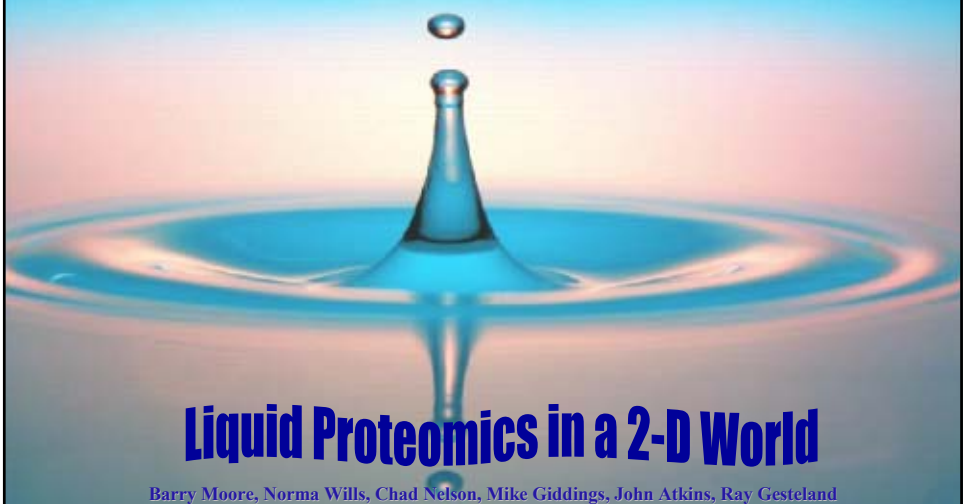
# Cast Thy Proteins Upon The Water



## Liquid Proteomics in a 2-D World

Barry Moore, Norma Wills, Chad Nelson, Mike Giddings, John Atkins, Ray Gesteland

Department of Human Genetics - University of Utah, Salt Lake City, UT

## Proteomics

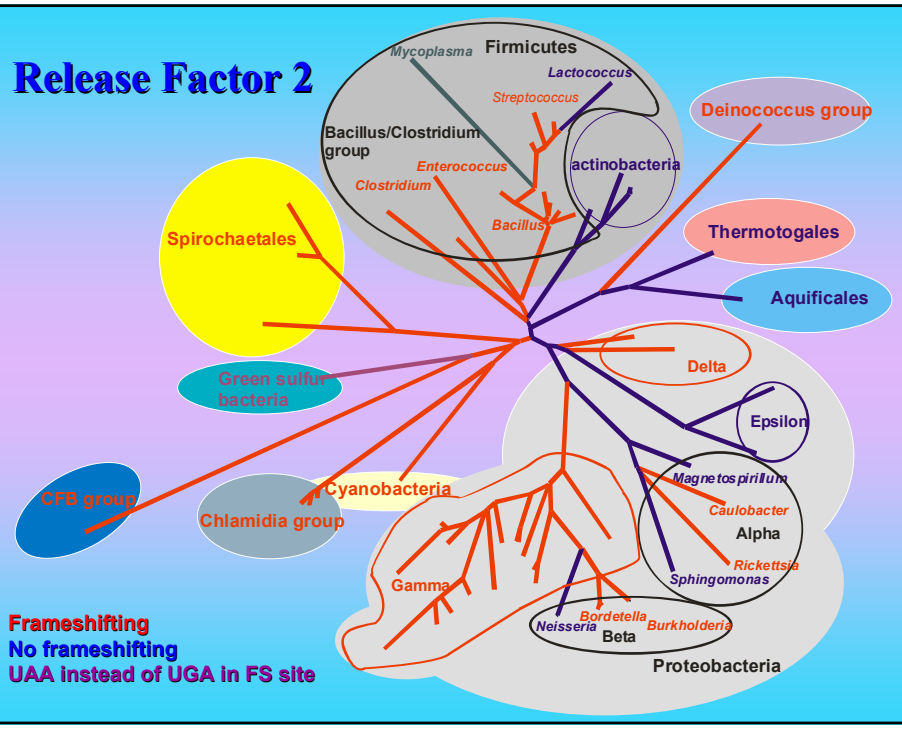
- Dutch Chemist Gerardus Johannes Mulder coined the word “Protein” in 1838 after Proteus the Greek sea god capable of assuming many forms.
- This etymology captures the difference between genomics and proteomics
  - Proteomics has assumed many forms from 2-D mapping to biochemical assays to structural studies.
  - A great deal of diversity can be added to the proteome co-translationally and post-translationally.

