

Clustering and scoring molecular interactions relying on community standards

Rafael C. Jimenez
rafael@ebi.ac.uk



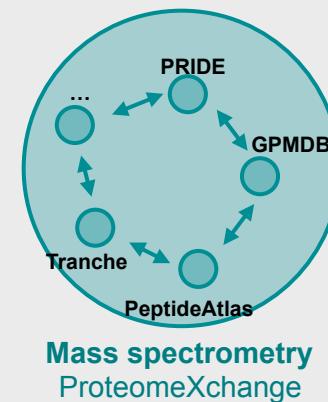
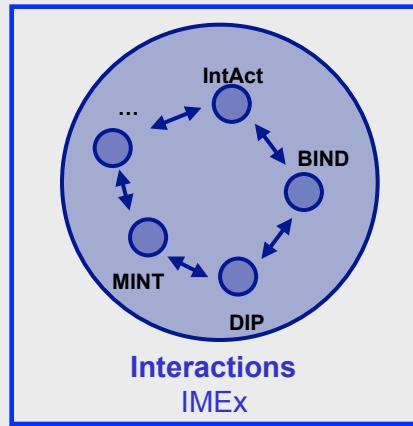
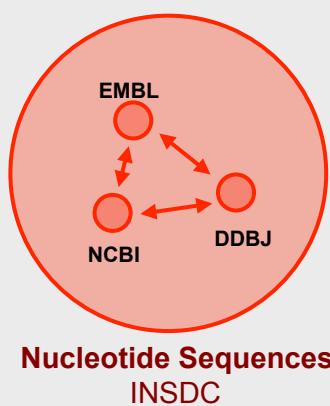
8th BSPR-EBI Meeting
From the visible to the hidden proteome
12th - 14th July 2011
Wellcome Trust Conference Centre, Cambridge, UK



EBI is an Outstation of the European Molecular Biology Laboratory.

Sharing infrastructures

- Collaboration among multiple repositories in one field
 - Adoption of standards
 - Coordination on data deposition



