions, called 'test' and 'base'. Initially, all nodes will be colored either gray - indicating that the dataset contains no relevant data concerning them (i.e., their corresponding gene is not present on the microarray used in the experiment), or yellow - indicating no significant fold change. This is because the test condition and the base condition are initially the same.

Change the "Test Condition" box value to another value. Now more colors should appear on the map. Use the tooltip to learn the fold change values of some of the nodes.

SPIKE provides two alternative coloring schemes, 'Basic' and 'Graded'. The coloring of the map can also be changed by setting the fold threshold values or the colors themselves.



For example, with the "Basic" coloring strategy, default values are: genes whose fold change was bigger than 1.5 will be colored red; genes whose fold change was smaller than -1.5 (that is, decrease in expression by more than 1.5) will be colored green; those with values in between will be colored yellow; and those with no available data are colored gray. The "Data contradiction" coloring will be explained below.

The "Graded" coloring strategy enables users to match a range of fold change values to a range of colors, thereby allowing them to differentiate between large and small fold changes according to a scale of their choice. E.g., in the default settings, up-regulated values will be colored from light-red (1.5) to dark-red, getting darker as the fold change increases, till 5.0 (values above 5.0 will be colored as 5.0).

The "Data" pane of the gene expression configuration tool dialog is composed of three main panels:

The first panel deals with the gene expression data sources (files). Note that you may work with multiple expression data files in a single session, switching back and forth between files.

The second panel deals with some preprocessing of the data. In case the file contains absolute values of expression levels (e.g., files generated by experiments that use Affymetrix chips) only coloring of nodes according to a comparison between two conditions (using the "Two Conditions" preprocessor) is possible. When working with a relative-type file (i.e., files that contain relative expression levels obtained by comparing expression in two conditions, e.g., files generated by experiments that use cDNA microarrays), coloring according to the values of a single condition (using the "One Condition" preprocessor) is also possible. The Threshold parameter is available for absolute-type files. Its default value is 40. This means that all values in the file that are less than 40 are considered to be 40. The threshold controls the sensitivity when comparing values one of which is very small.

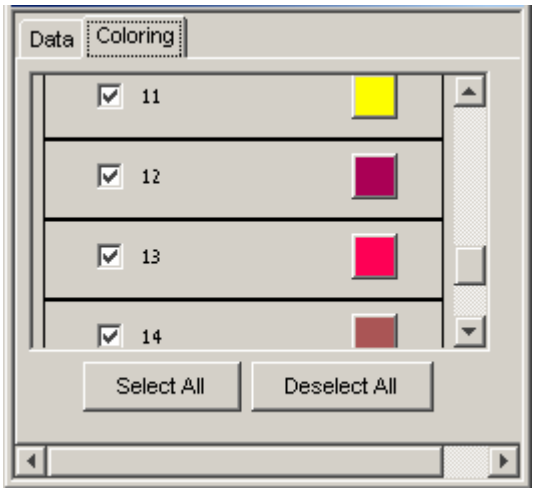
The third panel deals with analysis of the data. In case the chip contains multiple probes for certain genes, the program determines whether or not the data are contradictory. Data of a gene are considered contradictory if there are probes of the same gene with values (fold changes) on both "sides" of the "Prominence" range (that is, for the same gene, some probes indicate that the gene was up-regulated while some other indicate that it was down-regulated). With the default settings, this would mean values both above 1.5 and below -1.5 (Note that these thresholds are independent of the thresholds indicating up-regulation or down-regulation in the "Coloring" pane). For each gene whose values are not contradictory, a single average value is calculated from all its probes and the gene's node is colored accordingly. On the other hand, contradictory nodes are, by default, colored in black (see the "Coloring" pane, and the tooltip of the expression color box would supply all the contradictory values in that case.

Coloring Gene Clusters

SPIKE can superimpose any genes' partition information ("Clustering") on the map (see Appendix A for further details on the clustering file format). The mechanism is similar to the expression superimposition.

To activate the clustering view, select the "Clustering" button in the toolbar ( ). This opens the clustering tool dialog. Use the dialog to select the clustering file and type.

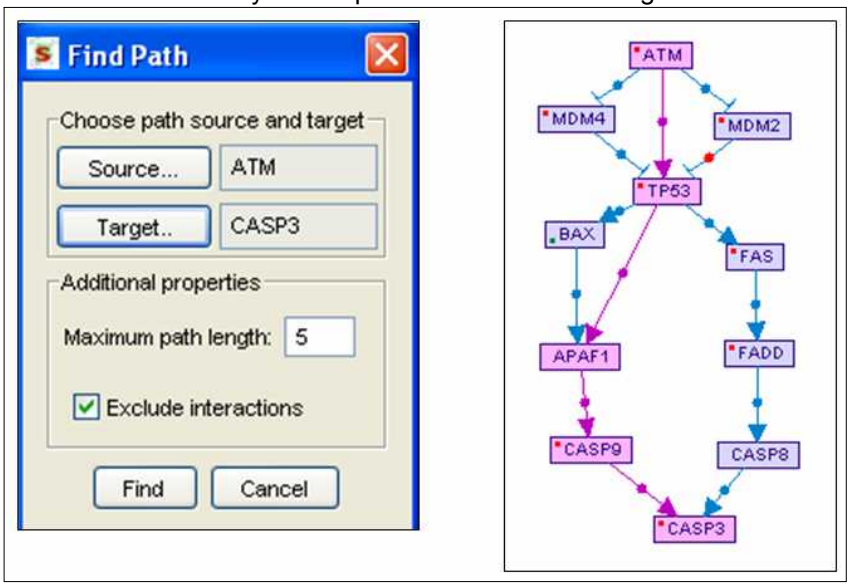
Move to the "Coloring" pane in the clustering configuration dialog. You can select which clusters to color, either with the checkbox on the left of each cluster, or by selecting or de-selecting all of them with the buttons on the bottom. You can also change the colors that are automatically applied to each cluster, by pressing on the color button and changing its settings.



