

ms-based proteomics data analysis

lennart martens

*lennart.martens@vib-ugent.be
computational omics and systems biology group
VIB / Ghent University, Ghent, Belgium*



Amino acids, peptides, proteins

Mass spectrometry basics

MS/MS spectra and identification

Database search algorithms

Sequencial search algorithms

Identification validation

Protein inference

Amino acids, peptides, proteins

Mass spectrometry basics

MS/MS spectra and identification

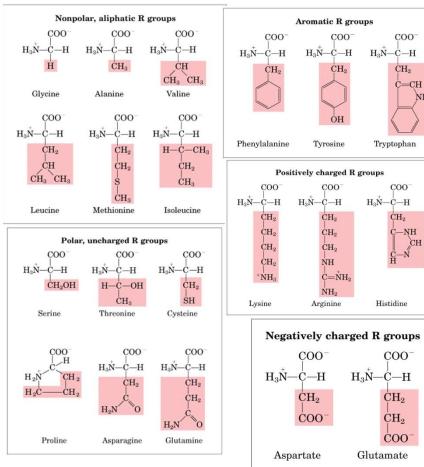
Database search algorithms

Sequencial search algorithms

Identification validation

Protein inference

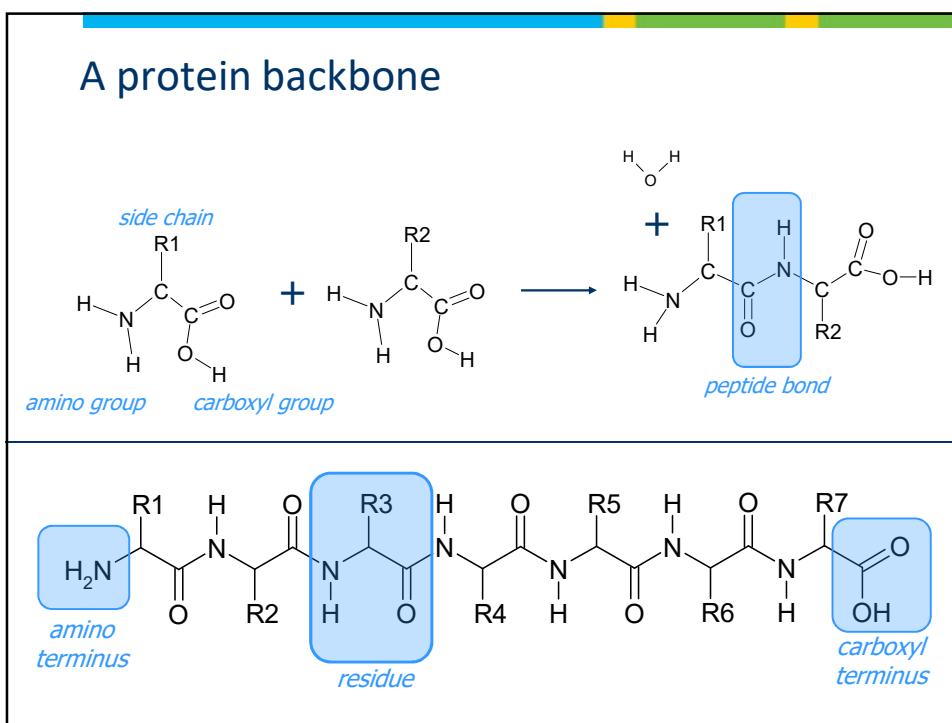
Amino Acids and their properties



Name	3-letter symbol	1-letter symbol	Molecular weight	Polarizability	Residue formula	Residue weight (C+H+O)	pKa ^a	pK ^b	δ _α
Alanine	Ala	A	70.10	C ₃ H ₇ NO ₂	C ₃ H ₇ NO	71.08	2.34	9.69	—
Arginine	Arg	R	174.20	C ₆ H ₁₃ N ₂ O ₂	C ₆ H ₁₃ N ₂ O	156.19	2.17	9.04	12.48
Asparagine	Asn	N	132.12	C ₄ H ₈ N ₂ O ₃	C ₄ H ₈ N ₂ O ₂	114.11	2.02	8.80	—
Aspartic acid	Asp	D	133.11	C ₄ H ₇ NO ₄	C ₄ H ₇ NO ₃	115.09	1.88	9.60	3.65
Cysteine	Cys	C	89.12	C ₃ H ₇ NO ₂ S	C ₃ H ₇ NO ₂ S	89.12	1.84	9.84	8.18
Glutamic acid	GLU	E	147.13	C ₅ H ₉ NO ₄ S	C ₅ H ₉ NO ₃	129.12	2.19	9.67	4.25
Glutamine	Gln	Q	146.15	C ₅ H ₁₁ N ₂ O ₃	C ₅ H ₁₁ N ₂ O ₂	128.13	2.17	9.13	5.65
Glycine	Gly	G	75.07	C ₂ H ₅ O ₂	C ₂ H ₅ O	57.05	2.34	9.60	—
Hydroxyproline	HP	P	131.13	C ₅ H ₉ NO ₂ O	C ₅ H ₉ NO ₂	113.11	1.82	9.65	—
Histidine	His	H	155.16	C ₆ H ₁₃ N ₂ O ₂	C ₆ H ₁₃ N ₂ O	137.14	2.36	9.69	7.59
Isoleucine	Ile	I	121.18	C ₆ H ₁₃ NO ₂	C ₆ H ₁₃ NO	113.16	2.36	9.60	—
Leucine	Leu	L	121.18	C ₆ H ₁₃ NO ₂	C ₆ H ₁₃ NO	113.16	2.36	9.60	—
Lysine	Lys	K	146.19	C ₆ H ₁₅ N ₂ O ₂	C ₆ H ₁₅ N ₂ O	126.18	2.16	8.95	10.53
Methionine	Met	M	149.21	C ₅ H ₁₁ NO ₂ S	C ₅ H ₁₁ NO ₂ S	131.20	2.28	9.21	5.74
Phenylalanine	Ph	F	165.19	C ₉ H ₁₁ NO ₂	C ₉ H ₁₁ NO	147.18	1.83	9.13	—
Proline	Pro	P	115.13	C ₅ H ₉ NO ₂	C ₅ H ₉ NO	97.12	1.99	10.60	—
Pyroglutamatic	GP	U	139.19	C ₅ H ₉ NO ₃	C ₅ H ₉ NO ₂	121.09	—	—	5.68
Serine	Ser	S	87.09	C ₃ H ₇ NO ₃	C ₃ H ₇ NO ₂	87.09	2.23	9.15	—
Threonine	Thr	T	119.12	C ₄ H ₉ NO ₃	C ₄ H ₉ NO ₂	101.11	2.09	9.10	—
Tryptophan	Trp	W	204.23	C ₁₁ H ₁₇ N ₂ O ₂	C ₁₁ H ₁₇ N ₂ O ₂	186.22	2.83	9.39	5.89
Tyrosine	Tyr	Y	181.13	C ₉ H ₁₁ NO ₃	C ₉ H ₁₁ NO ₂	163.18	2.01	9.11	10.07
Valine	Val	V	117.15	C ₅ H ₁₁ NO ₂	C ₅ H ₁₁ NO ₂	99.13	2.32	9.62	5.96

^a pKa is the negative of the logarithm of the association constant for the -COOH group.^b pKb is the negative of the logarithm of the dissociation constant for the -NH3+ group.^c pK_{a'} is the negative of the logarithm of the dissociation constant for any other group in the molecule.^d pI is the pH at the isoelectric point.References: D. R. Ude, *Handbook of Chemistry and Physics*, 72nd Edition, CRC Press, Boca Raton, FL, 1991.

<http://courses.cm.utexas.edu/jrobertus/ch339k/overheads-1/ch5-amino-acids.jpg>

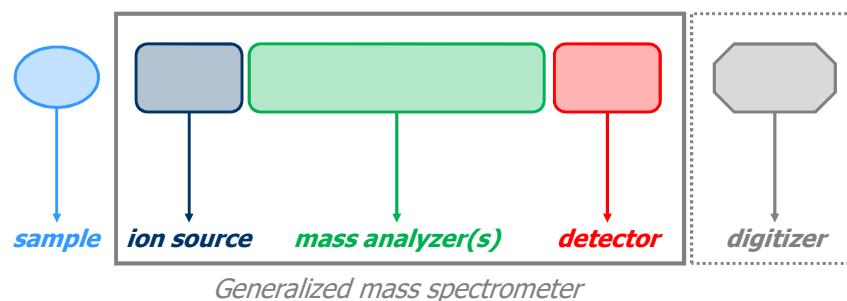


Amino acids, peptides, proteins

Mass spectrometry basics

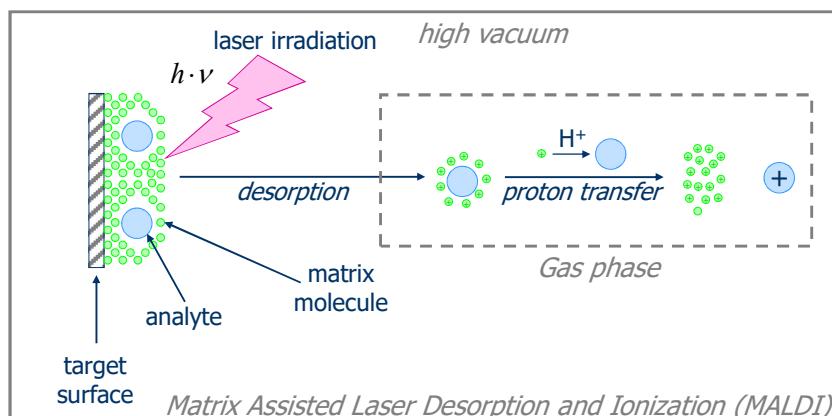
- MS/MS spectra and identification
- Database search algorithms
- Sequencial search algorithms
- Identification validation
- Protein inference

Schematic view of a generalized mass spec



All **mass analyzers** use electromagnetic fields to manipulate gas-phase ions. Results are plotted as a spectrum, with mass-over-charge (m/z) on the X-axis and ion intensity on the Y-axis. The latter can be absolute (counts) or relative. The **ion source** ensures that (a part of) the **sample molecules** are ionized and brought into the gas phase. The **detector** is responsible for actually recording the presence of ions. **Digitizers** (analog to digital converters; ADC) transform the continuous, analog detector signal into a digital, discretized spectrum.

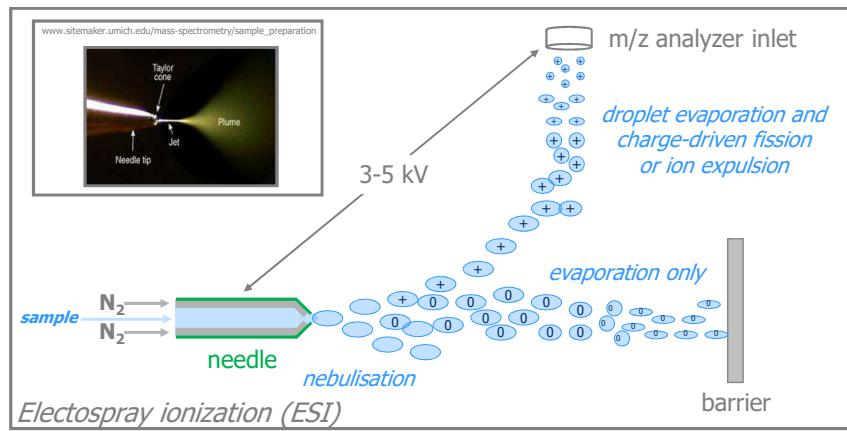
Ion sources: MALDI



MALDI sources for proteomics typically rely on a pulsed nitrogen UV laser ($\lambda = 337$ nm) and produce singly charged peptide ions. Competitive ionisation occurs.