Better accessibility

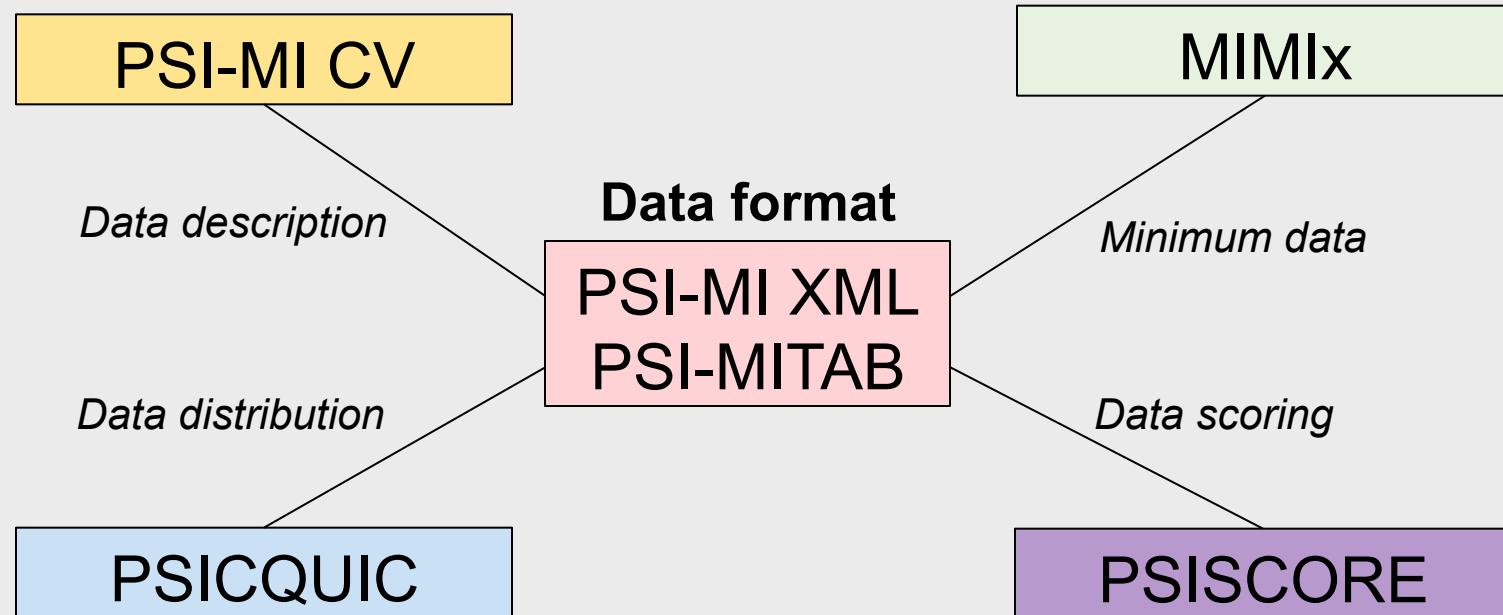
More data coverage

Less redundancy

PSI-MI standards



PSI-MI



PSICQUIC

nature methods
Techniques for life scientists and chemists

Full text access provided to Wellcome Trust Sanger Institute's *in vivo* *MICROGENOMICS* IMAGING AND SCREENING

[nature.com](#) ▶ [journal home](#) ▶ [current issue](#) ▶ [correspondence](#) ▶ [full text](#)

NATURE METHODS | CORRESPONDENCE

PSICQUIC and PSISCORE: accessing and scoring molecular interactions

Bruno Aranda, Hagen Blankenburg, Samuel Kerrien, Fiona S L Brinkman, Arnaud Ceol, Emilie Chautard, Jose M Dana, Javier De Las Rivas, Marine Dumousseau, Eugenia Galeota, Anna Gaulton, Johannes Goll, Robert E W Hancock, Ruth Isserlin, Rafael C Jimenez, Jules Kerssemakers, Jyoti Khadake, David J Lynn, Magali Michaut, Gavin O'Kelly, Keiichiro Ono, Sandra Orchard, Carlos Prieto, Sabry Razick, Olga Rrigina, et al.

[Affiliations](#) | [Corresponding author](#)

Nature Methods 8, 528–529 (2011) | doi:10.1038/nmeth.1637
Published online 29 June 2011

