

The PIMRider® platform V3.1

User's Guide

June, 2004

Table of contents

1) Introduction.....	4
2) General concepts.....	4
3) Specific concepts.....	4
3.1) SID® (Selected Interacting Domain).....	4
3.2) PBS® (PIM Biological Score).....	5
4) Basic Search Tool.....	5
4.1) Purpose.....	5
4.2) Features.....	5
4.3) Search criteria description.....	6
5) Advanced Search Tool.....	6
5.1) Purpose.....	6
5.2) Features.....	6
5.3) Search criteria description.....	7
5.4) Search criteria hierarchy.....	7
5.5) In-field logical operator.....	7
6) PIM program information.....	8
6.1) Browser description.....	8
6.2) Field description.....	8
7) Protein information.....	8
7.1) Browser description.....	8
7.2) Field description.....	8
7.3) Section description.....	9
8) Protein annotations.....	9
8.1) Browser description.....	9
8.2) Field description.....	9
9) The PIMViewer(TM) tool.....	11
9.1) Description.....	11
9.2) Symbol description.....	11
9.3) Contextual menus.....	13
9.4) Non trivial functions.....	13
10) The InteractionViewer(TM) tool.....	16
10.1) Description.....	16
10.2) Symbols description.....	16
10.3) Non trivial functions.....	17
11) The DomainViewer(TM) tool.....	18
11.1) description.....	18
11.2) Domain databases and algorithms.....	18
11.3) Symbol description.....	19
11.4) Interpro Domain size and location.....	19
11.5) Non trival functions.....	19

1) Introduction

PIMRider is a web application dedicated to the exploration of Protein-protein Interaction Maps (PIM®).

The application is composed of several viewers, which may be grouped in two classes.

The first class of viewers consists of a group of classical HTML viewers that give access to annotations of proteins and genes such as normalized names and ids, annotations, sequences, bibliographical references, and database entries.

The second class viewers comprises specific graphical viewers (*JAVA Applets*), each of these being specialised in one aspect of proteomics using *2-Hybrid technology (2H)*.

Two search tools were developed to explore network contents in terms of gene / proteins.

The '*Basic search tool*' seeks genes / proteins in the network on the basis of simple text criteria such as 'names' and 'annotations'.

The '*Advanced search tool*' seeks gene / protein on the basis of more sophisticated criteria such as 'functional category', 'functional domains', 'bibliographical references', etc.. that can be combined using logical operators.

The '*Basic search tool*' is embedded within HTML classical viewers as the '*Advanced search tool*' is an HTML viewer in itself.

2) General concepts

PIMRider basically gives access to a Protein-protein Interaction Map built from a series of experiments made using 2H technology.

In a 2H experiment a proteic fragment of interest (BAIT) is screened against a prey library, e.g. a proteome, for interacting partners.

In a PIM program, a set of baits is screened against one given prey proteome. PIMRider can access a PIM that was made from maximum 8 PIM programs. A PIM program can involve one or two Proteomes (two proteomes meaning that bait proteins and interacting partners are separated). Therefore PIMRider can access a PIM involving a maximum of 16 different proteomes.

A protein can belong to more than one proteom. Within the PIM and for two given interacting proteins, more than one interaction can be detected. But within the whole PIM, each 'atomic' interaction that was detected belongs to one and only one PIM program.

3) Specific concepts

3.1 SID® (Selected Interacting Domain)

SID corresponds to the smallest identified protein domain necessary for a particular protein-protein interaction.

3.2) PBS® (PIM Biological Score)

PBS is a score that quantifies the reliability of each interaction and allows the filtering of interactions based on their reliability. The aim of the PBS computation is to add value to the generated Protein Interaction Maps by allowing the rescue of false negatives while efficiently filtering out false positives.

PBS represents a confidence number assigned to each interaction identified. The PBS score (the result of Hybrigenics' bioinformatics algorithms) takes into account the characteristics of the libraries screened and the target organism as well as the results of the screens. The current PBS score is displayed as the probability that the interaction is found by chance, as a result of a two-hybrid artifact. The PBS score ranges from 0 (best) to 1 (worst).

