



Purdue-UAB Botanicals Center for Age-Related Disease

Peptide mass fingerprinting

Landon Wilson

MS operator, UAB

**Purdue-UAB Botanical Center Workshop 2002
Mass Spectrometry Methods in Botanicals Research**



Purdue-UAB Botanicals Center for Age-Related Disease

Session Overview

- **Introduction**
- **Sample Preparation**
- **Database Search**



Peptide mass fingerprinting

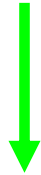
- This method has been developed because of the availability of predicted protein sequences from genome sequencing
- Proteins do not have to have been previously sequenced - only that the open reading frame in the gene is known - the rest is a virtual exercise in the hands of statisticians, bioinformaticists and computers



Purdue-UAB Botanicals Center for Age-Related Disease

From DNA to peptide fragments

.ATG.CTT.CCT.CAC.GGT.AAA.TCG.TAT.GCT....



NH₂-Met.Leu.Pro.His.Gly.Lys.Ser.Tyr.Ala....



↑
Trypsin

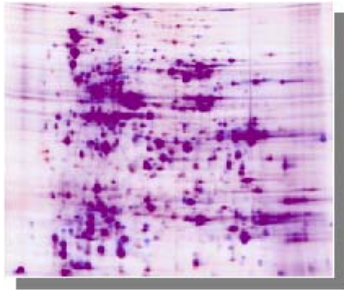
NH₂-Met.Leu.Pro.His.Gly.Lys-COOH



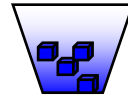
From Proteins to Sequence Tags

- If each protein (average 500 residues) had a cleavage site every 10 residues, then about 1.5 million peptides describe the expressed products of the human genome
- Each peptide has a molecular weight value that is its individual sequence tag
- Any modification will increase the peptide's molecular weight

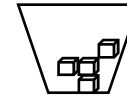
Peptide fingerprinting



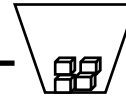
Eppendorf
tube



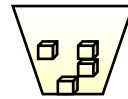
destain



Speed-Vac



trypsin
1:20



Choice of peptidase

- Analogous to DNA restriction enzymes
- Tryptic peptide fingerprinting may identify several related protein candidates (e.g., actins)
- Inspection of the sequences may reveal that there is a difference at one residue that distinguishes between two candidates.
- If for instance it is a glutamate, then use of Glu-C or V8-protease may enable the two proteins to be correctly identified
- INSPECT sequences carefully

