



PRIDE Inspector: a tool to visualize and validate MS proteomics data

Rui Wang

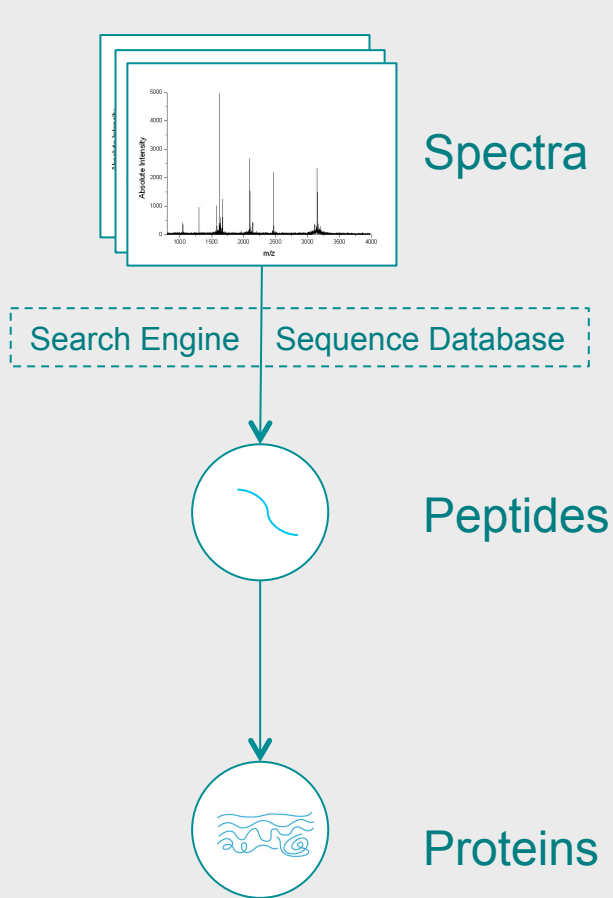
PRIDE Team, Proteomics Services Group



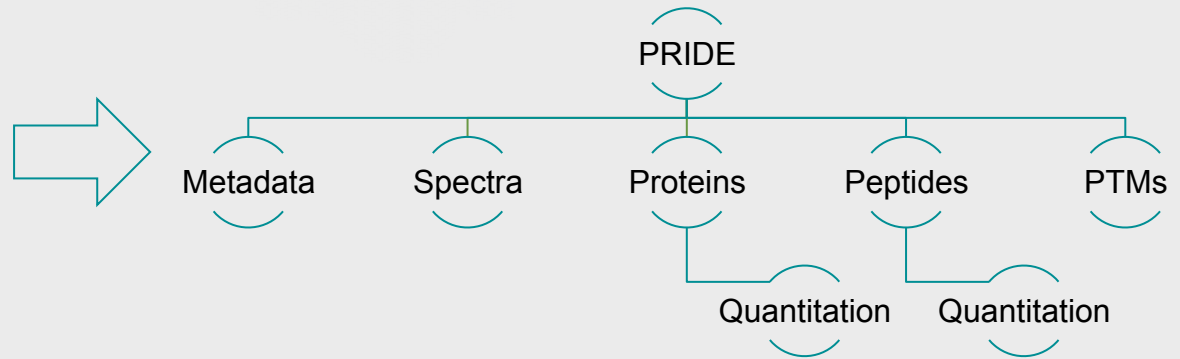
Agenda

- PRIDE Repository
- Challenges
- PRIDE Inspector
- Future Plans

PRoteomics IDEntifications Database (PRIDE)



Public mass spec repository

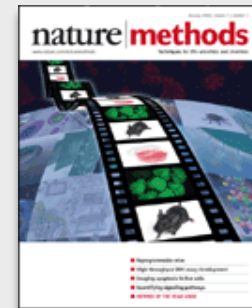
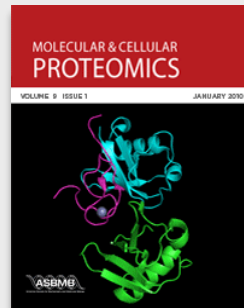
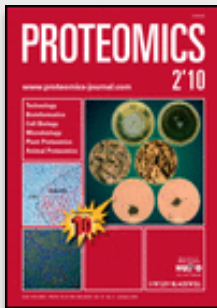


<http://www.ebi.ac.uk/pride>

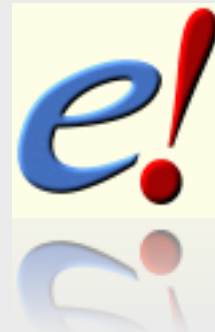
- 151+ M Spectra
- 28+ M Peptides
- 5.8+ M Proteins

PRoteomics IDEntifications Database (PRIDE)