Path Finding

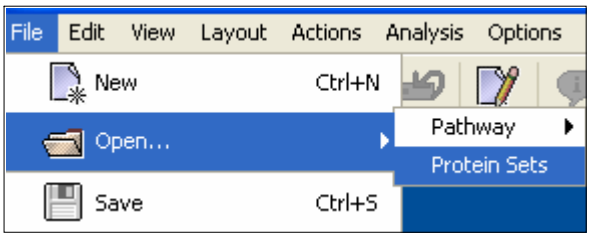
SPIKE implements a path finding utility that finds and displays all directed paths, up to a pre-defined maximal length, that connect source and target nodes selected by the user.

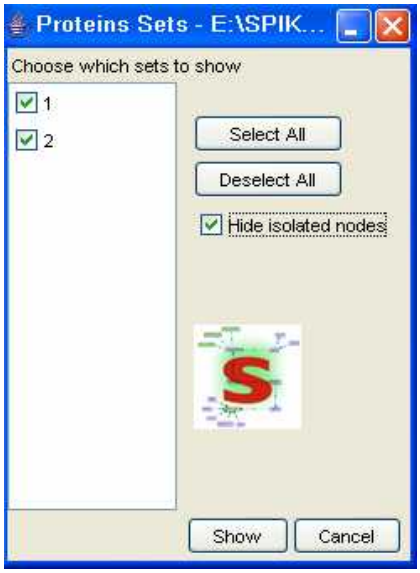
Selecting the Find Path button in the toolbar ( ) opens the corresponding tool dialog in which the user determines the source and target nodes, as well as the maximal length for the directed path connecting them. By default, un-directed edges (that is, protein-protein interactions) are excluded from the analysis. All paths that meet the length constraint are displayed with the shortest one highlighted.



Interconnections within genes/proteins in input target sets

SPIKE enables the user to upload target sets of genes/proteins and display all the connections among nodes in each set. Select 'File' → 'Open' in the toolbar and then "Protein sets" (see Appendix A for details on the gene/protein sets file format). In the tool dialog that opens check the protein sets you wish to analyze. Maps displaying the connections between nodes in each set will be generated (Note: a separate map is created for each set).