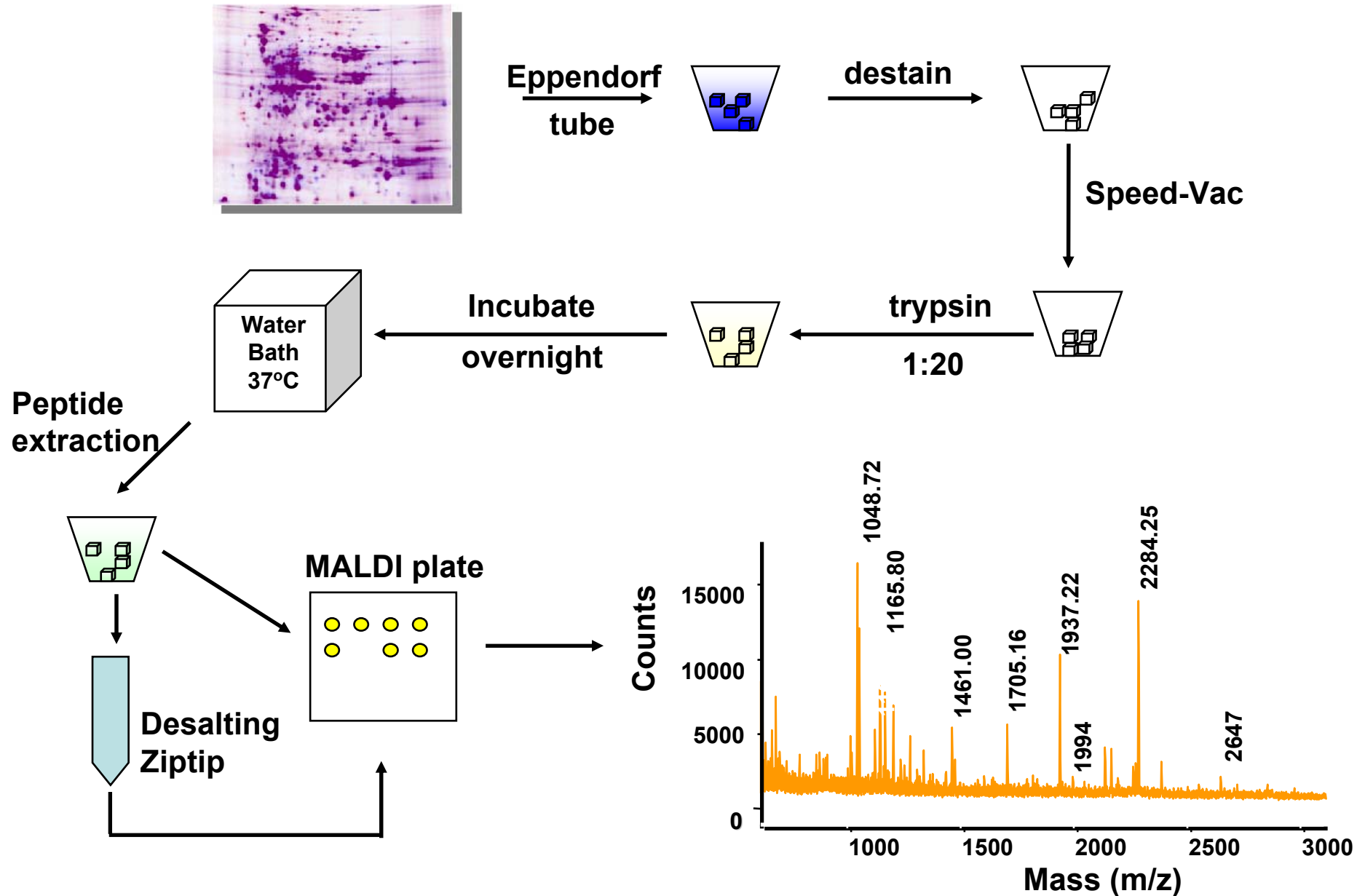
Proteolytic enzymes used to hydrolyze proteins

The choice of enzyme largely depends on the nature of the amino acid sequence and the specific issue that is being addressed

- Trypsin - *cleaves at arginine and lysine residues*
- Chymotrypsin - *cleaves hydrophobic residues*
- Arg-C - *cleaves at arginine residues*
- Glu-C - *cleaves at glutamic acid residues*
- Lys-C - *cleaves at lysine residues*
- V8-protease - *cleaves at glutamic acid residues*
- Pepsin - *cleaves randomly, but at acid pH*

See http://www.abrf.org/JBT/1998/September98/sep98m_r.html

Peptide fingerprinting





Genomics and proteins in 2002

- **The human genome consists of about 30,000 genes that are expressed as proteins**
- **Large Scale Biology Corp has cataloged 116,000+ protein forms from human tissues, representing the expressed products of 18,000 genes**
- **The expected number of protein forms is expected to be in excess of 200,000**



