

Protein mass spectrometry and proteomics

L. Dayon

Spring Semester 2026

EPFL

Course outline

1. Proteomics and mass spectrometry

Introduction to protein analysis and proteomics; Reminders in mass spectrometry; Why proteomics and mass spectrometry?; Ionization sources, analysers, and detectors used in proteomics; Latest generation of mass spectrometers used in proteomics

2. Mass spectrometry-based proteomic strategy and workflows

Bottom-up versus top-down strategies; Data-dependent acquisition (DDA) and data-independent acquisition (DIA) approaches; Sample preparation

Lab visit of the Proteomics Laboratory at Nestlé Research (EPFL Innovation Park)

3. Quantitative proteomic workflows

Label-free methods; Labelling-based techniques; Other quantitative techniques

4. Proteomic bioinformatics

Databases; Identification of protein; Quantification of proteins; Bioinformatics tools; Practical examples

5. Applications to biology, clinical research, and beyond

What strategy?; Experimental design & randomization; Biomarker discovery; Industrialized and population proteomics; Forensics; Targeted mass spectrometry-based approaches; Other biological applications of mass spectrometry; Advanced innovations (single-cells, 4D proteomics, multi-omics) and emerging technologies; Limitations and ethical consideration

Additional support/information on [Separations techniques in proteomics](#)

Course outline

- 4. Proteomic bioinformatics

Databases; Identification of protein; Quantification of proteins;
Bioinformatics tools; Practical examples

4.1. Databases

Sequence database

DDBJ

GeneCards

InterPro

NCBI

nr

OWL

PIR

PRF

UniProt

SYSTEMS

Web site

<http://www.ddbj.nig.ac.jp>

<http://bioinfo.weizmann.ac.il/cards/>

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

<http://www.ncbi.nlm.nih.gov>

<http://www.ncbi.nlm.nih.gov/BLAST/>

<http://www.bioinf.man.ac.uk/dbbrowser/OWL>

<http://pir.georgetown.edu>

http://www.genome.ad.jp/htbin/www_bfind?prf

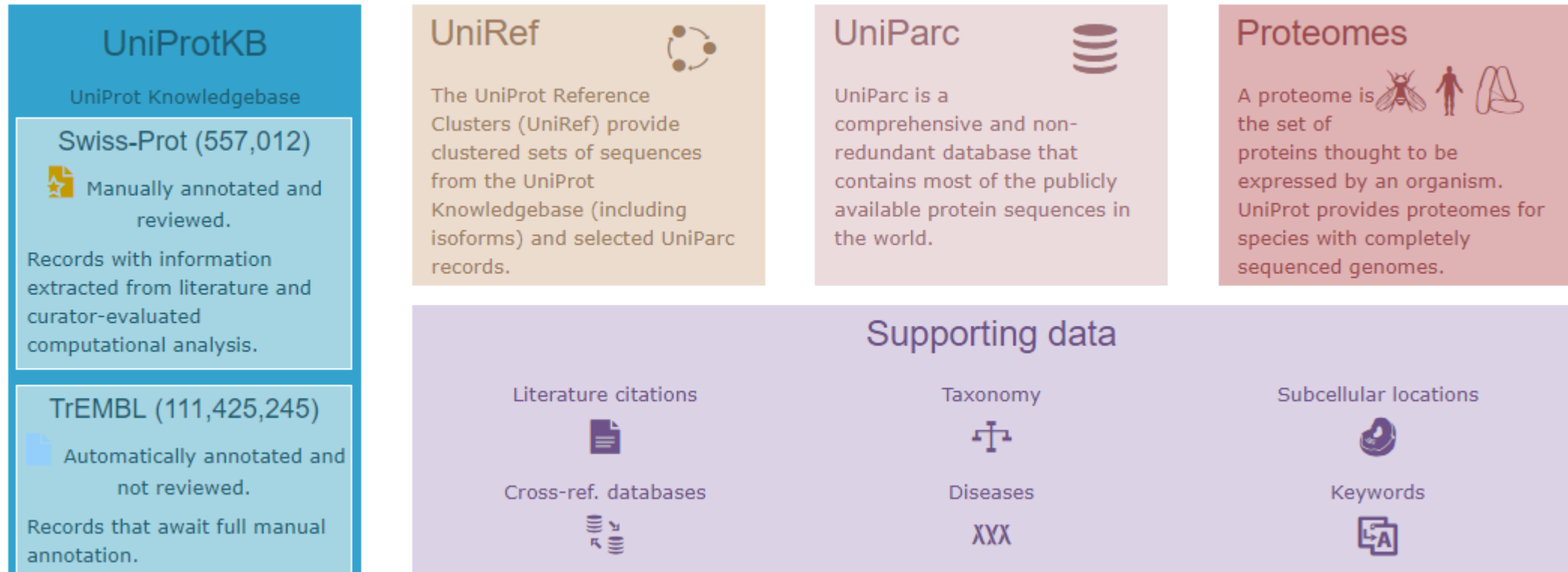
<http://www.uniprot.org>

<http://systems.molgen.mpg.de>

DOI: 10.1002/0471142727.mb1904s68



UniProtKB

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.



Information available at UniProtKB

UniProtKB - Q99497 (PARK7_HUMAN)

 Basket 






Display

- Entry
- Publications
- Feature viewer
- Feature table

All None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

[Top](#)

 BLAST  Align  Format  Add to basket  History

 Feedback  Help video  Other tutorials and videos




Protein | Protein/nucleic acid deglycase DJ-1

Gene | PARK7



Organism | *Homo sapiens (Human)*



Status |  Reviewed - Annotation score: ●●●●●● - Experimental evidence at protein levelⁱ



Functionⁱ

Protein and nucleotide deglycase that catalyzes the deglycation of the Maillard adducts formed between amino groups of proteins or nucleotides and reactive carbonyl groups of glyoxals (PubMed:25416785, PubMed:28596309). Thus, functions as a protein deglycase that repairs methylglyoxal- and glyoxal-glycated proteins, and releases repaired proteins and lactate or glycolate, respectively. Deglycates cysteine, arginine and lysine residues in proteins, and thus reactivates these proteins by reversing glycation by glyoxals. Acts on early glycation intermediates (hemithioacetals and aminocarbinals), preventing the formation of advanced glycation endproducts (AGE) that cause irreversible damage (PubMed:25416785, PubMed:28013050, PubMed:26995087). Also functions as a nucleotide deglycase able to repair glycated guanine in the free nucleotide pool (GTP, GDP, GMP, dGTP) and in DNA and RNA. Is thus involved in a major nucleotide repair system named guanine glycation repair (GG repair), dedicated to reversing methylglyoxal and glyoxal damage via nucleotide sanitization and direct nucleic acid repair (PubMed:28596309). Also displays an apparent glyoxalase activity that in fact reflects its deglycase activity (PubMed:22523093). Plays an important role in cell protection against oxidative stress and cell death acting as oxidative stress sensor and redox-sensitive chaperone and protease; functions probably related to its primary function (PubMed:17015834, PubMed:20304780, PubMed:18711745, PubMed:12796482, PubMed:19229105, PubMed:25416785, PubMed:26995087). It is involved in neuroprotective mechanisms like the stabilization of NFE2L2 and PINK1 proteins, male fertility as a positive regulator of androgen signaling pathway as well as cell growth and transformation through, for instance, the modulation of NF-kappa-B signaling pathway (PubMed:12612053, PubMed:15502874, PubMed:14749723, PubMed:17015834, PubMed:21097510, PubMed:18711745). Eliminates hydrogen peroxide and protects cells against hydrogen peroxide-induced cell death (PubMed:16390825). Required for correct mitochondrial morphology and function as well as for autophagy of dysfunctional mitochondria (PubMed:19229105, PubMed:16632486). Plays a role in regulating expression or stability of the mitochondrial uncoupling proteins SLC25A14 and SLC25A27 in dopaminergic neurons of the substantia nigra pars compacta and attenuates the oxidative stress induced by calcium entry into the neurons via L-type channels during pacemaking (PubMed:18711745). Regulates astrocyte inflammatory responses, may modulate lipid rafts-dependent endocytosis in astrocytes and neuronal cells (PubMed:23847046). In pancreatic islets, involved in the maintenance of mitochondrial reactive oxygen species (ROS) levels and glucose homeostasis in an age- and diet dependent manner. Protects pancreatic beta cells from cell death induced by inflammatory and cytotoxic setting (By similarity). Binds to a number of mRNAs containing multiple copies of GG or CC motifs and partially inhibits their translation but dissociates following oxidative stress (PubMed:18626009). Metal-binding protein able to bind copper as well as toxic mercury ions, enhances the cell protection mechanism against induced metal toxicity (PubMed:23792957). In macrophages, interacts with the NADPH oxidase subunit NCF1 to direct NADPH oxidase-dependent ROS production, and protects against sepsis (By similarity).  By similarity  24 Publications 

Catalytic activityⁱ

An N(omega)-(1-hydroxy-2-oxopropyl)-[protein]-L-arginine + H₂O = a [protein]-L-arginine + (R)-lactate.  1 Publication 

An N_ε-(1-hydroxy-2-oxopropyl)-[protein]-L-lysine + H₂O = a [protein]-L-lysine + (R)-lactate.  1 Publication 

An S-(1-hydroxy-2-oxopropyl)-[protein]-L-cysteine + H₂O = a [protein]-L-cysteine + (R)-lactate.  1 Publication 

<http://www.uniprot.org/uniprot/Q99497>

FASTA files (download from the database)

UniProtKB results

Filter byⁱ

Reviewed (20,316)
Swiss-Prot

Popular organisms

Human (20,316)

