Proteomics and Mass Spectrometry: Towards Comprehensive Proteomics of Cells and Tissues

Department of Cell Biology
The Scripps Research Institute
LaJolla, CA
http://fields.scripps.edu

Proteomics

The scientific discipline of characterizing and analyzing the proteins, protein interactions and protein modifications of an organism.

***Structural Proteomics*: refers to the development and application of experimental approaches to define the primary, secondary, and tertiary structures of proteins.**

***Functional Proteomics*: refers to the development and application of global (proteome-wide or system-wide) experimental approaches to assess protein function by making use of the information provided by structural genomics.**

Hieter and Boguski, *Science* 278, 601-602 (1997)

Desirable Analytical Capability in Functional Proteomic Studies

Comprehensive and Parallel Analysis of Experiments

Total protein analyses: Protein arrays are under development; Measure protein-protein interactions, protein ligand interactions, affinity arrays have been developed for focused use.

System-Wide Analyses: possible to comprehensively characterize and identify all proteins in a specific experiment, e.g. co-precipitation.

Functional Analysis

Who, What, Where, When

- **Who:** With Whom does a protein associate? Protein/Protein interactions: guilt by association
- **What:** What does it do: activity
- **Where:** Where does it do it: localization: nuclear, cytoplasmic, secreted, membrane
- **When:** When does it do it: response to environmental conditions, cell cycle dependent?

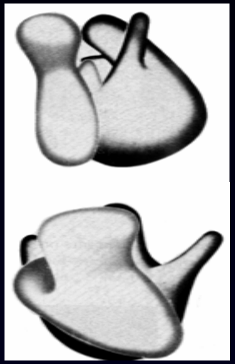
Discover the molecular pathways controlling physiological functions

Complete Genomic Information Does Not Include Post Translational Information

- What structural changes occur to create an active protein?
- How is a protein's activity regulated?
- Are modifications involved in regulation?