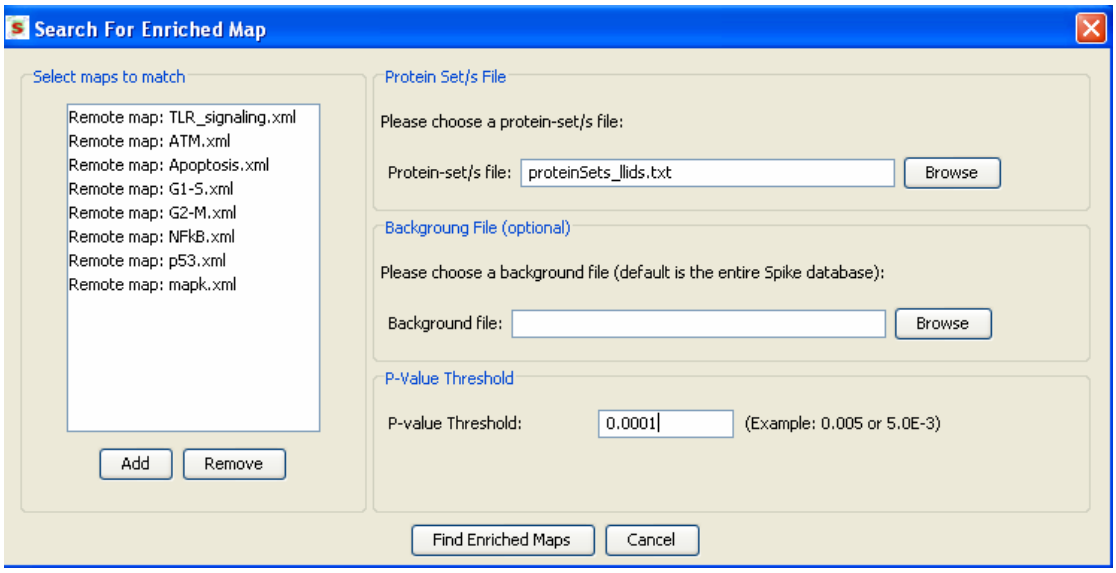




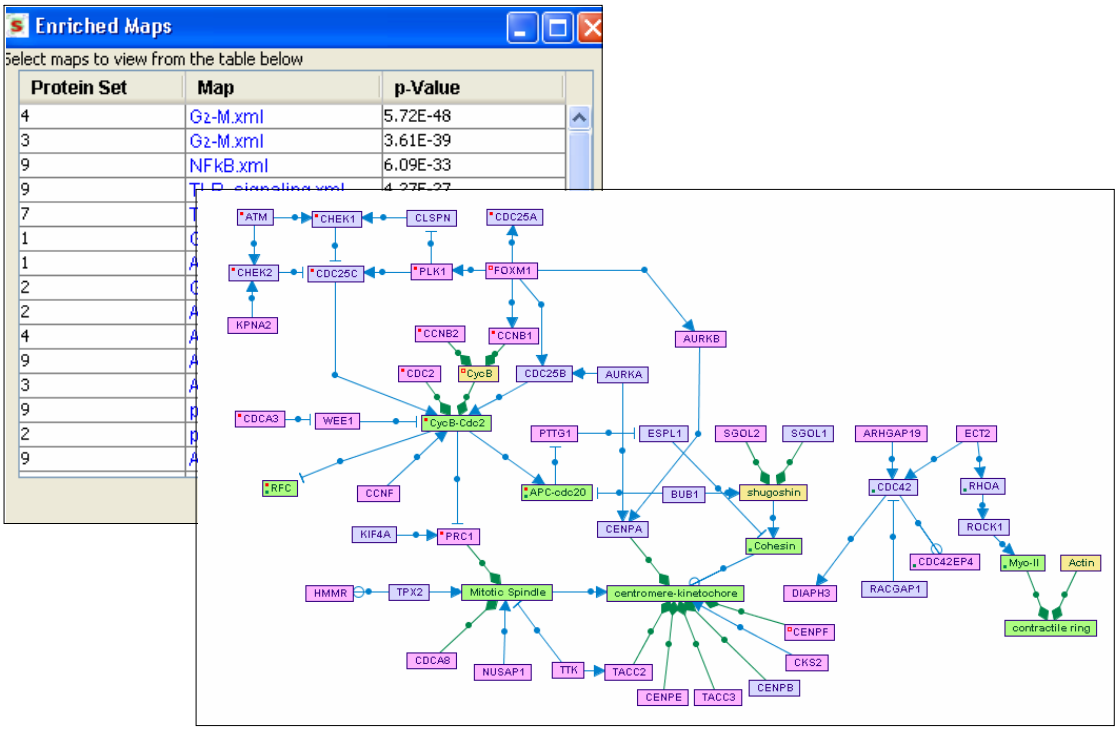
Identification of enriched maps

Given target sets of genes/proteins uploaded by the user, this analysis searches for signaling maps enriched for nodes included in one of the target sets. To run this analysis, select "Analysis" and then "Enriched Pathways".

In the window that pops-up, specify a file that contains the protein/gene sets to be analyzed (see Appendix A for details on gene/protein sets file format), a file that defines the background set ((e.g., all the genes present on a microarray platform; if not specified, the entire genes/proteins in SPIKE DB are used as the background), and a threshold p-value. Enrichment scores are calculated using the hypergeometric distribution.



Maps that are significantly enriched are listed in the results table. Clicking on a map's link in this table opens the enriched map, in which target set nodes are highlighted.



4. Additional display features

This section explores additional features of the SPIKE viewer.

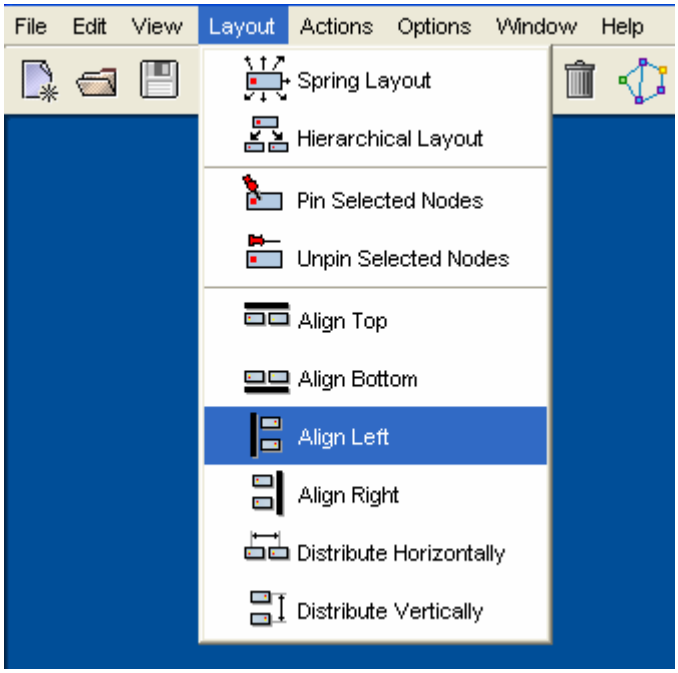
Hidden components

SPIKE allows hiding nodes or edges to reduce clutter in the view. Hiding is done by moving the mouse pointer on top of the node or edge, right-clicking the mouse and selecting the "Hide" option from the pop-up menu. To hide multiple nodes, select them all and press in the toolbar: "Actions -> Hide Selected Components".

Once a node or an edge was hidden, it remains hidden until it is restored from the hidden components list (which can be shown by selecting "Display list of hidden components" ( ) from the toolbar), and pressing the "Restore" button, from the hidden components list tool dialog.

Manual layout and alignment

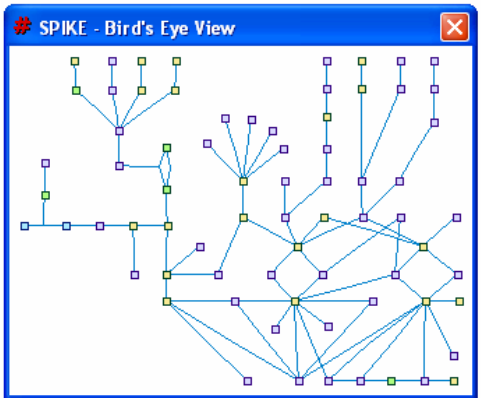
In addition to the automatic layout, SPIKE allows you to layout the map manually. Nodes and edges can be dragged and placed by the user. To improve this operation, the user can align or distribute multiple nodes either horizontally or vertically, by selecting the relevant nodes, and clicking on the alignment/distribution buttons in the layout menu.



Note that you can combine the auto-layout and manual layout. For example, after laying out a pathway manually, you can pin down all its nodes, and then expand some of them. The auto-layout will then lay out only the new nodes.

Bird's eye view

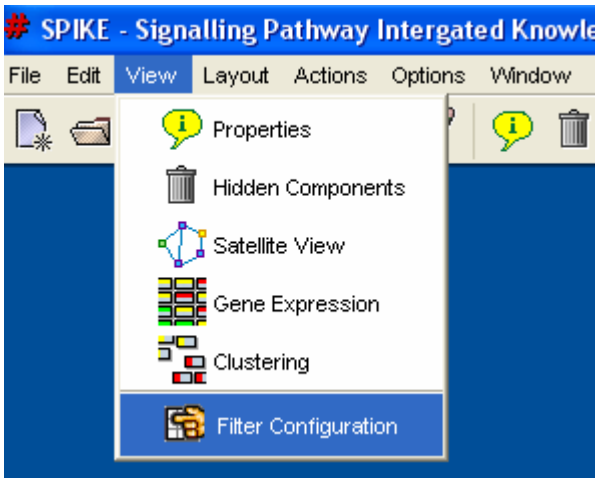
SPIKE allows a "Bird's eye" view of the pathway map, which can be very useful if you have a large, spread out map. This function is activated by pressing () from the menu.