Proteomes

UP000005640 (20,303)

View by

Results table

Taxonomy

Keywords

Gene Ontology

Enzyme class

Pathway

UniRef

Your results in sequence clusters
with identity of:
100%, 90% or 50%

Demo

Help video

BLAST Align Download Add to basket Columns

1 to 25 of 20,316 Show 25

Expand search

- Download selected (20316)
- Download all (20316)
- Format: FASTA (canonical)
- Compressed Uncompressed
- Preview first 10ⁱ

All 20,316 results selected. (or select only the results on this page)

Entry	Entry	Gene names	Organism	Length	Mass
<input checked="" type="checkbox"/>	L0R819	ASDURF	Homo sapiens (Human)	96	11,250
<input checked="" type="checkbox"/>	P98196	Probable prospaninid-transporting ...	Homo sapiens (Human)	1,134	129,756
<input checked="" type="checkbox"/>	P50993	Sodium/potassium-transporting ATPas...	Homo sapiens (Human)	1,020	112,265
<input checked="" type="checkbox"/>	Q13733	Sodium/potassium-transporting ATPas...	Homo sapiens (Human)	1,029	114,166
<input checked="" type="checkbox"/>	P05026	Sodium/potassium-transporting ATPas...	Homo sapiens (Human)	303	35,061
<input checked="" type="checkbox"/>	O14983	Sarcoplasmic/endoplasmic reticulum ...	Homo sapiens (Human)	1,001	110,252
<input checked="" type="checkbox"/>	Q93084	Sarcoplasmic/endoplasmic reticulum ...	Homo sapiens (Human)	1,043	113,977
<input checked="" type="checkbox"/>	Q01814	Plasma membrane calcium-transportin...	Homo sapiens (Human)	1,243	136,876
<input checked="" type="checkbox"/>	Q5VTU8	ATP synthase subunit epsilon-like p...	Homo sapiens (Human)	51	5,807
<input checked="" type="checkbox"/>	Q9ULK2	Ataxin-7-like protein 1	Homo sapiens (Human)	861	91,514
<input checked="" type="checkbox"/>	O60423	Phospholipid-transporting ATPase IK	Homo sapiens (Human)	1,300	146,752
<input checked="" type="checkbox"/>	Q8NBU5	ATPase family AAA domain-containing...	Homo sapiens (Human)	361	40,744
<input checked="" type="checkbox"/>	Q9ULI0	ATPase family AAA domain-containing...	Homo sapiens (Human)	1,458	164,914

Example of FASTA files

In bioinformatics, FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which nucleotides or amino acids are represented using single-letter codes.

```
>db|UniqueIdentifier|EntryName ProteinName OS=OrganismName OX=OrganismIdentifier [GN=GeneName  
]PE=ProteinExistence SV=SequenceVersion
```

```
>sp|Q99497|PARK7_HUMAN Protein DJ-1 OS=Homo sapiens GN=PARK7 PE=1  
SV=2  
MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTWAGLAGKDPVQC SRD VVICPDASLED  
AKKEGPDYVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG  
SKVTTHPLAKDKMMNGGHYTYSEN RV EKDGLILTSRGP GTSFEFALAI VEALNGKEVAAQ  
VKAPLV LKD
```

```
>sp|Q99497|PARK7_HUMAN Protein DJ-1 OS=Homo sapiens GN=PARK7 PE=1  
SV=2  
MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTWAGLAGKDPVQC SRD VVICPDASLED  
AKKEGPDYVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG  
SKVTTHPLAKDKMMNGGHYTYSEN RV EKDGLILTSRGP GTSFEFALAI VEALNGKEVAAQ  
VKAPLV LKD  
>sp|Q6UWI2|PARN1_HUMAN Prostate androgen-regulated mucin-like  
protein 1 OS=Homo sapiens GN=PARN1 PE=1 SV=1  
MVYKTLFALCILTAGWRVQSLPTSAPLSVSLPTNIWVPTTIWTS SPQNTDADTASPSNGT  
HNMSVLPVTASAPTSLLPKMISIESREEEITSPGSNWEGTNTDPSPSGFSSTSGGVHLTT  
TLEEHSSGTPEAGVAATLSQSAAEPTLISPQAPASSPSSLSTSPPEVFSASVTIHTSSST  
VTSTQPTGAPTAPESPT EESSSDHTPTSHATAEPVPQEKTPPTTVSGKVMCELIDMETTT  
TFPRVIMQEV EHALSSGSAIAITVTVI AVVLLVFGVAAYLKIRHSSYGRLLDDHDDYGSWG  
NYNMP LYDDS  
>sp|Q9H300|PARL_HUMAN Presenilins-associated rhomboid-like  
protein, mitochondrial OS=Homo sapiens GN=PARL PE=1 SV=2  
MAWRGWAQRGWGCGQAWGASVGGRSCEELTAVLTPQLLGRFRNFIIQQKCGFRKPRKV  
EPRRSDPGTSGEAYKRSALIPPVEETVVFYSPYPPIRSLIKPLFFTGVGTGCAFQSAAIWQ  
YESLKS RVQSYFDG IKADWLD S IRPQKEGD FRKEINKWNNLSDGQRTVTGITAANVLF  
CLWRVPSLQRTMIRYFTSNPASKVLCSPMLLSTFHFSLFHMAANMYVLSVSSSIWNIL  
GQEQFMAYLSAGVISNFVSYVGVK VATGRYGPLGASGAIMTVLAAVCTKIPEGR LAIIF  
LPMFTFTAGMALKAI IAMD TAGMILGWKFFDHA AHLGGALFGI WYVTYGHEL IWK NREPL  
VKIWEIR TNGPKKGGGSK  
>sp|095453|PARN_HUMAN Poly(A)-specific ribonuclease PARN OS=Homo  
sapiens GN=PARN PE=1 SV=1  
MEIIRS NFKSNLHKVYQAI EEA DFFAIDG EFGSISDGPVSALTNGFDTP EERYQK LKKH  
SMD FLLFQFGLCTFKYDYTD SKYITKSNFNYVFPKP FMRSSPDVWFVQCQSSSIDFLASQG  
FD FNVFRNGI PYL NQEEERQLREQYDEKRSQANGAGALSYVSPNTSKC PVTTIPEDQKFK  
IDQVVEKIEDLLQSEENK NLDLEPCTGFQRKLIYQTL SWKYPKG IHWLELETEKERYIV  
ISKVDEEERKRREQQKHAK EQEELND AVGFSRVIHA IANSGK LVI GHNMLLDVMHTVHQF  
YCP L PADLSEFKEMTT CVFPRLD LDKLMASTQP FKDIINNTSLAELEKRLKETFPNPKV  
ESAEGFP SYDTASEQLHEAGYDAYITGLCFI SMANYLGSFLSPPKIHVSARSK LIEPFFN  
KLF LMRVMDI PYL NLEGPDLQPKRDHVLHVTFPKEMKTS DLYQLFSAFGNIQISWIDDT S  
AFVSLSQPEQVKI AVNTSKYAE SYRIQTYAEYMRKQEEKQIKRKWTD SWKEADSKRLN  
PQCIPYTLQNHYYRNNSFTAPSTVGRNLSPSQEEAGLEDGVS GEISDTELEQTDSCAEP  
LSEGRKAKKLRMKKELSPAGSISKMS PATLFEVPDTW  
>sp|P09874|PARP1_HUMAN Poly [ADP-ribose] polymerase 1 OS=Homo  
sapiens GN=PARP1 PE=1 SV=4  
MAESSDKLYRVEYAKS GRASCKKCS ESIPKDSL RMAIMVQSPMFDGKVP HWHYFSCFWKV  
GHSIRHPDVEVDG FSELWDDQ QKVKTA EAGGVTKGQD GIGSKAEKTLGDFAAEYAKS  
NRSTCKGCMKIEKGQVRLSKKMVDPEK PQLGMIDRWYHPGCFVKNREELGFRPEYSASQ  
LKGFSLLATEDKEALKKQLPGVKSEGKRKGEVDG VDEVAKKSKKEDKDKSLKALKALKA  
QNDLIWNIKDELK KVCSTNDLKE LLI FNKQVPSGESAILDRVADGMVFGALLPCEEC SG  
QLVFKSDAYCTGDVTAWTKCMVKTQTPNRKEWVTPKEFREISYLKRLKVKKQDRIFPPE  
TSASVAATPPPSTASAPA AVNNSASADKPLSNMKILTLGKLSRNKDEVKAMIEKLGKLT  
GTANKASLCISTKKEVEKMMKMEVKEANIRVVSDFLQDVSASTKSLQELFLAHLIS P  
WGAEVKAE PVEVVA PRGKSGAALSKKSKGQVKEEGINKSEKRMKLT LKGGAAVDPDSGLE  
HSAHVLEKGGKVF SATLGLVDIVKGTNSYKQLQLLEDDKENRYMIFRSWGRVGTVIGSNK  
LEQMP SKEDAIEHFMKLYE EKTGNAWHSKNMFTKYPKKFPYPLEIDYGQDEAVKLT VNPQ  
TKSKLPKPVQDLIKMIFDVE SMKKANWEYEIDLQKMP LGLSKRQLQAAYSILSEVQQAV  
SQGSSDSQILDLSNRFYTLIPHDFGMKPP LLNNADSVQAKVEMLDNLLDIEVAYS LLRG  
GSDDSSKDPIDVNYEKLKTDIKVVD RDEAEAIIRKYVKNTHATTHNAYDLEVIDIFKIE
```