Comprehensive Analysis of Complex Protein Mixtures

Cells/Tissues



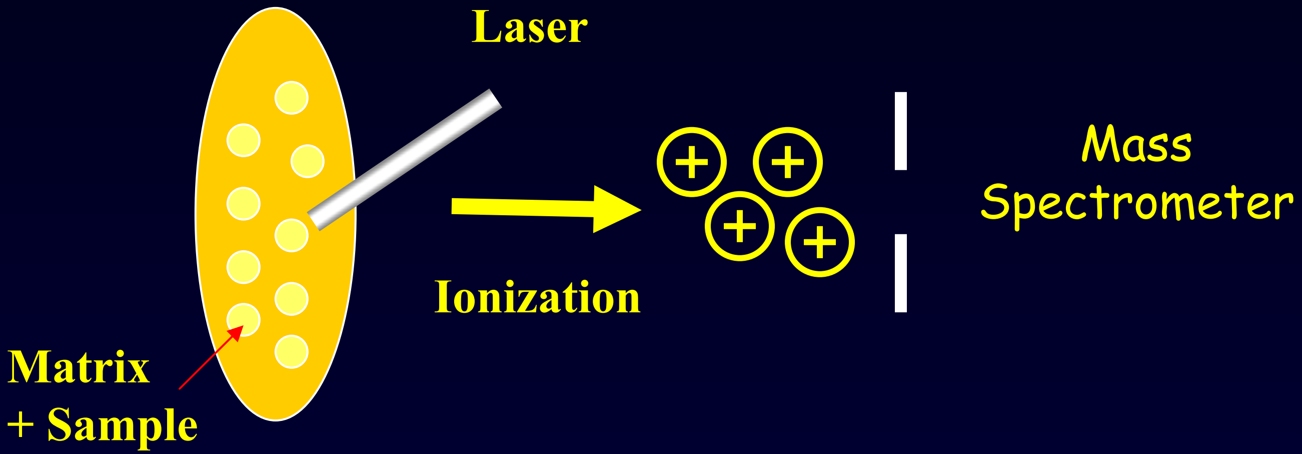
Multiprotein Complex/Organelle

Total Protein Characterization

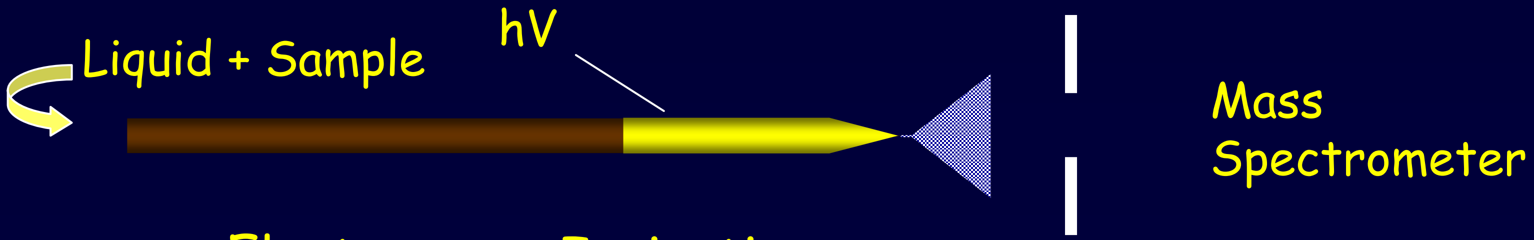
- Protein Identification: *What's there*
- Post Translational Modifications: *Regulation*
- Quantification: *Dynamics*
- Validation of targets: *Genetics, RNAi, siRNA*

Translation of technology development into biological discovery

Ionization Techniques
“The horse that pulls the cart”