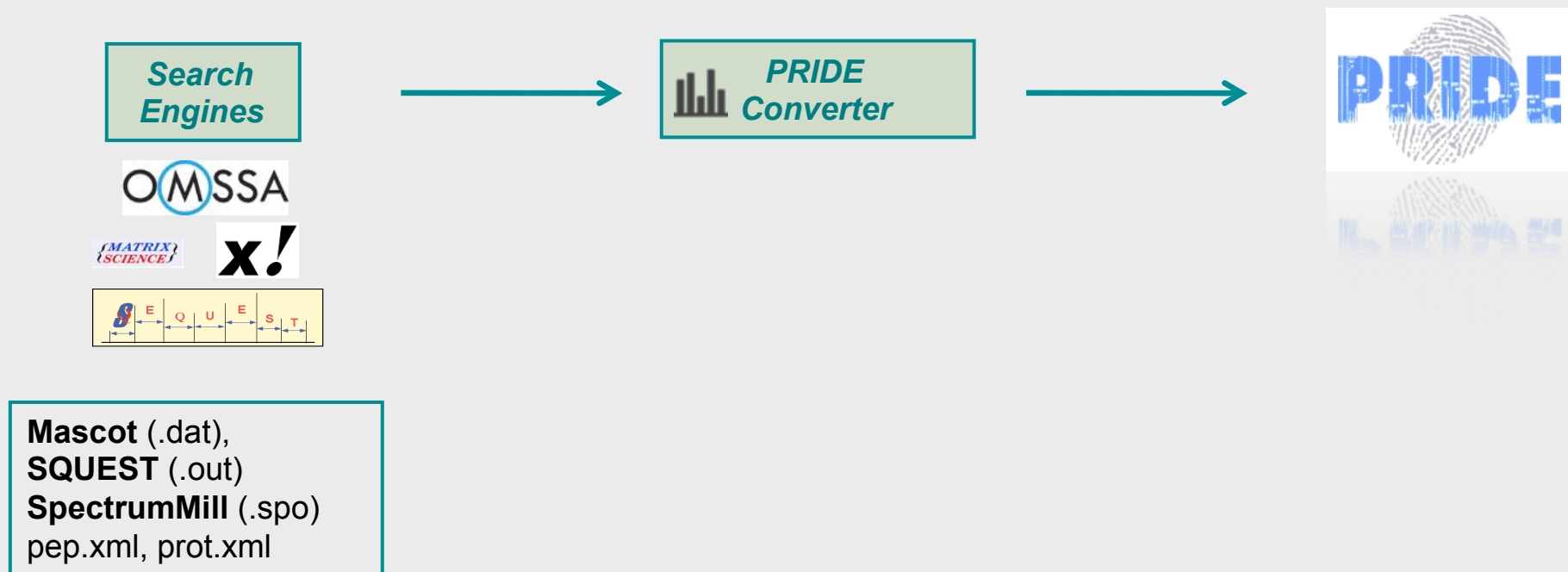
- Support for publications



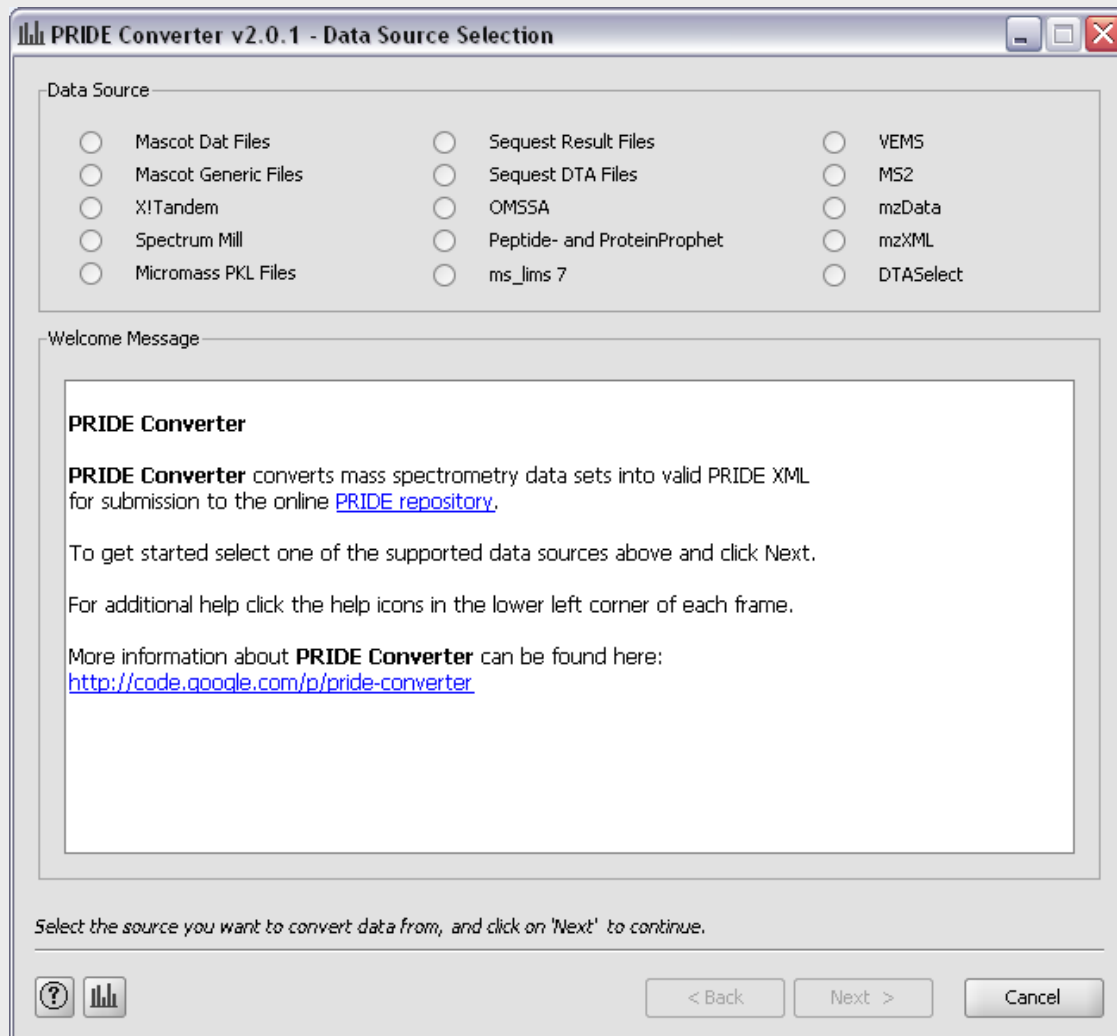
- Reliable source of MS proteomics data for other resources