For practical use, the PBS score is displayed by four main categories: A (best), B, C, and D (worst), and two special distinct categories E and 'N/A'. Each PBS category possesses its own color to emphasize its meaning and enhance its visualization in the different viewers:



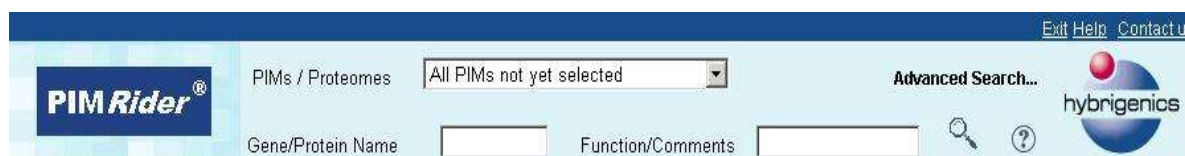
	Interactions with different levels of reliability, corresponding to PBS scores ranging from 0 to <1. The A category represents very reliable interactions while the 'D' category contains much less reliable interactions and could correspond to two-hybrid artifacts (for instance autoactivating baits). PBS thresholds between categories fixed by Hybrigenics' biologist experts depending upon the organisms libraries.
	Interactions with highly connected proteins (very frequently found as preys in Hybrigenics' screens and thus probably a well known two-hybrid artifact: 'sticky' preys). The PBS score of these interactions is set to 1.
N/A	No score can be assigned by Hybrigenics' algorithms (for example when the interaction only comes from the scientific literature).

Chart 1: PBS categories

4) Basic Search Tool

4.1) Purpose



Screenshot 1: basic search tool

To select proteins within the PIM corresponding to the given criteria.

4.2) Features

Search criteria are case-insensitive.

If a word is entered, it is used as a pattern in the search. (example: 'kin' will match for 'kinase')

Multiple Criteria are processed as a single boolean expression where criteria are combined with logical operator 'and'. (Example: seek a protein whose name is 'name' and comments is 'comments' and belongs to proteome 'proteome').

4.3) Search criteria description

Criterion 'PIM/Proteome'

restrict search to proteins that are within in a given PIM program or belong to a given PROTEOME.

Criterion 'Gene/Protein Name'

select a protein when the criterion matches either 'Original name' or 'Biological name' or 'Alias' property.

Criterion 'Function/Comments'

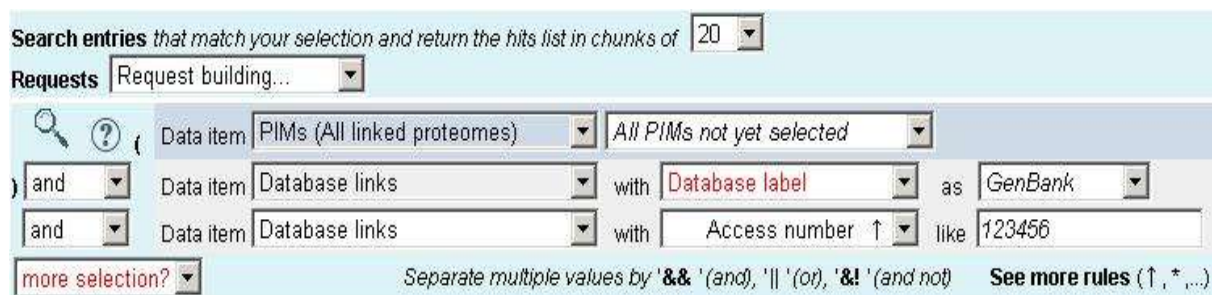
select a protein when the criterion matches either 'Synthetic description' or 'Function' property.

5) Advanced Search Tool

5.1) Purpose

To select proteins within the PIM via a composite request that may include the following criteria:

- PIMs / proteomes combinations,
- protein annotations / classifications,
- database / bibliographical links,
- protein domain related data,
- fragment (SID,BAIT, preys) information.



Screenshot 2: advanced search tool

5.2) Features

Search criteria are case-insensitive.

If a word is entered, it is used as a pattern in the search (example: 'kin' will match for 'kinase').

Request is built from clauses linked with a logical operator ('And', 'OR', 'But Not').

Clauses are of 2 types:

'PIM/PROTEOME',
'PROTEIN/GENE'.

Operator 'OR' can link only two 'PIM/PROTEOME' clauses.