The term 'MALDI' was coined by **Karas and Hillenkamp** (*Anal. Chem.*, 1985) and **Koichi Tanaka** received the 2002 Nobel Prize in Chemistry for demonstrating MALDI ionization of biological macromolecules (*Rapid Commun. Mass Spectrom.*, 1988)

Ion sources: ESI

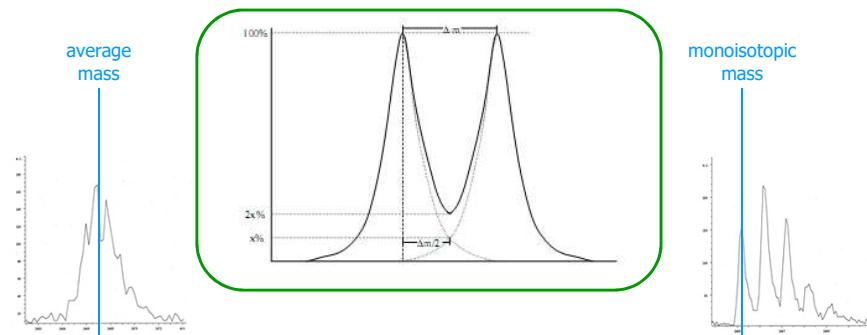


ESI sources typically heat the needle to 40° to 100° to facilitate nebulisation and evaporation, and typically produce multiply charged peptide ions (2⁺, 3⁺, 4⁺)

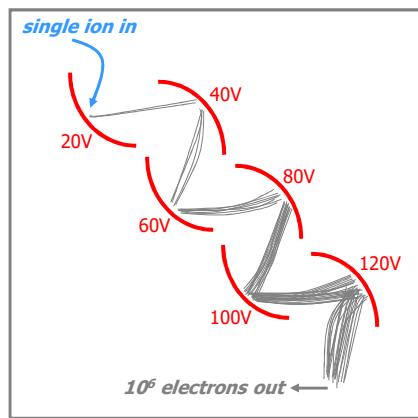
John B. Fenn received the 2002 Nobel Prize in Chemistry for demonstrating ESI ionization of biological macromolecules (*Science*, 1989) – ESI is also used in fine control thrusters on satellites and interstellar probes...

Resolution and why it matters

Resolution in mass spectrometry is usually defined as the width of a peak at a given height (there is an alternative definition based on percent valley height). This width can be recorded at different heights, but is most often recorded at 50% peak height (FWHM).

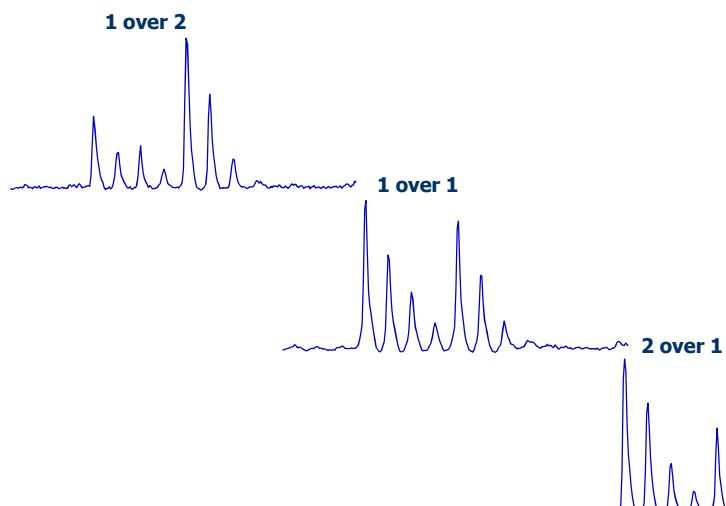


Detectors: electron multiplier



Different variations of electron multiplier (EM) detectors are used, and these are the most common type of detector. An EM relies on several Faraday cup dynodes with increasing charges to produce an electron cascade from a few incident ions.

The detector signal strength is assumed to represent the abundance of the analyte

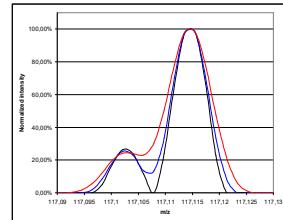


Errors on the order of 10% are common

Mass spectrometer specific processing required

Sets the dynamic range lower limit (S/N)

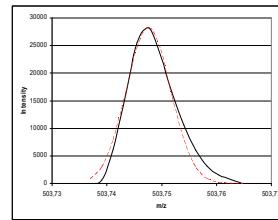
10-15% error in the final ratios with the instrument vendor peak-picker



Black: 0,02 Da

Blue: 0,04 Da

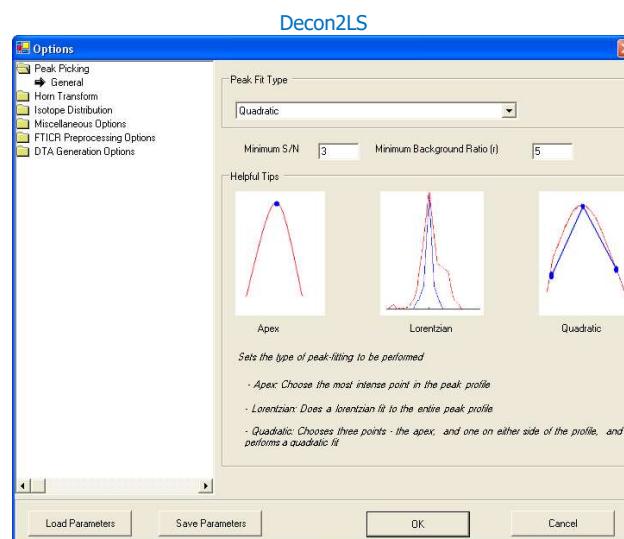
Red: 0,08 Da



Non-adapted shape -> +10% error

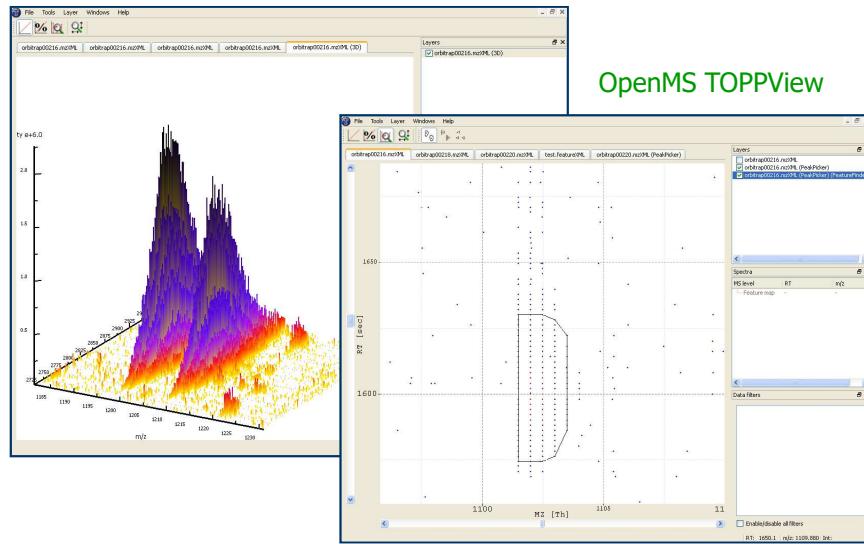
Vaudel, Proteomics, 2010

Different options for peak detection



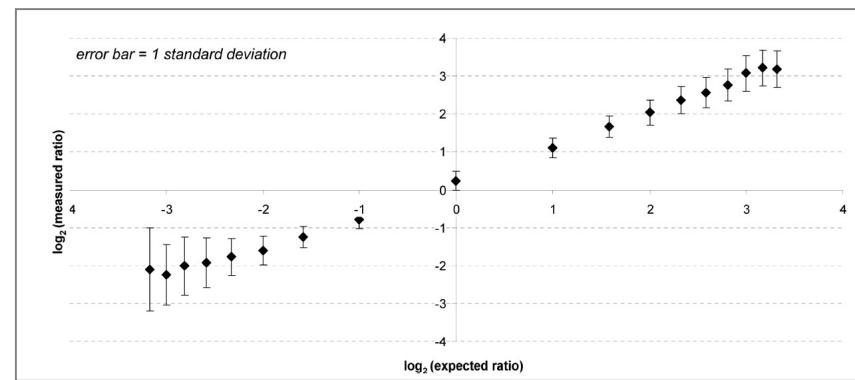
Vaudel, Proteomics, 2010

But there's more to a peak than m/z



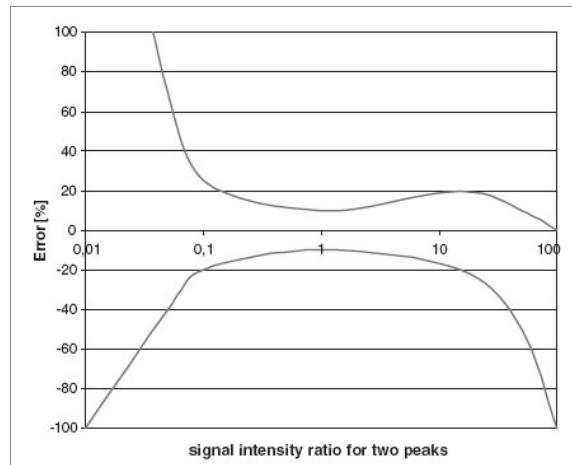
Vaudel, Proteomics, 2010

While 1/1 ratios are usually quite reliable, outlying ratios quickly deviate from a line



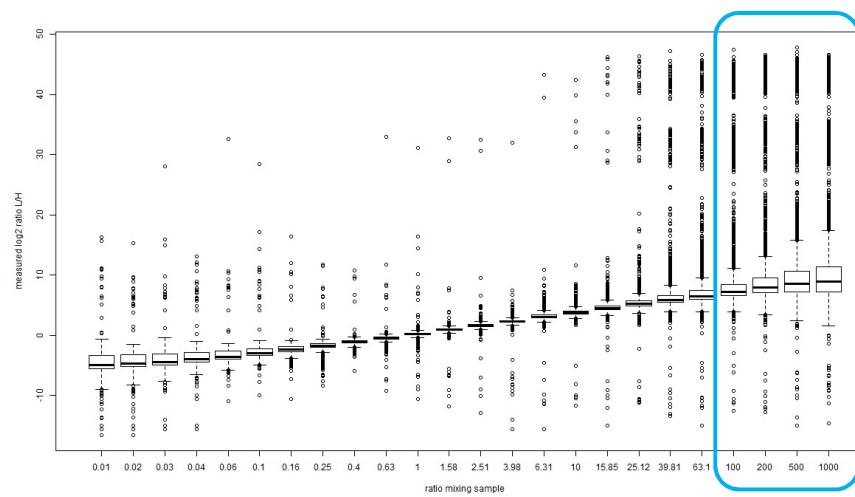
Gevaert, Proteomics, 2007

The errors also rise quickly for outlying ratios



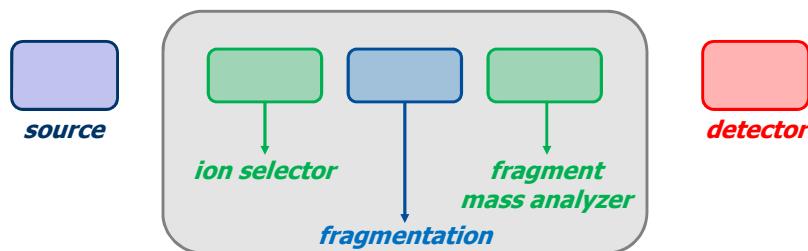
Vaudel, Proteomics, 2010

And the effects remain quite visible,
even on modern instruments (LTQ-OrbiTrap)



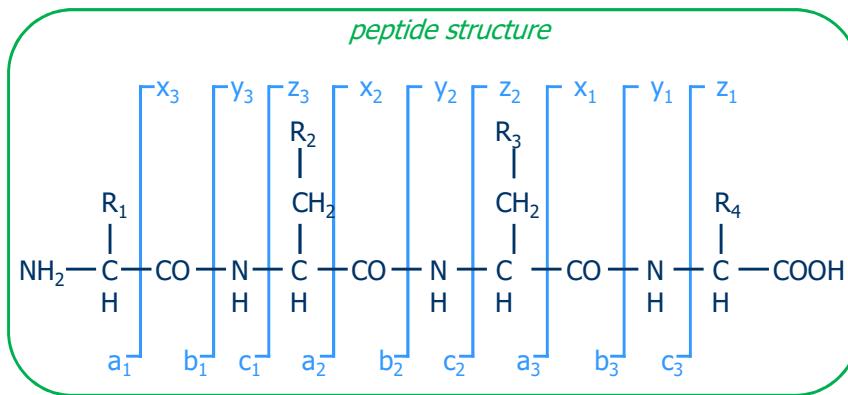
Amino acids, peptides, proteins
Mass spectrometry basics
MS/MS spectra and identification
Database search algorithms
Sequencial search algorithms
Identification validation
Protein inference

Tandem-MS: the concept



Tandem-MS is accomplished by using two mass analyzers in series (tandem). A single ion trap can also perform tandem-MS. The first mass analyser performs the function of ion selector, by selectively allowing only ions of a given m/z to pass through. The second mass analyzer is situated after fragmentation is triggered (see next slides) and is used in its normal capacity as a mass analyzer for the fragments.