PRIDE Submission Workflow



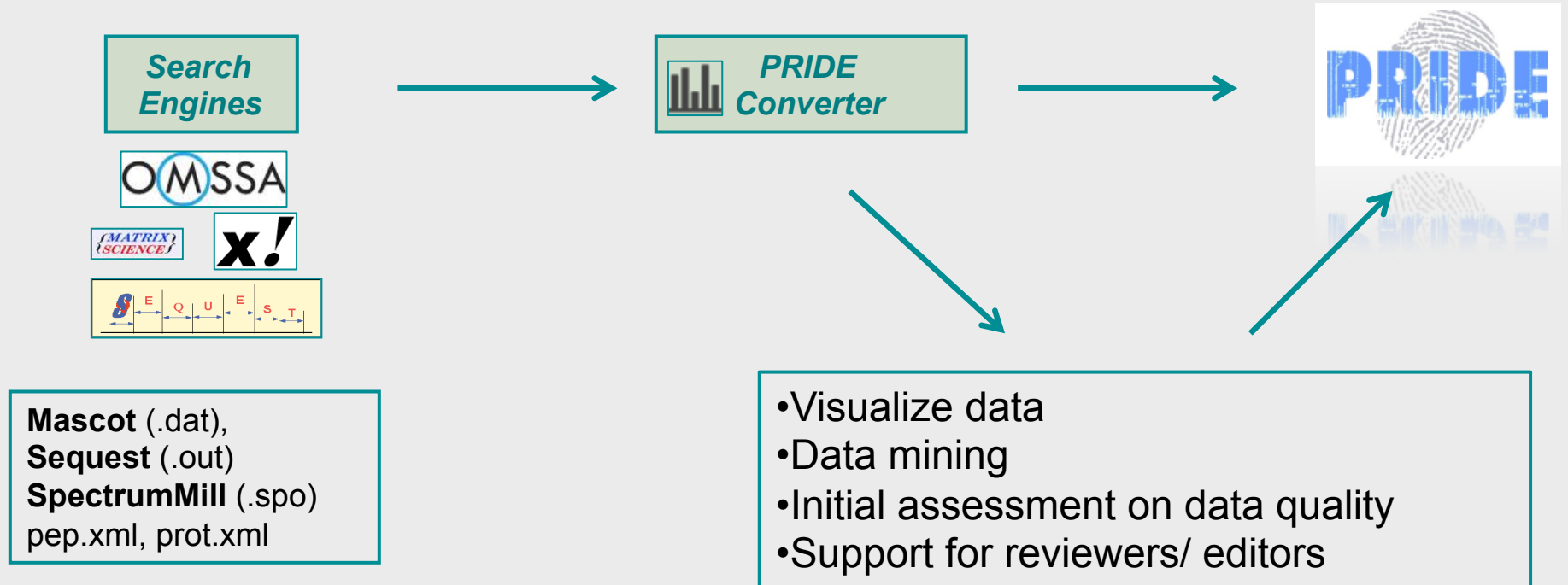
PRIDE Converter



- Stand-alone tool for creating PRIDE XML files
- A variety of popular proteomics data formats
- Metadata using CVs/ Ontologies.
- GUI based tool, easy to use
- New version in development

<http://code.google.com/p/pride-converter>

The need for new tool




PRIDE Inspector

- Standard-alone tool
- 100% Java
- Open Source and Free





<http://code.google.com/p/pride-toolsuite/>

Java Web Start



- Home
- Search
 - Advanced Search
 - Browse
 - PRIDE BioMart
 - Search PTMs by Mass
- Submission Guidelines
- Validate XML
- Documentation
- Publications
- Developer Resources
- Register
- Log in
- External Links
- Contact

News 

8 July 2011 - Ongoing database maintenance
PRIDE is currently undergoing unplanned but necessary database maintenance and normal service should resume by Wednesday, July 13. This means that no new submissions are going to be processed until that time and users are encouraged not to create new user accounts as there might be some disruptions during this time. Thank you for your

[More >>](#)

24 February 2011 - PRIDE Inspector is now available

PRoteomics IDentifications database (PRIDE)

Introduction

The PRIDE PRoteomics IDentifications database is a centralized, standards compliant, public data repository for proteomics data. It has been developed to provide the proteomics community with a public repository for protein and peptide identifications together with the evidence supporting these identifications. PRIDE is also able to capture details of post-translational modifications coordinated relative to the peptides in which they have been found.

Search PRIDE

You can use the search box to perform simple keyword queries. Allowed values for the simple search are:

- PRIDE Experiment accession number (ex: 8500).
- Protein accession: This will search the database and return the experiments where this protein was identified. You can use protein accessions from IPI, UniProtKB, ENSEMBL or RefSeq, and you can also use UniProtKB IDs and NCBI gi numbers.
For example:
 - UniProt accession: P29375
 - UniProt ID: KDM5A_HUMAN
 - IPI: IPI00619935
 - ENSEMBL: ENSP00000382688
 - RefSeq: NP_005047
 - NCBI gi: gi1435778 (please note that the gi prefix is mandatory)

UniProtKB, ENSEMBL and RefSeq accessions be can versioned (e.g. IPI00619935.1).

The [Browse](#) page provides access to datasets via their associated project name or categories. Categories include species, tissue, cell type, GO terms and disease.

More complex searches can be performed using the [Advanced Search](#).

Please also consider using the [PRIDE BioMart](#) for powerful sophisticated searches (also allows searches across data resources, for example linking PRIDE identifications to REACTOME pathways).

Submitting Data to PRIDE

PRIDE encourages and welcomes direct user submissions of protein and peptide identification data to be published in peer-reviewed publications.