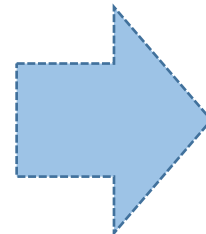
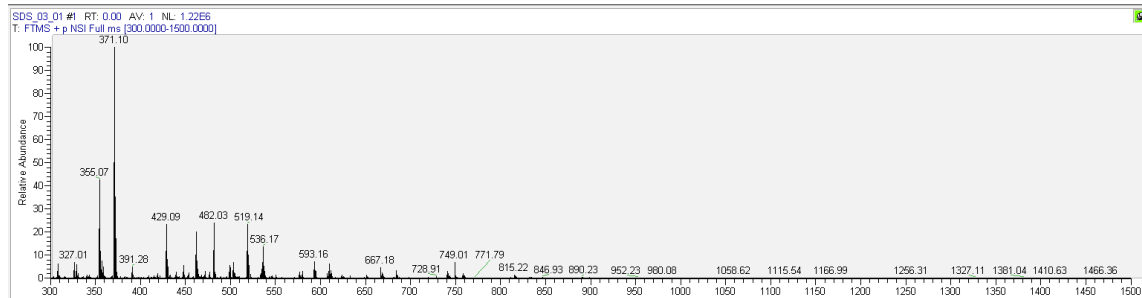
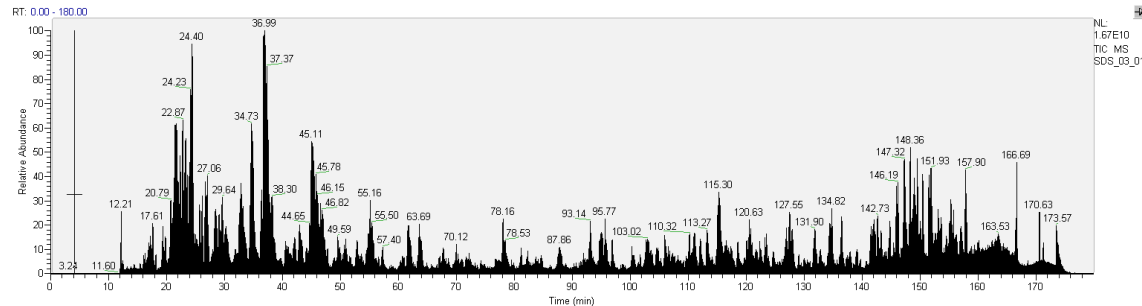
4.2. Identification of protein

Search Engine

Name	Type	Description
Andromeda (part of MaxQuant)	freeware	Andromeda is a peptide search engine based on probabilistic scoring. On proteome data, Andromeda performs as well as Mascot, a widely used commercial search engine, as judged by sensitivity and specificity analysis based on target decoy searches. It can handle data with arbitrarily high fragment mass accuracy, it is able to assign and score complex patterns of post-translational modifications, such as highly phosphorylated peptides, and accommodates extremely large databases.
Byonic	proprietary	Database search algorithm released in 2011 by Protein Metrics Inc. with original developments at PARC that searches MS/MS data from all types of instruments and internally employs the program Combyne, which combines peptide identifications to produce protein scores and identification probabilities.
Comet	open source	Database search algorithm developed at the University of Washington available for Windows and Linux. Note that Comet is just a single command line binary that does MS/MS database search. It takes in spectra in some supported input format and writes out .pep.xml, .pin.xml, .sqt and/or .out files. You will need some other support tool(s) to actually make use of Comet results (A GUI for Windows only is available).
Mascot	proprietary	Performs mass spectrometry data analysis through a statistical evaluation of matches between observed and projected peptide fragments.
OMSSA	freeware	The Open Mass Spectrometry Search Algorithm (OMSSA) is an efficient search engine for identifying MS/MS peptide spectra by searching libraries of known protein sequences. OMSSA scores significant hits with a probability score developed using classical hypothesis testing, the same statistical method used in BLAST. It is developed at the National Center for Biotechnology Information.
ProteinPilot Software	proprietary	Uses Paragon database search algorithm that combines the generation of short sequence tags ('taglets') for computation of sequence temperature values and estimates of feature probabilities to enable the peptide identification considering hundreds of modifications, non-tryptic cleavages and amino acid substitutions. Uses the Pro Group Algorithm for protein inference analysis to report the minimal set of proteins justified based on the peptide evidence. Supports quantification for label-based workflows (iTRAQ reagents, mTRAQ reagents and SILAC labeling). A translation layer translates user interface controls in the language of the proteomics experimental scientist to underlying complex informatics parameters.
Protein Prospector	open source	Protein Prospector is a package of about twenty proteomic analysis tools developed at the University of California San Francisco. The tandem mass spectrometry searching software is Batch-Tag / Batch-Tag Web, with the results processed and displayed using Search Compare. It uses scoring systems tailored to instrument and fragmentation mode to optimize analysis of different types of fragmentation data.
SEQUEST	proprietary	Identifies collections of tandem mass spectra to peptide sequences that have been generated from databases of protein sequences.
X!Tandem	open source	Matches tandem mass spectra with peptide sequences.

Conversion of mass spectra into peak lists (1)

MS raw files



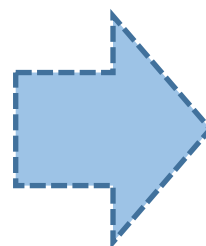
Conversion tool

1. General Settings	
Precursor Selection	Use MS1 Precursor
2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
Lowest Charge State	0
Highest Charge State	0
Min. Precursor Mass	350 Da
Max. Precursor Mass	5000 Da
Total Intensity Threshold	0
Minimum Peak Count	1
3. Scan Event Filters	
Mass Analyzer	Any
MS Order	Is MS2
Activation Type	Any
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Is Full
Ionization Source	Any
Polarity Mode	Any
4. Peak Filters	
S/N Threshold (FT-only)	1.5
5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacements	Automatic
Unrecognized Mass Analyzer Replacements	ITMS
Unrecognized MS Order Replacements	MS2
Unrecognized Activation Type Replacements	CID
Unrecognized Polarity Replacements	+

Conversion of mass spectra into peak lists (2)

Conversion tool

1. General Settings	
Precursor Selection	Use MS1 Precursor
2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
Lowest Charge State	0
Highest Charge State	0
Min. Precursor Mass	350 Da
Max. Precursor Mass	5000 Da
Total Intensity Threshold	0
Minimum Peak Count	1
3. Scan Event Filters	
Mass Analyzer	Any
MS Order	Is MS2
Activation Type	Any
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Is Full
Ionization Source	Any
Polarity Mode	Any
4. Peak Filters	
S/N Threshold (FT-only)	1.5
5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacements	Automatic
Unrecognized Mass Analyzer Replacements	ITMS
Unrecognized MS Order Replacements	MS2
Unrecognized Activation Type Replacements	CID
Unrecognized Polarity Replacements	+



Peak list

```
BEGIN IONS
TITLE=Spectrum41093 scans:49959,
PEPMASS=314.71024002909854
CHARGE=2+
INSTRUMENT=Default
107.049 814.9
110.07123 25070.0
110.07507 891.1
111.05541 3336.0
112.07589 3249.0
112.08685 868.8
115.08643 758.8
116.07055 1225.0
117.07008 1346.0
120.08083 1692.0
123.11691 1327.0
124.07595 939.9
126.12762 19310.0
127.12467 19000.0
127.13026 1033.0
128.12878 2244.0
128.13458 8418.0
129.07013 812.0
129.13162 5651.0
130.14146 4147.0
131.08568 1146.0
131.13841 1642.0
136.0756 1564.0
138.09106 3231.0
141.10226 1699.0
143.08368 1916.0
145.05014 921.6
146.93361 915.2
149.05997 943.9
150.0914 1632.0
155.11803 3612.0
155.9746 1289.0
156.09131 7085.0
156.11272 942.8
158.09264 1781.0
158.76318 952.6
159.09698 1776.0
159.11717 1217.0
162.06616 1309.0
1093.6023 16590.0
1107.6099 64190.0
1108.6106 194100.0
1109.6106 119800.0
1110.6208 171500.0
1111.6179 126700.0
1112.6206 5077.0
1178.5533 1390.0
END IONS

BEGIN IONS
TITLE=Spectrum40214 scans:49000,
PEPMASS=633.4154700290985
CHARGE=2+
INSTRUMENT=Default
101.07073 2256.0
102.05476 2600.0
110.06744 2277.0
110.0713 54330.0
112.07542 7924.0
120.08075 19940.0
121.08384 1454.0
126.12772 244400.0
127.12476 280500.0
128.12831 13070.0
128.13437 205900.0
129.13147 147000.0
129.13731 11300.0
130.06488 1983.0
130.13487 7022.0
130.1411 121000.0
131.08524 1335.0
131.13818 82400.0
131.14351 2647.0
136.07568 19370.0
137.07887 2764.0
138.0912 9590.0
142.08595 8976.0
152.10661 8178.0
156.12227 2307.0
159.09155 14260.0
...
```

A common search software: Mascot



Search this site

Home Mascot database search Products Technical support Training News Blog Newsletter Contact
Access Mascot Server | Database search help

Mascot database search > Access Mascot Server > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name Email

Search title

Database(s)
Amino acid (AA)
cRAP
SARS-CoV-2
SwissProt
UP186698_X_iaevis
UP1940_C_elegans
UP2195_D_discoideum
UP219602_F_oxysporum
UP2311_S_cerevisiae
UP241690_T_harzianum

Taxonomy

Enzyme Allow up to missed cleavages

Quantitation

Crosslinking

Fixed modifications
Display all modifications

Variable modifications
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Carbamidomethyl (C)
Carbamidomethyl (N-term)
Carbamyl (K)
Carbamyl (N-term)
Carboxymethyl (C)