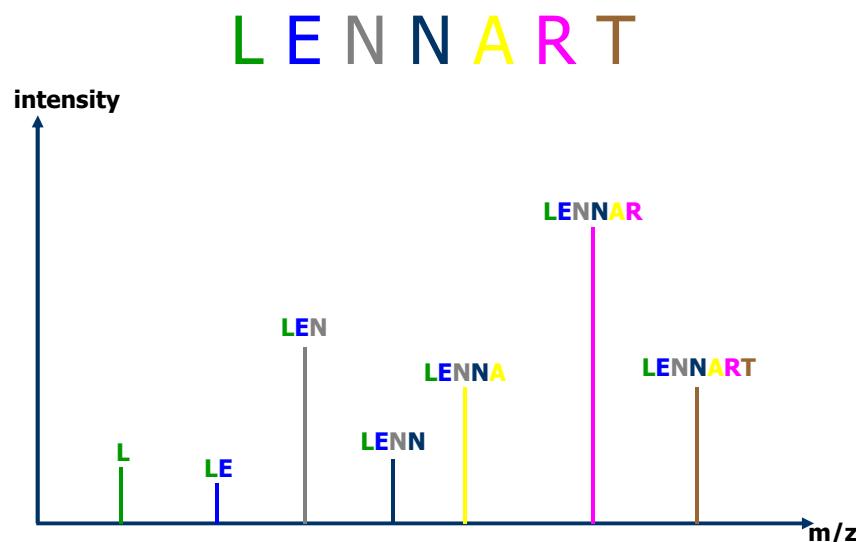
Why tandem-MS?



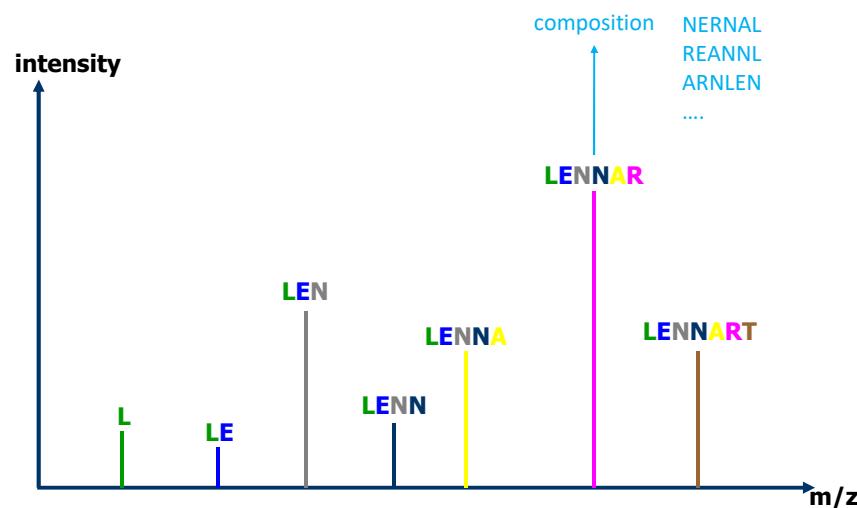
There are several other ion types that can be annotated, as well as 'internal fragments'. The latter are fragments that no longer contain an intact terminus. These are harder to use for 'ladder sequencing', but can still be interpreted.

This nomenclature was coined by **Roeperstorff and Fohlmann** (*Biomed. Mass Spec.*, 1984) and **Klaus Biemann** (*Biomed. Environ. Mass Spec.*, 1988) and is commonly referred to as 'Biemann nomenclature'. Note the link with the Roman alphabet.

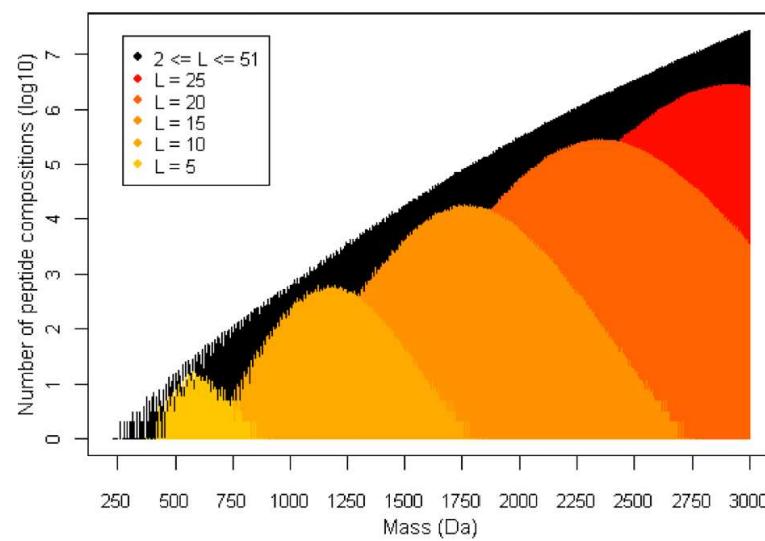
These fragment ions map to the spectrum



Individual m/z values however,
provide only composition information

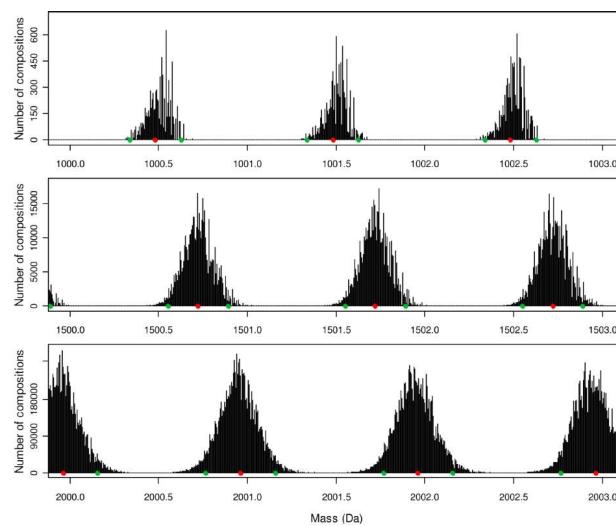


Many peptides share similar masses



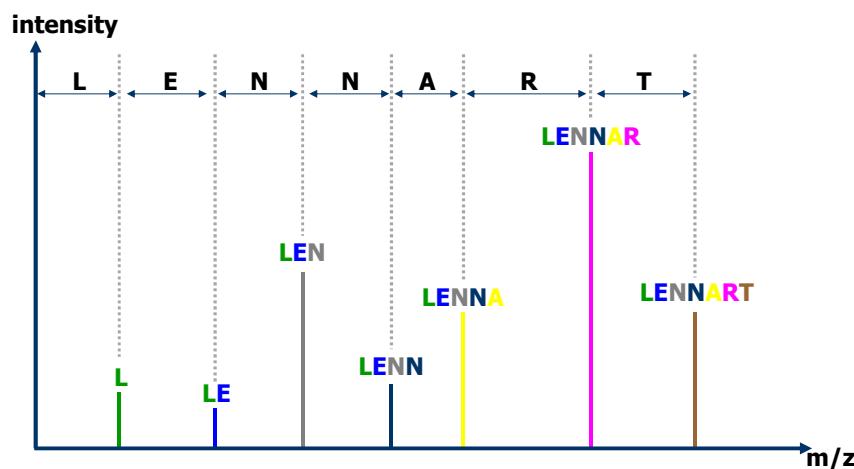
Nefedov, Journal of Proteome Research, 2011

Peptides masses occur in delineated islands, requiring sub-pmm accuracy for 1k Da masses

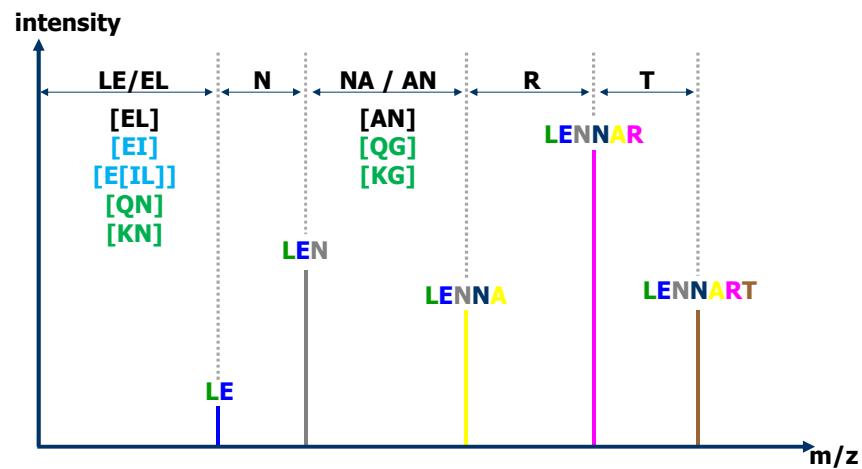


Nefedov, Journal of Proteome Research, 2011

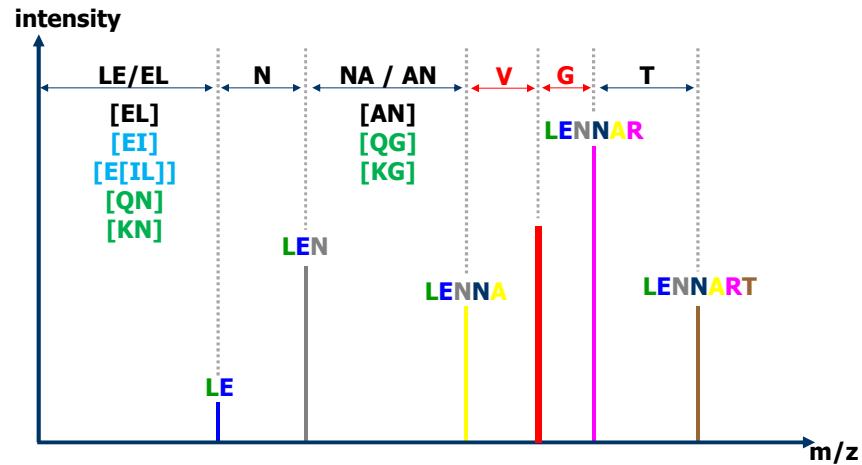
The sequence information is inferred through ladder sequencing



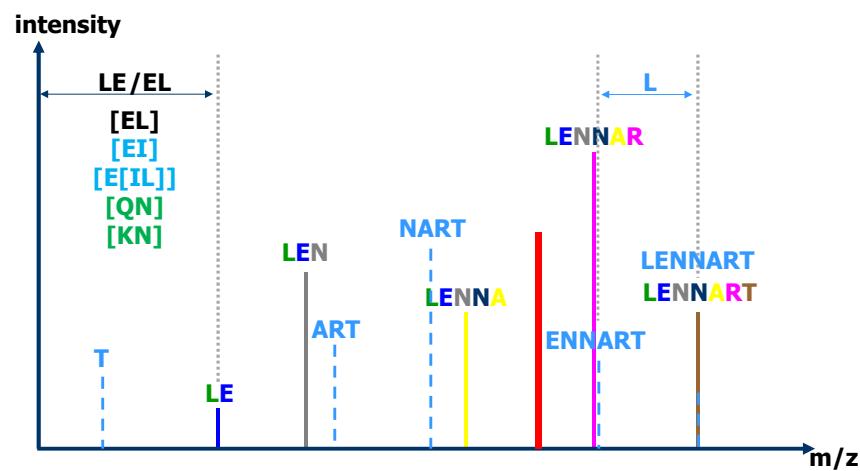
When peaks are missing, things get unsure