PRIDE Basic Statistics

The PRIDE database currently contains:

- 17,066 Experiments
- 5,839,338 Identified Proteins
- 28,527,467 Identified Peptides
- 3,662,382 Unique Peptides
- 151,690,006 Spectra

PRIDE core version: 2.8.8
PRIDE web version: 2.8.8

PRIDE Inspector




Key features: [\[More\]](#)

- Visualize MS data in **PRIDE XML** and **mzML** format
- View ALL public experiments in PRIDE
- Assess **data quality** using plots and charts
- Download and view locally ALL private experiments in PRIDE

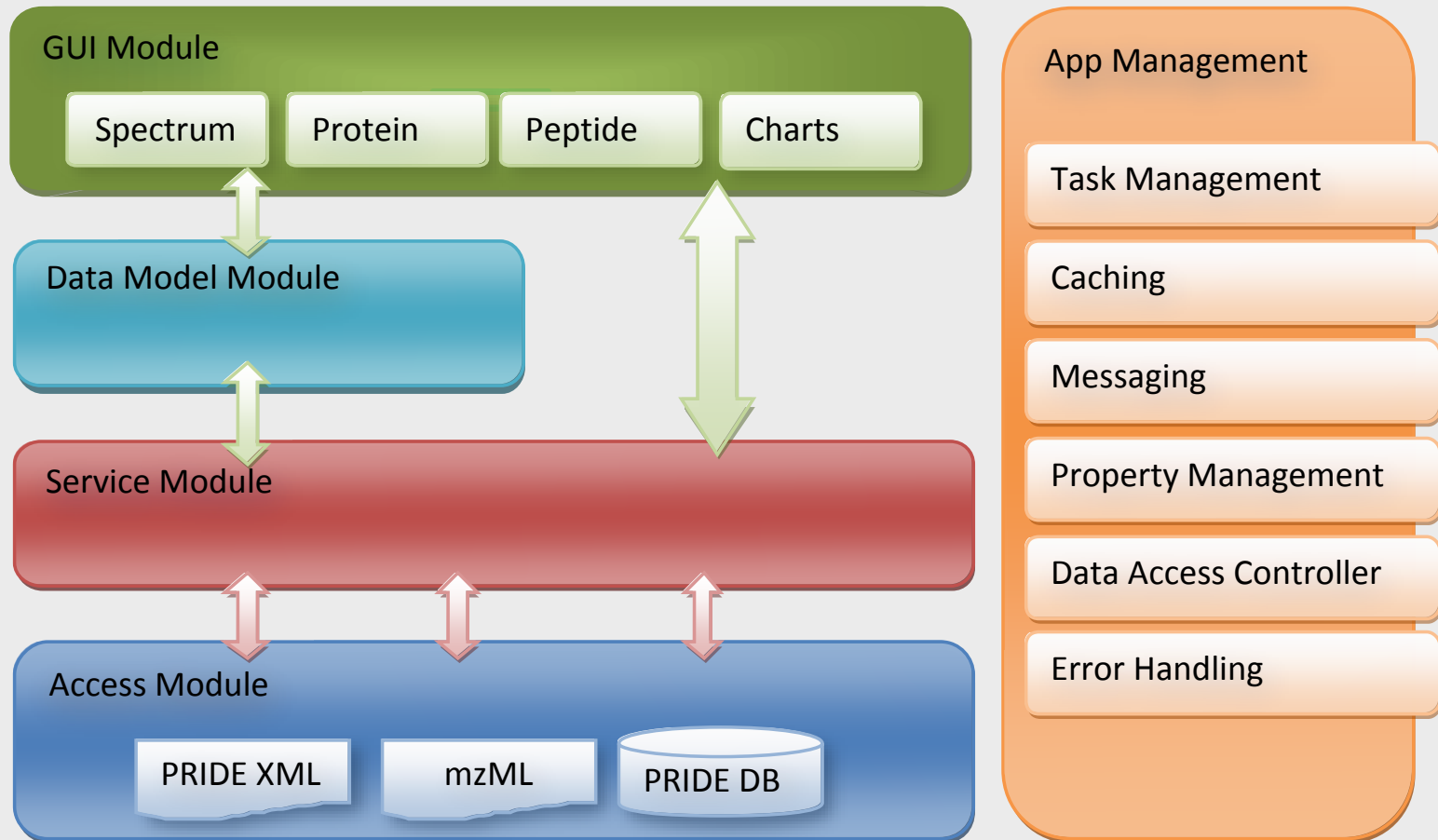
Get it from:

[Download](#) or click the button below to run directly



Java Web Start

Design



Core Features

The screenshot shows the PRIDE Inspector 0.8.5 application window. The interface includes a menu bar (File, Examples, Export, Help), a toolbar, and a 'Welcome' tab. The main content area is divided into sections: 'Quick Start', 'Learn More About', and 'Feedback'. The 'Quick Start' section contains four buttons: 'Open mzML or PRIDE XML Files', 'Connect to PRIDE Database', 'Private Download', and 'Getting Help'. The 'Learn More About' section contains six links: 'PRIDE', 'PRIDE XML', 'PRIDE Public Database', 'mzML', 'Private Download', and 'Quality Assessment Charts'. The 'Feedback' section contains one link: 'Give Us Your Feedback'. Three green callout boxes with arrows point to specific features: 'Fast loading of large *mzML* and *PRIDE XML* files.' points to the 'Open mzML or PRIDE XML Files' button; 'Direct access to all PRIDE public database experiments.' points to the 'PRIDE XML' link; and 'Private experiment download facility.' points to the 'Private Download' button. A large 'PRIDE' logo with a fingerprint background is visible in the bottom right corner of the application window.

PRIDE Inspector 0.8.5

File Examples Export Help

Welcome

Quick Start

- Open mzML or PRIDE XML Files
- Connect to PRIDE Database
- Private Download
- Getting Help

Try Examples

Fast loading of large *mzML* and *PRIDE XML* files.

PRIDE XML

Direct access to all PRIDE public database experiments.

Private experiment download facility.