5.3) Search criteria description

Field '*Data Item*'

Identifies clause's subject. ('Database' criteria for instance).

Field '*with*'

Restricts clause's subject. ('Database' 'Database label' criteria for instance).

Field '*like/as*'

Defines clause's value. ('Database' 'Database label' equals '*GenBank*' for instance).

Label is:

- '*like*' when value is a free alpha-numeric string,
- '*as*' when value is restrained to a pre-defined list.

5.4) Search criteria hierarchy

A hierarchy can be defined between particular clauses.

Some clauses once defined act as 'master' for following 'slave' clauses.

Red color emphasizes each 'master' clause.

↑ Up arrow emphasizes a clause that will refer to the corresponding 'master' clause if one was specified before.

↑* Up arrow with a star emphasizes a clause that will refer to the corresponding 'master' clause that MUST have been specified before.

Examples:

1) Search for a protein with a given database id.

```
CLAUSE 'Database links' 'access number' like '12435'
```

(it will search for an existing access number in all databases).

2) Search for a protein with a given database id within a given database.

```
(MASTER CLAUSE) 'Database links' 'Database label' as 'GenBank'  
AND (SLAVE CLAUSE) 'Database links' 'Access number' like '12435'
```

3) Search for a protein with a given domain id without out specifying domain database name is NOT PERMITTED.

4) Search for a protein with a given domain id within a given domain database.

```
(MASTER CLAUSE) 'Domains' 'Database label' as 'InterPro'
```

```
AND (SLAVE CLAUSE) 'Domains' 'Access number' like '12435'
```

5.5) In-field logical operator

Three logical operators are available:

-'&&': 'and',

-'||': 'or',

-'&!': 'and not'.

Example: Search for a protein with two possible given domain ids within a given domain database.

```
(MASTER CLAUSE) 'Domains' 'Database label' as 'InterPro'  
AND (SLAVE CLAUSE) 'Domains' 'Access number' like '12435 || 678910'
```

6) PIM program information

6.1) Browser description

Two HTML pages display PIM program information:

- *PIM program list*,
- *PIM program details*.

6.2) Field description

Field '*Name*'

Usually this name is built from bait and prey proteome names (example '*Mammals-Placenta*' where '*Mammals*' is the bait proteome name and '*placenta*' the prey proteome name).

Field '*SName*'

Usually this name is built from bait and prey proteome short names (example '*Mammals-Pla*' where '*Mammals*' is the bait proteome name and '*pla*' the prey proteome name).

Field '*Nb elements*'

For a given proteome (there can be 2 here), gives the number of proteins contained in the PIM program. This number can be lower than the total number of proteins of the whole proteome.

7) Protein information

7.1) Browser description

Two HTML pages display protein information:

- *protein list*,
- *protein details*.

7.2) Field description

Following fields can occur in 'protein list' or 'protein details' or 'annotation details' pages.

Field '*Original Name*'

Internal alphanumeric code that identifies the protein.

Field '*Biological Name*'

Usual gene/protein name.

Field '*Nb interaction total*'

Gives the total number of interactions found within the whole PIM for a given protein . This number is equal to or greater than the total actual number of protein partners within the whole PIM for this given protein (there can be more than one interaction found between two interacting proteins).

Field '*Synthetic description*'

Summary of the protein.

Field '*Alias*'

Alternative names to '*Biological*' if any.

Field '*Product*'

Usual protein names. They can differ from 'Biological name' and 'Alias'.

Field '*PIM-based annotation*'

Hybrigenics' specific annotation.

Field '*Partners*'

See section '*partners*'. Gives number of interacting protein partners found either within the whole PIM or for one given PIM program.

7.3) Section description

Section '*Proteomes*'

Occurs in '*protein details*' window.

Gives the list of proteomes containing this protein.

Section '*Partners*'

Occurs in '*protein details*' window.

For each interacting protein partners gives for each PIM program (whenever) the interaction PBS score and category. PBS score is 'clickable' and gives access to the InteractionViewer tool.

8) Protein annotations

8.1) Browser description

Two HTML pages display protein annotations:

- *protein details*, section *annotation* (whenever),
- *annotation details*.

8.2) Field description

Field '*Type*'
annotation type
'role' when the annotation describes protein's biological role.