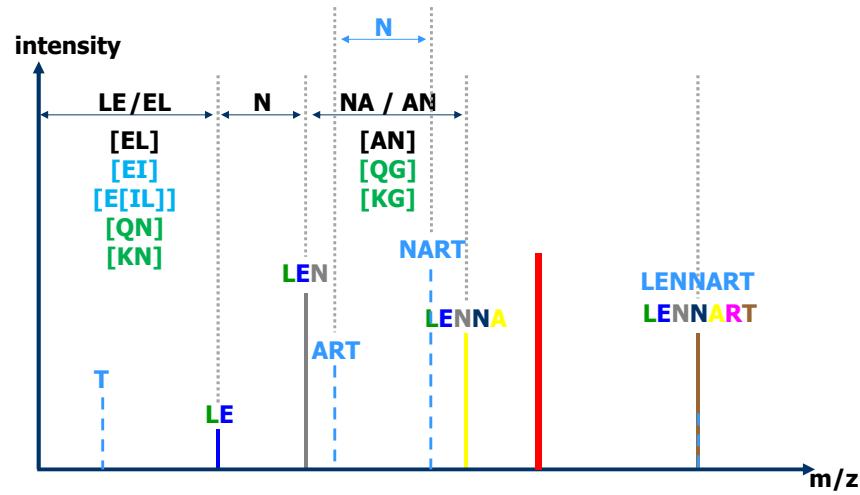
When there are interfering noise peaks,
things also get trickier

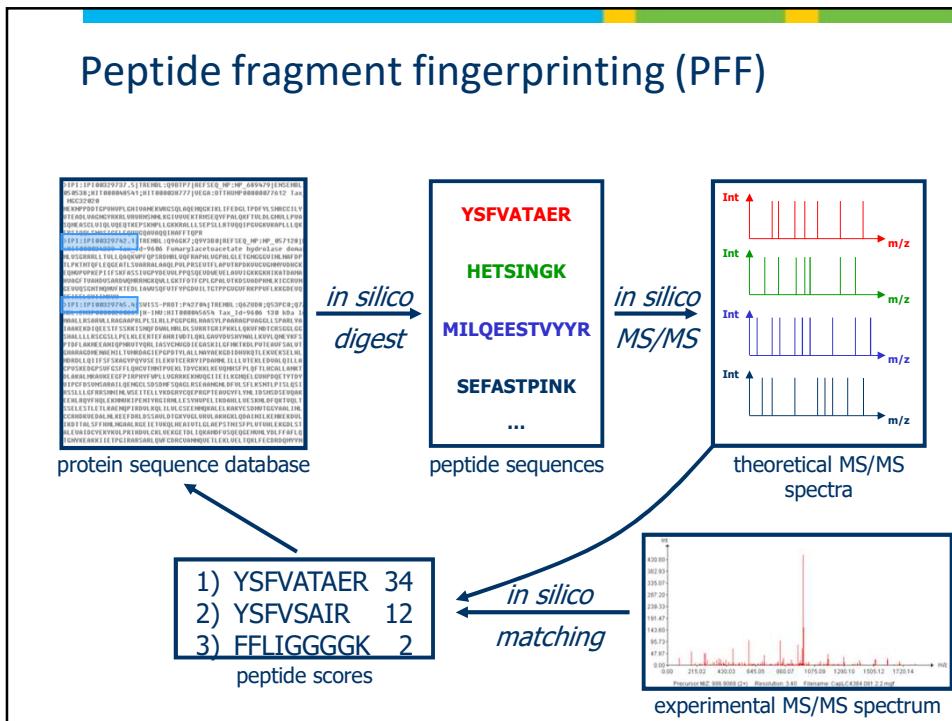
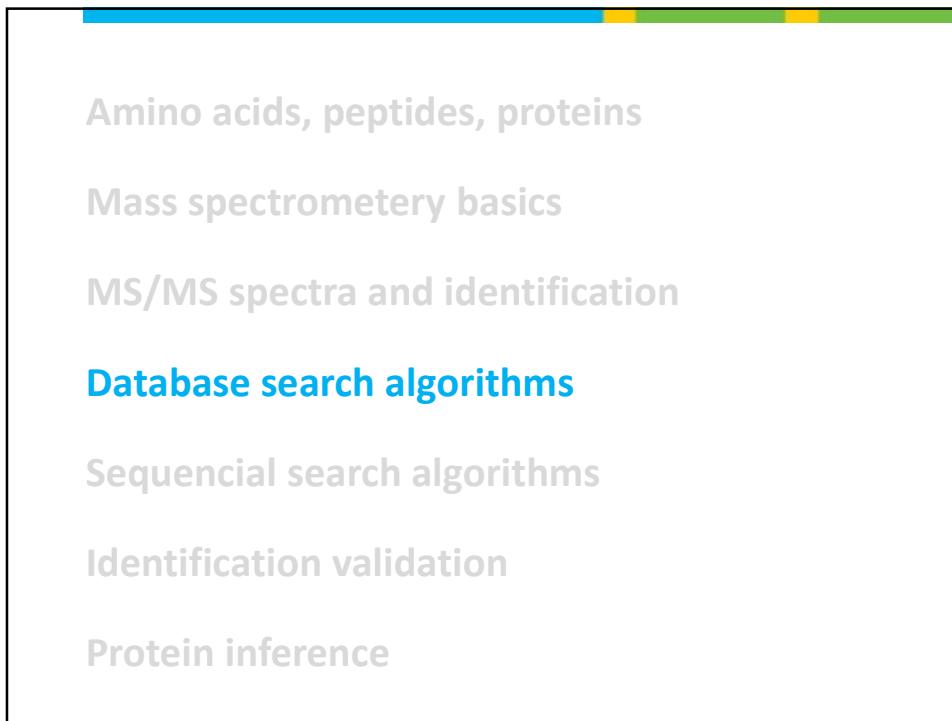


Complementary ions can be a burden, but can also help with the sequencing



Often, only a limited length of sequence can be read reliably from the spectrum





The most popular algorithm types

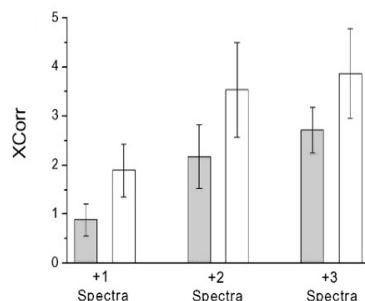
- SEQUEST (Scripps, Thermo Fisher Scientific)
<http://fields.scripps.edu/sequest>
- MASCOT (Matrix Science)
<http://www.matrixscience.com>
- X!Tandem (The Global Proteome Machine Organization)
<http://www.thegpm.org/TANDEM>

SEQUEST is the original search engine, but not that much used these days anymore

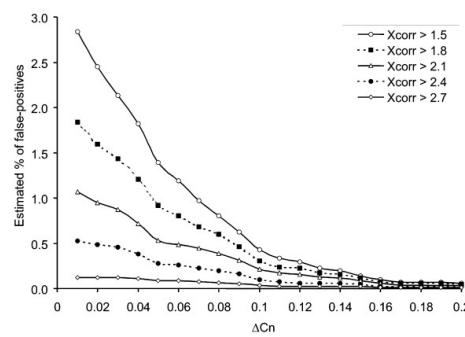
- Can be used for MS/MS (PFF) identifications
- Based on a cross-correlation score (includes peak height)
- Published core algorithm (patented, licensed to Thermo), Eng, JASMS 1994
- Provides preliminary (Sp) score, rank, cross-correlation score (XCorr), and score difference between the top two ranks (deltaCn, ΔC_n)
- Thresholding is up to the user, and is commonly done *per* charge state
- Many extensions exist to perform a more automatic validation of results

$$R_i = \sum_{j=1}^n x_j \cdot y_{(j+i)}$$
$$XCorr = R_0 - \frac{1}{151} \left(\sum_{i=-75}^{+75} R_i \right)$$
$$\text{deltaCn} = \frac{XCorr_1 - XCorr_2}{XCorr_1}$$

SEQUEST reveals the problems with scoring different charges, and using different scores



From: MacCoss et al., Anal. Chem. 2002



From: Peng et al., J. Prot. Res.. 2002

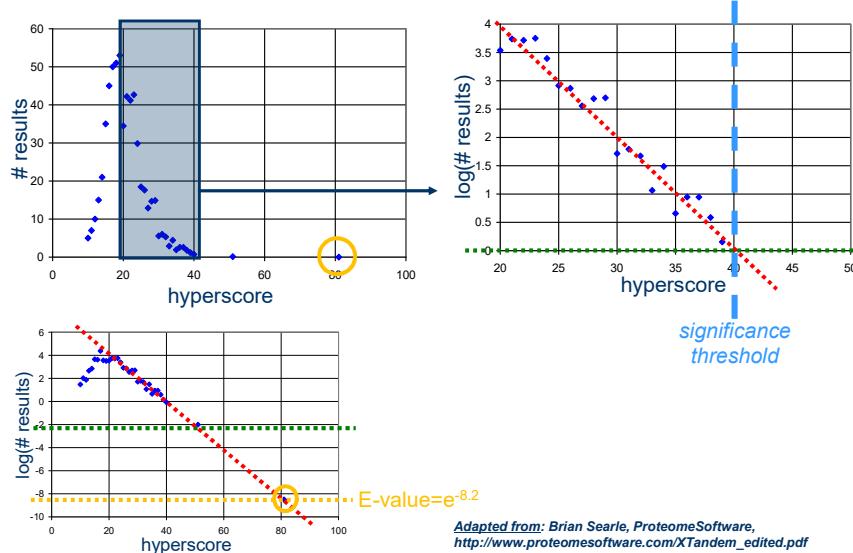
Mascot is probably the most recognized search engine, despite its secret algorithm

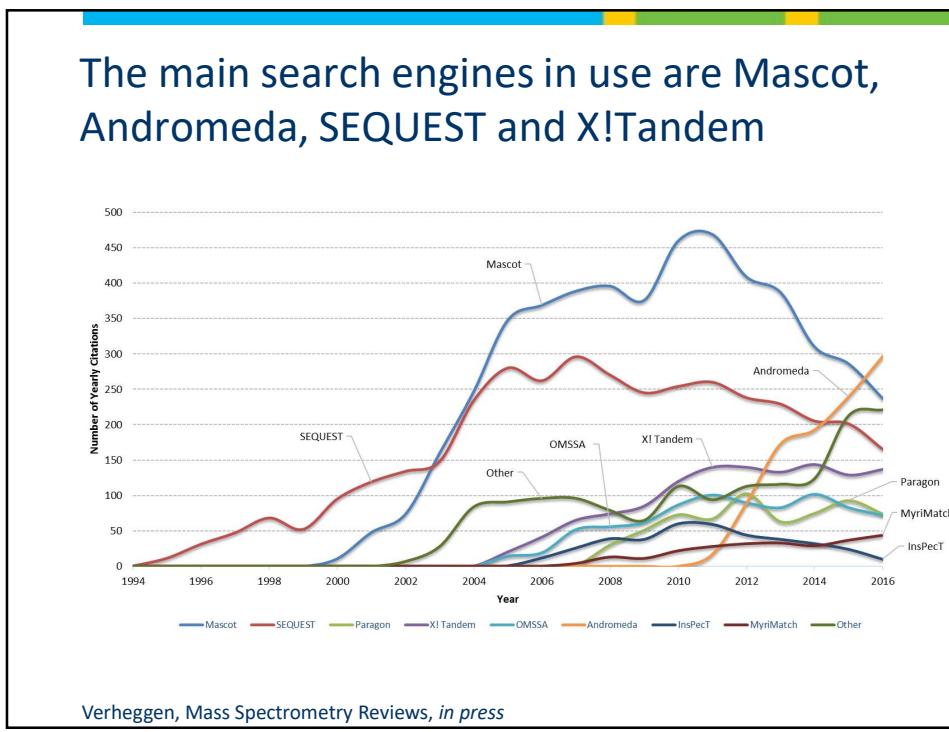
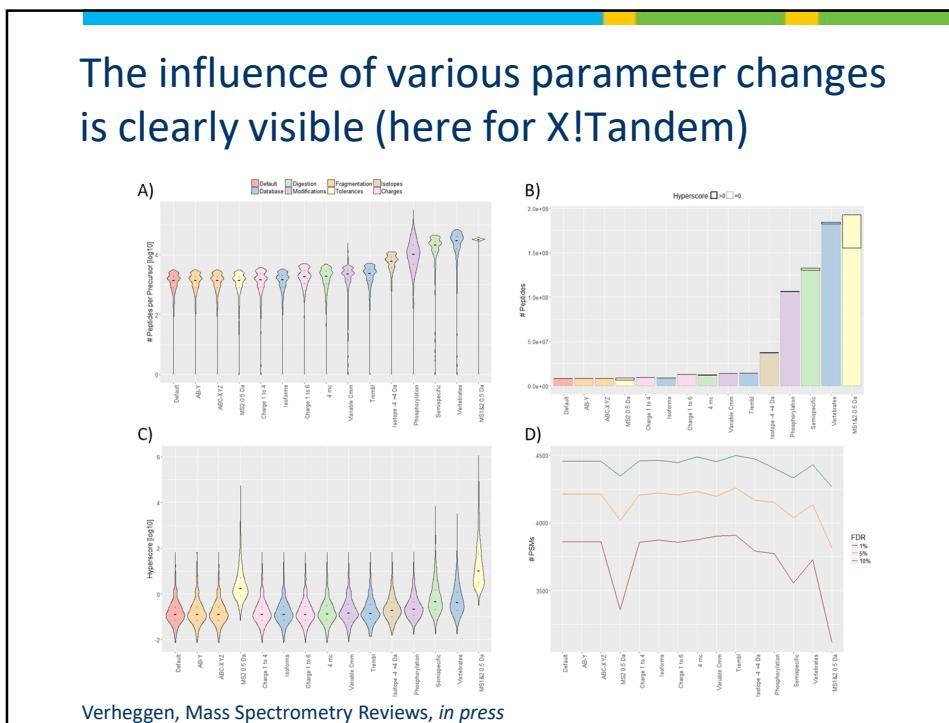
- Very well established search engine, Perkins, Electrophoresis 1999
- Can do MS (PMF) and MS/MS (PFF) identifications
- Based on the MOWSE score,
- Unpublished core algorithm (trade secret)
- Predicts an *a priori* threshold score that identifications need to pass
- From version 2.2, Mascot allows integrated decoy searches
- Provides rank, score, threshold and expectation value per identification
- Customizable confidence level for the threshold score