Learn More About

- PRIDE
- PRIDE XML
- PRIDE Public Database
- mzML
- Private Download
- Quality Assessment Charts

Feedback

- Give Us Your Feedback

PRIDE

Experiment Metadata View

PRIDE Inspector 0.8.5

File Examples Export Help

Data Sources

- PRIDE_Exp_Complete_Ac_1...
- 10018.xml

General Protein (56) Peptide (400) Spectrum (19012) Quality Charts (8)

Contact

Contact name	Ciara McManus
Contact organization	UCD Conway Institute
Contact information	ciara.mcmanus@ucd.ie

General

Accession	10018
Version	2.1
Title	Human Brain 6-11 reference map
Short Label	brain dlpfc 6-11 map
XML generation software	PRIDE Converter v2.0.1
Original MS data file format	Mascot Dat File
Project	Human Brain 6-11 reference map (dlpfc brain region)

Sample

ID	sample1
Name	brain
Parameter	brain
Parameter	human (CS20)
Parameter	Homo sapiens (Human)

Protocol

ID	protocol1
Name	In Gel Protein Digestion
Step 1 - Reduction	DTT
Step 2 - Alkylation	iodoacetamide
Step 3 - Enzyme	Trypsin

Instrument Configuration

ID	LTQ-Orbitrap
Software ID	unknown
Source - Order	1
Source - Parameter	Electrospray Ionization
Analyzer - Order	2
Analyzer - Parameter	LTQ Orbitrap
Detector - Order	3

Experiments

Jump To:

experiment accession

Public Experiments

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24

General

Sample

Protocol

Instrument

Protein Centric View

PRIDE Inspector 0.8.5

File Examples Export Help

Data Sources

- PRIDE_Exp_Complete_Ac_1...
- 10018.xml
- 10885.xml

General Protein (252) Peptide (3099) Spectrum (3099) Quality Charts (8)

Protein Details Type: Gel Free Search Engine: Mascot Search Database: IPI_human

Download Protein Names

	Submitted Protein Accession	Mapped Protein Accession	Score	Threshold	# Peptides	# Distinct Peptides	# PTMs
1	IP100028888	IP100028888	54.57	54.57	1	1	3
2	IP100007423	IP100007423	42.03	42.03	2	2	4
3	IP100418262	IP100418262	70.25	70.25	1	1	1
4	IP100429191	IP100429191	54.5	54.5	2	1	2
5	IP100073772	IP100073772	51.57	51.57	28	6	45
6	IP100012493	IP100012493	45.25	45.25	2	1	2
7	IP100219871	IP100219871	49.12	49.12	4	1	4
8	IP100290462	IP100290462	50.78	50.78	4	2	4
9	IP100470791	IP100470791	49.16	49.16	1		
10	IP100024915	IP100024915	47.8	47.8	7		
11	IP100000816	IP100000816	53.04	53.04	53		
12	IP100026781	IP100026781	59.93	59.93	15	3	15

Selected Protein

Peptide Details PTM: [T - 144.1059]

	Peptide Sequence	Submitted ...	Mapped Pr ...	Precursor C...	Delta m/z	Precursor m/z	# PTMs	Length	Start	Stop	Spectrum	# Ions	Mascot Score
1	TPVEPEVAIHR	IP100012493	IP100012493	3	-0.0591	464.5393	1	11	9	19	837	18	51.67
2	TPVEPEVAIHR	IP100012493	IP100012493	3	-0.0603	464.5381	1	11	9	19	836	10	38.82