Field '*Relevance*'
'clinical'

Field '*ABS*'
Annotation Biological Score. Annotation confidence score category similar to PBS score.

Field '*Conflict*'
Indicates whether this annotation conflicts with one or more other annotations.

Field '*Annotation*'
Annotation (comments) on this annotation.

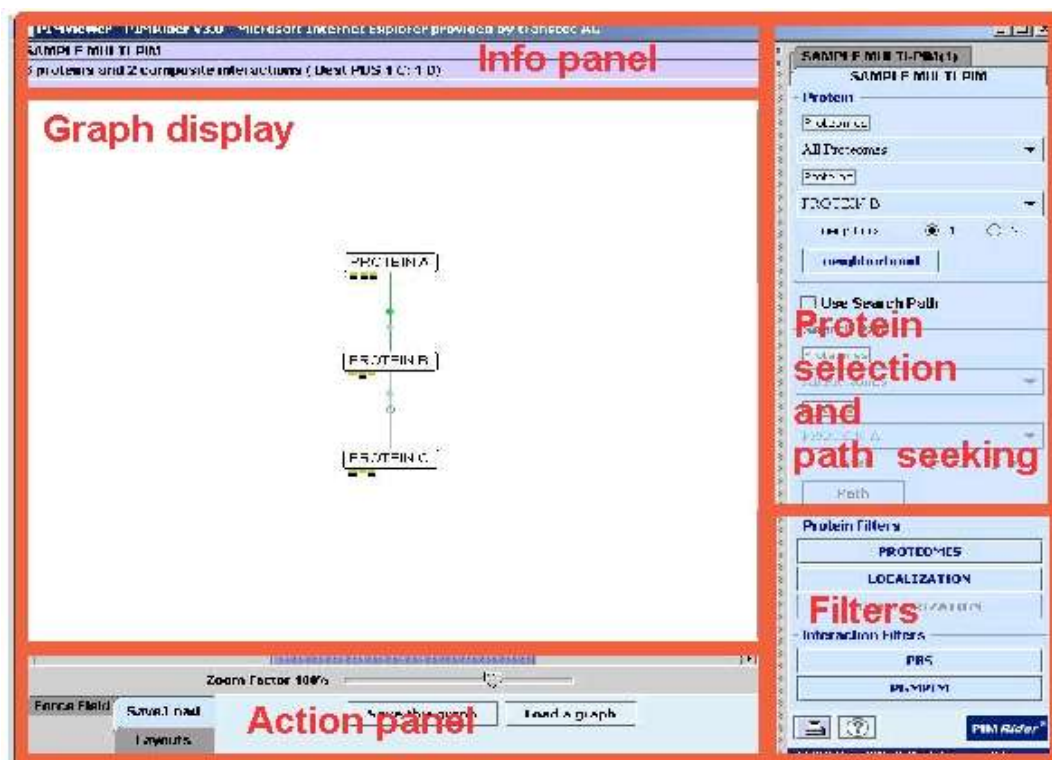
Field '*Summary*'
Annotation summary.

9) The PIMViewer(TM) tool

9.1) Description

The PIMViewer tool displays the proteins and their interactions as an interactive graph.

A protein of interest is first displayed with its immediate interacting partners. The map can be enlarged step by step. Remember that large graphs (hundreds of proteins or thousands) are difficult to manipulate and read, so it is advised to display the whole graph at once except for a specific purpose.



Screenshot 3: PIMViewer

9.2) Symbol description

Symbol 'Protein' 

Each protein is represented as a rounded rectangle and is named according to its 'usual' name.

Whenever a protein contains partners that are not currently displayed within the map, a small red cross '+' is drawn on the top right of the rectangle.

Symbol 'Interaction'

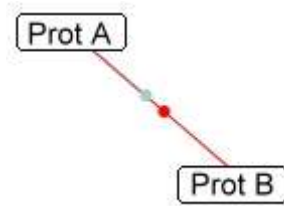
Aggregated interaction

Interactions here are considered as an aggregation of all the 'atomic' interactions found within the different PIM programs.

An aggregated interaction is figured as an edge linking two interacting proteins. Edges are colored according to the best PBS category found among 'atomic' interactions. Interactions here are not oriented e.g. there no way to know directly which protein acted as prey or bait in the experiment.