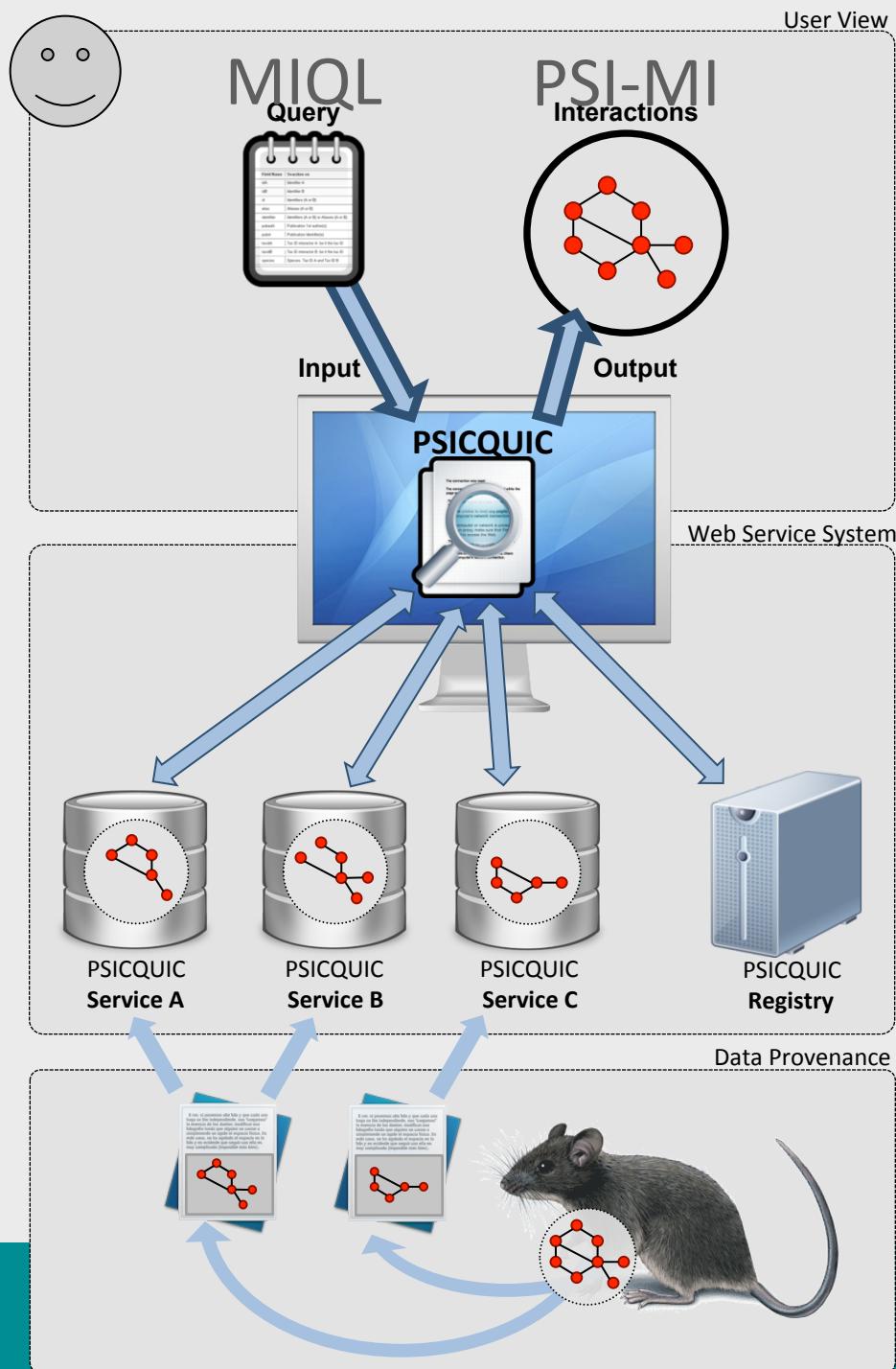
Subject terms: Bioinformatics • Proteomics

To the Editor:

To study proteins in the context of a cellular system, it is essential that the molecules with which a protein interacts are identified and the functional consequence of each interaction is understood. A plethora of resources now exist to capture molecular interaction data from the many laboratories generating such information, but whereas such databases are rich in information, the sheer number and variability of such databases constitutes a substantial challenge in both data access and quality assessment to the researchers interested in a specific biological domain.

[print](#)
[email](#)
[download pdf](#)
[download citation](#)
[order reprints](#)
[rights and permissions](#)
[share/bookmark](#)