Peptide tol. \pm Da ¹³C MS/MS tol. \pm Da

Peptide charge Monoisotopic Average

Data file No file chosen

Data format Precursor m/z

Instrument Error tolerant

Decoy Report top hits

http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS

Mascot search results (1)

MASCOT Search Results

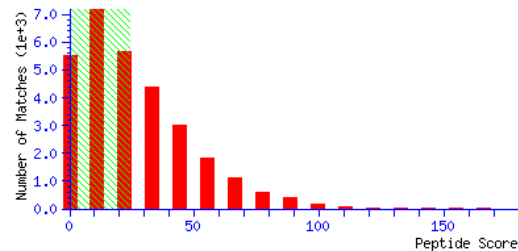
User : Discoverer_LabPC
E-mail :
Search title : Control_Hela_Helios_03_Node:2
MS data file : File Name: Control_Hela_Helios_03.raw
Database : SP2017_07_Human 20171018 (20,224 sequences; 11,332,467 residues)
Timestamp : 4 Apr 2018 at 05:55:46 GMT

Re-search All Non-significant Unassigned [\[help\]](#) Export As XML

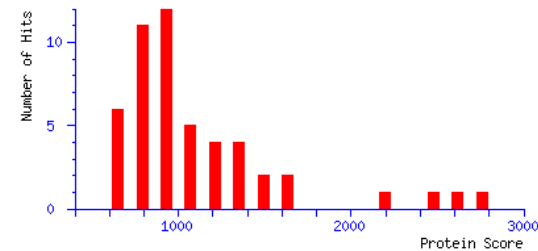
Not what you expected? Try [the select summary](#).

Search parameters

Score distribution



Peptide score distribution. Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 24 indicate **identity** or **extensive homology** ($p < 0.05$).



[Deprecated] Score distribution for family members in the first 50 proteins. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein families.

Legend

Protein Family Summary

Filter Significance threshold $p <$ Max. number of families [\[help\]](#)
Ions score or expect cut-off Dendrograms cut at
Show Percolator scores
Preferred taxonomy

Decoy search summary (reversed protein sequences)

Proteins (2960) [Report Builder](#) [Unassigned \(17030\)](#)

Mascot search results (2)

MATRIX Mascot Search Results

User : Discoverer_LabPC
Email :
Search title : Control_Hela_Helios_03_Node:2
MS data file : File Name: Control_Hela_Helios_03.raw
Database : SP2017_07_Human_07_Human_20171018 (20224 sequences; 11332467 residues)
Timestamp : 4 Apr 2018 at 05:55:46 GMT
Enzyme : Trypsin
Fixed modifications : Carbamidomethyl (C)
Variable modifications : Oxidation (M), Deamidated (NQ)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 10 ppm
Fragment Mass Tolerance : ± 0.6 Da
Max Missed Cleavages : 2
Instrument type : ESI-TRAP
Number of queries : 40012
Protein hits :

- [Q09666](#) Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2
- [P21333](#) Filamin-A OS=Homo sapiens GN=FLNA PE=1 SV=4
- [P78527](#) DNA-dependent protein kinase catalytic subunit OS=Homo sapiens GN=PRKDC PE=1 SV=3
- [Q15149](#) Plectin OS=Homo sapiens GN=PLEC PE=1 SV=3
- [Q75369](#) Filamin-B OS=Homo sapiens GN=FLNB PE=1 SV=2
- [P31327](#) Carbamoyl-phosphate synthase [ammonia], mitochondrial OS=Homo sapiens GN=CPS1 PE=1 SV=2
- [P06733](#) Alpha-enolase OS=Homo sapiens GN=ENO1 PE=1 SV=2
- [Q9Y490](#) Talin-1 OS=Homo sapiens GN=TLN1 PE=1 SV=3
- [Q14204](#) Cytoplasmic dynein 1 heavy chain 1 OS=Homo sapiens GN=DYNC1H1 PE=1 SV=5
- [P63261](#) Actin, cytoplasmic 2 OS=Homo sapiens GN=ACTG1 PE=1 SV=1
- [Q00610](#) Clathrin heavy chain 1 OS=Homo sapiens GN=CLTC PE=1 SV=5
- [Q13813](#) Spectrin alpha chain, non-erythrocytic 1 OS=Homo sapiens GN=SPTAN1 PE=1 SV=3
- [Q01082](#) Spectrin beta chain, non-erythrocytic 1 OS=Homo sapiens GN=SPTBN1 PE=1 SV=2
- [P46940](#) Ras GTPase-activating-like protein IQGAP1 OS=Homo sapiens GN=IQGAP1 PE=1 SV=1
- [P68371](#) Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1
- [P07437](#) Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2
- [P35579](#) Myosin-9 OS=Homo sapiens GN=MYH9 PE=1 SV=4
- [Q43707](#) Alpha-actinin-4 OS=Homo sapiens GN=ACTN4 PE=1 SV=2
- [P12814](#) Alpha-actinin-1 OS=Homo sapiens GN=ACTN1 PE=1 SV=2
- [P14618](#) Pyruvate kinase PKM OS=Homo sapiens GN=PKM PE=1 SV=4
- [P08238](#) Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4
- [P49327](#) Fatty acid synthase OS=Homo sapiens GN=FASN PE=1 SV=3
- [P04406](#) Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3
- [P00558](#) Phosphoglycerate kinase 1 OS=Homo sapiens GN=PGK1 PE=1 SV=3
- [P22626](#) Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2
- [P02545](#) Prelamin-A/C OS=Homo sapiens GN=LMNA PE=1 SV=1
- [Q14980](#) Nuclear mitotic apparatus protein 1 OS=Homo sapiens GN=NUMA1 PE=1 SV=2
- [P10809](#) 60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2
- [P38646](#) Stress-70 protein, mitochondrial OS=Homo sapiens GN=HSPA9 PE=1 SV=2
- [P62937](#) Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens GN=PPIA PE=1 SV=2
- [Q92616](#) eIF-2-alpha kinase activator GCN1 OS=Homo sapiens GN=GCN1 PE=1 SV=6
- [P11021](#) 78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2
- [P07814](#) Bifunctional glutamate/proline-tRNA ligase OS=Homo sapiens GN=EPRS PE=1 SV=5
- [Q13885](#) Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1
- [Q13263](#) Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRIM28 PE=1 SV=5
- [P06576](#) ATP synthase subunit beta, mitochondrial OS=Homo sapiens GN=ATP5B PE=1 SV=3
- [P05787](#) Keratin, type II cytoskeletal 8 OS=Homo sapiens GN=KRT8 PE=1 SV=7
- [P34932](#) Heat shock 70 kDa protein 4 OS=Homo sapiens GN=HSPA4 PE=1 SV=4
- [P13639](#) Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4
- [Q04637](#) Eukaryotic translation initiation factor 4 gamma 1 OS=Homo sapiens GN=EIF4G1 PE=1 SV=4

1. [Q09666](#) Mass: 629213 Score: 2756 Matches: 223(102) Sequences: 195(100) emPAI: 0.81

Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
41	310.6918	619.3690	619.3693	-0.47	1	6	0.97	5	U	K.TPKFK.M
99	313.6680	625.3215	625.3224	-1.43	0	15	0.14	1	U	K.FGFGAK.S
172	316.6920	631.3694	631.3727	-5.27	1	9	4.5	3		K.MPKIK.M
256	321.1864	640.3582	640.3584	-0.31	0	22	0.01	1		K.FSLFK.S
740	342.2260	682.4374	682.4378	-0.51	0	20	0.025	1	U	K.IGIPGVK.M 741
848	345.7108	689.4070	689.4072	-0.27	0	3	6.1	6	U	R.TITVTR.R
913	348.1897	694.3649	694.3650	-0.12	0	27	0.011	1	U	K.FSVSGAK.G
1119	356.1950	710.3753	710.3785	-4.47	1	6	2.4	4	U	K.KTMFGK.M
1146	357.6816	713.3487	713.3497	-1.34	0	7	0.95	1	U	K.FTFSGR.E
1472	367.7030	733.3915	733.3970	-7.49	1	4	4.3	1	U	K.GDLKSSK.A
1816	379.2083	756.4019	756.4018	0.24	0	10	0.57	1	U	K.GPEVDIK.G
2076	386.2157	770.4169	770.4174	-0.64	0	14	0.26	1	U	K.APEVDIK.G
2077	386.2158	770.4170	770.4174	-0.57	0	22	0.041	1		K.GPEIDIK.G
2231	390.2098	778.4050	778.4047	0.35	0	25	0.072	1	U	K.ISMPGFK.G 2232
2268	391.2159	780.4172	780.4170	0.22	0	8	1.6	1	U	K.FSVPGFK.A 2271
2336	393.2113	784.4080	784.4079	0.10	0	23	0.031	1	U	K.GPEVDIR.G
2339	393.2241	784.4337	784.4331	0.78	0	30	0.006	1		K.VPDVDIK.G
2631	400.2314	798.4483	798.4487	-0.48	0	26	0.013	1	U	K.VPDVEIK.G
2811	405.2147	808.4148	808.4153	-0.58	0	29	0.019	1	U	K.FSMPSLK.G
2883	407.2016	812.3887	812.3891	-0.48	0	33	0.004	1	U	K.FSMPGFK.A
2886	407.2268	812.4391	812.4392	-0.19	0	23	0.035	1	U	K.VPDVDIR.G
2888	407.2287	812.4429	812.4392	4.53	0	6	2	1	U	K.VPEVDVR.G
3229	415.1991	828.3836	828.3840	-0.52	0	(11)	0.4	1	U	K.FSMPGFK.A 3228
3552	422.2318	842.4490	842.4498	-0.92	0	41	0.00064	1	U	K.ADLGVSGPK.V
3633	423.2527	844.4907	844.4840	7.94	1	9	1.5	1	U	K.LNVPMK.V
3652	423.7395	845.4644	845.4680	-4.26	1	(3)	12	4	U	K.LNVPMK.V 3653
4018	431.7116	861.4087	861.4088	-0.10	0	25	0.056	1	U	K.MPEMNIK.V
4145	434.2413	866.4681	866.4684	-0.32	0	26	0.028	1	U	K.IPEMHLK.A
4180	435.2347	868.4549	868.4477	8.30	0	4	2.1	1	U	K.MPDLHLK.A 4181
4282	437.2174	872.4203	872.4215	-1.30	0	3	4.3	1	U	K.MPDVHFK.S
4287	437.2373	872.4600	872.4603	-0.41	0	29	0.024	1	U	K.GDISTSQPK.V
4427	440.1979	878.3811	878.3877	-7.51	0	(10)	1	3	U	K.MPEMNIK.V
4615	444.2271	886.4397	886.4396	0.12	0	54	3.6e-05	1	U	K.ADVDSGPK.M
4813	449.7537	897.4928	897.4920	0.93	0	16	0.17	1	U	K.GPGVDVNLK.G
4864	450.7431	899.4717	899.4712	0.52	0	37	0.0018	1	U	K.GNVDSAPK.I
4888	451.2349	900.4552	900.4553	-0.04	0	36	0.0024	1	U	K.ADIVDSGPK.V
4899	451.2532	900.4919	900.4917	0.28	0	3	5.9	5	U	K.VPGIDATTK.L
5199	457.2530	912.4914	912.4916	-0.30	0	13	0.39	1	U	R.EVDVNLPK.A
5201	457.2585	912.5025	912.5029	-0.41	0	34	0.0033	1	U	K.GGVDTVLPK.V
5255	458.2425	914.4705	914.4709	-0.42	0	26	0.022	1	U	K.GDQVDSVPK.V
5356	460.2120	918.4095	918.4092	0.33	0	9	0.98	1	U	K.MPEMHFK.T
5625	465.2326	928.4506	928.4502	0.44	0	4	2.7	5	U	K.GEGPEVDVK.L
5633	465.2505	928.4864	928.4866	-0.22	0	36	0.0025	1	U	K.GDQVDSLPLK.L