Atomic interaction

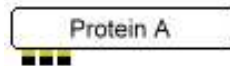
'Atomic' interactions are figured as circles drawn along the edge. Circles are colored according to their interaction PBS value.



Screenshot 4: Interaction

(interaction A-B found in 2 PIM programs: 1 PBS 'A' and 1 PBS 'D', best PBS of A therefore edge's color is red).

Symbol 'Proteome'



Screenshot 5: protein and proteomes

(protein A belongs to 3 proteomes).

Proteomes are figured within the map as little yellow squares on the bottom right of each protein rectangle. The order of the proteomes is always the same from left to right within the map.

Symbol for 'Protein functional category'

Caspases	Cas
Kinases	K
Phosphatases	Pho
Proteases	Pro
Matrix metalloproteinases	Mtp

Caspases	Cas
Membrane receptor	Rm
Nuclear receptors	Rnu
Secreted proteins	S

Chart 2: Protein functional category

Symbol for 'Protein localisation'





Nucleus	
Membrane	
Intracellular	
Extracellular	

Chart 3: Protein localisation

9.3) Contextual menus

Contextual menu 'PIM'

Right-click on the graph background to get global actions available on the map.

Contextual menu 'Protein'

Right-click within protein rectangle to get actions available on the protein.

Contextual menu 'Interaction'

Right-click on each interaction edge to get actions available. If an 'atomic' interaction is found then program name 'Mammals-placenta' for instance and PBS category are displayed; this menu entry gives access to the InteractionViewer.

Contextual menu 'Multiple selection'

'CTRL' + Right-click on multiple selection of edges or nodes to get actions available on the selection.

9.4) Non trivial functions

Function 'protein interacting partners display'

In the right panel:

- 1) select the proteome within the corresponding drop down list box,
- 2) select the protein of interest within the corresponding drop down list box,
- 3) select the neighborhood level '1' or '2',
- 4) click on 'neighborhood' button.

Function 'protein pathway seeking'

In the right panel:

- 1) see function 'display protein interacting partners' in order to select protein #1,
- 2) check 'Use search path' box,

- 3) select protein #2 (within 'search path box'),
- 4) select the neighborhood level '1' or '2',
- 5) click on 'path' button.

Function '*protein/interaction filtering and highlighting*'

Important: whenever a filter is set (for instance '*keep only protein belonging to placenta*', it remains active during the session (consequence would be in this case: if you display a protein with its immediate neighbors later in the same session you would get only 'placenta' protein partners).

In the filtering palettes:

Click on the 'eye' symbol to filter protein/interaction.

Click on the check box to highlight protein/interaction.

Function '*save/load graph*'

Everything on the screen is saved in an archive file whose extension must be '.jar' (for it is the actual file format).

Function '*protein/interaction multiple selection*'

You can perform multiple selections of edges or nodes using 'CTRL click' with the mouse or by using the highlight functionality that comes in filtering palettes.

Function '*graph layout*'