Filtering

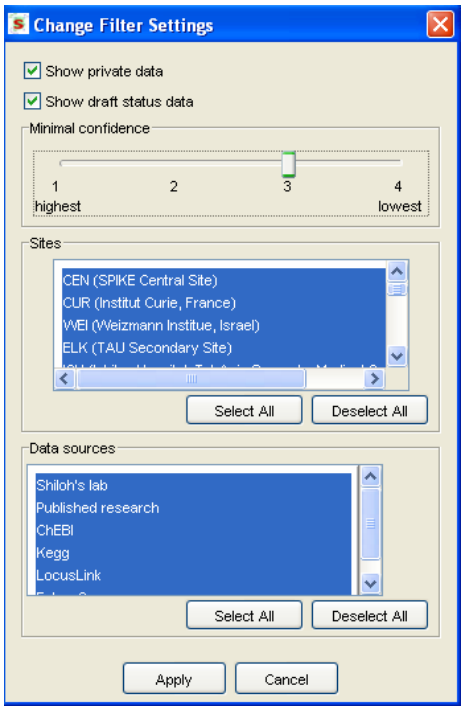
The SPIKE Database contains data at diverse quality levels - from carefully curated data, through data imported en masse from public databases (KEGG, Reactome) to highly noisy data obtained from high-throughput protein interaction screens. Data items in SPIKE database are assigned reliability levels according to their quality. The filtering option allows selection of certain data only according to quality and other criteria.

To configure the filter choose "Filter Configuration" from the View option in the menu:



The parameters for the filtering are:

- Private data - whether you want to view your private data on the map
- Draft data - whether you want to view data which was not curated yet
- Minimal quality - will not show data whose quality level is below the chosen value
- Sites - will show data only from the chosen sites
- Data sources - will show data only from the chosen data sources



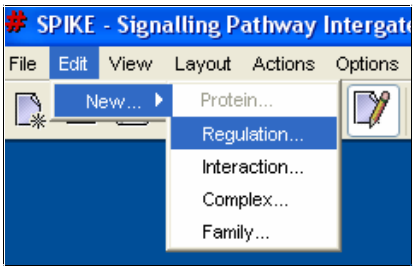
5. Editing the data

This section describes how to modify the database (that is, to upload new data or edit/remove existing data). All users can browse and view the data in the database, but only registered users are allowed to modify it. If you are not a registered user, contact your SPIKE system administrator.

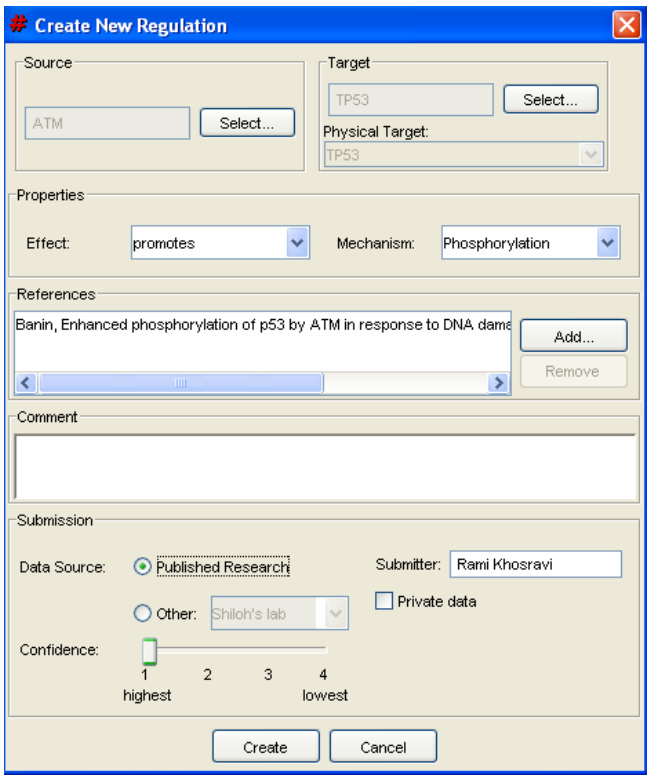
To start modifying the database, press "Edit Mode" ( ). This will only be possible for registered users. The options of the Edit menu are now active, and you may access the various forms for modifying the database (below).

Creating New Regulations

To define a new regulation, make sure that the Edit Mode is switched on and select the "Edit -> New -> Regulations…" option from the menu.

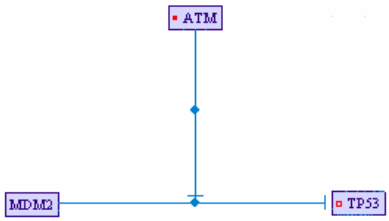


The following form appears:



A regulation is defined by a source node, a target, a physical target (see below) and the regulatory effect that the source exerts on the target. The dialog enables you to enter the following details:

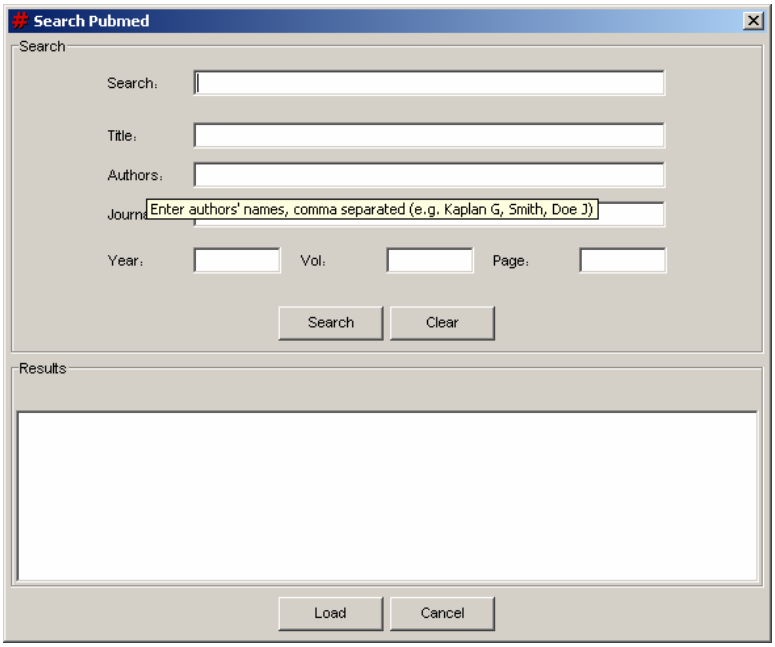
- The source of the regulation: its type (protein, protein family, protein complex or signaling molecule), and name (note that you can type the first letters of the name on the filter field in order to locate it).
- The target of the regulation: its type and name. Note that the type may be a regulation as well as the four entities possible for the source (protein, protein family, etc.). For example, ATM phosphorylates MDM2, thereby disrupting its inhibition of TP53. In the SPIKE data model, this is represented by an inhibitory regulation with ATM as the source, and the regulation of TP53 by MDM2 as the target.



- The form includes a "*Physical Target*" pull-down menu as well. The physical target is the molecule that is physically involved in the interaction with the source. When the target is a protein or a signaling molecule, the physical target will be automatically set to the same node. However, when the target is a regulation or a complex, the physical target must be specified by the user. In the above example, MDM2 is the physical target. When the target is a protein family, the family would also be the physical target, since all its members are assumed to be similarly affected by the source node.

  When a family is selected as the source or target of a regulation, this means that all its members have the same functionality of participating in such a regulation. This helps to make the pathway description parsimonious, since a single regulation suffices for all the family members.
- The regulatory effect (activation, inhibition or unknown) and biochemical mechanism (e.g, phosphorylation, transcription) of the regulation.
- The comment field enables addition of free text (e.g. specifying the amino acid that is modified when the mechanism is phosphorylation).
- A reference article to support the regulation. In order to add new references, select the "Add…" button to open a dialog that enables you to search Pubmed for the relevant article/s.

This dialog allows to search for existing PubMed articles, and to select the articles you need. Define the search criteria and press "Search". The articles that satisfy the criteria will then appear. Select the relevant results (hold down the ctrl key to choose more than one result) and press "Load" to load the selected references. The articles will then appear in the "References" panel in the "Create New Regulation" form.



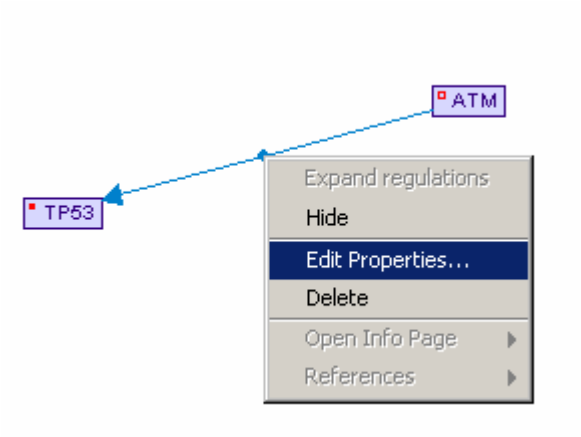
- Data Source – "Published Research" for PubMed-supported regulations; in other cases, press "Other" and select the source from the list, or type in a new data source (which will appear on the list from now on).
- Submitter – your name.

- Private Data – whether the entered data is private. In that case, only your user will be able to view it.
- Quality – the reliability of the entered data. The scale is from 1 to 4, with 1 as the highest quality level. The quality level reflects the reliability the user ascribes to the uploaded data; in general, relations derived from highly focused biochemical studies are assigned high quality level, while those derived from high-throughput experiments are assigned low quality level.

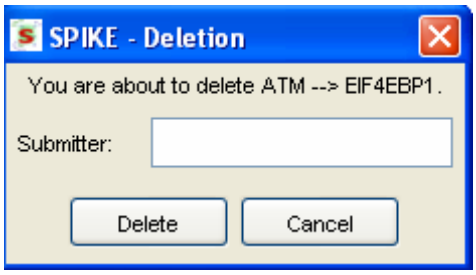
Before pressing "Create" to create the new regulation in the database, make sure that you have filled in all the necessary details, and that they are correct.

Editing and Deleting Regulations

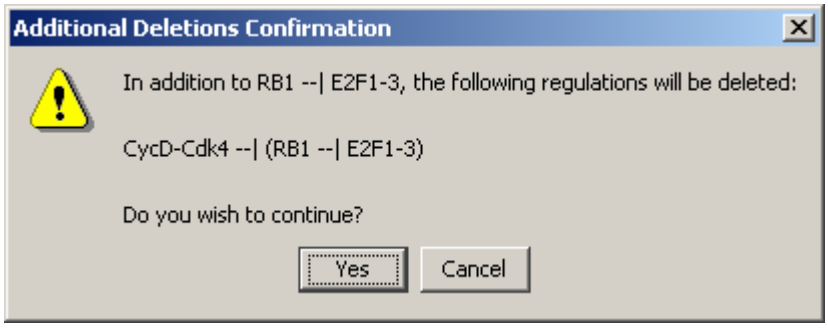
To edit or delete an existing regulation, first make sure that Edit Mode is switched on and then place the mouse pointer on the circle in the middle of the regulation's edge; right-click it and select "Edit Properties…" or "Delete" in the pop-up menu that appears.



Before deleting, a confirmation dialog will appear.