Mascot search results (3)

MS/MS Fragmentation of VTAYTVDVTGR

Found in **Q09666** in **SP2017_07_Human**, Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2

Match to Query 13429: 1180.609228 from(591.311890,2+) index(13392)

Title: Spectrum13394 scans:17257,

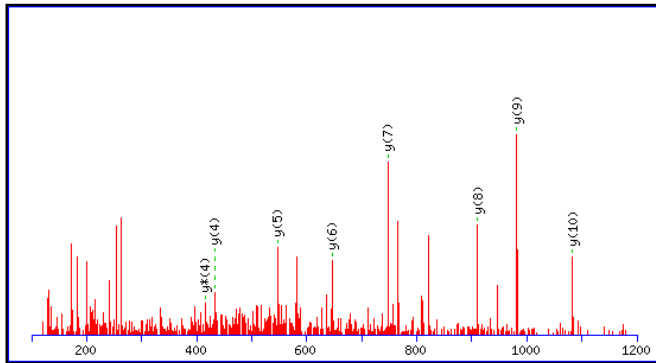
Data file File Name: Control_Hela_Helios_03.raw

Click mouse within plot area to zoom in by factor of two about that point

Or, 100 1200

Label all possible matches Label matches used for scoring

Show Y-axis



Monoisotopic mass of neutral peptide Mr(calc): 1180.6088

Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)

Ions Score: 57 Expect: 3.6e-05

Matches : 8/94 fragment ions using 11 most intense peaks ([help](#))

#	b	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	100.0757	50.5415			V							11
2	201.1234	101.0653	183.1128	92.0600	T	1082.5477	541.7775	1065.5211	533.2642	1064.5371	532.7722	10
3	272.1605	136.5839	254.1499	127.5786	A	981.5000	491.2536	964.4734	482.7404	963.4894	482.2483	9
4	435.2238	218.1155	417.2132	209.1103	Y	910.4629	455.7351	893.4363	447.2218	892.4523	446.7298	8
5	536.2715	268.6394	518.2609	259.6341	T	747.3995	374.2034	730.3730	365.6901	729.3890	365.1981	7
6	635.3399	318.1736	617.3293	309.1683	V	646.3519	323.6796	629.3253	315.1663	628.3413	314.6743	6
7	750.3668	375.6871	732.3563	366.6818	D	547.2835	274.1454	530.2569	265.6321	529.2729	265.1401	5
8	849.4353	425.2213	831.4247	416.2160	V	432.2565	216.6319	415.2300	208.1186	414.2459	207.6266	4
9	950.4829	475.7451	932.4724	466.7398	T	333.1881	167.0977	316.1615	158.5844	315.1775	158.0924	3
10	1007.5044	504.2558	989.4938	495.2506	G	232.1404	116.5738	215.1139	108.0606			2
11					R	175.1190	88.0631	158.0924	79.5498			1

MS/MS Fragmentation of VNVEAPDVLNLEGLGK

Found in **Q09666** in **SP2017_07_Human**, Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2

Match to Query 25878: 1609.830668 from(805.922610,2+) index(27850)

Title: Spectrum27858 scans:34059,

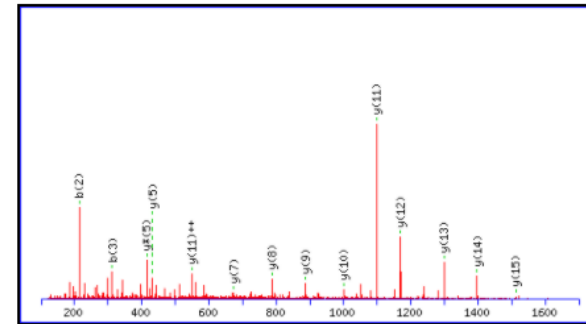
Data file File Name: Control_Hela_Helios_03.raw

Click mouse within plot area to zoom in by factor of two about that point

Or, 100 1700

Label all possible matches Label matches used for scoring

Show Y-axis



Monoisotopic mass of neutral peptide Mr(calc): 1609.8311

Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)

Ions Score: 95 Expect: 6.7e-09

Matches : 14/162 fragment ions using 15 most intense peaks ([help](#))

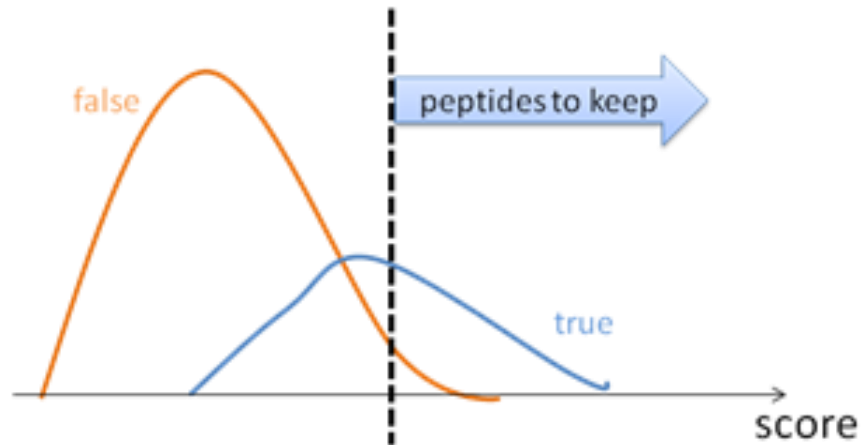
#	b	b ⁺⁺	b [*]	b ⁺⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	100.0757	50.5415					V							16
2	214.1186	107.5629	197.0921	99.0497			N	1511.7700	756.3886	1494.7435	747.8754	1493.7594	747.3834	15
3	313.1870	157.0972	296.1605	148.5839			V	1397.7271	699.3672	1380.7005	690.8539	1379.7165	690.3619	14
4	442.2296	221.6184	425.2031	213.1052	424.2191	212.6132	E	1298.6587	649.8330	1281.6321	641.3197	1280.6481	640.8277	13
5	513.2667	257.1370	496.2402	248.6237	495.2562	248.1317	A	1169.6161	585.3117	1152.5895	576.7984	1151.6055	576.3064	12
6	610.3195	305.6634	593.2930	297.1501	592.3089	296.6581	P	1098.5790	549.7931	1081.5524	541.2798	1080.5684	540.7878	11
7	725.3464	363.1769	708.3199	354.6636	707.3359	354.1716	D	1001.5262	501.2667	984.4997	492.7535	983.5156	492.2615	10
8	824.4149	412.7111	807.3883	404.1978	806.4043	403.7058	V	886.4993	443.7533	869.4727	435.2400	868.4887	434.7480	9
9	938.4578	469.7325	921.4312	461.2193	920.4472	460.7272	N	787.4308	394.2191	770.4043	385.7058	769.4203	385.2138	8
10	1051.5418	526.2746	1034.5153	517.7613	1033.5313	517.2693	L	673.3879	337.1976	656.3614	328.6843	655.3774	328.1923	7
11	1180.5844	590.7959	1163.5579	582.2826	1162.5739	581.7906	E	560.3039	280.6556	543.2773	272.1423	542.2933	271.6503	6
12	1237.6059	619.3066	1220.5794	610.7933	1219.5953	610.3013	G	431.2613	216.1343	414.2347	207.6210			5
13	1350.6900	675.8486	1333.6634	667.3353	1332.6794	666.8433	L	374.2398	187.6235	357.2132	179.1103			4
14	1407.7114	704.3594	1390.6849	695.8461	1389.7009	695.3541	G	261.1557	131.0815	244.1292	122.5682			3
15	1464.7329	732.8701	1447.7063	724.3568	1446.7223	723.8648	G	204.1343	102.5708	187.1077	94.0575			2
16							K	147.1128	74.0600	130.0863	65.5468			1