X!Tandem is a clear front-runner among open source search engines

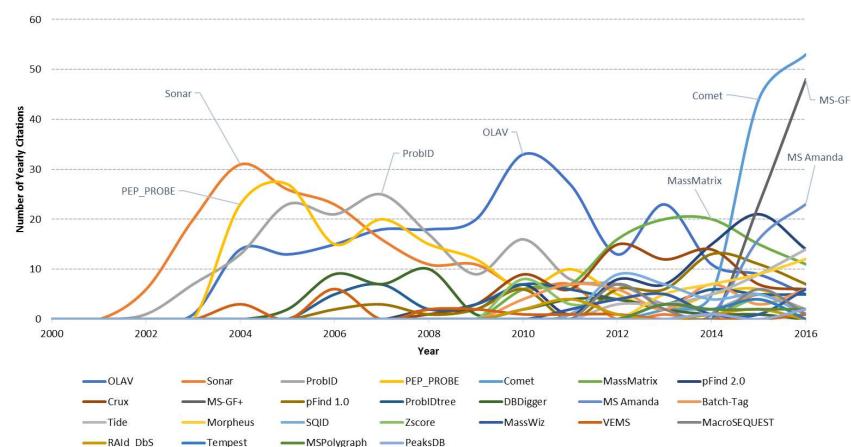
- A successful open source search engine, Craig and Beavis, *RCMS* 2003
- Can be used for MS/MS (PFF) identifications
- Based on a hyperscore (P_i is either 0 or 1): $\text{HyperScore} = \left(\sum_{i=0}^n I_i * P_i \right) * N_b! * N_y!$
- Relies on a hypergeometric distribution (hence hyperscore)
- Published core algorithm, and is freely available
- Provides hyperscore and expectancy score (the discriminating one)
- X!Tandem is fast and can handle modifications in an iterative fashion
- Has rapidly gained popularity as (auxiliary) search engine

X!Tandem's significance calculation for scores can be seen as a general template





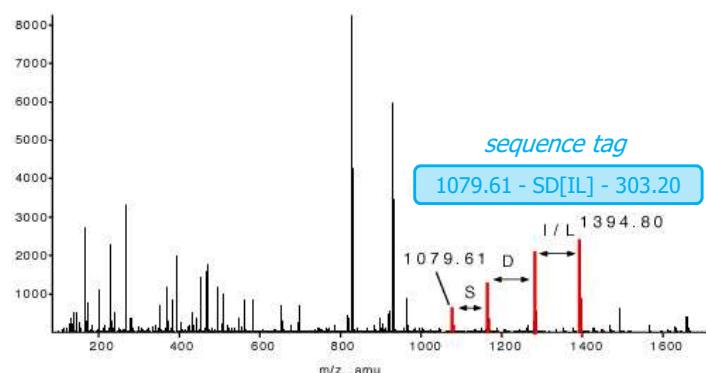
Among the up-and-coming engines, Comet, MS-GF+ and MS-Amanda are most notable



Verheggen, Mass Spectrometry Reviews, *in press*

Amino acids, peptides, proteins
Mass spectrometry basics
MS/MS spectra and identification
Database search algorithms
Sequencial search algorithms
Identification validation
Protein inference

Sequence tags are as old as SEQUEST, and these still have a role to play today

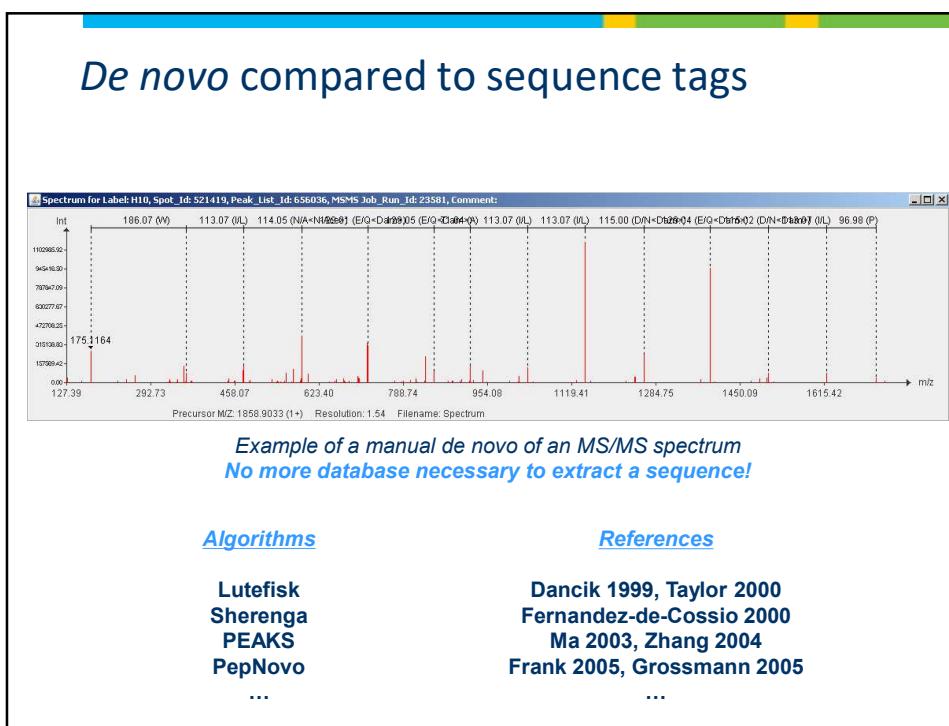
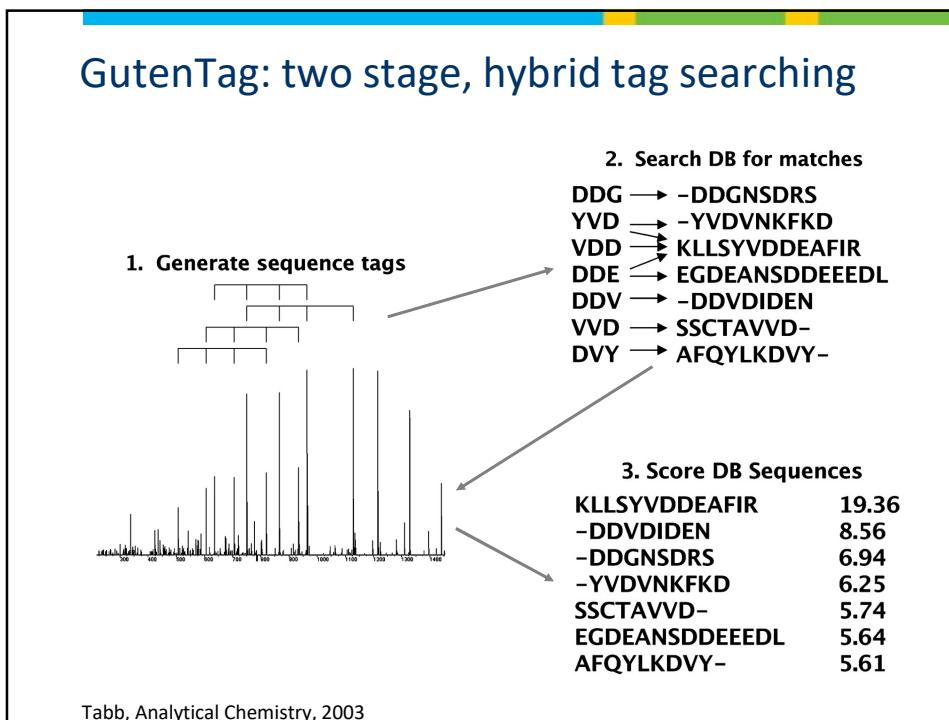


The concept of sequence tags was introduced by Mann and Wilm

Mann, Analytical Chemistry, 1994

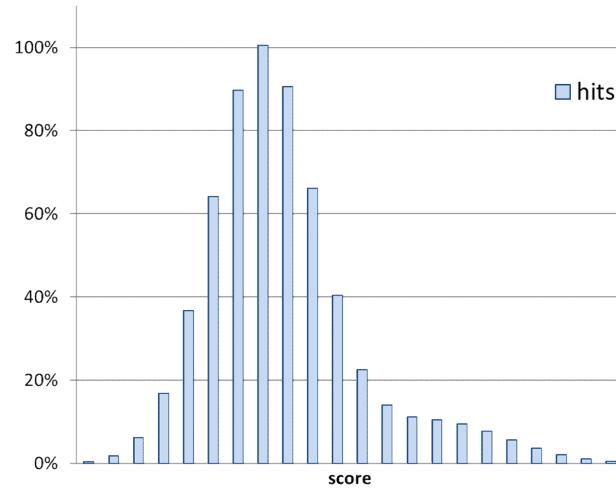
GutenTag, DirecTag, TagRecon

- Tabb, *Anal. Chem.* 2003, Tabb, *JPR* 2008, Dasari, *JPR* 2010
- Recent implementations of the sequence tag approach
- Refine hits by peak mapping in a second stage to resolve ambiguities
- Rely on an empirical fragmentation model
- Published core algorithms, DirecTag and TagRecon freely available
- GutenTag and DirecTag extract tags,
- TagRecon matches these to the database
- Very useful to retrieve unexpected peptides (modifications, variations)
- Entire workflows exist (e.g., combination with IDPicker)



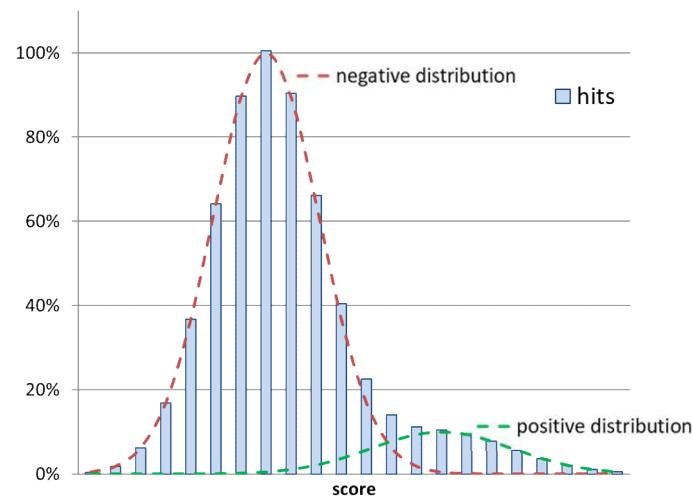
Amino acids, peptides, proteins
Mass spectrometry basics
MS/MS spectra and identification
Database search algorithms
Sequential search algorithms
Identification validation
Protein inference