# Understanding the Increase in Diversity Between the Genome and the Proteome

- Alternate mRNA Splicing
- Post Translational Modifications and Processing

- RNA Editing
- Transcriptional Slippage
- Recoding
  - Frameshifting
  - Bypassing
  - Redefinition

## Release Factor 2




Frameshifting
No frameshifting
UAA instead of UGA in FS site

Proteomics offers the opportunity to further our understanding of how proteins are made beyond the standard genetic code, and ORF prediction.

# 2-D Gel Proteomics

- *E. coli* Release Factor 2

- 2-D gel proteomics doesn't provide full length masses

## Another Approach to Proteomics

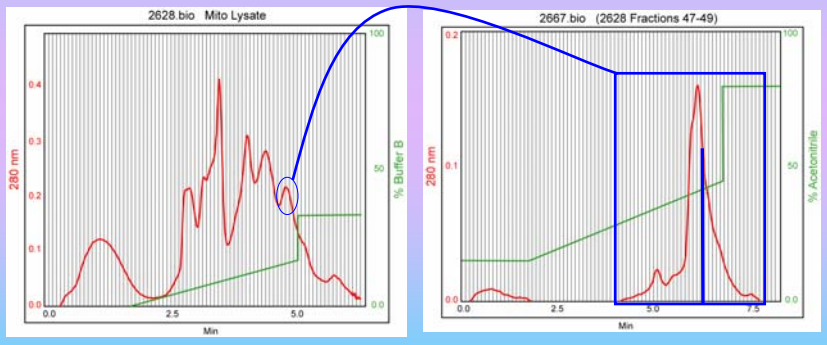
- Combine peptide mapping with full length protein analysis.
  - 2-D Liquid Chromatography
  - Electrospray and MALDI Mass Spectrometry
  - Complete Genome Bioinformatics

## Yeast Mitochonrdia

***Saccharomyces cerevisiae***

- A model without the full complexity of RNA splicing.
- Provides a convenient "bag" of proteins about 1/10$^{th}$ the complexity of an entire genome.

# 2-D Liquid Chromatography




**Anion Exchange or Size Exclusion Chromatography**

**Reversed Phase Chromatography**

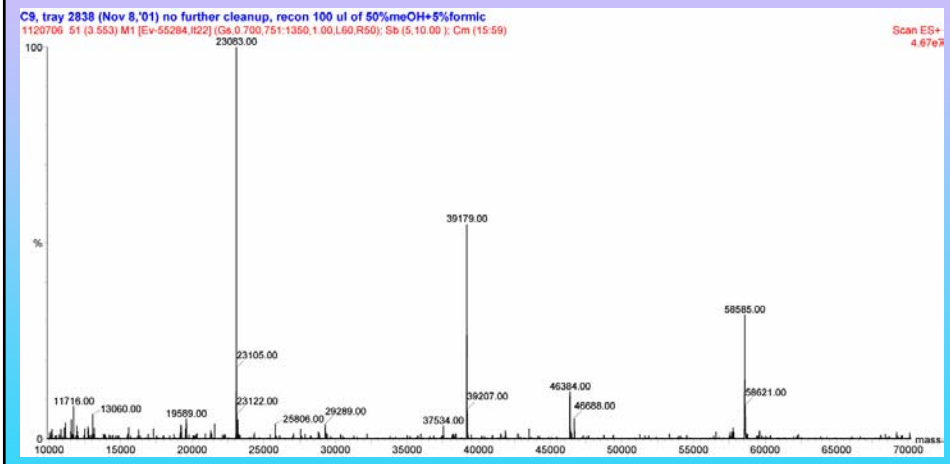
# Electrospray Ionization Mass Spectrometry

- Triple Quadrapole (Micromass Quattro II )
  - Mass Accuracy of 0.01% (100 ppm)