Force directed graph layout

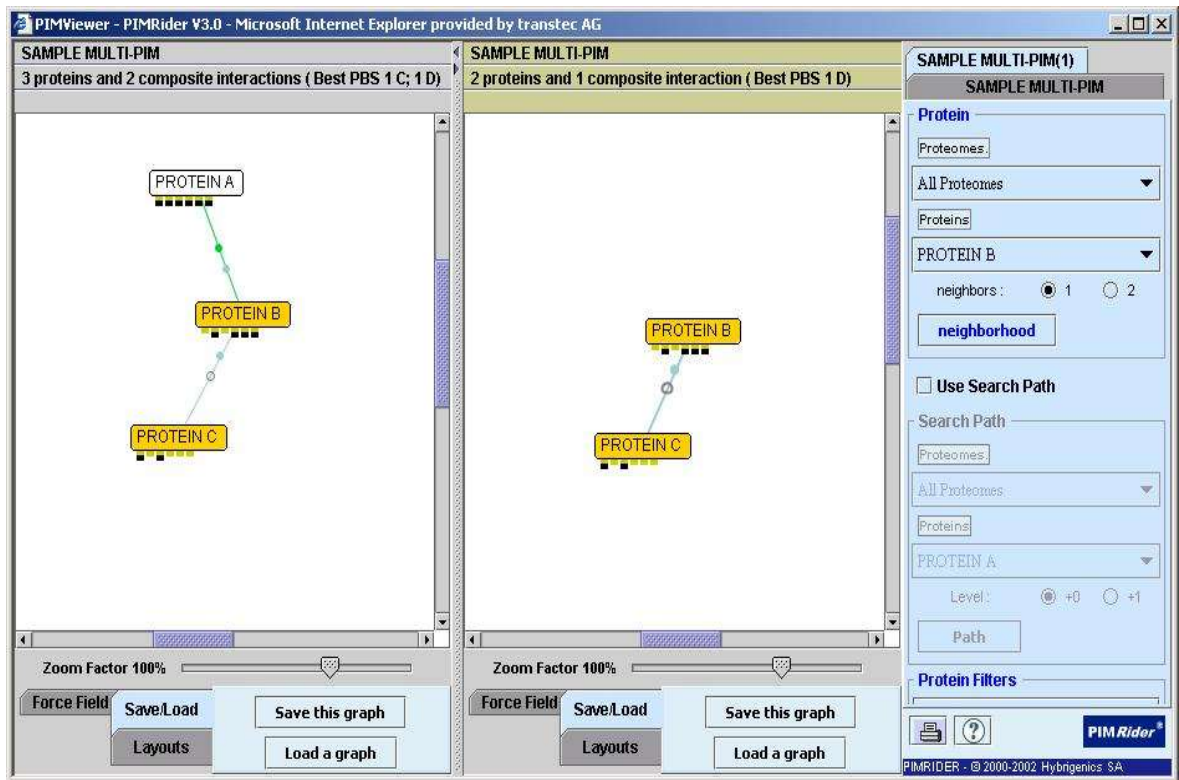
Default graph layout. Use this algorithm with small graphs (up to 500 proteins). Go to bottom tabbed panel 'layout', and use apply/stop button to start/stop simulation.

Radial tree graph layout

Use it for graphs of all sizes. Do not forget to select a protein (which will act as a root node) before applying this algorithm.

Function '*simultaneous two-graph display*'

You can split map a panel in two. This allows the possibility of displaying at the same time two views of the same map. Go to right panel which is actually a 2-tabbed panel, click-on right tab and then select a protein and show it with its neighbors, the map panel will split in two.



Screenshot 5: PIMViewer - split view

10) The InteractionViewer(TM) tool

10.1) Description

This tool is dedicated to 2-H technology.

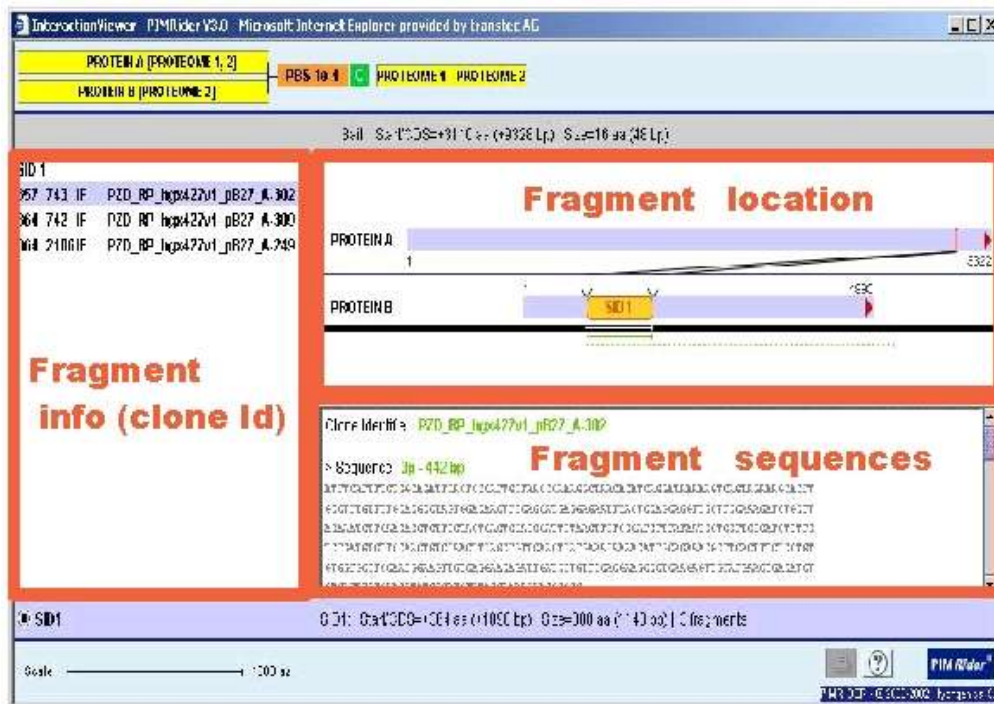
For two interacting proteins and a given PIM program, it displays the protein primary sequence, BAIT fragments, preys fragments and SIDs. All these fragments are located relative to the proteins' primary sequences.

