



**Purdue-UAB Botanicals Center for Age-Related Disease**

# **Bioinformatics for proteomics**

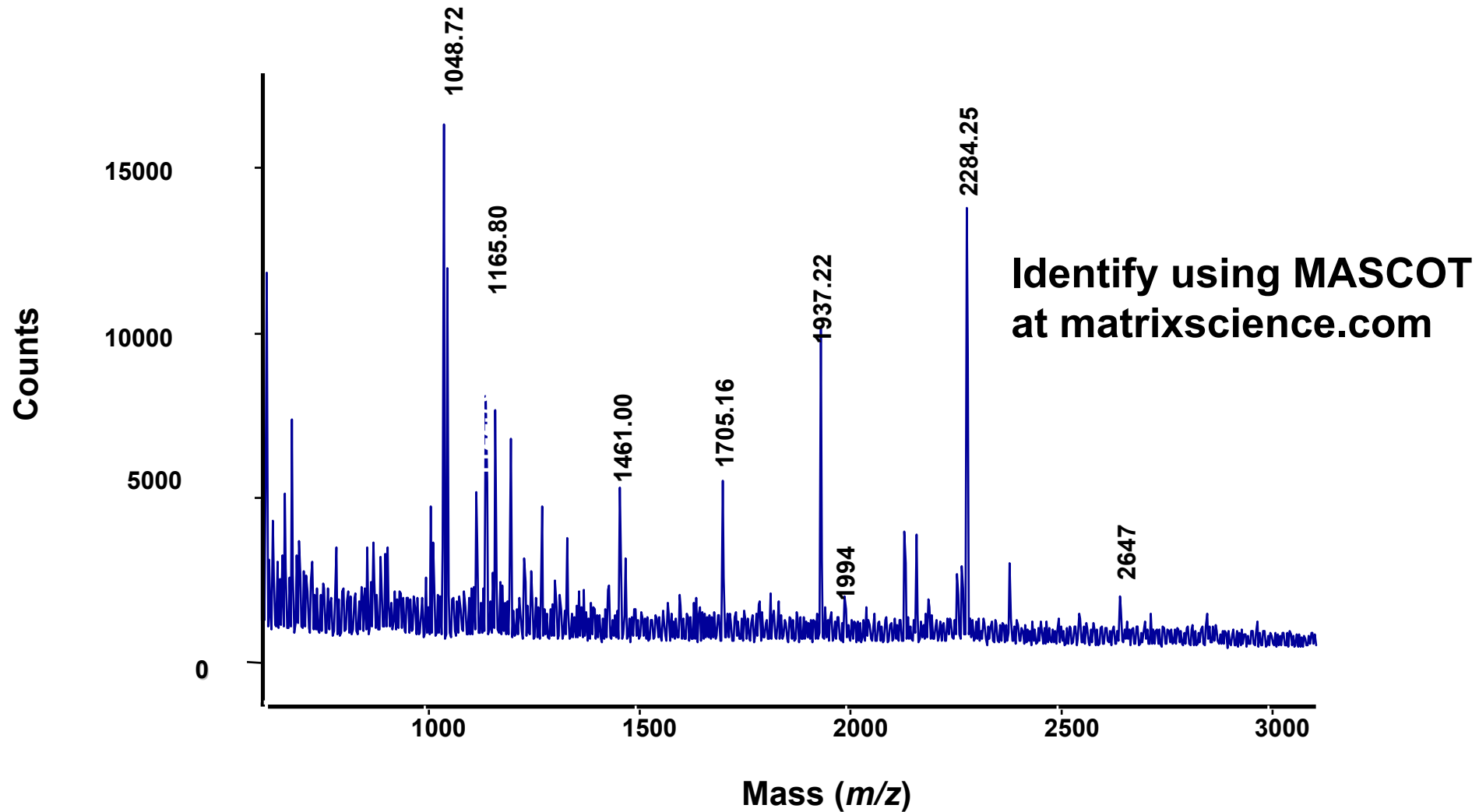
**Stephen Barnes, PhD**

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

# **Objectives for bioinformatics**

- **Principles of peptide mass fingerprinting with Mascot**
- **Linking proteomics data to other databases**
- **Transport through informatics space**