Purdue-UAB Botanicals Center for Age-Related Disease

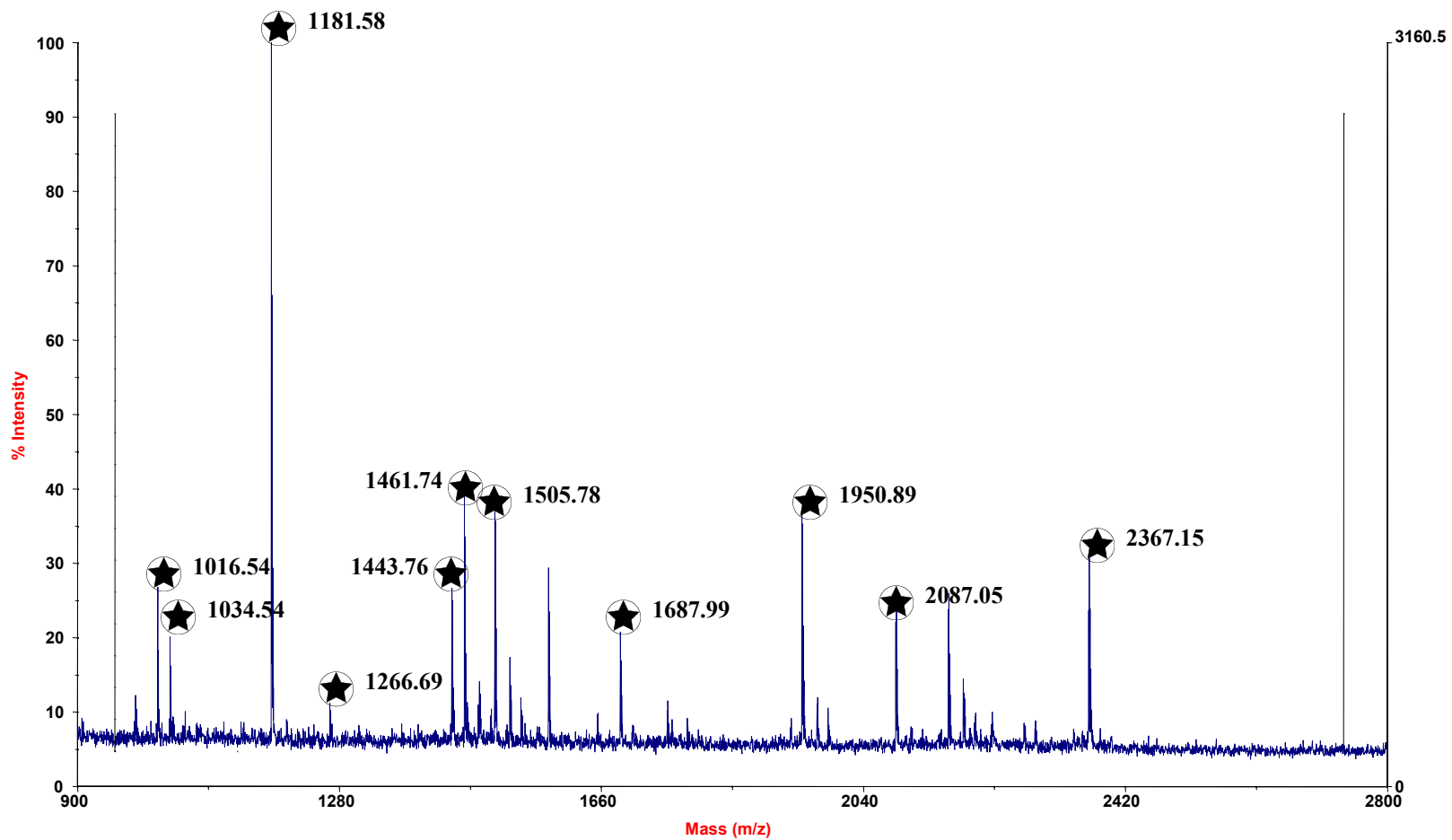
Searching databases with peptide masses to identify proteins

Best site is at www.matrixscience.com

The program (MASCOT) can search the OWL or NCBI databases using a set of tryptic peptide masses, or the fragment ions (specified or unspecified) of peptides

Presents the expected set of tryptic peptides for each matched protein

MALDI-TOF mass spectrum of tryptic digest of porin-P1 Voltage-dependent anion-selective channel



MASCOT Search Query

Mascot: Peptide Mass Fingerprint

Your name Email

Search title

Database

Taxonomy

Enzyme Allow up to missed cleavages

Fixed modifications: Acetyl (K), Acetyl (N-term), Amide (C-term), Biotinylated (K), Biotinylated (N-term)

Variable modifications: Acetyl (K), Acetyl (N-term), Amide (C-term), Biotinylated (K), Biotinylated (N-term)

Protein mass kDa Peptide tol. ppm

Mass values MH⁺ M_r Monoisotopic Average

Data file

Query
NB Contents of this field are ignored if a data file is specified.