Validation and false discovery rate (FDR)

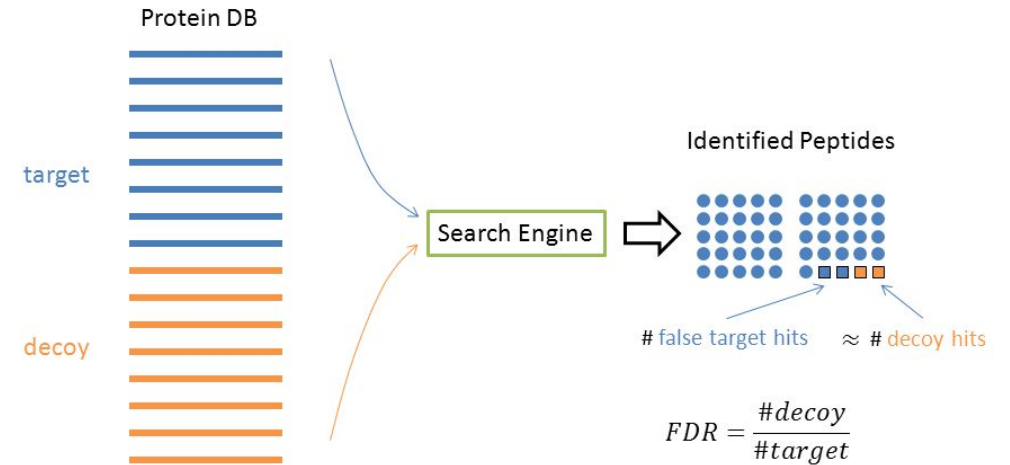
PSM refers to peptide-spectrum match

PSMs are sorted by their scores

FDR, is defined to be the ratio between the false PSMs and the total number of PSMs above the score threshold



Estimating FDR with the Target-Decoy Method



In this method, the software is used to search the concatenation of a target database and a decoy database with the same size. If the decoy is constructed properly, the software's false identifications will be evenly distributed in the target and decoy databases. Since all the decoy identifications are false, FDR can be estimated by $FDR = (\# \text{ Decoy Hits}) / (\# \text{ target hits})$

Q1: Exercise

Here you will find the search engine: http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS

Here you have the peak list (*i.e.*, mgf file) (also on the Moodle):

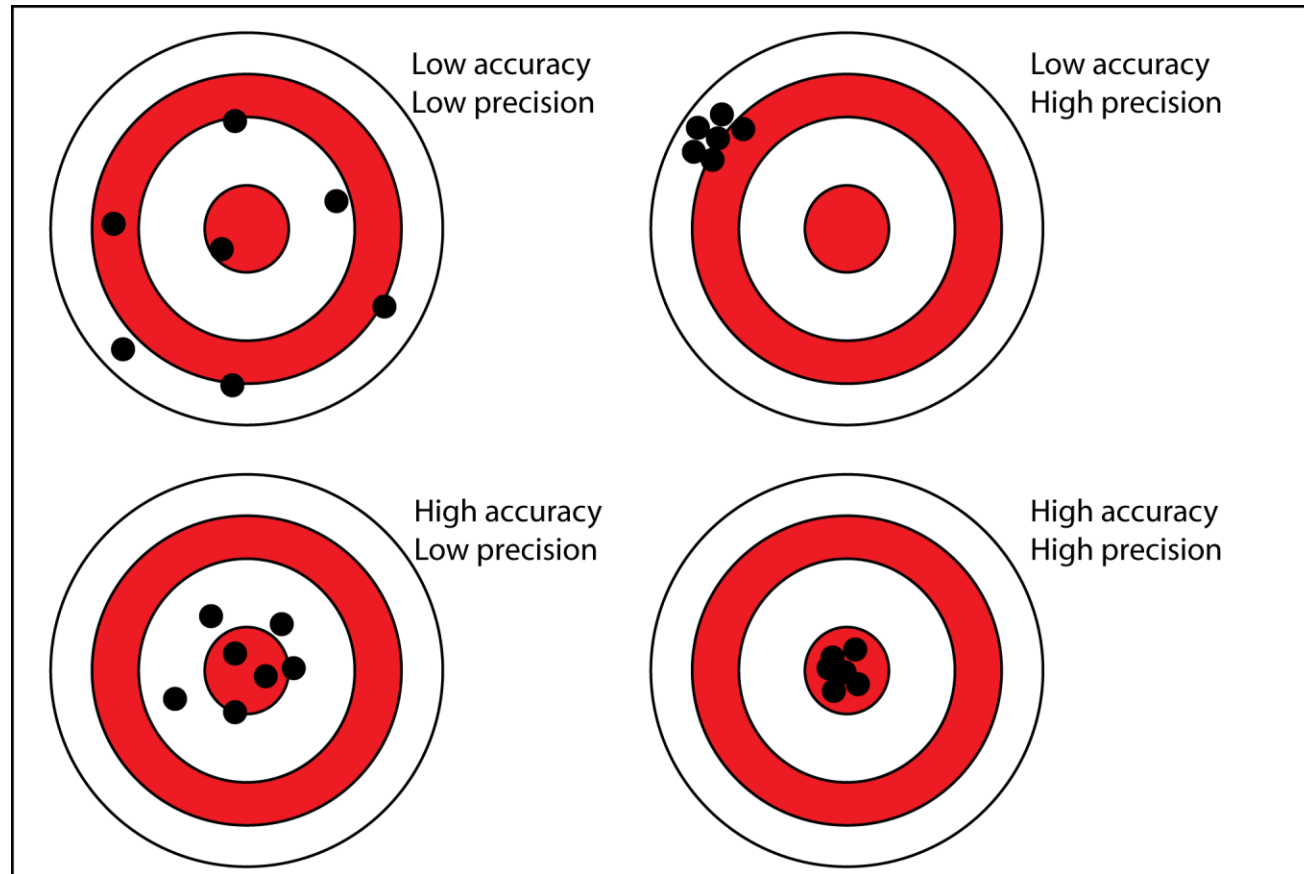


SDS_03_02_shorted.MGF.txt

Let's try a database search!

You can try with more data available at: <http://proteomecentral.proteomexchange.org/cgi/GetDataset>

4.3. Quantification of proteins



<http://www.antarcticglaciers.org>

Figures of merit and useful metrics

Fold change (FC)

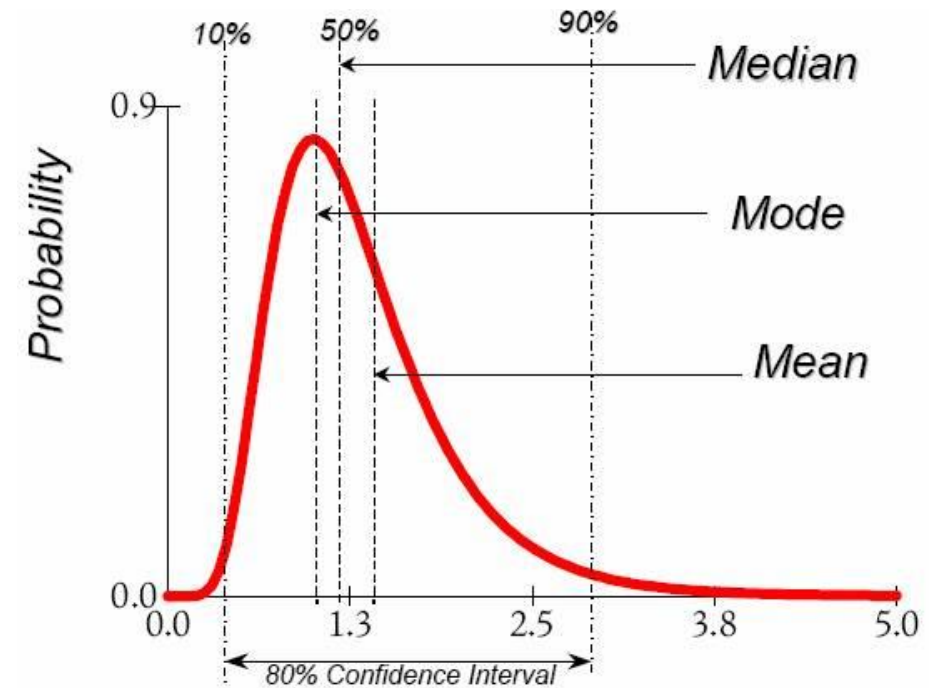
Standard deviation (SD)

Coefficient of variation (CV%)