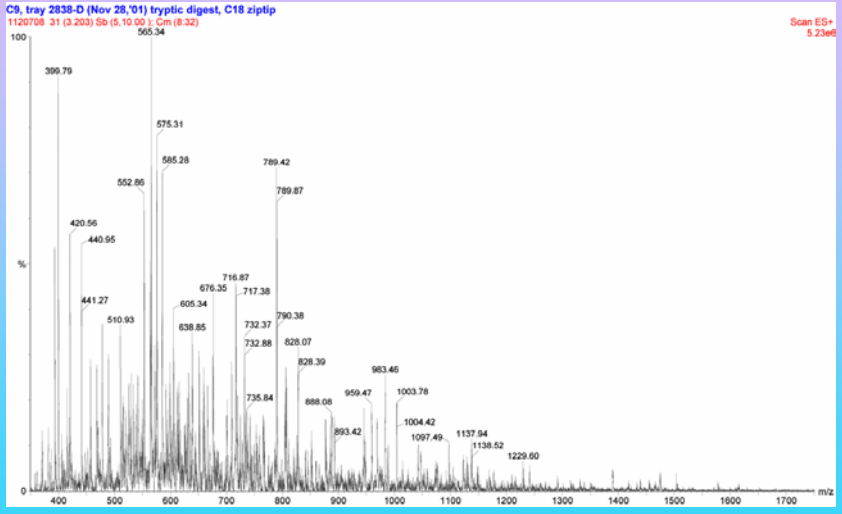
# MALDI Mass Spectrometry

- MALDI-TOF (Perseptive Biosystems DESTR)
  - Mass Accuracy of 0.01% (100 ppm)

# Intact Protein Spectrum




# Trypsin Digest Spectrum



Matrix Science - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://www.matrixscience.com/cgi/index.pl?page=/search_form_select.html What's Related

Instant Message WebMail Calendar Radio People Yellow Pages Download Customize...

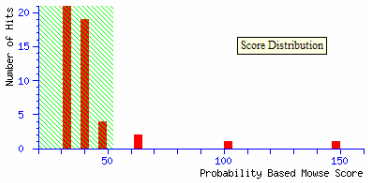
HOME
MASCOT
HELP
WHAT'S NEW
PRODUCTS
SUPPORT
SITE SEARCH
LINKS
EMPLOYMENT
CONTACT US

# Mascot Search Results

```
User            : Barry Moore
Email           : barry.moore@genetics.utah.edu
Search title    : 2838-C9
Database        : MSDB 20010813 (708181 sequences; 219889859 residues)
Taxonomy        : Saccharomyces Cerevisiae (baker's yeast) (9431 sequences)
Timestamp       : 22 Jan 2002 at 19:04:50 GMT
Top Score       : 148 for S28549, acetyl-CoA hydrolase (EC 3.1.2.1) [validated] - yeast (Saccharomyces cerevi
```

## Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 52 are significant (p<0.05).




## Concise Protein Summary Report

Document: Done

start Microsoft PowerPoint ... 5 Microsoft Excel MASSLYNX Matrix Science - Nets... 3 MassLynx NT 11:07 AM

Matrix Science - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://www.matrixscience.com/cgi/index.pl?page=/search_form_select.html What's Related

Instant Message WebMail Calendar Radio People Yellow Pages Download Customize...

HOME
MASCOT
HELP
WHAT'S NEW
PRODUCTS
SUPPORT
SITE SEARCH
LINKS
EMPLOYMENT
CONTACT US

## Protein View

Match to: S28549; Score: 148
acetyl-CoA hydrolase (EC 3.1.2.1) [validated] - yeast (Saccharomyces cerevisiae)

Nominal mass ($M_r$): 58675; Calculated pI value: 6.30
NCBI BLAST search of S28549 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Saccharomyces cerevisiae
Links to retrieve other entries containing this sequence from NCBI Entrez:
CAA48570 from Saccharomyces cerevisiae
CAA84834 from Saccharomyces cerevisiae

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 74
Number of mass values matched: 17
Sequence Coverage: 51%