If the selected regulation is a target of other regulations, a list of all such dependent regulations will be shown. If you choose to delete the regulation, these regulations will be deleted too.

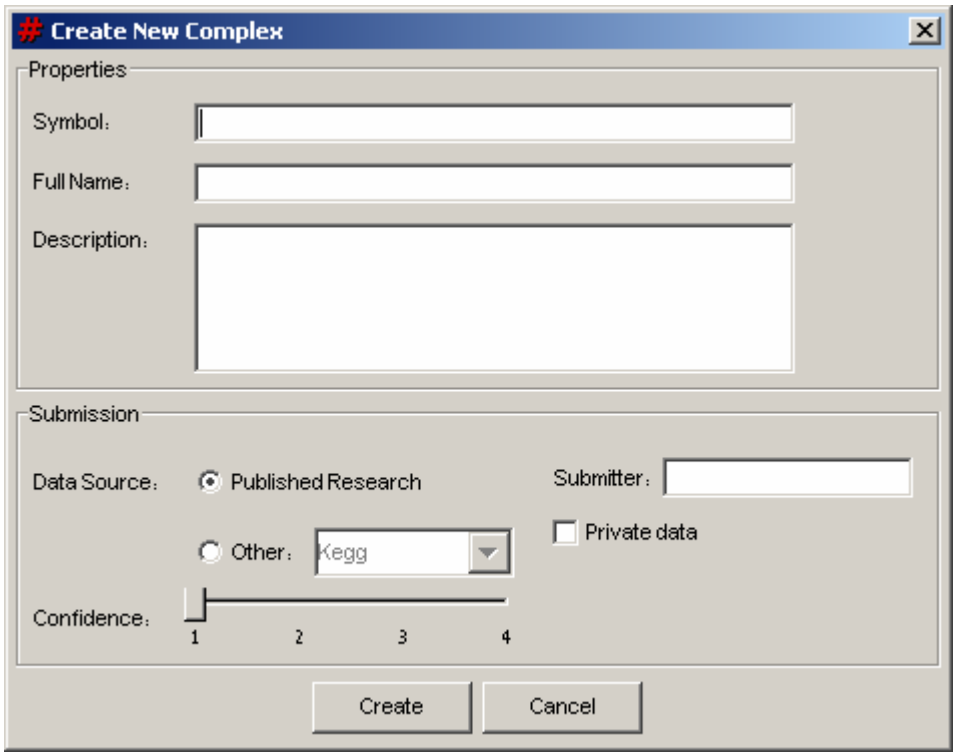


Creating, Editing and Deleting Complexes and Families

The processes of creating, editing and deleting complexes and families are very similar. This section will describe both. To define new complexes, edit existing ones or delete them, Edit Mode must be switched on.

To create a new complex or family, select the "Edit -> New -> Complex/Family" option from the menu.

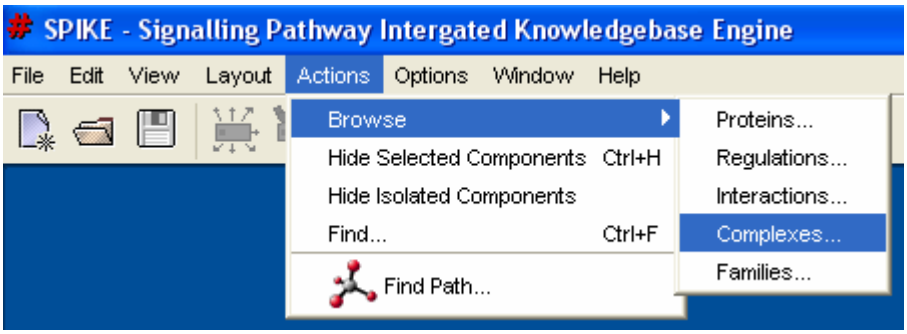
The “Create new complex” form appears:



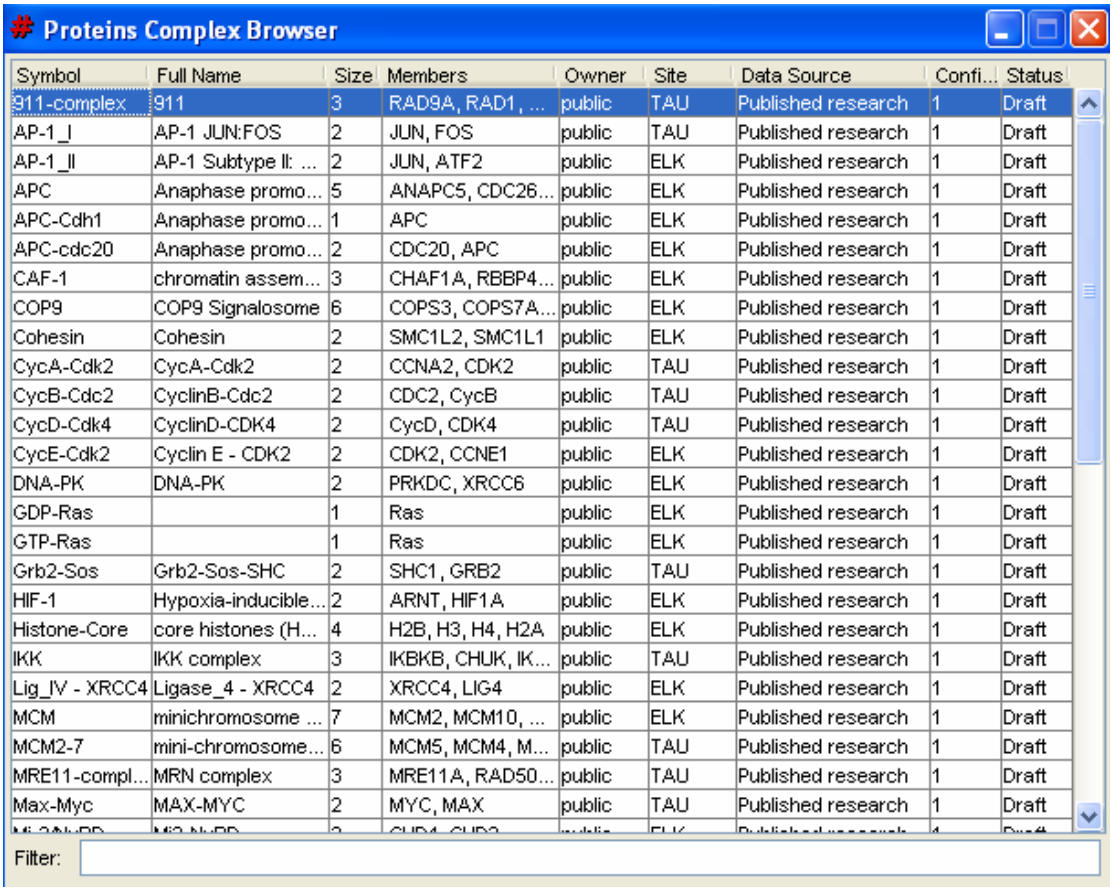
In this dialog, you are asked to type in the symbol that SPIKE will use to identify your complex/family on the pathway, as well as a fuller name and a textual description (optional).