Overview Report top hits

Taxonomy

Enzyme

of Allowed Missed Cleavages

Modifications

Peptide Tolerance

Query

Top Search Queries

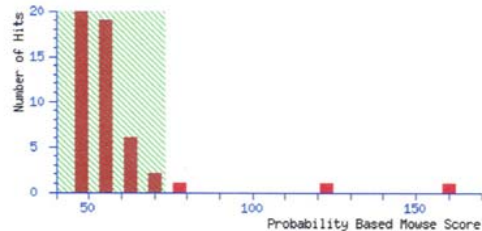
MASCOT Search Query Results

{MATRIX} Mascot Search Results *{SCIENCE}*

User : Landon Wilson
Email : landon.wilson@ccc.uab.edu
Search title : Sample ID (name)
Database : NCBI nr 20020830 (1042297 sequences; 329709346 residues)
Timestamp : 5 Sep 2002 at 19:30:42 GMT
Top Score : 160 for [gi|17136632](#), (NM_057465) porin-P1; Voltage-dependent anion-selective channel [D:

Probability Based Mowse Score

Score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
Protein scores greater than 73 are significant ($p < 0.05$).



Statistical Protein Score for Top Matches

Protein Summary Report

[Switch to Concise Protein Summary Report](#)

To create a bookmark for this report, right click this link: [Protein Summary Report \(Sample ID \(name\)\)](#)

Re-Search All

Search Unmatched

[gi| 17136632](#)

Index

Accession	Mass	Score	Description
gi 17136632	30551	160	(NM_057465) porin-P1; Voltage-dependent anion-selective channel
gi 15246677	30587	123	(X92408) mitochondrial porin [Drosophila melanogaster]
gi 21300816	49448	77	(AAAB01008966) agCP14446 [Anopheles gambiae str. FES1]
gi 14520230	34784	70	(NC_000868) QUINOLINATE SYNTHETASE A [Pyrococcus abyssi]
gi 15642770	36538	67	(NC_002505) O-sialoglycoprotein endopeptidase [Vibrio cholerae]
gi 1364073	65674	66	genome polyprotein - border disease virus (isolate L83/84) (fr
gi 1346637	222678	61	Myosin heavy chain, fast skeletal muscle, embryonic
gi 186355	222834	61	Myosin heavy chain, fast skeletal muscle, adult [validated] - ct
gi 13432175	223006	61	Myosin heavy chain, skeletal muscle, adult
gi 11842051	223040	61	(087231) myosin heavy chain [Gallus gallus]

Top Protein Matches

***E. coli*: FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (EC 5.2.1.8)**

Nominal mass of protein (Mr): 20840

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Matched peptides shown in **Bold brown**

Histidine residues are in **Bold Cerise**

1 **MKVAKDLVVS** **LAYQVRTEDG** **VLVDESPVSA** **PLDYLHGHGS**
41 **LISGLETALE** **GHEVGDKFDV** **AVGANDAYGQ** **YDENLVQRVP**
81 **KDVFMGVDEL** **QVGMRFLAET** **DQGPVPVEIT** **AVEDDHVVVD**
121 **GNHMLAGQNL** **KFNVEVVAIR** **EATEEELAHG** **HVHGAHDHHH**
161 **DHDHDGCCGG** **HGHDHGHEHG** **GEGCCGGKGN** **GGCGCH**

This 21 kDa protein was purified by the Ni-NTA column because of its unusually high histidine content, i.e., it will be present in most purifications of 6xHis-tag recombinant proteins expressed in bacteria.

Mascot search