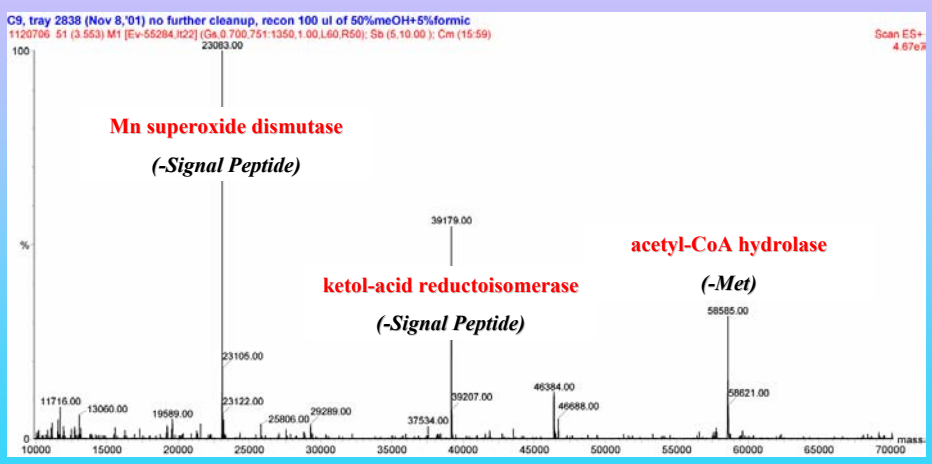
Matched peptides shown in Bold Red

```
  1 MTISNLLKQR VRYAPYLKKV KEAHELIPLF KNGQYLGWSG FTGVGTPKAV
 51 PEALIDHVEK NNLQGKLRFN LFVGASAGPE ENRWAEHDMI IKRAPHQVGK
101 PIAKAINQGR IEFFDKHLSM FPQDLTYGFY TRERKDNKIL DYTIIEATAI
151 KEDGSIVPGP SVGGSPEFIT VSDKVIIEVN TATPSFEGIH DIDMPVNPPF
201 RKPYPYLKVD DKCGVDSIPV DPEKVVAIVE STMRDQVPPN TPSDDMSRAI
251 AGHLVEFFRN EVKHGRLPEN LLPLQSGIGN IANAVIEGLA GAQFKHLTVW
301 TEVLQDSFLD LFENGSLDYA TATSVRLTEK GFDRAFANWE NFKHRLCLRS
351 QVVSNNPEMI RRLGVIAMNT PVEVDIYAHA NSTNVNGSRM LNGLGGSADF
401 LRNAKLSIMH APSARPTKVD PTGISTIVPM ASHVDQTEHD LDILVTDQGL
451 ADLRGLSPKE RAREIINKCA HPDYQALLTD YLDRAEHYAK KHNCLHEPHM
501 LKNAFKFHTN LAEKGTMKVD SWEPVD
```

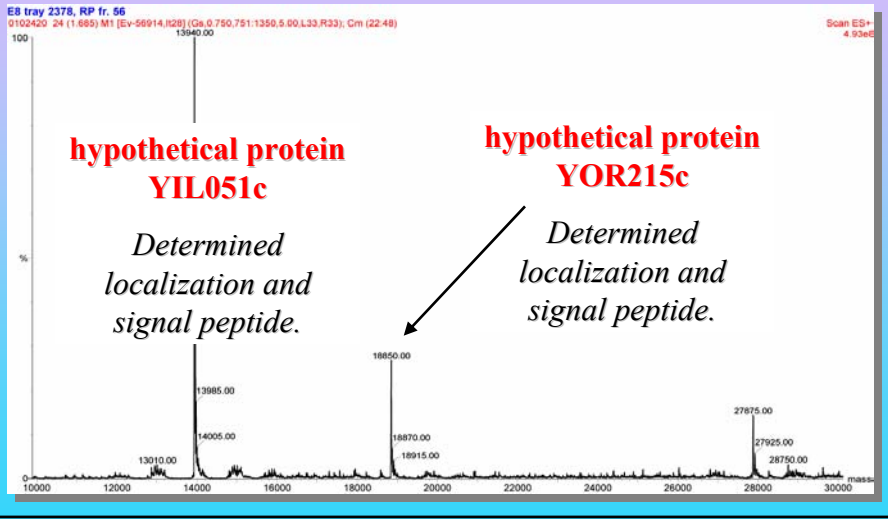
Show predicted peptides also

start Microsoft PowerPoint ... 5 Microsoft Excel MASSLYNX Matrix Science - Nets... 3 MassLynx NT 11:08 AM

# Proteins found by peptide mapping match full length masses with modifications




# Other cases were unknown in the database



# What have we found

- 183 unique samples analyzed
- 46 proteins are significant matches ($p<0.05$)
- 12 of those have full length matches
- 5 of those reveal information not predicted by in the databases