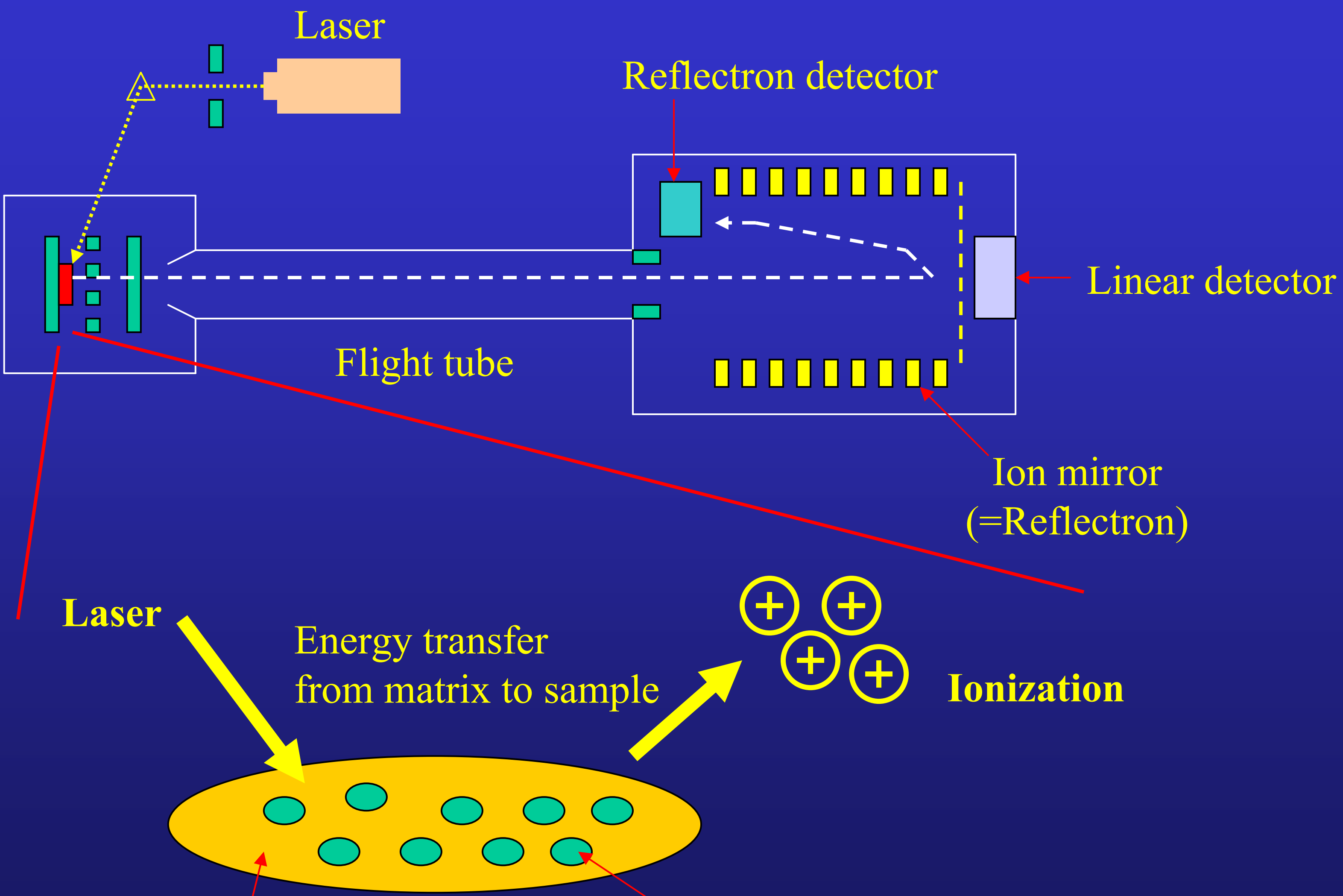
Matrix-Assisted Laser Desorption



Electrospray Ionization

2003 Nobel Prize in Chemistry: Fenn and Tanaka

MALDI-TOF Mass Spectrometry



Laser

Reflectron detector

Linear detector

Flight tube

Ion mirror
(=Reflectron)