As in creating new regulations, additional data are required: choose a Data Source, fill your name in the Submitter field, check whether the data is private and choose the appropriate quality level. After verifying all the data, press "Create".

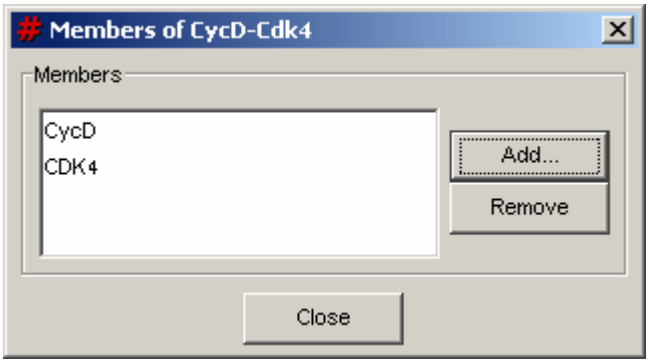
To add or remove family or complex members, right-click on the group's (complex or family) node on the map and press "Edit members…" on the pop-up menu that appears. Another way is to select from the menu: Actions -> Browse -> Complexes (or Families).



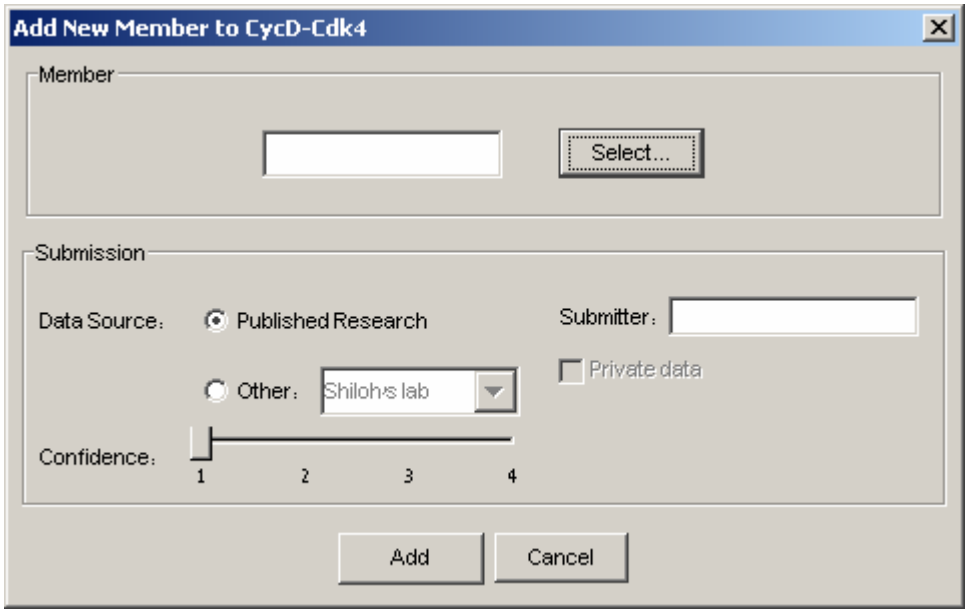
This will open a window that lists all the complexes defined in the database.



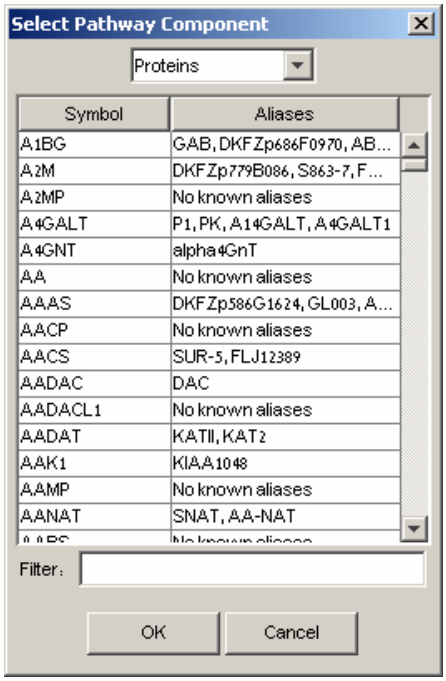
To add a member to a complex, identify the complex in the list, right click it and press "Edit members…" on the pop-up menu that appears. This will open the Members dialog. This dialog will show a list of existing members of the chosen group.



To remove a member, choose it from the list and click "Remove". To add a new member, click "Add…". A new dialog will appear:



This dialog allows you to choose the type and the name of the new member by clicking "Select…":



As previously, additional data are required: choose a Data Source, fill your name in the Submitter field, check whether the data is private and choose the appropriate quality level.