Peptides identify the same protein

Intensity

m/z

Spectrum for selected peptide

Fragment Ion

- ☒ immonium ion
- ☒ y ion
- ☒ b ion

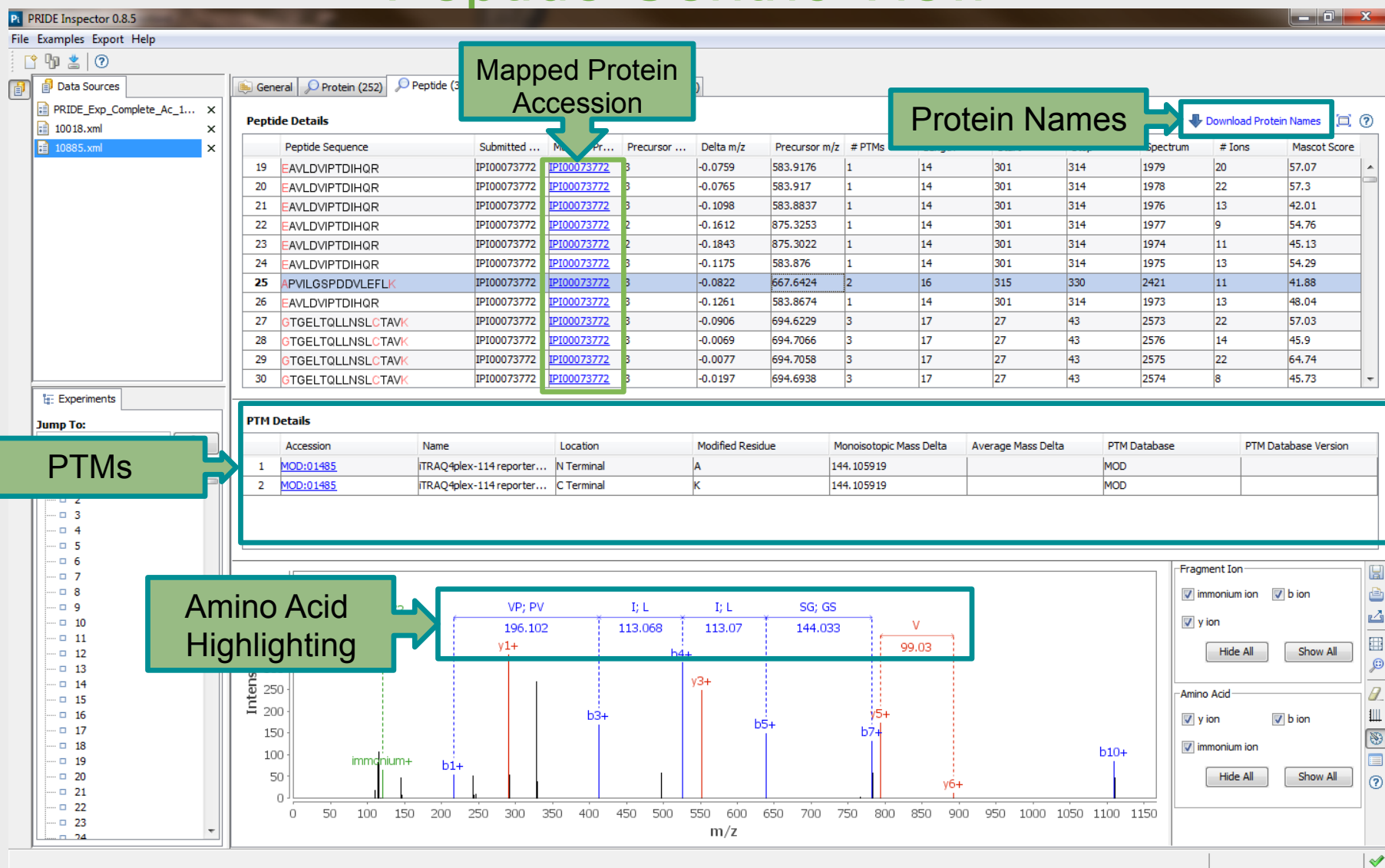
Hide All Show All

Amino Acid

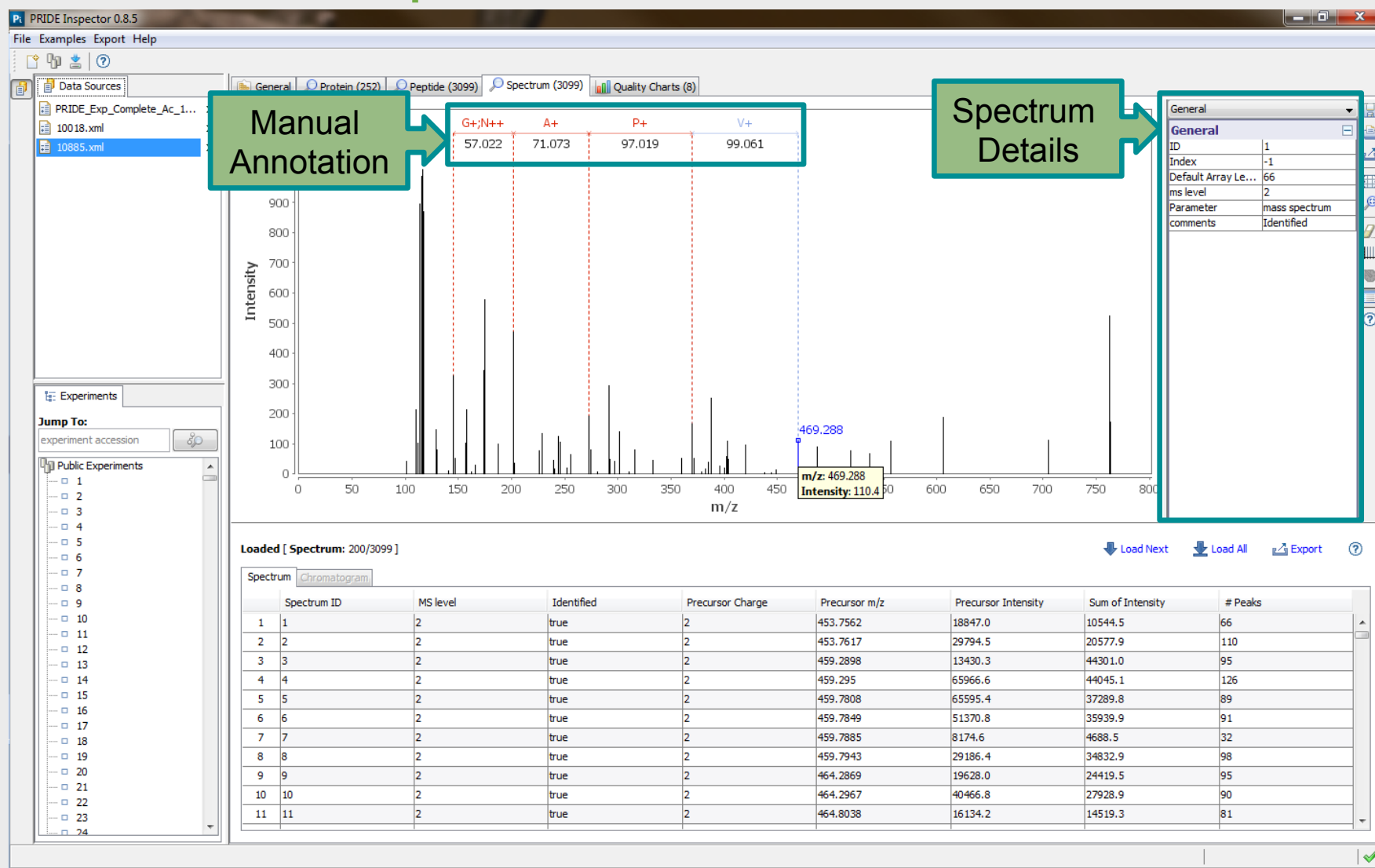
- ☒ y ion
- ☒ b ion
- ☒ immonium ion

Hide All Show All

Peptide Centric View



Spectrum Centric View



Summary Chart View

PRIDE Inspector 1.0.6

File Examples Export Help

PRIDE Inspector - Help

Frequency

- PRIDE Inspector
 - Familiarize Yourself with User Interface
 - Familiarize Yourself with Browsing Data
 - Visualizing Spectra and Chromatogram
 - Understanding Summary Charts
 - Peak Intensity Distribution Chart
 - Precursor Ion Charge Distribution Chart
 - MS2 m/z Distribution Chart
 - Distribution of Precursor Ion Masses
 - Number of Peptides Identified per Protein Chart
 - Number of Peaks per Spectrum Chart
 - Delta m/z Chart
 - Number of Missed Tryptic Cleavages
 - Opening mzML or PRIDE XML Files
 - Trying mzML or PRIDE XML Example Files
 - Connecting to PRIDE Public Database Instance
 - Downloading Your Experiments from PRIDE
 - Exporting Data from PRIDE Inspector
 - FAQ
 - Feedback Page

Number of Peptides Identified per Protein Chart

[Previous](#) | [Next](#) | [See Also](#)

Displays a bar chart with the number of peptides identified per protein for a single PRIDE experiment.