Laser

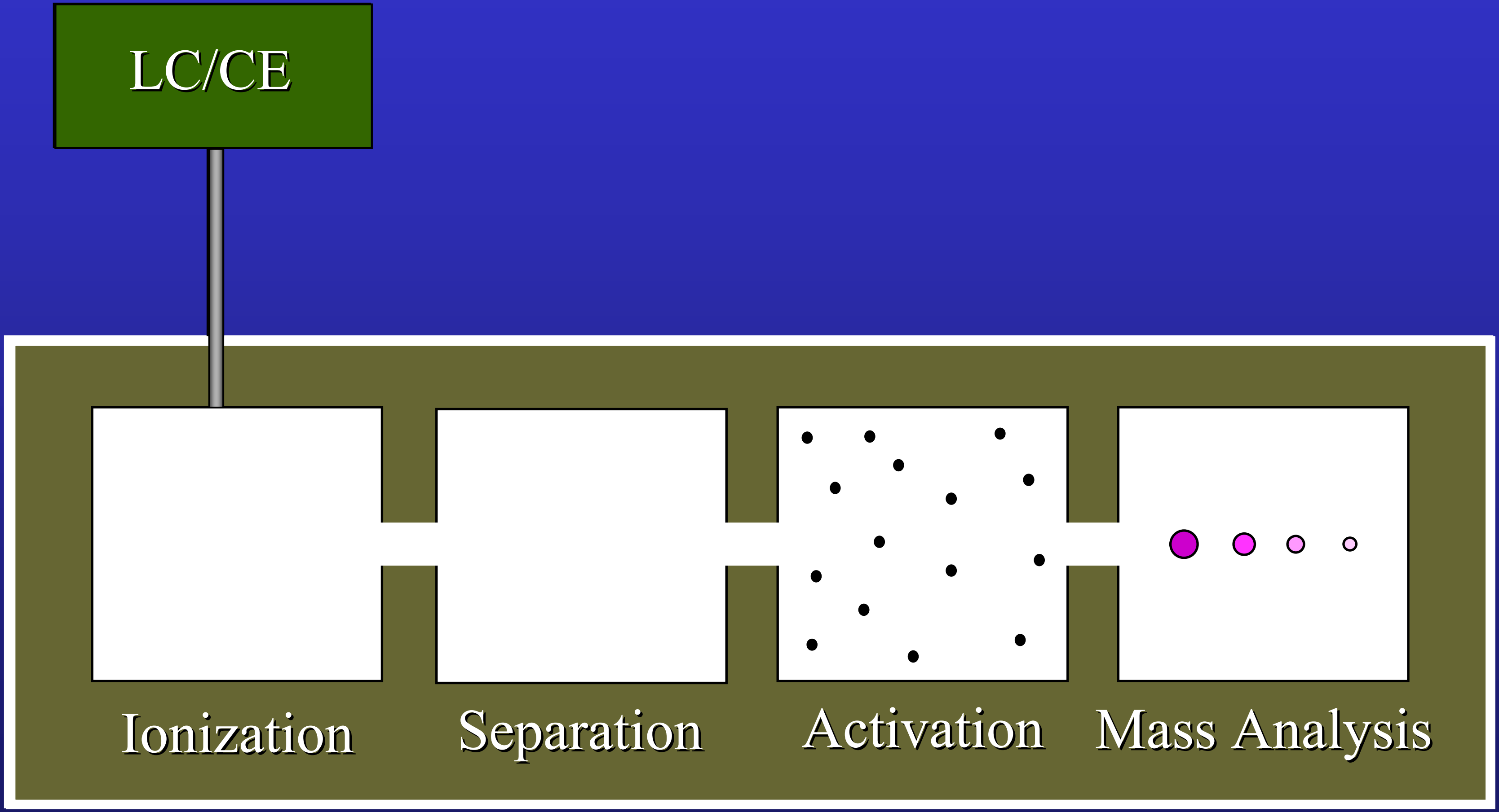
Energy transfer
from matrix to sample

Ionization

Matrix

Sample

TANDEM MASS SPECTROMETRY



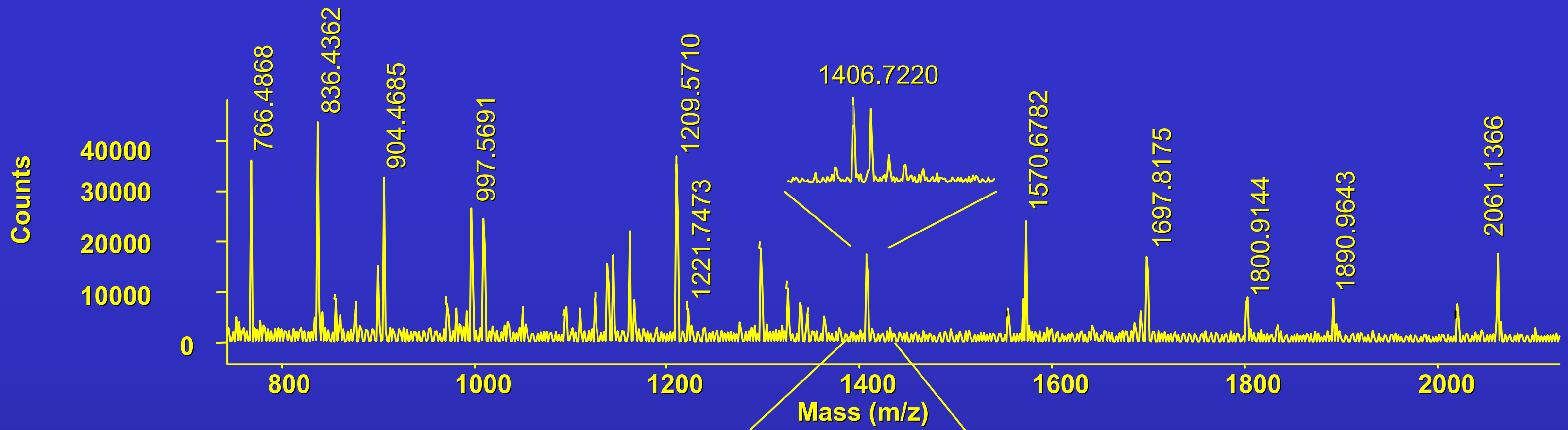
LC/CE