It displays prey fragments sequence information in fasta format:

- '5p' experimental nucleic sequence (when any),
- '3p' experimental nucleic sequence (when any),
- theoretical nucleic sequence,
- theoretical proteic sequence.

It displays prey fragment clone HGX identifiers.

In terms of nucleic translation, '5p-3p' direction is from left to right in the layout.




10.2) Symbols description



Prey fragments color and styles :

- green / red
green: nucleic fragment was translated in frame ('IF'),
red: nucleic fragment was translated out of frame ('OOF1/2').
- plain / dashed segment
plain: '5p' and '3p' experimental fragments are present,
dashed: an experimental fragment is missing ('5p', '3p' or both).
- normal / highlighted tips color
normal: translation occurred inside start and stop codons.
highlighted: translation started before start codon and/or stopped after stop codon.

'BAIT' fragment 

'SID' fragment 

10.3) Non trivial functions

Function '*display fragment (SID, BAIT, prey) location*'

Click on the fragment. Check the info bar at the window bottom/top.

Function '*prey fragment sequence information*'

Click on the prey segment to display sequence information in the bottom panel.

Function '*prey fragment clone information*'

Click on the prey segment to select clone information in the left panel.

Function '*prey fragment sequence information export*'

With the mouse pointer select sequence to export and copy it to clipboard using 'CTRL-C'.

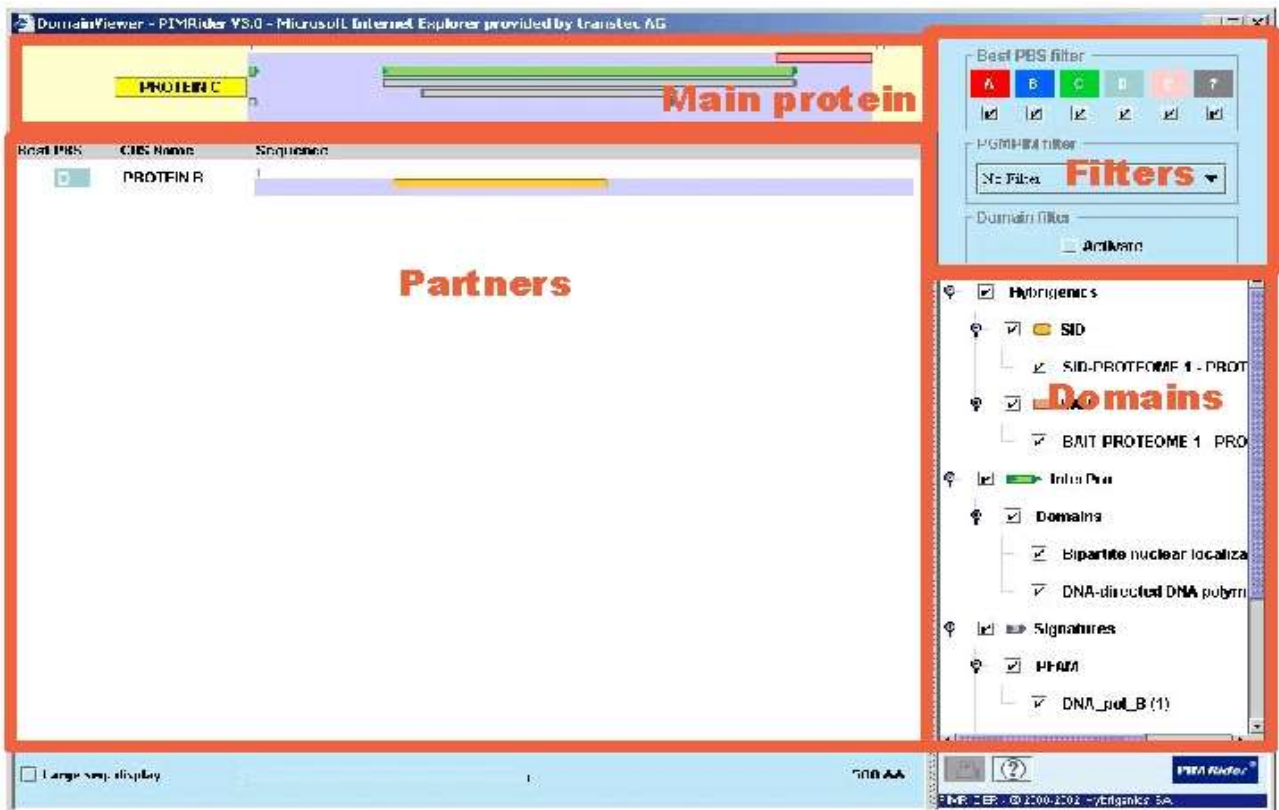
Function '*reciprocal interaction display*'