PMID: 21716279



PSICQUIC Registry

PSICQUIC Registry					
Name	Version	URL	Description	SOAP	Comments
APID	YES	http://cicblade.dep.usal.es/	PSICQUIC service provided by APID	NO	protein-protein protein-DNA protein-RNA protein-nucleic acid protein-lipid protein-carbohydrate protein-peptide protein-glycoprotein protein-glycolipid protein-protein complex protein-DNA complex protein-RNA complex protein-nucleic acid complex protein-lipid complex protein-carbohydrate complex protein-peptide complex protein-glycoprotein complex protein-glycolipid complex protein-protein interaction protein-DNA interaction protein-RNA interaction protein-nucleic acid interaction protein-lipid interaction protein-carbohydrate interaction protein-peptide interaction protein-glycoprotein interaction protein-glycolipid interaction
ChEMBL	YES	http://www.ebi.ac.uk/Tools	PSICQUIC service provided by ChEMBL	NO	protein-protein protein-DNA protein-RNA protein-nucleic acid protein-lipid protein-carbohydrate protein-peptide protein-glycoprotein protein-glycolipid protein-protein complex protein-DNA complex protein-RNA complex protein-nucleic acid complex protein-lipid complex protein-carbohydrate complex protein-peptide complex protein-glycoprotein complex protein-glycolipid complex protein-protein interaction protein-DNA interaction protein-RNA interaction protein-nucleic acid interaction protein-lipid interaction protein-carbohydrate interaction protein-peptide interaction protein-glycoprotein interaction protein-glycolipid interaction
BioGrid	YES	http://tyerslab.bio.ed.ac.uk	PSICQUIC service provided by BioGrid	NO	protein-protein protein-DNA protein-RNA protein-nucleic acid protein-lipid protein-carbohydrate protein-peptide protein-glycoprotein protein-glycolipid protein-protein complex protein-DNA complex protein-RNA complex protein-nucleic acid complex protein-lipid complex protein-carbohydrate complex protein-peptide complex protein-glycoprotein complex protein-glycolipid complex protein-protein interaction protein-DNA interaction protein-RNA interaction protein-nucleic acid interaction protein-lipid interaction protein-carbohydrate interaction protein-peptide interaction protein-glycoprotein interaction protein-glycolipid interaction
IntAct	YES	http://www.ebi.ac.uk/Tools	PSICQUIC service provided by IntAct	NO	protein-protein protein-DNA protein-RNA protein-nucleic acid protein-lipid protein-carbohydrate protein-peptide protein-glycoprotein protein-glycolipid protein-protein complex protein-DNA complex protein-RNA complex protein-nucleic acid complex protein-lipid complex protein-carbohydrate complex protein-peptide complex protein-glycoprotein complex protein-glycolipid complex protein-protein interaction protein-DNA interaction protein-RNA interaction protein-nucleic acid interaction protein-lipid interaction protein-carbohydrate interaction protein-peptide interaction protein-glycoprotein interaction protein-glycolipid interaction

- 16 PSICQUIC services
- **16,335,145** interactions

Name	Active Interactions	Version	SOAP URL
APID	YES	416,124 1.1.5	http://cicblade.dep.usal.es/
ChEMBL	YES	612,785 1.1.6	http://www.ebi.ac.uk/Tools
BioGrid	YES	337,957 1.1.6-SNAPSHOT	http://tyerslab.bio.ed.ac.uk
IntAct	YES	267,114 1.1.7-SNAPSHOT	http://www.ebi.ac.uk/Tools



PSICQUIC client



PSICQUIC View

Search: P37173

[Search](#) [Clear](#) [Fields »](#)

PSICQUIC View														
Search results for query P37173														
Number of results: 1														
Cluster this query														
Detailed view of the results														

Click on the links below to display the results for each selected service ([refresh list](#))

Use the check boxes to include or excludes services from the search and cluster operations - Select: [All](#) , [None](#)

- [APID](#) - 77
- [BIND](#) - 0
- [BioGrid](#) - 0
- [ChEMBL](#) - 16

- [DIP](#) - 12
- [DrugBank](#) - 2
- [InnateDB](#) - 24
- [IntAct](#) - 10

- [Interporc](#) - 1
- [iRefIndex](#) - 144
- [MatrixDB](#) - 1
- [MINT](#) - 4

- [MPIDB](#) - 0
- [Reactome](#) - 13
- [Reactome-Fls](#) - 125
- [STRING](#) - 226

[Cluster this query](#)

IntAct

View: [Table](#) | [Graph](#)

PSI-MI TAB														PSI-MI 2.5	
	Name molecule A	Links molecule A	Name molecule B	Links molecule B	Alt. Identifiers molecule A	Alt. Identifiers molecule B	Aliases molecule A	Aliases molecule B	Species molecule A	Species molecule B	First Author	PubMed Identifier	Interaction ID	Interaction Type	
1	P37173; EBI-296151		Q81LN0; EBI-2820887		TGF-beta type II receptor; Transforming growth factor-beta receptor type II; tgfr2_human	BA_4584; BAS4252; GBAA_4584; q81ln0_bacan	TGFB2	vpR	Human (9606)	Bacillus anthracis (1392)	Dyer et al. (2010)	20711500 IM-13779		physical association	
2	Q35613; EBI-77304		P37173; EBI-296151		Daxx; daxx_mouse	TGF-beta type II receptor; Transforming growth factor-beta receptor type II; tgfr2_human	Daxx	TGFB2	Mouse (10090)	Human (9606)	Perlman et al. (2001)	11483955		physical association	
3	P37173; EBI-296151		P07200; EBI-907660		TGF-beta type II receptor; Transforming growth factor-beta receptor type II; tgfr2_human	tgfb1_pig	TGFB2	TGFB1	Human (9606)	Pig (9823)	Zacchigna et al. (2006)	16530041 IM-11825		association	