Ionization

Separation

Activation

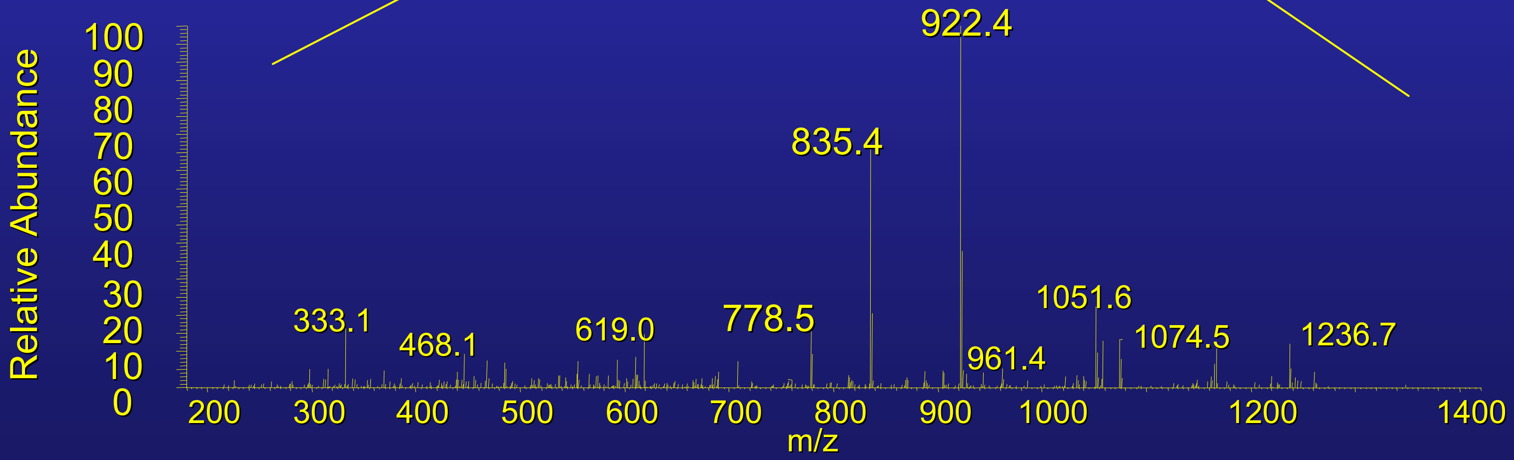
Mass Analysis



Molecular Weight

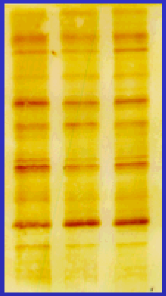
…PPGTGKTLLAK AVANESGANFISVK FYVINGPEIM...