Explanation

A bar chart displaying the percentage of protein identifications in the whole experiment according to the total number of peptides used to report the identification. Proteins supported by more peptide identifications can constitute more confident results.

Note: To investigate further, in the Protein view, one can sort the proteins by number of peptide identifications.

Number of Peptides	Percentage
18	50%
7	19%
1	17%

PRIDE Toolsuite

- PRIDE-Chart: Generating statistical chart
- PRIDE-mzGraph-Browser: Visualizing spectra or chromatogram
- PRIDE-JAXB: Reading PRIDE XML
- jmzML: Reading mzML
- XXIndex: Indexing large XML

Future Plan

- Better support for submitters and reviewers
 - Integration with PRIDE Converter
 - Quantitative data
 - Semantic validation
 - Summary report
- Better support for MS community
 - Additional formats: mzIdentML, MGF
 - Search PRIDE Database

Search PRIDE

PRIDE Inspector 1.0.6

File Examples Export Help

Search for:

☐ Search within results

10359 results found

Row	Selected	View	PRIDE Experiment Accession	Experiment Title	Experiment Short Label	Taxonomy Term (NEWT / NCBI Ta...	Taxonomy ID (NEWT / NCBI Taxon)	Tissue Ontology Term (BRENDA)	BRENDA ID (Tissue)
1	<input checked="" type="checkbox"/>	View	99	Golgi Intact TOTAL N=1	Golgi N=1	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
2	<input type="checkbox"/>	View	100	Golgi Intact TOTAL N=2	Golgi N=2	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
3	<input checked="" type="checkbox"/>	View	101	Golgi Intact TOTAL N=3	Golgi N=3	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
4	<input type="checkbox"/>	View	102	COPI Vesicles N=1	COPI N=1	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
5	<input type="checkbox"/>	View	103	COPI Vesicles N=2	COPI N=2	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
6	<input type="checkbox"/>	View	104	COPI Vesicles N=3	COPI N=3	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
7	<input type="checkbox"/>	View	105	COPI GTP gamma S N=1	COPIGTPgS N=1	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
8	<input type="checkbox"/>	View	106	COPI GTP gamma S N=2	COPIGTPgS N=2	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
9	<input type="checkbox"/>	View	107	COPI GTP gamma S N=3	COPIGTPgS N=3	Rattus norvegicus (Rat)	10116	liver	BTO:0000759
10	<input type="checkbox"/>	View	13	"HUPO Plasma Proteome Project,...	HUPO PPP Lab #2 Expt #17	Homo sapiens (Human)	9606	blood plasma	BTO:0000131
11	<input type="checkbox"/>	View	65	"HUPO Plasma Proteome Project,...	HUPO PPP Lab #28 Expt #7	Homo sapiens (Human)	9606	blood plasma	BTO:0000131
12	<input type="checkbox"/>	View	98	"HUPO Plasma Proteome Project,...	HUPO PPP Lab #55 Expt #35	Homo sapiens (Human)	9606	blood plasma	BTO:0000131
13	<input type="checkbox"/>	View	108	Anonymous data set GPM32000...	GPM32000001463	Homo sapiens (Human)	9606		
14	<input type="checkbox"/>	View	109	Plasma Proteome (GPM1010000...	GPM10100000612	Homo sapiens (Human)	9606		
15	<input type="checkbox"/>	View	110	Aurum spectra T10467 (GPM87...	GPM87400002009	Homo sapiens (Human)	9606		
16	<input type="checkbox"/>	View	111	Aurum spectra T10475 (GPM87...	GPM87400002010	Homo sapiens (Human)	9606		
17	<input type="checkbox"/>	View	112	Aurum spectra T10622 (GPM87...	GPM87400002011	Homo sapiens (Human)	9606		
18	<input type="checkbox"/>	View	113	Aurum spectra T10445 (GPM87...	GPM87400002012	Homo sapiens (Human)	9606		
19	<input type="checkbox"/>	View	114	Aurum spectra T10707 (GPM87...	GPM87400002013	Homo sapiens (Human)	9606		
20	<input type="checkbox"/>	View	115	Aurum spectra T10739 (GPM87...	GPM87400002014	Homo sapiens (Human)	9606		
21	<input type="checkbox"/>	View	116	Aurum spectra T10761 (GPM87...	GPM87400002015	Homo sapiens (Human)	9606		
22	<input type="checkbox"/>	View	117	Plasma Proteome (GPM1010000...	GPM10100000613	Homo sapiens (Human)	9606		
23	<input type="checkbox"/>	View	118	Anonymous data set (GPM10100...	GPM10100000614	Homo sapiens (Human)	9606		
24	<input type="checkbox"/>	View	119	Anonymous data set (GPM10100...	GPM10100000615	Homo sapiens (Human)	9606		
25	<input type="checkbox"/>	View	120	Anonymous data set (GPM10100...	GPM10100000616	Homo sapiens (Human)	9606		
26	<input type="checkbox"/>	View	121	Plasma Proteome (GPM1010000...	GPM10100000617	Homo sapiens (Human)	9606		
27	<input type="checkbox"/>	View	122	Plasma Proteome (GPM1010000...	GPM10100000618	Homo sapiens (Human)	9606		
28	<input type="checkbox"/>	View	123	Plasma Proteome (GPM1010000...	GPM10100000619	Homo sapiens (Human)	9606		
29	<input type="checkbox"/>	View	124	Plasma Proteome (GPM1010000...	GPM10100000620	Homo sapiens (Human)	9606		
30	<input type="checkbox"/>	View	125	Plasma Proteome (GPM1010000...	GPM10100000621	Homo sapiens (Human)	9606		
31	<input type="checkbox"/>	View	126	Plasma Proteome (GPM1010000...	GPM10100000622	Homo sapiens (Human)	9606		
32	<input type="checkbox"/>	View	127	Plasma Proteome (GPM1010000...	GPM10100000623	Homo sapiens (Human)	9606		
33	<input type="checkbox"/>	View	128	Plasma Proteome (GPM1010000...	GPM10100000624	Homo sapiens (Human)	9606		
34	<input type="checkbox"/>	View	129	Plasma Proteome (GPM1010000...	GPM10100000625	Homo sapiens (Human)	9606		