# MALDI-TOF of a tryptic digest



*Courtesy of Mindan Sfakianos*

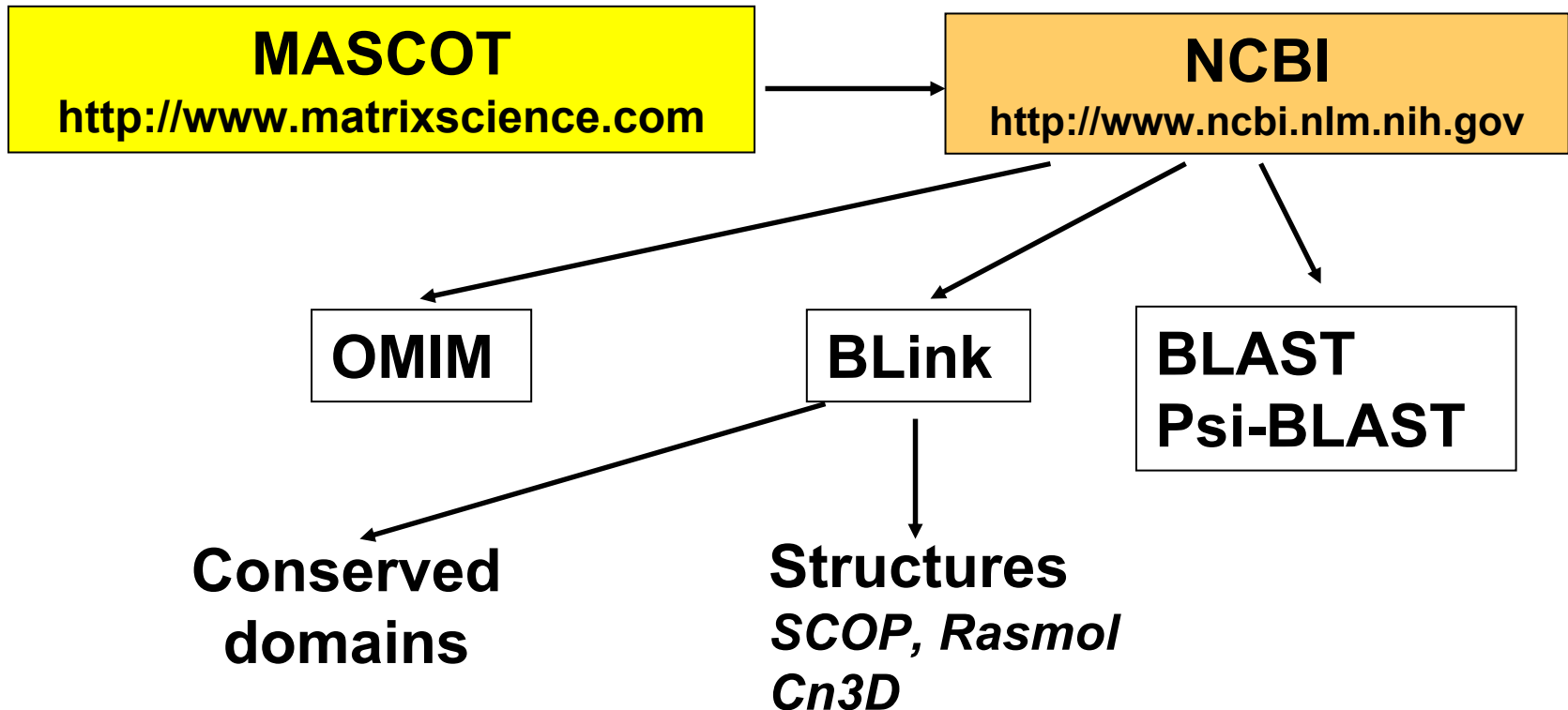


## Steps for identifying a protein from a tryptic digest

- De-isotope MALDI-TOF data and remove trypsin peaks - save as .txt file
- Go to <http://www.matrixscience.com> and click on MASCOT
- Choose first option - enter .txt file in data box and complete name and e-mail address
- Set following: protease = *trypsin*; missed cleavages = 1; M+H; mass accuracy = 100 ppm
- Submit request - answer in 5-10 secs



# Steps from MASCOT to information space





Purdue-UAB Botanicals Center for Age-Related Disease

## Bioinformatics sites

- **NCBI** - <http://www.ncbi.nlm.nih.gov>

**National Center for Biotechnology Information** - contains suites of programs based on proteins, nucleotides, related sequences (BLink, BLAST), conserved domains and structures

- **ExpASy** - <http://www.expasy.org/>

**Expert Protein Analysis System** - contains protein and DNA sequence data, as well as 2D-gels.



# Purdue-UAB Botanicals Center for Age-Related Disease

## BLAST

- This program, available through the NCBI site, aligns the sequence of the protein you've just “discovered” with all other known sequences
- Sequences in *Crocodylus palustris* related to aa 1-60 of human alpha enolase

```
- gi|21325980|gb|AAM47551.1|AF428105_1 (AF428105) tau-crystallin protein
- gi|21325982|gb|AAM47552.1|AF428106_1 (AF428106) alpha-enolase
- gi|21325984|gb|AAM47553.1|AF428107_1 (AF428107) alpha-enolase
- gi|21325986|gb|AAM47554.1|AF428108_1 (AF428108) alpha-enolase
- Length = 434
```

```
- Score = 120 bits (302), Expect = 3e-27
- Identities = 57/60 (95%), Positives = 59/60 (98%)
```

```
- Query: 1 MSILKIHAREIFDSRGNPTVEVDLYTAKGLFRAAVPSGASTGIYEALELRDNDKTRFMGK 60
- MS+LK+HAREIFDSRGNPTVEVDLYT KGLFRAAVPSGASTGIYEALELRDNDKTRFMGK
- Sbjct: 1 MSVLKVHAREIFDSRGNPTVEVDLYTNKGLFRAAVPSGASTGIYEALELRDNDKTRFMGK 60
```