Fragmentation



General Strategy for Proteome Characterization

Purification



1-DE

2-DE

Solution

↓

Peptides

↓

Mass Spectrometry <
- MALDI-TOF MS
- μ-(LC)-ESI-MS/MS

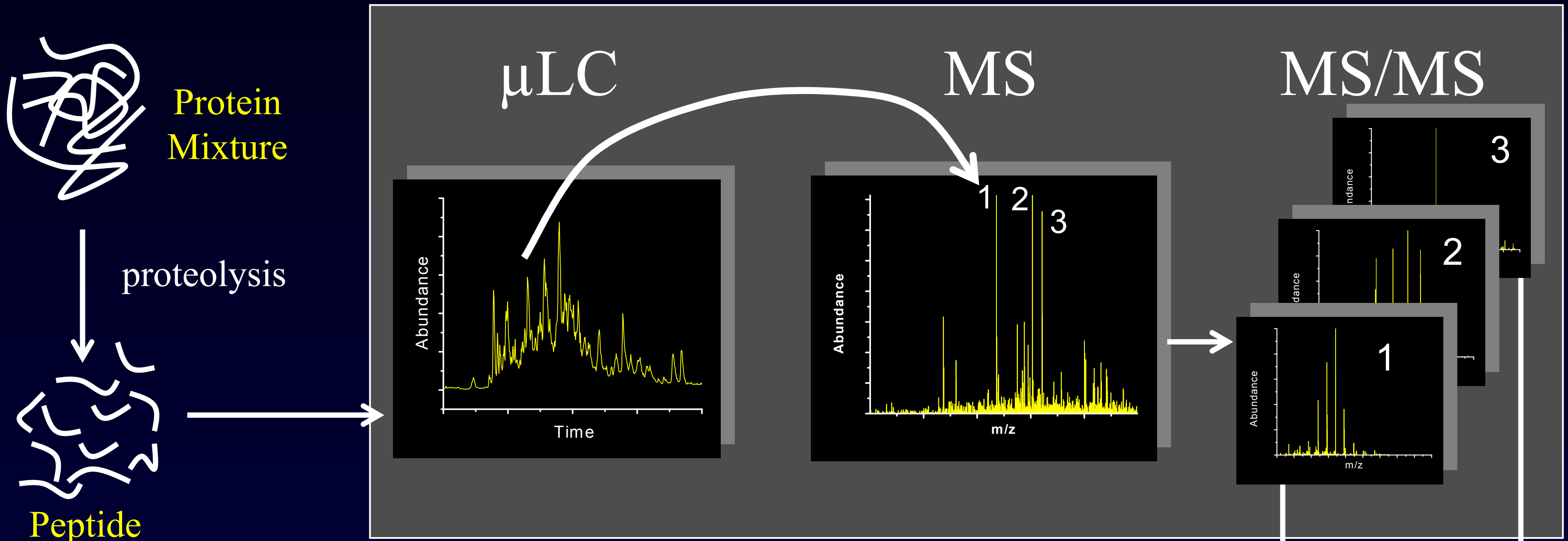
↓

Database Search

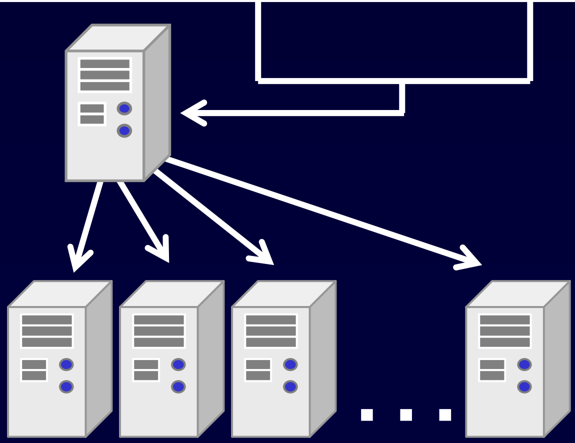
Characterization

- Identification
- PTM
- Quantification

“Shotgun Proteomics”