All hits, good and bad together,
form a distribution of scores

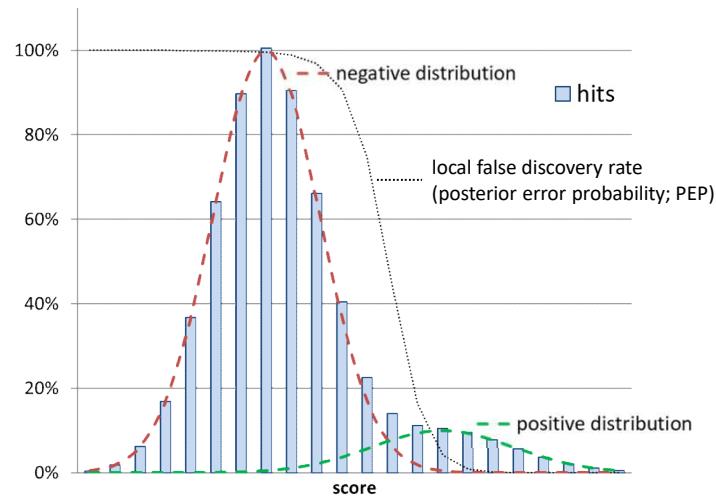


Nesvizhskii, J Proteomics, 2010

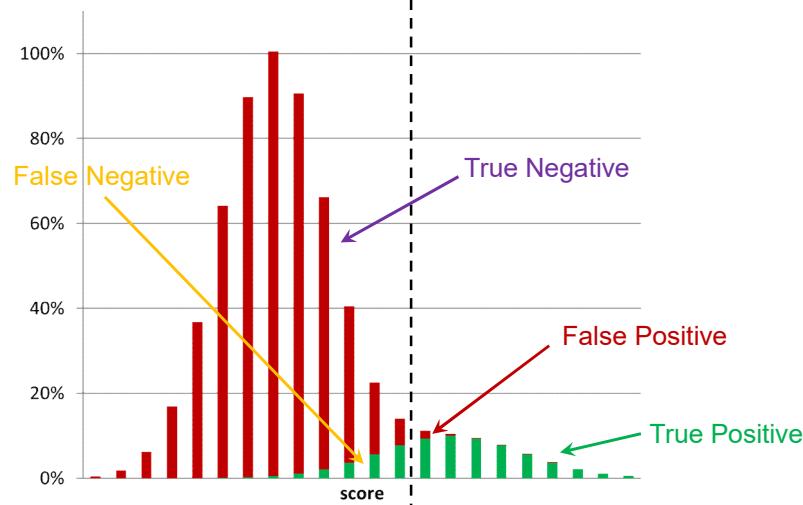
If we know how scores for bad hits distribute, we can distinguish good from bad by score



The separation is not perfect, which leads to the calculation of a local false discovery rate



Setting a threshold classifies all hits as either bad or good, which inevitably leads to errors



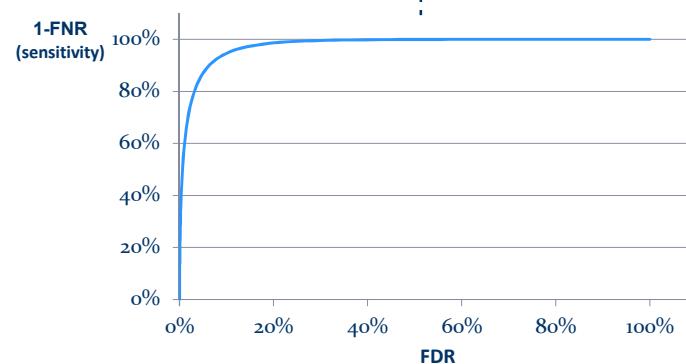
We can evaluate the effect of these errors by plotting the effect of moving the threshold

False Positive Rate \ominus

$$FDR = \frac{n_{FP}}{n_{FP} + n_{TP}}$$

False Negative Rate \oslash

$$FNR = \frac{n_{FN}}{n_{FN} + n_{TP}}$$



Decoy databases are false positive factories that are assumed to deliver reliably bad hits

Three main types of decoy DB's are used:

- Reversed databases (*easy*)

LENNARTMARTENS → SNETRAMTRANNEL

- Shuffled databases (*slightly more difficult*)

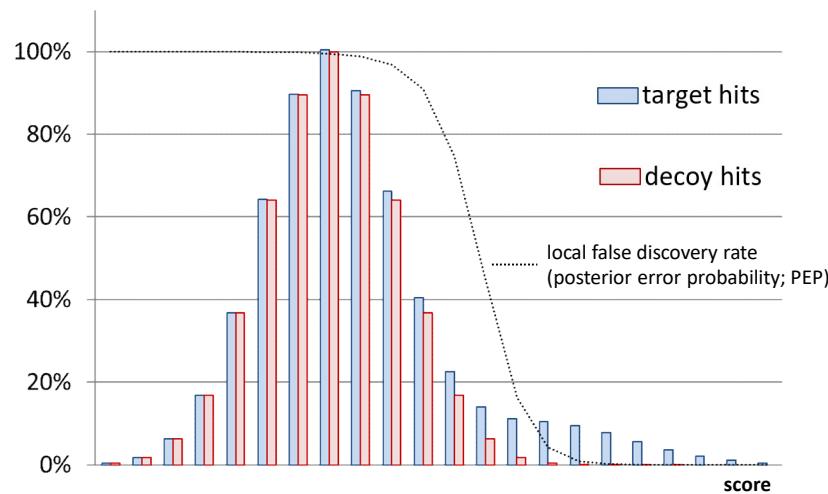
LENNARTMARTENS → NMERLANATERTTN (for instance)

- Randomized databases (*as difficult as you want it to be*)

LENNARTMARTENS → GFVLAEPHSEAITK (for instance)

The concept is that each peptide identified from the decoy database is an incorrect identification. By counting the number of decoy hits, we can estimate the number of false positives in the original database, **provided that the decoys have similar properties as the forward sequences.**

With the help of the scores of decoy hits, we can assess the score distribution of bad hits



Käll, Journal of Proteome Research, 2008

Popular automatic ID validation algorithms

- **PeptideProphet**

- published in 2002
- mainly used (and useful) for SEQUEST data
- uses on-the-fly mixture modeling based on built-in (*later versions: semi-configurable*) assumptions

- **Percolator**

- published in 2007
- developed for SEQUEST, but now also implemented in Mascot (*v2.3 and up*)
- uses on-the-fly trained support vector machine (SVM) based on decoy database hits

The Percolator workflow relies on decoys as the means to profile bad identifications

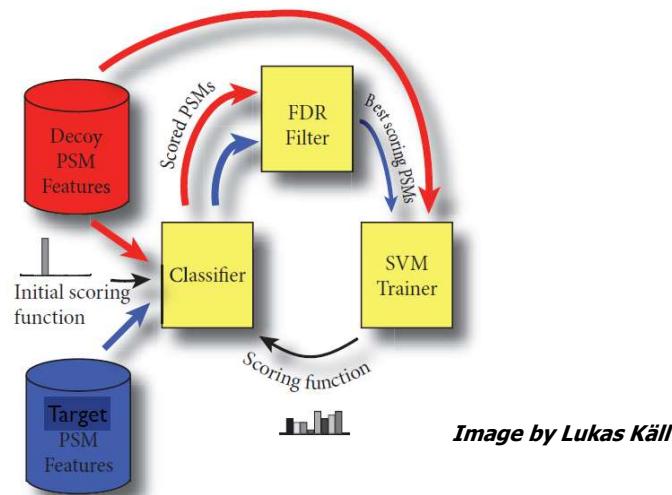


Image by Lukas Käll

Käll, Nature Methods, 2007

Unfortunately, the first ‘big’ paper
got the FDR formula wrong

$$FDR = \frac{2 \times nbr_decoy_hits}{nbr_forward_hits + nbr_decoy_hits}$$

FDR is the False Discovery Rate – it is a metric that gives you an indication of how many (percent) of your identifications are potentially incorrect. Note that we multiply the number of decoy hits by 2, because we should not only count the actual decoy hits, but also the ‘hidden’ false positives that are present in the forward identifications. The assumption here is that we expect one forward false positive hit *per* decoy false positive hit, hence the doubling term.

From: Elias and Gygi, Nature Methods 2007, 4(3): 207-214

The correct way to calculate a top-down FDR
is fortunately both simpler and more elegant

$$FDR = \frac{nbr_decoy_hits}{nbr_forward_hits}$$

This metric was proposed by Storey and Tibbs for genomics data, and further investigated by Lukas Käll for proteomics. It provides a more accurate (and simpler!) estimate of the FDR.

See: Storey and Tibbs, PNAS 2003, 100(16): 9440-9445
See: Käll et al., JPR 2008, 7(1): 29-34

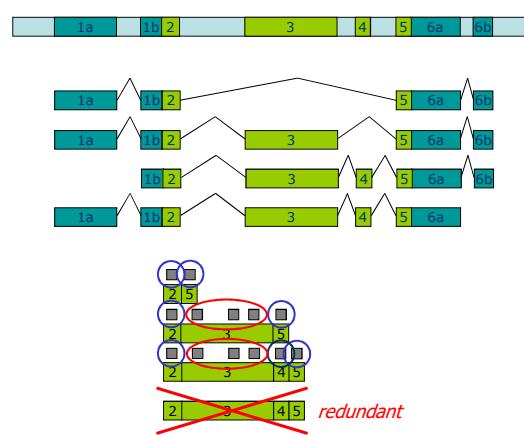
Amino acids, peptides, proteins
Mass spectrometry basics
MS/MS spectra and identification
Database search algorithms
Sequential search algorithms
Identification validation
Protein inference

Not all peptides are created equal

Gene
↓
Transcripts
↓
Translations
↓
Peptides

matching all transcripts
matching a transcript subset
matching exactly 1 translation

■ Intron ■ Exon UTR ■ Exon CDS ■ Peptide



Protein inference: a question of conviction

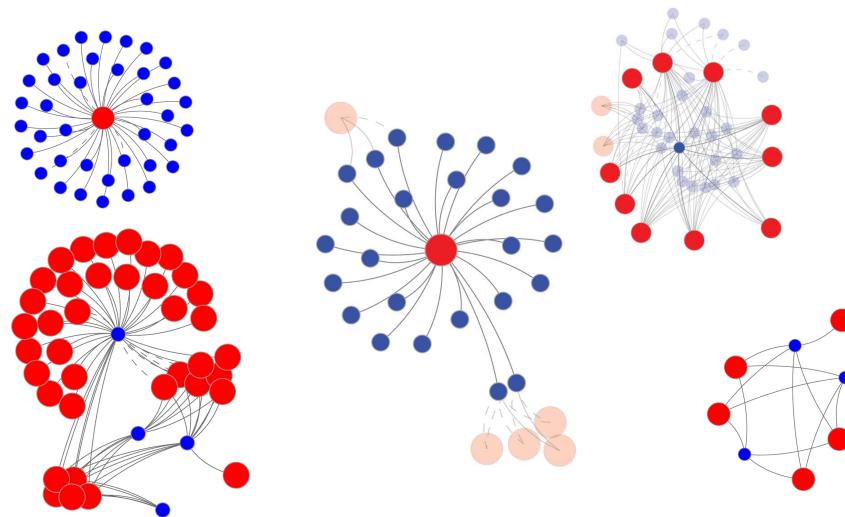
	peptides	a	b	c	d
	proteins				
Minimal set <i>Occam</i>	prot X	x		x	
	prot Y	x			
	prot Z		x	x	x

	peptides	a	b	c	d
	proteins				
Maximal set <i>anti-Occam</i>	prot X	x		x	
	prot Y	x			
	prot Z		x	x	x

	peptides	a	b	c	d
	proteins				
Minimal set with maximal annotation <i>true Occam?</i>	prot X (-)			*	*
	prot Y (+)	x			
	prot Z (0)		x	x	x

See: Martens and Hermjakob, Molecular BioSystems, 2007

In real life, protein inference issues will be mainly bad, often ugly, and occasionally good



A few algorithms for protein inference

- IDPicker

Zhang et al, Journal of Proteome Research, 2007

- ProteinProphet

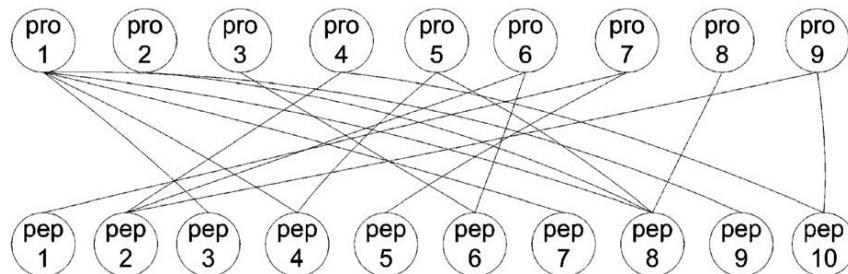
Nesvizhskii AI et al, Analytical Chemistry, 2003