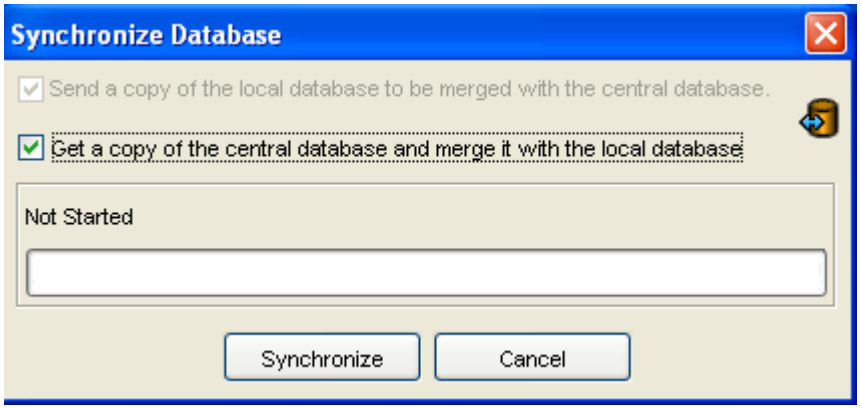
To delete an existing complex or family, right-click it (from the browser or the pathway map) and select "Delete" from the pop-up menu that appears. Before deleting, a confirmation dialog will appear. By deleting an existing complex or family, all its associations will be deleted too.

6. Getting synchronized with other research labs

SPIKE is based on a decentralized database architecture that supports automatic database synchronization: A local copy of the database is installed along with the software in each research lab, and these databases are periodically synchronized with the central database. Database synchronization allows SPIKE users to share data and benefit from the collaborative effort of populating the database and keeping it up-to-date.

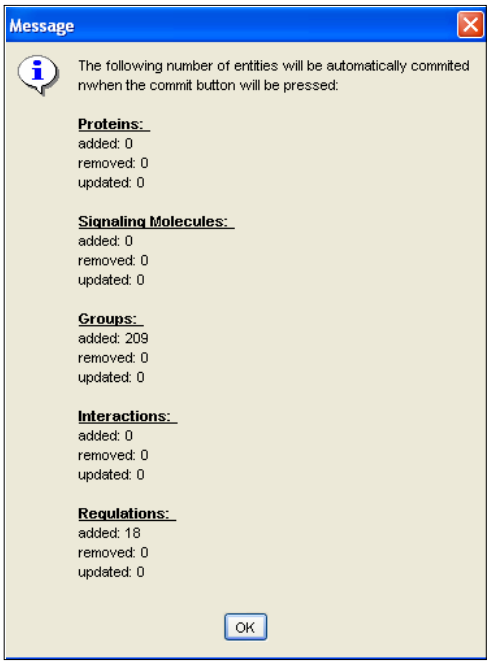
The synchronization process includes 2 steps: first, the local database is packed and sent to the central site; and second, the most updated version of the database at the central site is sent and merged with the database at the local site.

To synchronize your database select: "Options -> Synchronize Database ..." from the menu, and then in the opened tool dialog select 'Synchronize':



Note: database synchronization is only possible if SPIKE was installed in the standard mode (and not in the evaluation mode).

When the synchronization process completes, a summary of the DB changes is displayed.



Acknowledgements

The development of SPIKE, formerly called SHARP, was supported by grants from the A-T Children Project (ATCP), and by the European Union 6th framework grant ESBIC.

The work is also supported in part by the Wolfson Foundation.

Appendix A - File Formats

Note: files that do not follow the required format may be rejected or presented incorrectly.

- SPIKE identifies genes according to their **Human** Entrez Gene ID. **In case your microarray experiment was done using a model animal, you should first replace the gene IDs with the Entrez Gene IDs of their human homologues**.
- Files are tab-delimited text files. Each line should contain exactly the same number of fields, separated by tabs. Therefore the files are actually table files, and we can speak of columns as well as rows.
- Example files are posted on SPIKE website.

Expression File Format

SPIKE currently supports two kinds of experimental gene expression data - "absolute" values (as generated by Affymetrix chips) and "relative" values (as generated by cDNA microarrays).

The first column should contain the specific id of each probe (as it is identified on the chip).

The second column should contain the Human Entrez Gene ID of the gene whose probe it is. If no Entrez Gene ID exists, keep it empty (thereby leaving two consecutive tabs), and SPIKE will ignore this row. There may of course be more than one line with the same Entrez Gene ID (for different probes of the same gene). SPIKE can handle this situation (see above).

The next one or more columns should include the expression levels measured in the various biological conditions that were profiled in the experiment.

The first row should contain a header for each column ("probeID", "GeneID", and then labels for the various conditions; SPIKE will name the conditions according to these labels). The other rows should contain the data for each probe.

Example:

AffyId	Entrez-GeneId	wt0	wt30	wt120
100005_at	9618	228.80	236.05	236.50
100009_r_at	6657	150.06	83.25	108.35
100011_at	51274	80.50	81.60	71.15

Gene/Protein sets File Format

This tab-delimited text file should contain 2 columns: the first contains (Human) Entrez GeneIDs and the second contains the set ids.

Gene/Protein Background Set File Format

This file has only one column, listing the Entrez-Gene IDs of the genes/proteins in the background set.

Appendix B – Spike Default System Users

Initially SPIKE is preloaded with default users, shown in the following table:

SPIKE Username	SPIKE password	Used database user
spike_guest	<no password>	spike_guest
spike_user	Spike	spike_user
spike_curator	Spike	spike_curator
spike_admin	spike	spike_admin

It is highly recommended not to allow users to access the application with these default usernames. Instead, define new usernames for them, using the SPIKE application:

- Login to SPIKE as the SPIKE user "spike_admin" (the password is "spike2004").
- From the Options > Admin menu select "Create SPIKE User…". Choose a username and password for the user, select "SPIKE_user" as the role and fill in the rest of the details. Select "SPIKE_curator" only for the user responsible for data maintenance and data merging on your remote site.