MI cluster

- Clusters binary interactions providing non redundant information
- It looks at all the interactor accessions
 - Primary identifiers, aliases and cross-references
- Rely on **standards** (PSI-MI format and PSI-MI CV)
- Results back in PSI-MI format
 - Also mapping information between the data provider and the original data.
- Used by ... PSICQUIC view, IntAct, EnCore, Reactome and the EBI search
- Freely available:
 - Java API (Maven dependency)
http://code.google.com/p/enfin-core/wiki/wp1_encore_tools_micluster

Clustered query

Search: P37173

Search

Clear

Fields »

← Back to all services

Clustered query: 'P37173' from APID, ChEMBL, DIP, DrugBank, InnateDB, IntAct, Interoporc, M, Reactome-Fls, STRING, iRefIndex

View: [Table](#) | [Graph](#)

Protein View	
This page displays the results of a search for the protein P37173 across various databases. The results are clustered by source database. You can click on any row to view detailed information about the interaction.	
APID	1
ChEMBL	1
DIP	1
DrugBank	1
InnateDB	1
IntAct	2
Interoporc	1
Reactome-Fls	1
STRING	1
iRefIndex	1

Export: [PSI-MI TAB](#)

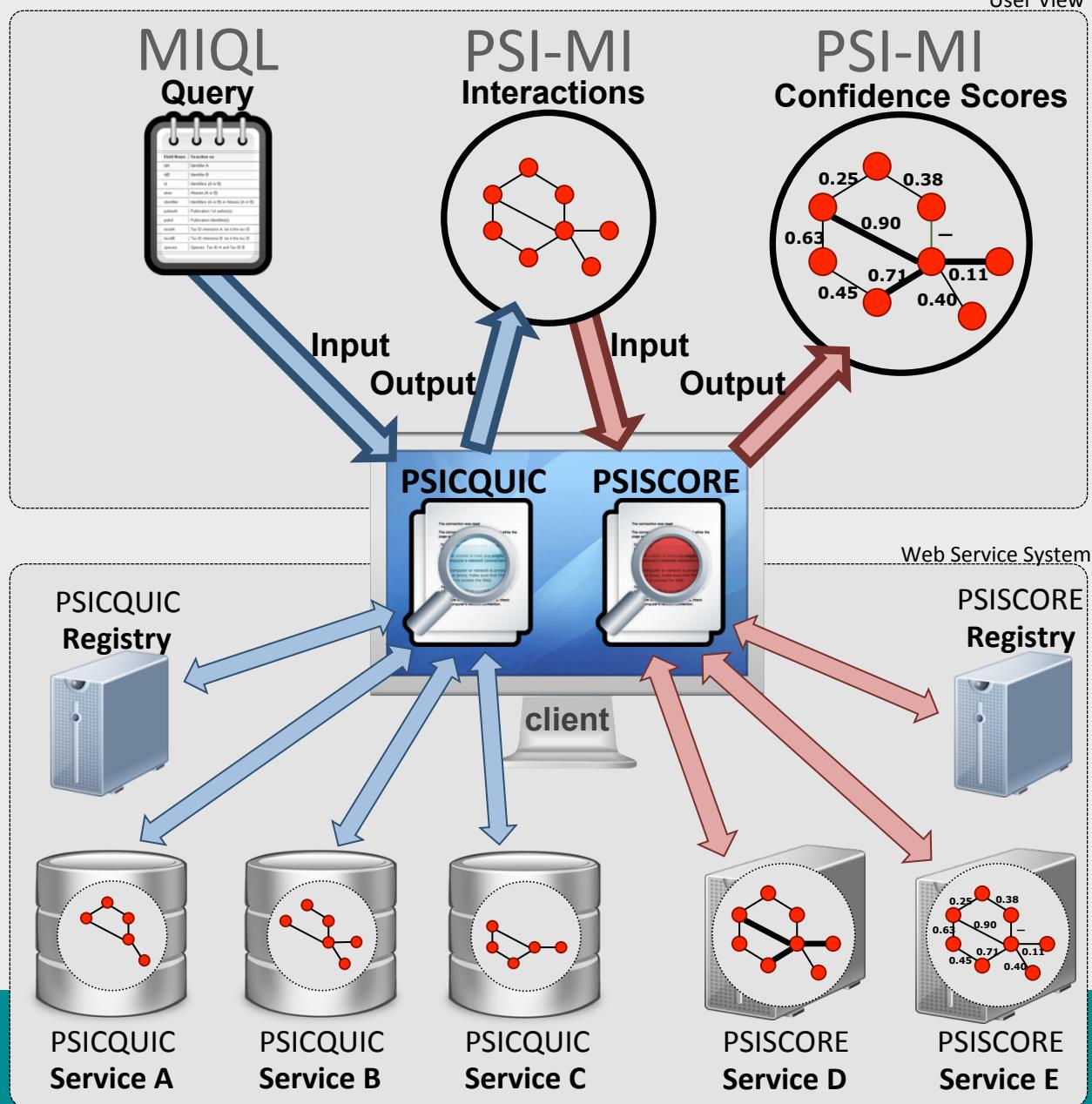
	Name molecule A	Name molecule B	PubMed Identifier	Interaction Type	Interaction Detection Method	Source Database
1	P37173	P37173	7890683 11279102 10026191 9169454 12015308 17367534	association physical association phosphorylation reaction	coimmunoprecipitation two hybrid in-gel kinase assay inferred by curator	GRID HPRD DIP InnateDB reactome
2	P37173	P07200	16530041 IM-11825	direct interaction association	anti tag coimmunoprecipitation	IntAct
3	P37173	Q99K41	16530041 IM-11825	association	anti tag coimmunoprecipitation	IntAct