Average

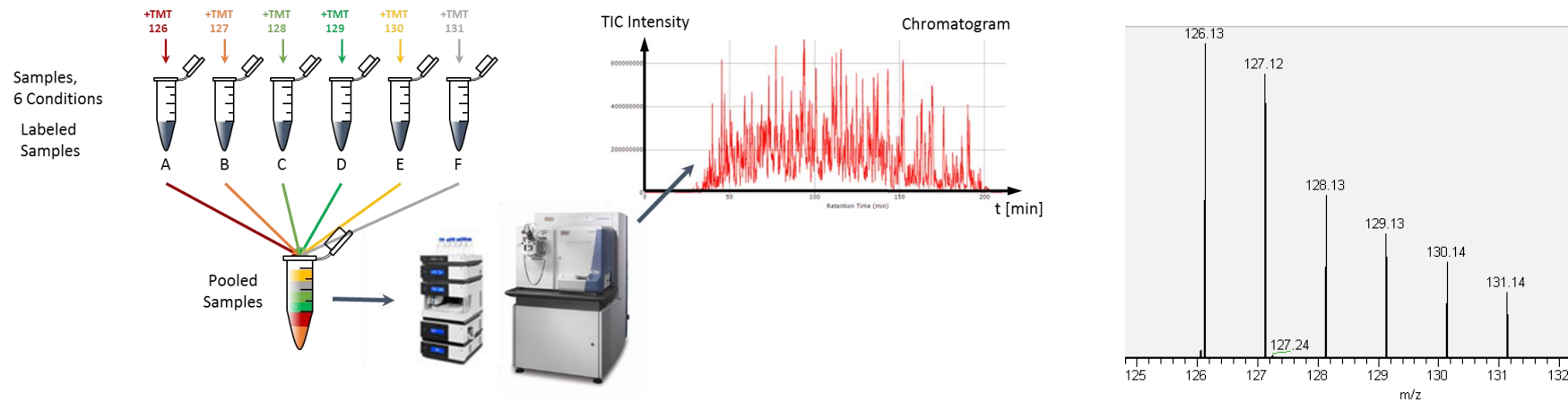
Median

Mode



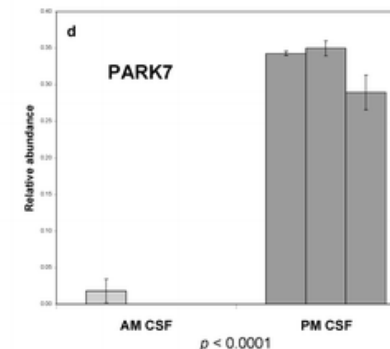
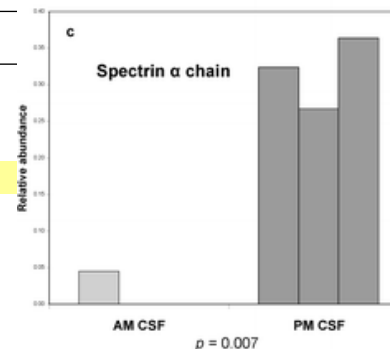
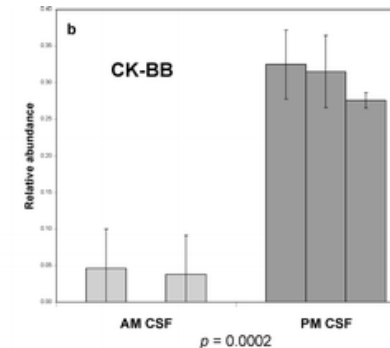
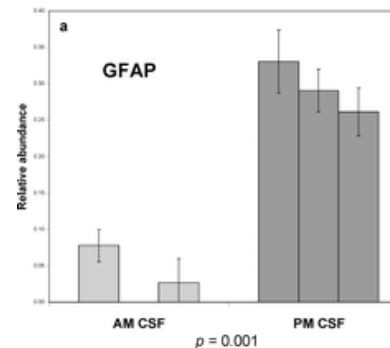
<http://herdingcats.typepad.com>

An example of relative quantification from isobaric tagging data



AC	Description	Score	#valid pept	% Cov	sequence	NORM i126.1	NORM i127.1	NORM i128.1	NORM i129.1	NORM i130.1	NORM i131.1	Norm i126.1/SUM	Norm i127.1/SUM	Norm i128.1/SUM	Norm i129.1/SUM	Norm i130.1/SUM	Norm i131.1/SUM	
Q16270	Insulin-like growth factor-binding protein 7 precursor	15.7	3	3%	RGHYGVQR	8 695.13	9 675.97	6 606.49	8 740.77	9 448.77	9 806.96	0.16	0.18	0.12	0.17	0.18	0.19	
Q16270					RGHYGVQR	7 792.05	8 225.70	5 312.67	7 928.39	7 951.24	8 620.03	0.17	0.18	0.12	0.17	0.17	0.19	
Q16270					GHYGVQR	7 922.98	8 977.85	8 627.91	9 517.74	9 157.37	10 038.93	0.15	0.17	0.16	0.18	0.17	0.19	
												AVG	0.16	0.18	0.13	0.17	0.17	0.19
												STD	0.01	0.01	0.02	0.01	0.00	0.00
												% STD	7.80	5.19	17.11	3.20	2.75	0.93

Quantitative protein report (FC)



AC	ID	Description						# peptides for quantification	Ratio PM/AM
P01024	CO3_HUMAN	Complement C3 precursor						58	1.48
P0C0L4	CO4A_HUMAN	Complement C4-A precursor (or Complement C4-B precursor, CO4B_HUMAN, P0C0L5)						34	1.24
P02763	A1AG1_HUMAN	Alpha-1-acid glycoprotein 1 precursor						29	4.35
P41222	PTGDS_HUMAN	Prostaglandin-H2 D-isomerase precursor						51	0.34
P00450	CERU_HUMAN	Ceruloplasmin precursor						21	1.55
P02790	HEMO_HUMAN	Hemopexin precursor						24	1.05
P05067	A4_HUMAN	Amyloid beta A4 protein precursor						22	0.54
P19652	A1AG2_HUMAN	Alpha-1-acid glycoprotein 2 precursor						20	2.83
P02766	TTHY_HUMAN	Transthyretin precursor	64.7	7	13	15	25%	9	0.74
P01023	A2MG_HUMAN	Alpha-2-macroglobulin precursor	60.1	7	13	18	5%	13	1.02
Q14515	SPRL1_HUMAN	SPARC-like protein 1 precursor	52.4	7	9	13	11%	8	0.60
P00918	CAH2_HUMAN	Carbonic anhydrase 2	49.2	5	29	30	16%	29	17.80
Q13822	ENPP2_HUMAN	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2 precursor	47.5	6	10	13	6%	7	0.81
P01034	CYTC_HUMAN	Cystatin-C precursor	42.3	5	10	11	38%	10	0.88
P15311	EZRI_HUMAN	Ezrin	41.7	5	12	15	8%	11	5.08
P00751	CFAB_HUMAN	Complement factor B precursor	41.4	5	9	13	7%	8	1.63
P02774	VTDB_HUMAN	Vitamin D-binding protein precursor	41.0	5	11	15	8%	8	1.24
P02751	FINC_HUMAN	Fibronectin precursor	40.6	5	7	9	2%	6	0.59
P51693	APLP1_HUMAN	Amyloid-like protein 1 precursor	37.9	4	14	16	7%	14	0.85
P05060	SCG1_HUMAN	Secretogranin-1 precursor	37.0	4	14	16	7%	14	0.24
P06396	GELS_HUMAN	Gelsolin precursor	36.3	4	10	15	5%	10	1.11
P68032	ACTC_HUMAN	Actin, alpha cardiac muscle 1 (or Actin, alpha skeletal muscle, ACTS_HUMAN, P68133)	32.1	3	6	6	9%	5	10.41
P02671	FIBA_HUMAN	Fibrinogen alpha chain precursor	30.9	4	6	12	6%	6	7.02
Q6NR85	Q6NR85_HUMAN	Superoxide dismutase [Cu-Zn]	30.2	3	30	32	12%	30	6.20

4.4. Bioinformatics tools

The screenshot shows the ExpASY Bioinformatics Resource Portal homepage. At the top left, there is a logo for SIB (Swiss Institute of Bioinformatics) and the text "ExpASY Bioinformatics Resource Portal". To the right of the logo is a search bar with a dropdown menu set to "Query all databases", a search button, and a "help" link. In the top right corner, there are links for "Home", "About", and "Contact".

On the left side, there is a vertical navigation menu with several categories: "Visual Guidance", "Categories", "Resources A..Z", and "Links/Documentation". Under "Categories", there is a list of sub-categories: proteomics, genomics, structure analysis, systems biology, evolutionary biology, population genetics, transcriptomics, biophysics, imaging, IT infrastructure, medicinal chemistry, and glycomics.

The main content area features a "Featuring today" section with a card for "Sequence Similarity Maps (SSM)". The card includes a small tree diagram and the text: "Visualize relationships between viral isolates" and a "[details]" link. Below the card are four dots, with the second one filled, and a double arrow icon.

Below the "Featuring today" section is a "How to use this portal?" section with a question mark icon and a list of bullet points: "Features and updates", "New to ExpASY", and "Experienced ExpASY users: what is different". To the right of this section is a large blue link: <https://expasy.org/>.

On the right side, there are two sections: "Popular resources" and "Latest News". The "Popular resources" section lists: UniProtKB, SWISS-MODEL, STRING, and PROSITE. The "Latest News" section has a sub-header "UniProt Knowledgebase release 2018_03 - 2018-03-28" and includes "Release notes" with statistics: "557,012 UniProtKB/Swiss-Prot entries (More..)" and "111,425,245 UniProtKB/TrEMBL entries (More..)". Below this is a "Protein Spotlight: It's a thin line - 2018-03-23" with a short paragraph and a "[More news] [SIB news]" link.

At the bottom center, there is a circular logo celebrating "20 YEARS SIB" (Swiss Institute of Bioinformatics).