Open selected

Close

Additional Features

PRIDE Inspector 1.0.6

File Examples Export Help

Open File(s)
Connect PRIDE
Private Download
Help

Data Sources
Pride Experiment 101
Pride Experiment 99

Summary Report

- Spectra not found
- Protein identifications four
- Peptides found
- Reference found
- Contacts found
- Sample found
- Protocol found
- Instrument found
- Data processing found
- Software found

Experiment Details Protein (410) Peptide (2186) Spectrum (0) Summary Charts (8)

Protein Details Type: Gel Free Search Engine: Mascot v1.9.0.3 Search Database: UniProt Download Protein Details

Row	Submitted Protein Accession	Mapped Protein Accession	Protein Name	Sequence Coverage	Score	Threshold	# Peptides	# Distinct Peptides	# PTMs
4	P00884	P00884	Fructose-bisphosphate aldolase B (E...	15.9%	158.0	0.0	6	4	0
5	P02651	P02651	Apolipoprotein A-IV (Apo-AIV) (ApoA...	17.6%	435.0	0.0	6	6	0
6	P02706	P02706	Asialoglycoprotein receptor 1 (ASGP-...	11.6%	98.0	0.0	2	2	0
7	P04638	P04638	Apolipoprotein A-II (Apo-AII) (ApoA-II...	10.8%	40.0	0.0	1	1	0
8	P04916	P04916	Retinol-binding protein 4 (Plasma reti...	15.4%	201.0	0.0	4	4	0
9	P05183	P05183	Cytochrome P450 3A2 (EC 1.14.14.1...	2.4%	39.0	0.0	1	1	0
10	P05545	P05545	Serine protease inhibitor A3K (Serp...	39.9%	479.0	0.0	40	11	0
11	P07895	P07895	Superoxide dismutase [Mn], mitochon...	6.3%	75.0	0.0	1	1	0
12	P09875	P09875	UDP-glucuronosyltransferase 2B1 (U...	3%	51.0	0.0	1	1	0
13	P10536	P10536	Ras-related protein Rab-1B	35.3%	443.0	0.0	7	6	0
14	P10867	P10867	L-gulonolactone oxidase (LGO) (EC 1...	5.9%	123.0	0.0	2	2	0
15	P14740	P14740	Dipeptidyl peptidase 4 (EC 3.4.14.5)...	3.8%	130.0	0.0	2	2	0
16	P15999	P15999	ATP synthase subunit alpha, mitoch...	22.4%	118.0	0.0	10	8	0
17	P17046	P17046	Lysosome-associated membrane glyc...	4.9%	58.0	0.0	1	1	0
18	P17764	P17764	Acetyl-CoA acetyltransferase, mitoch...	7.1%	152.0	0.0	2	2	0
19	P22734	P22734	Catechol O-methyltransferase (EC 2...	4.5%	46.0	0.0	1	1	0

Peptide Details [P02651] PTM: NONE

Row	Peptide Sequence	Fit In Sequence	Precursor Charge	Delta m/z	Precursor m/z	# PTMs	PTM List	# Ions	Length	Start	Stop	Spectrum
1	ATIDQNLEDLR	Fit				0		0	11	210	220	
2	FNQNMGLK	Fit				0		0	9	190	198	
3	LGNINTYADDLQNK	Fit				0		0	14	66	79	
4	TDVTQQLNTLFQDK	Fit				0		0	14	52	65	
5	VSQMFQDNVQK	Fit								120	130	
6	VSTNIDQLQK	Fit								256	265	

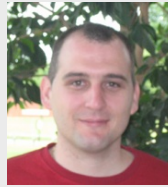
Sequence View

Accession: P02651, Name: Apolipoprotein A-IV (Apo-AIV) (ApoA-IV) (Apolipoprotein A4)
6 peptides (6 matched, 6 distinct), 69/391 amino acids (17.6% coverage)

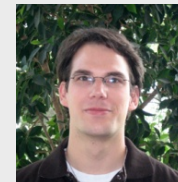
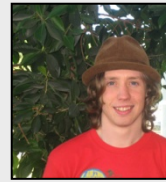
Selected	PTM	Fit	Fuzzy Fit	Overlap						
MFLKAVVLTV	ALVAITGTQA	EVTSDQVANV	MWDYFTQLSN	NAKEAVEQLQ	KTDVTQQLNT	LFQDKLGNIN	TYADDLQNK	VPPFAVQLSGH	LTKETERVRE	100
EIQKELEDLR	ANMMPHANKV	SQMFGDNVQK	LQEHLPYAT	DLQAQINAGT	QDMKRQLTPY	IQRMTTIQD	NVENLQSSMV	PFANELKEKF	NONMEGLKGG	200
LTPRANELKA	TIDQNLEDLR	SRLAPLAEGV	QEKLNHQMEG	LAFQMKKNAE	ELQTKVSTNI	DQLQKNLAPL	VEDVQSKLKG	NTEGLQKSL	DLNKQLDQGV	300
EVFRRAVEPL	GDKFNMAVLQ	QMEKFRQQLG	SDSGDVESHL	SFLEKNLREK	VSSFMSTLQK	KGSPDQPLAL	PLPEQVQEQV	QEQVQPKPLE	S	391

Acknowledgements

Antonio Fabregat
Daniel Ríos



Florian Reisinger
Joe Foster
David Ovelleiro
Johannes Griss

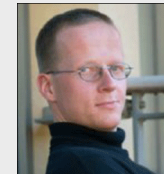


Attila Csordas
Richard Côté



Lennart Martens

Juan Antonio Vizcaíno
Henning Hermjakob



Funding Sources

Funding



Wellcome Trust Grant (PRIDE)



ProteomeXchange (EU FP7 grant number 260558)



SLING (EU FP7 grant number 226073)

Questions & Feedbacks



Contact us:

pride-support@ebi.ac.uk