- DBToolkit

Martens et al, Bioinformatics, 2005

IDPicker parsimonious protein assembly

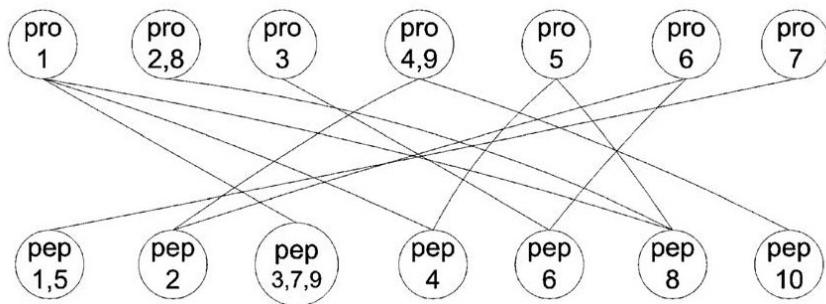
(I) Initialize



Zhang, Journal of Proteome Research, 2007

IDPicker parsimonious protein assembly

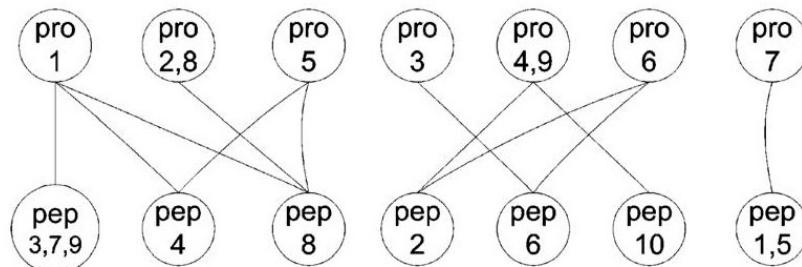
(II) Collapse



Zhang, Journal of Proteome Research, 2007

IDPicker parsimonious protein assembly

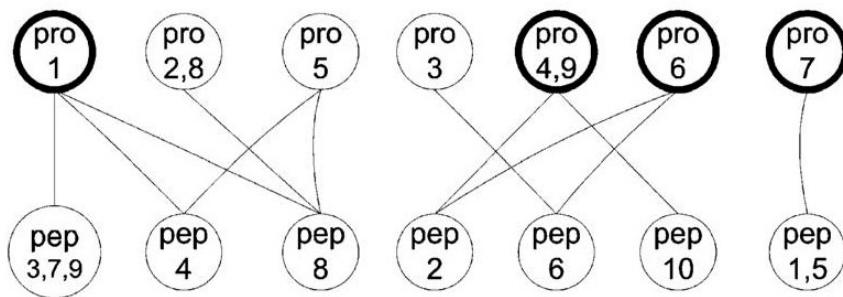
(III) Separate



Zhang, Journal of Proteome Research, 2007

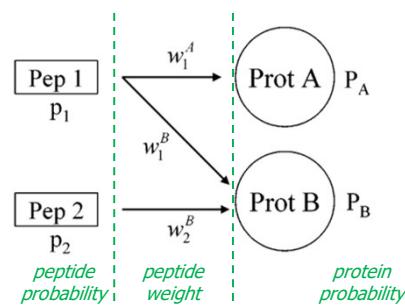
IDPicker parsimonious protein assembly

(IV) Reduce



Zhang, Journal of Proteome Research, 2007

ProteinProphet: the simplified view



$$P_n = 1 - \prod_i (1 - w_i^n p(+|D_i))$$

$$w_i^n = \frac{P_n}{\sum_{s=1 \dots N_s} P_s}$$

In iteration 1, all weights w start off as $1/n$, with n the degeneracy count for the peptide

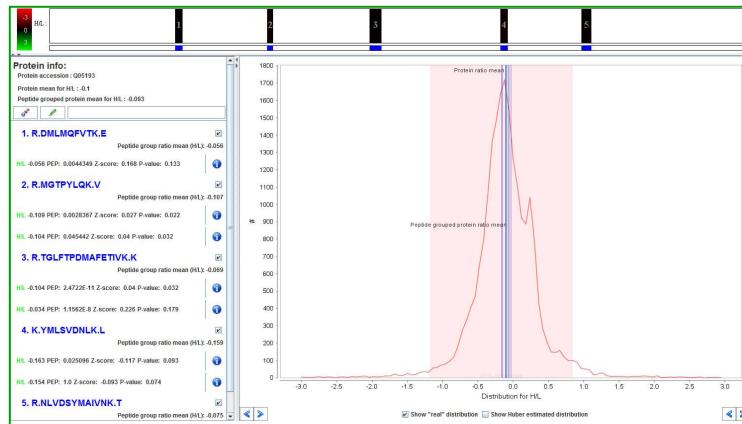
Nesvizhskii, Analytical Chemistry, 2003

DBToolkit protein inference

Accession	Start	Stop	Previous (?) Sequence	Isoforms	peptides	a	b	c	d
					prot X (-)	*	*	*	*
Minimal set with maximal annotation {									
IP0001914	105	117 K	DDTDVDPN	SWISS-PROT P0001914 P0001914.1 (105-117)* A P D O 00796349.1 (105-117)					
IP0000616	148	157 R	DDETMYY	SWISS-PROT P0000616 P0000616.1 (93-102)					
IP0002273	756	766 K	VEMFAQY	SWISS-PROT P0002273 P0002273.1 (756-766)* A P D O 00816182.1 (105-117)					
IP0003203	429	441 R	SEDPOTS	SWISS-PROT P0003203 P0003203.1 (429-441)					
IP0001956	1504	1513 R	TEMEDLM	SWISS-PROT P0001956 P0001956.1 (107-116)					
IP00021975	83	101 K	DQGEAAAL	SWISS-PROT P00021975 P00021975.1 (83-101)					
IP0002203C	109	125 R	ELAILLGN	SWISS-PROT P0002203C P0002203C.1 (125-126)* A P D O 0079108.1 (423-441)					
IP0001395	265	281 K	VLVAVNQES	SWISS-PROT P0001395 P0001395.1 (265-281)					
IP000096C	21	30 K	AGFAGACD	SWISS-PROT P000096C P000096C.1 (21-30)* A P D O 0021439.1 (19-28)					
IP00032812	180	189 R	DMTTGIVS	SWISS-PROT P00032812 P00032812.1 (180-189)* A P D O 00742780.1 (107-116)					
IP0047776	656	665 K	QMESELE	SWISS-PROT P0047776 P0047776.1 (656-665)* A P D O 00000355237 REFSEQ NP_006026 H-INV HITO					
IP000039C	478	490 K	DGVMEMI	SWISS-PROT P000039C P000039C.1 (478-490)					
IP0003354	2622	2631 K	DKGEYTL	SWISS-PROT P0003354 P0003354.2 (2614-2623)* A P D O 00644576.1 (2582-2591)					
IP0001959	1182	1191 K	THEAQIQE	SWISS-PROT P0001959 P0001959.1 (1182-1191)* A P D O 0079579 TREMBL A8J7R2;Q8ETR8;Q8EZQ3 ENSEMBL ENSP00000355237 REFSEQ NP_006026 H-INV HITO					
IP0000657	186	198 R	MSLFYAE	SWISS-PROT P0000657 P0000657.2 (187-198)* A P D O 0065170.1 (143-156)* A P D O 00789253.1 (187-199)					
IP0029777	157	170 R	QDLUMNIA	SWISS-PROT P0029777 P0029777.1 (170-170)					
IP0000896	154	153 R	NATVGEQES	SWISS-PROT P0000896 P0000896.1 (154-153)* A P D O 0074519.3 (154-153)* A P D O 008162.1 (154-163)					
IP0003645	248	264 K	G-S-ETEG	SWISS-PROT P0003645 P0003645.1 (248-264)* A P D O 00726763.03 REFSEQ NP_005132 H-INV HITO					
IP0003726	581	593 K	SQPIPMP	SWISS-PROT P0003726 P0003726.4 (602-619)* A P D O 00235412.5 (681-693)* A P D O 0446579.4 (607-72)* A P D O 0473086.3 (570-582)* A P D O 0565					
IP0003554	354	365 K	NAMSGSLASW	SWISS-PROT P0003554 P0003554.5 (359-365)* A P D O 00733415.2 (611-72)* A P D O 00853603.1 (170-181)					
IP0001321	132	139 K	YGEMPV	SWISS-PROT P0001321 P0001321.1 (132-139)* A P D O 007162.2 (271-283)* A P D O 0072343.1 (478-490)					
IP0002917	200	210 K	ELLPVVLIS	SWISS-PROT P0002917 P0002917.2 (200-210)* A P D O 00394973.1 (127-137)					
IP0003238	211	222 K	LDYDEDA	SWISS-PROT P0003238 P0003238.1 (211-222)* A P D O 0088339.1 (211-222)* A P D O 0982521.1 (124-135)					
IP0021952	435	447 R	HEMLPAS	SWISS-PROT P0021952 P0021952.2 (468-478)					
IP0029786	116	127 K	QTLNQLC	TREMBL Q8N7N2 ENSEMBL ENSP00000317768 Tax_Id:9606 Gene_Symbol=:cDNA FLJ40807 fis, clone TRACH2009268					
IP0000511	750	761 K	MISGMYI	SWISS-PROT P0000511 P0000511.2 (744-755)* A P D O 00720663.3 (743-754)* A P D O 00720664.6 (745-759)* A P D O 00720665.7 (746-757)* A P D O 00720666.8 (747-758)* A P D O 00720667.9 (748-759)* A P D O 00720668.0 (749-759)* A P D O 00720669.1 (750-759)* A P D O 00720670.2 (751-759)* A P D O 00720671.3 (752-759)* A P D O 00720672.4 (753-759)* A P D O 00720673.5 (754-759)* A P D O 00720674.6 (755-759)* A P D O 00720675.7 (756-759)* A P D O 00720676.8 (757-759)* A P D O 00720677.9 (758-759)* A P D O 00720678.0 (759-759)* A P D O 00720679.1 (760-759)* A P D O 00720680.2 (761-759)* A P D O 00720681.3 (762-759)* A P D O 00720682.4 (763-759)* A P D O 00720683.5 (764-759)* A P D O 00720684.6 (765-759)* A P D O 00720685.7 (766-759)* A P D O 00720686.8 (767-759)* A P D O 00720687.9 (768-759)* A P D O 00720688.0 (769-759)* A P D O 00720689.1 (770-759)* A P D O 00720690.2 (771-759)* A P D O 00720691.3 (772-759)* A P D O 00720692.4 (773-759)* A P D O 00720693.5 (774-759)* A P D O 00720694.6 (775-759)* A P D O 00720695.7 (776-759)* A P D O 00720696.8 (777-759)* A P D O 00720697.9 (778-759)* A P D O 00720698.0 (779-759)* A P D O 00720699.1 (780-759)* A P D O 00720600.2 (781-759)* A P D O 00720601.3 (782-759)* A P D O 00720602.4 (783-759)* A P D O 00720603.5 (784-759)* A P D O 00720604.6 (785-759)* A P D O 00720605.7 (786-759)* A P D O 00720606.8 (787-759)* A P D O 00720607.9 (788-759)* A P D O 00720608.0 (789-759)* A P D O 00720609.1 (790-759)* A P D O 00720610.2 (791-759)* A P D O 00720611.3 (792-759)* A P D O 00720612.4 (793-759)* A P D O 00720613.5 (794-759)* A P D O 00720614.6 (795-759)* A P D O 00720615.7 (796-759)* A P D O 00720616.8 (797-759)* A P D O 00720617.9 (798-759)* A P D O 00720618.0 (799-759)* A P D O 00720619.1 (800-759)* A P D O 00720620.2 (801-759)* A P D O 00720621.3 (802-759)* A P D O 00720622.4 (803-759)* A P D O 00720623.5 (804-759)* A P D O 00720624.6 (805-759)* A P D O 00720625.7 (806-759)* A P D O 00720626.8 (807-759)* A P D O 00720627.9 (808-759)* A P D O 00720628.0 (809-759)* A P D O 00720629.1 (810-759)* A P D O 00720630.2 (811-759)* A P D O 00720631.3 (812-759)* A P D O 00720632.4 (813-759)* A P D O 00720633.5 (814-759)* A P D O 00720634.6 (815-759)* A P D O 00720635.7 (816-759)* A P D O 00720636.8 (817-759)* A P D O 00720637.9 (818-759)* A P D O 00720638.0 (819-759)* A P D O 00720639.1 (820-759)* A P D O 00720640.2 (821-759)* A P D O 00720641.3 (822-759)* A P D O 00720642.4 (823-759)* A P D O 00720643.5 (824-759)* A P D O 00720644.6 (825-759)* A P D O 00720645.7 (826-759)* A P D O 00720646.8 (827-759)* A P D O 00720647.9 (828-759)* A P D O 00720648.0 (829-759)* A P D O 00720649.1 (830-759)* A P D O 00720650.2 (831-759)* A P D O 00720651.3 (832-759)* A P D O 00720652.4 (833-759)* A P D O 00720653.5 (834-759)* A P D O 00720654.6 (835-759)* A P D O 00720655.7 (836-759)* A P D O 00720656.8 (837-759)* A P D O 00720657.9 (838-759)* A P D O 00720658.0 (839-759)* A P D O 00720659.1 (840-759)* A P D O 00720660.2 (841-759)* A P D O 00720661.3 (842-759)* A P D O 00720662.4 (843-759)* A P D O 00720663.5 (844-759)* A P D O 00720664.6 (845-759)* A P D O 00720665.7 (846-759)* A P D O 00720666.8 (847-759)* A P D O 00720667.9 (848-759)* A P D O 00720668.0 (849-759)* A P D O 00720669.1 (850-759)* A P D O 00720670.2 (851-759)* A P D O 00720671.3 (852-759)* A P D O 00720672.4 (853-759)* A P D O 00720673.5 (854-759)* A P D O 00720674.6 (855-759)* A P D O 00720675.7 (856-759)* A P D O 00720676.8 (857-759)* A P D O 00720677.9 (858-759)* A P D O 00720678.0 (859-759)* A P D O 00720679.1 (860-759)* A P D O 00720680.2 (861-759)* A P D O 00720681.3 (862-759)* A P D O 00720682.4 (863-759)* A P D O 00720683.5 (864-759)* A P D O 00720684.6 (865-759)* A P D O 00720685.7 (866-759)* A P D O 00720686.8 (867-759)* A P D O 00720687.9 (868-759)* A P D O 00720688.0 (869-759)* A P D O 00720689.1 (870-759)* A P D O 00720690.2 (871-759)* A P D O 00720691.3 (872-759)* A P D O 00720692.4 (873-759)* A P D O 00720693.5 (874-759)* A P D O 00720694.6 (875-759)* A P D O 00720695.7 (876-759)* A P D O 00720696.8 (877-759)* A P D O 00720697.9 (878-759)* A P D O 00720698.0 (879-759)* A P D O 00720699.1 (880-759)* A P D O 00720600.2 (881-759)* A P D O 00720601.3 (882-759)* A P D O 00720602.4 (883-759)* A P D O 00720603.5 (884-759)* A P D O 00720604.6 (885-759)* A P D O 00720605.7 (886-759)* A P D O 00720606.8 (887-759)* A P D O 00720607.9 (888-759)* A P D O 00720608.0 (889-759)* A P D O 00720609.1 (890-759)* A P D O 00720610.2 (891-759)* A P D O 00720611.3 (892-759)* A P D O 00720612.4 (893-759)* A P D O 00720613.5 (894-759)* A P D O 00720614.6 (895-759)* A P D O 00720615.7 (896-759)* A P D O 00720616.8 (897-759)* A P D O 00720617.9 (898-759)* A P D O 00720618.0 (899-759)* A P D O 00720619.1 (900-759)* A P D O 00720620.2 (901-759)* A P D O 00720621.3 (902-759)* A P D O 00720622.4 (903-759)* A P D O 00720623.5 (904-759)* A P D O 00720624.6 (905-759)* A P D O 00720625.7 (906-759)* A P D O 00720626.8 (907-759)* A P D O 00720627.9 (908-759)* A P D O 00720628.0 (909-759)* A P D O 00720629.1 (910-759)* A P D O 00720630.2 (911-759)* A P D O 00720631.3 (912-759)* A P D O 00720632.4 (913-759)* A P D O 00720633.5 (914-759)* A P D O 00720634.6 (915-759)* A P D O 00720635.7 (916-759)* A P D O 00720636.8 (917-759)* A P D O 00720637.9 (918-759)* A P D O 00720638.0 (919-759)* A P D O 00720639.1 (920-759)* A P D O 00720640.2 (921-759)* A P D O 00720641.3 (922-759)* A P D O 00720642.4 (923-759)* A P D O 00720643.5 (924-759)* A P D O 00720644.6 (925-759)* A P D O 00720645.7 (926-759)* A P D O 00720646.8 (927-759)* A P D O 00720647.9 (928-759)* A P D O 00720648.0 (929-759)* A P D O 00720649.1 (930-759)* A P D O 00720650.2 (931-759)* A P D O 00720651.3 (932-759)* A P D O 00720652.4 (933-759)* A P D O 00720653.5 (934-759)* A P D O 00720654.6 (935-759)* A P D O 00720655.7 (936-759)* A P D O 00720656.8 (937-759)* A P D O 00720657.9 (938-759)* A P D O 00720658.0 (939-759)* A P D O 00720659.1 (940-759)* A P D O 00720660.2 (941-759)* A P D O 00720661.3 (942-759)* A P D O 00720662.4 (943-759)* A P D O 00720663.5 (944-759)* A P D O 00720664.6 (945-759)* A P D O 00720665.7 (946-759)* A P D O 00720666.8 (947-759)* A P D O 00720667.9 (948-759)* A P D O 00720668.0 (949-759)* A P D O 00720669.1 (950-759)* A P D O 00720670.2 (951-759)* A P D O 00720671.3 (952-759)* A P D O 00720672.4 (953-759)* A P D O 00720673.5 (954-759)* A P D O 00720674.6 (955-759)* A P D O 00720675.7 (956-759)* A P D O 00720676.8 (957-759)* A P D O 00720677.9 (958-759)* A P D O 00720678.0 (959-759)* A P D O 00720679.1 (960-759)* A P D O 00720680.2 (961-759)* A P D O 00720681.3 (962-759)* A P D O 00720682.4 (963-759)* A P D O 00720683.5 (964-759)* A P D O 00720684.6 (965-759)* A P D O 00720685.7 (966-759)* A P D O 00720686.8 (967-759)* A P D O 00720687.9 (968-759)* A P D O 00720688.0 (969-759)* A P D O 00720689.1 (970-759)* A P D O 00720690.2 (971-759)* A P D O 00720691.3 (972-759)* A P D O 00720692.4 (973-759)* A P D O 00720693.5 (974-759)* A P D O 00720694.6 (975-759)* A P D O 00720695.7 (976-759)* A P D O 00720696.8 (977-759)* A P D O 00720697.9 (978-759)* A P D O 00720698.0 (979-759)* A P D O 00720699.1 (980-759)* A P D O 00720600.2 (981-759)* A P D O 00720601.3 (982-759)* A P D O 00720602.4 (983-759)* A P D O 00720603.5 (984-759)* A P D O 00720604.6 (985-759)* A P D O 00720605.7 (986-759)* A P D O 00720606.8 (987-759)* A P D O 00720607.9 (988-759)* A P D O 00720608.0 (989-759)* A P D O 00720609.1 (990-759)* A P D O 00720610.2 (991-759)* A P D O 00720611.3 (992-759)* A P D O 00720612.4 (993-759)* A P D O 00720613.5 (994-759)* A P D O 00720614.6 (995-759)* A P D O 00720615.7 (996-759)* A					