Purdue-UAB Botanicals Center for Age-Related Disease

# Protein structures

- **SCOP** - <http://scop.mrc-lmb.cam.ac.uk/scop/index.html>  
a compilation of protein structures organized into groups by Alexy Murzin at the University of Cambridge
- **Rasmol and Cn3D** - available through NCBI
- Each uses PDB-defined protein crystal structures



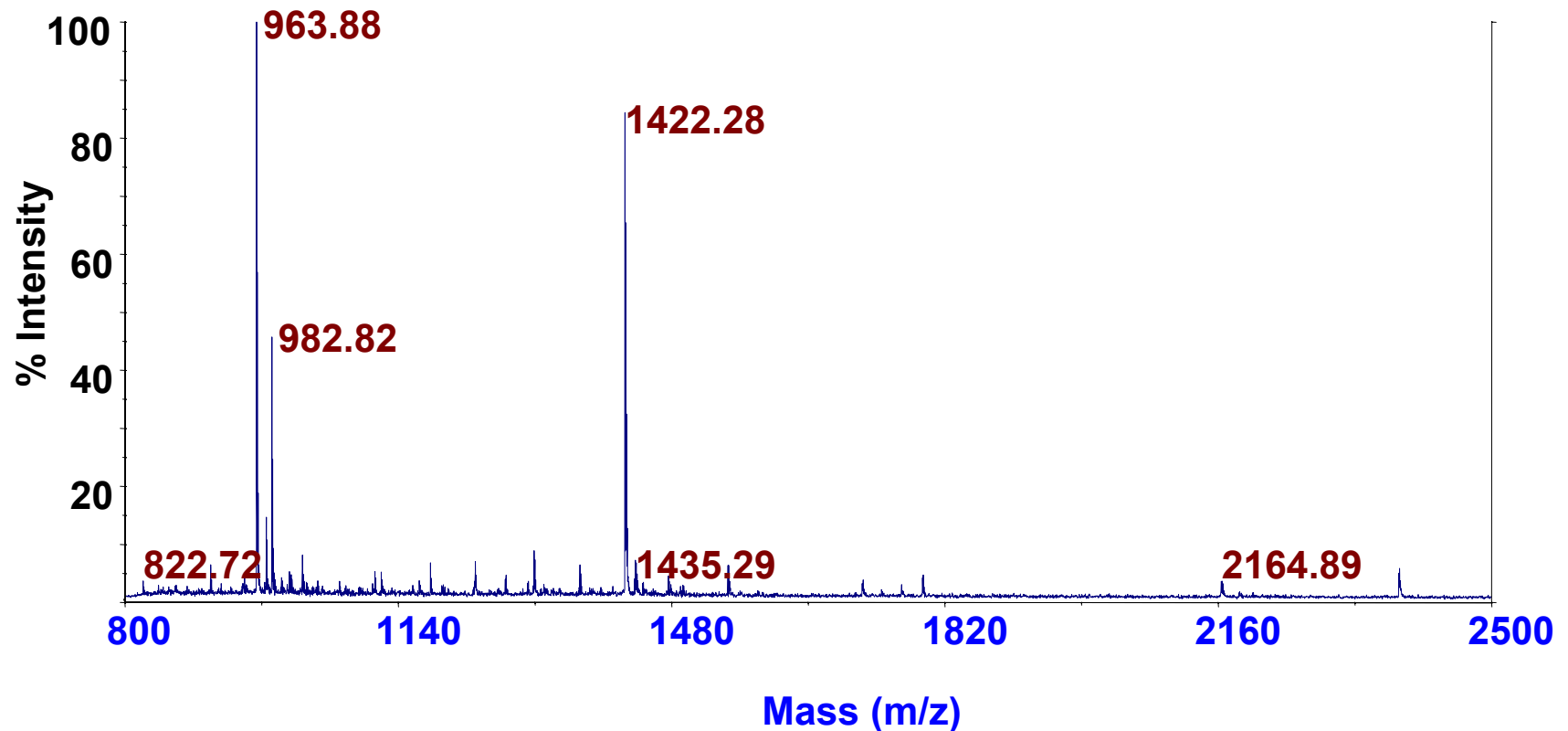
# **OMIM - Online Mendelian inheritance in man**

- **Also available at the NCBI site is a curated source of information about each protein and its corresponding gene**
- **The mutations known to occur in humans are documented and references to published work are provided**

# **A tour through information space**

- **2D-gel analysis of the proteome of DMSO-differentiated HL-60 cancer cells led to many observed changes**
- **The largest change was to a low molecular weight protein which following alkylation with iodoacetamide and trypsinization gave rise to the following ions (m/z): 822.72 963.88 982.82, 1422.28 1435.29**

# MALDI-TOF analysis of trypsinized spot from 2D-IEF/SDS-PAGE analysis of DMSO-differentiated HL-60 cells



# Compiling and organizing data

- **Biomedical science is creating a need to efficiently store data in a rapidly searchable format**
- **Databases such as the one used at UAB for Botanicals Center projects <http://www.uab.edu/proteomics> can be used to compare data from different types of experiments as well as to create links to the rest of the bioinformatics network**