MI score

- Designed to calculate **annotation evidence**
- Based on common and minimum experimental information
 - Number of **publications**
 - experimental **detection methods** (CV)
 - **interaction types** (CV)
- Formula based on **curation experience**
 - Default parameters are **configurable**
- Rely on **standards** (PSI-MI format, PSI-MI CV, MIMIx)
- Scores between 0-1
- One method to score interactions from **different resources**
- Used by ... the EBI search, EnCore and UniProt
- Freely available:
 - **Java API** (Maven dependency) & **PSISCORE** web service
http://code.google.com/p/enfin-core/wiki/wp1_encore_tools_micluster

Example Scoring clusters					
			TGFR2_HUMAN	TGFB1_HUMAN	
PSICQUIC Services	Binary interaction evidences	Publications	Interaction Type	Detection Method	MI score
String	4	1	-	experimental interaction detection predictive text mining interologs mapping inferred by curator	0.06
IntAct	2	1	direct interaction	cross-linking study	0.44
iRefIndex	17	9	physical association	confirmational text mining coimmunoprecipitation surface plasmon resonance interologs mapping	0.54
DIP InnateDB IntAct MINT Reactome	14	10	direct interaction physical association	pull down anti tag coimmunoprecipitation coimmunoprecipitation surface plasmon resonance two hybrid cross-linking study inferred by curator	0.94
APID DIP InnateDB IntAct MINT Reactome Reactome-Fls STRING iRefIndex	43	16	direct interaction association physical association	confirmational text mining anti tag coimmunoprecipitation coimmunoprecipitation surface plasmon resonance two hybrid experimental knowledge based pull down experimental interaction detection predictive text mining interologs mapping cross-linking study inferred by curator	0.97

PSICQUIC & PSISCORE



PSISCORE client



1) Upload data.
Upload your interaction data in the form of a valid PSI-MI file

H:\public_html\temp\inte [Browse...](#) [Upload](#)

File uploaded and parsed successfully.