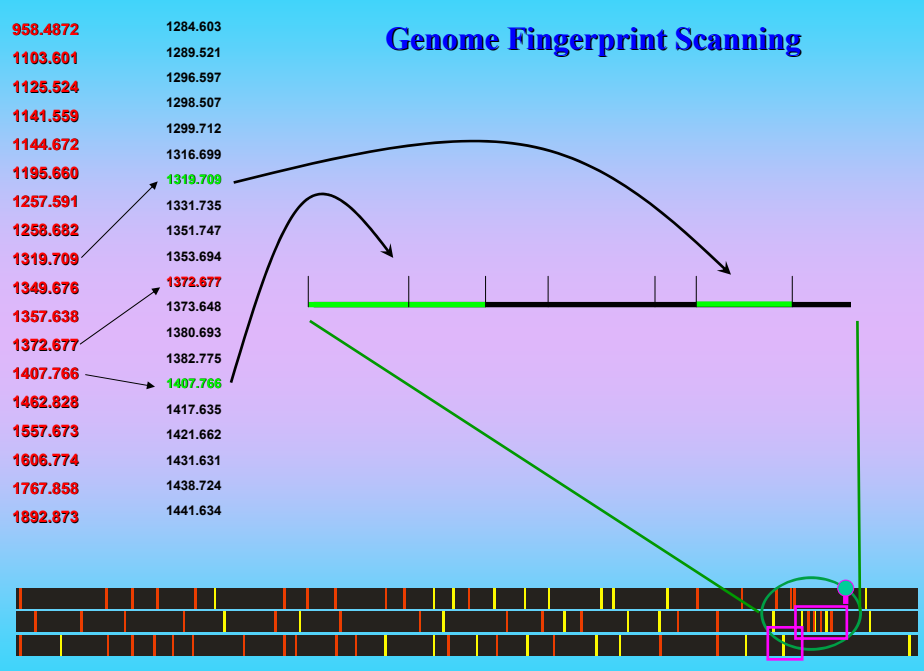
| Protein | Mass | | | | | |
|---|---|---|---|---|---|---|
| aldehyde dehydrogenase (NAD+) | 53,967 | M | I | | S | |
| malate dehydrogenase | 33,834 | M | I | | S | |
| ATP synthase beta chain precursor | 51,109 | M | | | | |
| citrate (si)-synthase | 49,216 | M | I | | S | |
| acetyl-CoA hydrolase | 58,583 | | I | | | D |
| aconitate hydratase | 82,624 | M | I | | | |
| phosphoglycerate kinase | 44,648 | M | I | | | D |
| hypothetical protein YOR215c | 18,849 | M | I | | S | D |
| NADH-cytochrome b5 reductase | 34,140 | M | | | | |
| dihydrolipoamide dehydrogenase | 51,547 | M | | | | |
| cytochrome C peroxidase | 33,418 | M | | | | |
| heat shock protein MGE1 precusor | 21,009 | M | I | | S | D |
| ketol-acid reductoisomerase | 39,174 | M | I | | S | |
| thiol peroxidase | 29,500 | M | | | | |
| dnaK-type molecular chaperone | 68,081 | M | | | | |
| heat shock protein | 85,111 | M | | | | |
| Cu-Zn superoxide dismutase | 15,719 | | I | | | |
| serine hydroxymethyltransferase | 60,584 | M | | | | |
| peripheral membrane protein | 95,867 | M | | | | |
| ATP synthase alpha chain precursor | 55,481 | M | | | | |
| protein essential for cell division | 37,041 | | | | | |
| manganese superoxide dismutase | 23,083 | M | I | | S | |
| hypothetical protein YHR148w | 21,765 | | | | | |
| hypothetical protein YIL051c | 13,939 | M | I | | S | D |

# Bioinformatics

- PTM were known in the database, or found by manual searching
- Signal peptides were known in the database or identified manually
- No recoding events were found

**We need better bioinformatics**

## Genome Fingerprint Scanning



# Challenges…

- Separation
  - Maximizing separation
  - Avoiding proteolytic cleavages
- Mass spectrometer sensitivity
  - Identifying non-abundant proteins
- Bioinformatics
  - Computational complexity builds rapidly

# …and Opportunities

- Affinity Tagging
  - Going after the proteins one at a time as tagged proteomes become available
  - The tagging has to be done carefully with proteome diversity in mind, or you will miss things.


# Acknowledgements

Ray Gesteland &
John Atkins

- Laboratory
  - Norma Wills
  - Jason Simmons
  - AJ Bauchum
  - Rukshana Chowdry

- Mass Spectromety
  - Chad Nelson

- Bioinformatics
  - Mike Giddings
  - Mark Holmes
  - Melissa Kimball
  - Andy Hammer

Department of Energy