Q2: Let's play with few bioinformatic tools

Q99497-1 [UniParc] [FASTA](#) [Add to basket](#)

« Hide

Length: 189

Mass (Da): 19,891

Last modified: July 5, 2004 - v2

Checksum: 4B21661B3A76BC67

```
      10      20      30      40      50
MASKRALVIL AKGAEEMETV IPVDVMRRAG IKVTVAGLAG KDPVQCSRDV
      60      70      80      90     100
VICPDASLED AKKEGPYDVV VLPGGNLGAQ NLSESAAVKE ILKEQENRKG
     110     120     130     140     150
LIAAICAGPT ALLAHEIGFG SKVTTHPLAK DKMMNGGHYT YSENVEKDG
     160     170     180
LILTSRGPQT SFEFALAIVE ALNGKEVAAQ VKAPLVVKD
```

<https://www.uniprot.org/uniprot/Q99497>

Give the *pI* and *MW* of protein/nucleic acid deglycase DJ-1

Give the peptide sequences and masses after digestion with trypsin

Other useful tools: Retrieve/ID mapping

UniProt

BLAST Align Retrieve/ID mapping Peptide search Help Contact

Retrieve/ID mapping

How to use this tool

Enter or upload a list of identifiers to do one of the following:

Retrieve the corresponding UniProt entries to download them or work with them on this website.
Convert identifiers which are of a different type to UniProt identifiers or vice versa and download the identifier lists.

1. Enter identifiers, separated by spaces or new lines, into the form field, for example:

```
P31946 P62258  
ALBU_HUMAN  
EFTU_ECOLI
```

2. If you need to convert to another identifier type, select the source and target type from the dropdown menus.
3. Click the Go button.

[Help](#) [Help video](#) [Other tutorials and videos](#) [Downloads](#)

1. Provide your identifiers

e.g. P31946 P62258 ALBU_HUMAN EFTU_ECOLI

OR upload your own file: No file chosen

Run in a new window.

2. Select options

From To

<https://www.uniprot.org/uploadlists/>

Other useful tools: Blast

The image shows the UniProt BLAST search interface. At the top left is the UniProt logo. A search bar contains 'UniProtKB' and a search button labeled 'Advanced Search'. Below the search bar is a navigation menu with 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. On the right of the menu are 'Help' and 'Contact' links. The main heading is 'BLAST'. A 'How to use this tool' section provides instructions: 1. Enter a protein or nucleotide sequence or a UniProt identifier (e.g. P00750 or A4_HUMAN or UPI0000000001) into the form field. 2. Optionally, change the program parameters with the dropdown menus under the form. 3. Click the Run BLAST button. Below this are links for 'Help', 'BLAST help video', 'Other tutorials and videos', and 'Downloads'. A large text input field is provided for the sequence, with a placeholder 'Protein sequence, Nucleotide sequence or UniProt identifier'. Below the input field are several dropdown menus for parameters: 'Target database' (UniProtKB), 'E-Threshold' (0.001), 'Matrix' (Auto), 'Filtering' (None), 'Gapped' (yes), and 'Hits' (250). There is also a checkbox for 'Run BLAST in a separate window.' At the bottom are 'Clear' and 'Run BLAST' buttons.

UniProt

UniProtKB

Advanced Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

BLAST

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g. P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

[Help](#) [BLAST help video](#) [Other tutorials and videos](#) [Downloads](#)

Protein sequence, Nucleotide sequence or UniProt identifier

Target databaseⁱ UniProtKB E-Thresholdⁱ 0.001 Matrixⁱ Auto Filteringⁱ None Gappedⁱ yes Hitsⁱ 250

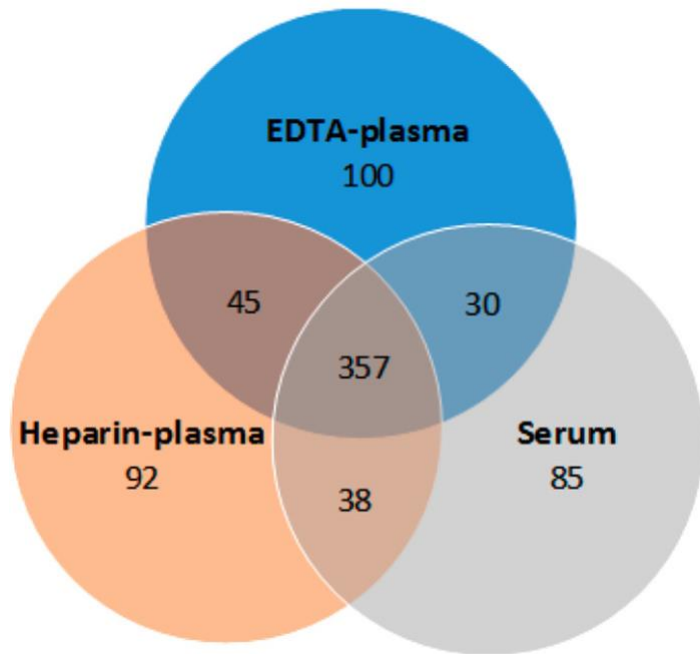
Run BLAST in a separate window.

Clear Run BLAST

<https://www.uniprot.org/blast/>

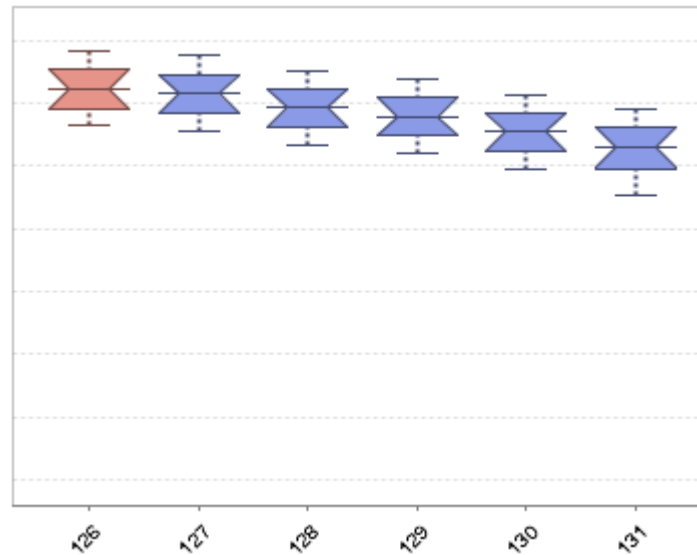
4.5. Practical examples

Compare proteomes

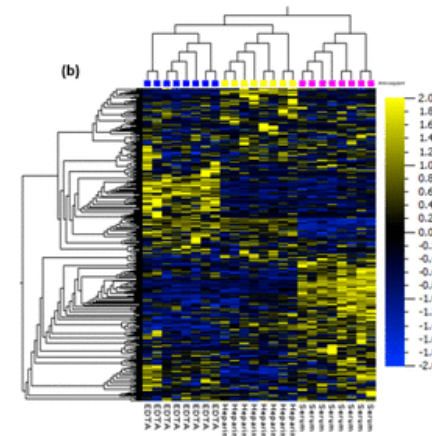
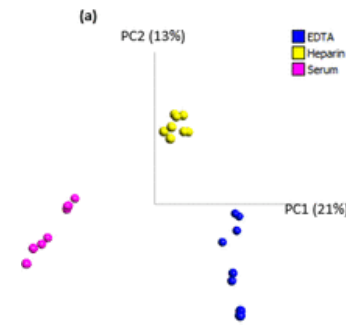
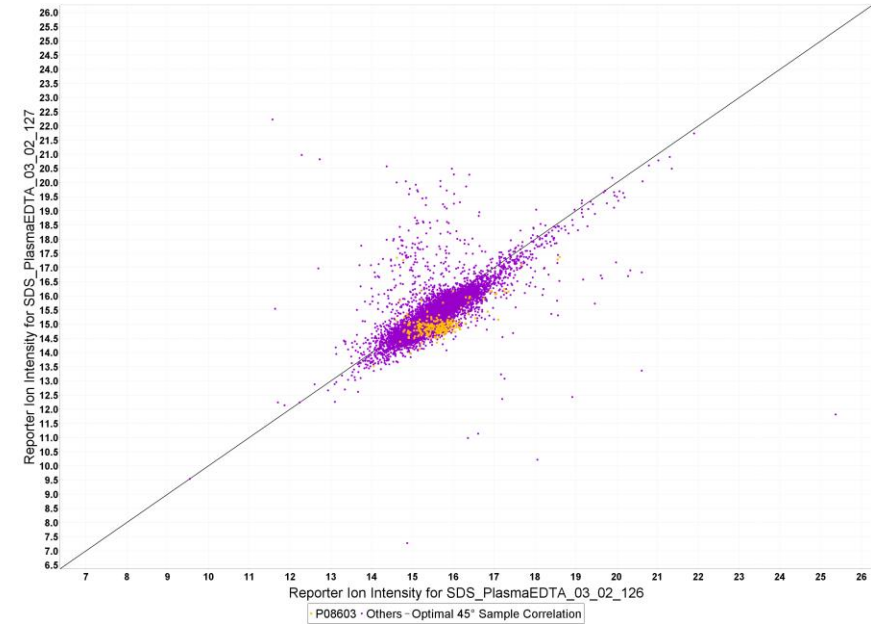


DOI: 10.1021/acs.jproteome.7b00788

Evaluate technical performances



Assess reproducibility



Decipher differences

Summary

- In proteomics, bioinformatics is a key pillar to process the large MS datasets comprising tens of thousands of spectra and tandem mass spectra
- Bioinformatic tools enable to identify peptides/proteins and quantify them
- Additional bioinformatics will be needed to analyze data in-depth and interpret the results. Examples will be given in the next chapter
- Next, let's see some examples of projects where we apply mass spectrometry and proteomics...