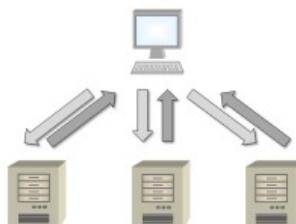


2) Select scoring methods.

Select the scoring methods you want to run on your data. You can get more information on each method by hovering the mouse over its name.

- <http://mint.bio.uniroma2.it/psiscore-ws-0.9.7-SNAPSHOT/webservices/psiscore>
 - MINT-score
 - HomoMINT-score
- <http://psiscore.bioinf.mpi-inf.mpg.de/psiscorews/webservices/psiscore>
 - BPScore
 - CCScore
 - MFScore
 - Domain support, inferred
 - Domain support, structural
 - np-number_pmids
 - lpr-lowest_pmid_reuse
 - hpr-highest_pmid_reuse
- <http://www.ebi.ac.uk/enfin-srv/miscore/webservices/psiscore>
 - Miscore - psicquic sources
 - Miscore - intact
 - Miscore - imex curation
 - Miscore - intact plus imex curation

Please note: Due to institutional security restrictions, PSISCOREweb is currently unable to communicate with servers that are running on ports other than the standard http port 80.



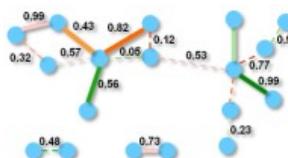
3) Start scoring.

Press the button to the right to send your data to the selected scoring servers. The confidence scores will be included into your input file as soon as they are calculated.

[Start scoring!](#)

started job c76y94y7j4dm on server <http://mint.bio.uniroma2.it/psiscore-ws-0.9.7-SNAPSHOT/webservices/psiscore>, estimated duration 1 sec
started job gamrq1n2t1qw on server <http://psiscore.bioinf.mpi-inf.mpg.de/psiscorews/webservices/psiscore>, estimated duration 1 sec
started job 14qb8bt815ehu on server <http://www.ebi.ac.uk/enfin-srv/miscore/webservices/psiscore>, estimated duration 1 sec

Scoring job started. Press the button below to retrieve the results.



4) Download results.

Download your PSI-MI file with all confidence scores by pressing the button to the right.

[Get Job](#)

[scored_intact.txt](#)

Proteomics Services Team

			
Henning Hermjakob Head of Proteomics Services	Sandra Orchard Senior Scientific Database Curator	Juan A Vizcaino PRIDE coordinator	Bruno Aranda Software Engineer
			
Richard Côté Senior Software Engineer	David Croft Bioinformatician	Attila Csordas Bioinformatician	Bernard de Bono T.B.A.
			
Margaret Duesbury Scientific Database Curator	Marine Dumousseau Bioinformatician	Antonio Fabregat Mundo Visitor	Joe Foster PhD Student
			
Phani Garapati Curator	Laurent Gatto Visitor	Nelson Ildegwa Gichora Trainee	Pierre Grenon Bioinformatician
			
Johannes Griss Database Application Developer	Bijay Jassal Curator	Andrew Jenkinsion Information Integration Specia	Rafael Jimenez Database Application Developer: ENFIN
			
Steven Jope Knowledge Database Curator/Developer	Samuel Kerrien Senior Software Engineer	Jyoti Khadake Senior Database Curator	Gavin O'Kelly Bioinformatician
			
David Ovellerio Bioinformatician	Daniel Rios Pride Software Developer	Florian Reisinger Software Developer	Gustavo A. Salazar Visitor
			
Chris Taylor Senior Software Engineer	Jose Villaveces Trainee	Matthieu Visser Visitor	Rui Wang Bioinformaticien
			
Sarala Wimalaratne Software Engineer			

Thank you!
Questions?

Acknowledgements



Henning Hermjakob
Sandra Orchard
Marine Smousseau
Samuel Kerrien
Bruno Aranda



Hagen Blankenburg



Work supported by the Enfin grant
(EU contract number LSHG-CT-2005-518254)