Click on the radio button (usually SID2) of the protein used as bait.

11) The DomainViewer(TM) tool

11.1) description

The purpose of this tool is, for a protein of interest and all its partners, to graphically compare 2-H results in terms of interacting domains (Bait/SID) to known proteic domain annotations. For instance the tool may answer a question like 'does a given SID includes or overlap with a known 'kinase' Interpro domain?'.



Remark: the same BAIT fragment may have been used in more than one PIM program, but a SID 'belongs' to one and only one PIM program.

Screenshot 7: DomainViewer

11.2) Domain databases and algorithms

Some protein domain annotations were obtained from specific algorithms with no database entries associated:








- Transmembrane region (TMHMM algorithm),
- Peptide signal sequence (SignalP algorithm).
- Coiled coil (NCoils algorithm).

Other protein domain annotations were obtained using InterProScan algorithm (which contains different protein signature recognition methods) which gave entries with the following databases:

- InterPro,
- SMART,

-PROSITE,
-PFAM,
-PRINTS.

11.3) Symbol description

'SID' fragment	
'BAIT' fragment	
'Interpro' domain	
'Signature' domain	
'Transmembrane' domain	
'Peptide signal' domain	
'Coiled coil' domain	

11.4) Interpro Domain size and location

For a given protein, coordinates of an Interpro domain are deduced from all *'signature'* domains found attached to it. The size of the Interpro domain represents the maximum extent of all the underlying database domains and therefore is not an actual size.

11.5) Non trivial functions

Function *'domain name display'*

In the layout move cursor over domain to get its name as a tooltip.

Function *'check for SID/BAIT partners'*

Right-click on the SID/BAIT domain, interacting domains will be highlighted within the protein list.

Please remember that:

- a SID interacts with one and only one BAIT,
- a BAIT can have more than SID partners.

Function *'domain Highlighting'*

In the right panel right-click on a given domain name. The domain selected is then highlighted in the list. (the number after the name shows the number of occurrences of this domain the protein list).

Function *'proteic Partners Filtering'*

Best PBS Filter

Shows only partners which best PBS interaction is one of those specified as a filter.

PIM Program Filter

Shows only partners that contain at least one SID that belongs to the specified PIM program.

Domain Filter

Shows only partners that contain at least one domain that belongs to the domain filter list.

Click on the check box '*activate*' to switch to domain filtering mode. All domains below are then automatically unselected, and the protein partner list is erased. Use then the domain list as filtering criteria, and recheck the domain of interest to display protein that contains this particular domain within the partner list.

Function '*protein sorting*'

Click on partners table header to sort partners on:

- Best PBS,
- Protein name,
- Protein size.

Function '*Domain annotations display*'

Double-click a domain symbol to open an external web browser and display its annotations in the corresponding database (Interpro, SMART, PROSITE ...).

Function display '*Interaction details*'

Double-click in a SID area in the layout to open the InteractionViewer tool.