Some inference examples (i)

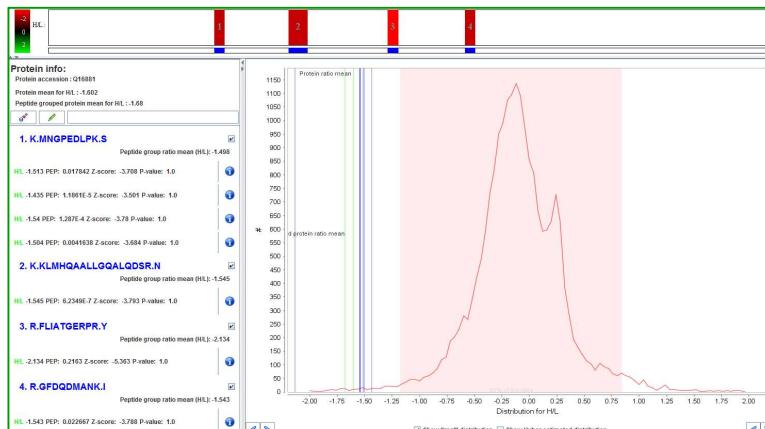
<http://genesis.ugent.be/rover>



Nice and easy, 1/1, only unique peptides (blue) and a narrow distribution

Colaert, Proteomics, 2010

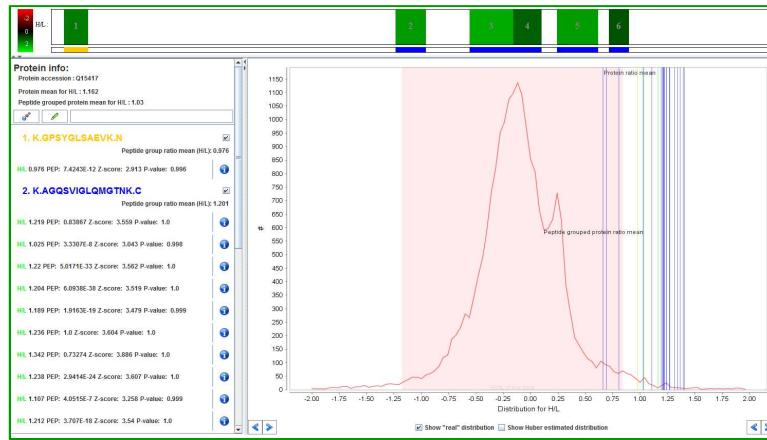
Some inference examples (ii)



Nice and easy, down-regulated

Colaert, Proteomics, 2010

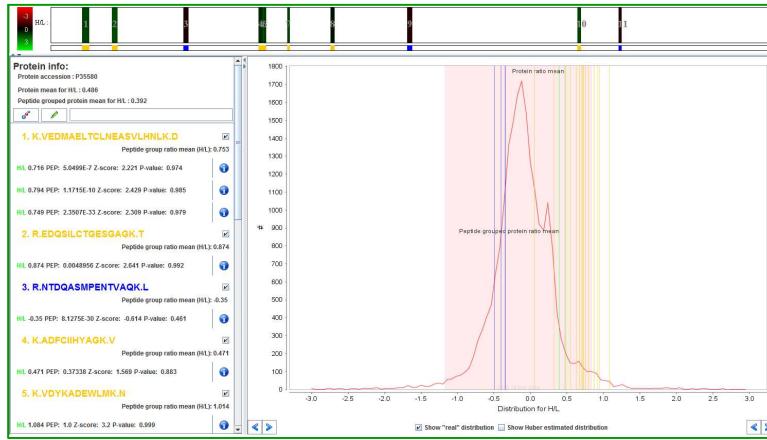
Some inference examples (iii)



A little less easy, up-regulated

Colaert, Proteomics, 2010

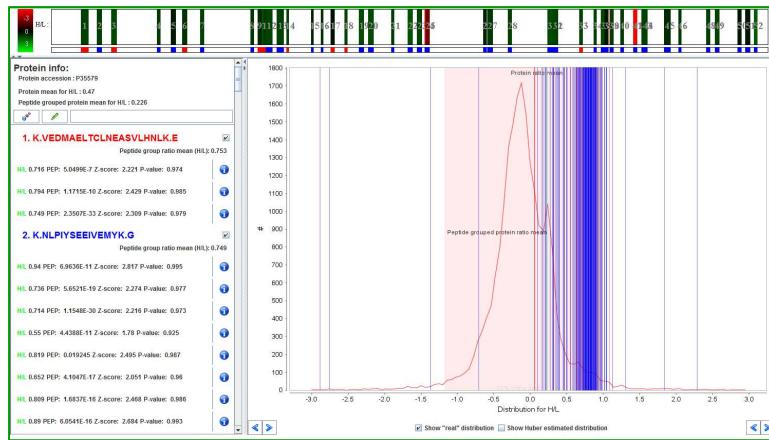
Some inference examples (iv)



A nice example of the mess of degenerate peptides

Colaert, Proteomics, 2010

Some inference examples (v)



A bit of chaos, but a defined core distribution

Colaert, Proteomics, 2010