Peptide Mixture

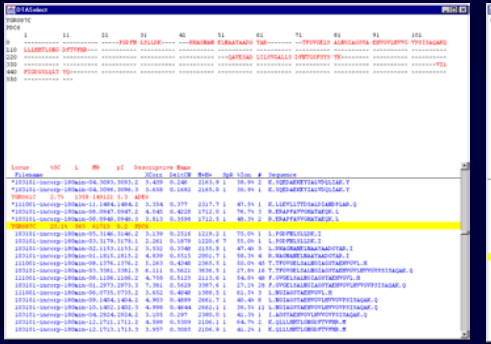


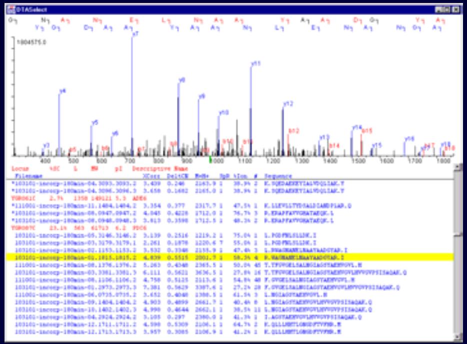
52 node Beowulf Computer Cluster

SEQUEST

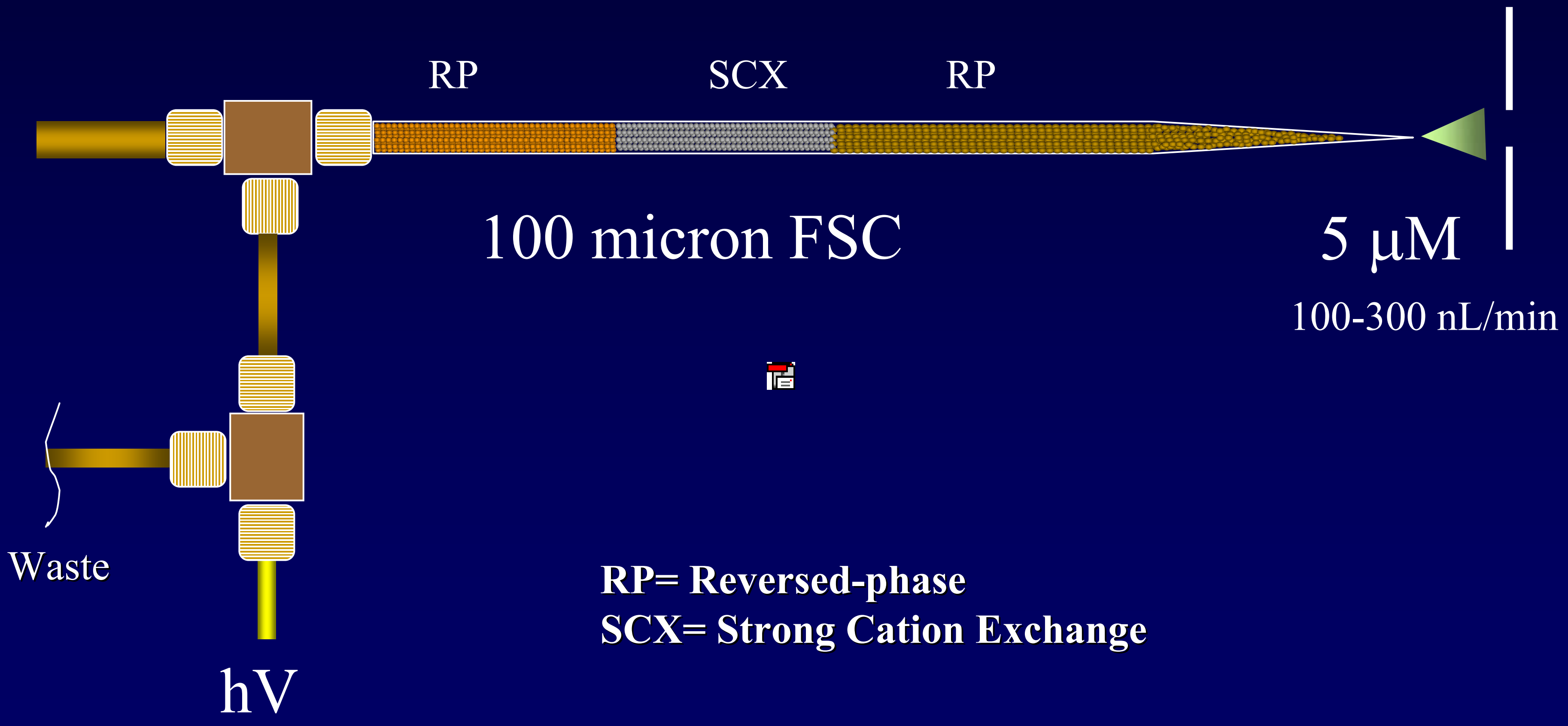
Output Filtering and Re-Assembly

DTASelect





Integrated Multi-Dimensional Liquid Chromatography



RP

SCX

RP

100 micron FSC

5 µM

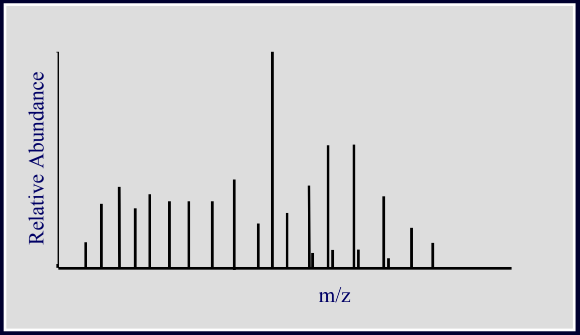
100-300 nL/min

Waste

hV

RP= Reversed-phase
SCX= Strong Cation Exchange

Protein Identification Data Analysis for Tandem Mass Spectra:



LIBQUEST

MS/MS Comparison

→ Library Searching
Comparative Analysis
Subtractive Analysis

Yates et al. *Anal.Chem.* 70, 3557 (1998)

SEQUEST & Pep_Prob

Yates et al., *JASMS* 5, 976 (1994)
Yates et al. *Anal. Chem* 67, 1426 (1995)
Yates et al. *Anal. Chem.* 67, 3205 (1995)
MacCoss et al. *Anal. Chem* (2002)
Sadygov and Yates, *Anal. Chem* (in press)

Database Search

→ Existing Sequence
PTM

SEQUEST-SNP

Variant Search

→ Related Sequence
SNP Analysis
Mutation Analysis

Gatlin et al. *Anal.Chem.* 72, 757 (2000).

GutenTag

De Novo Sequencing

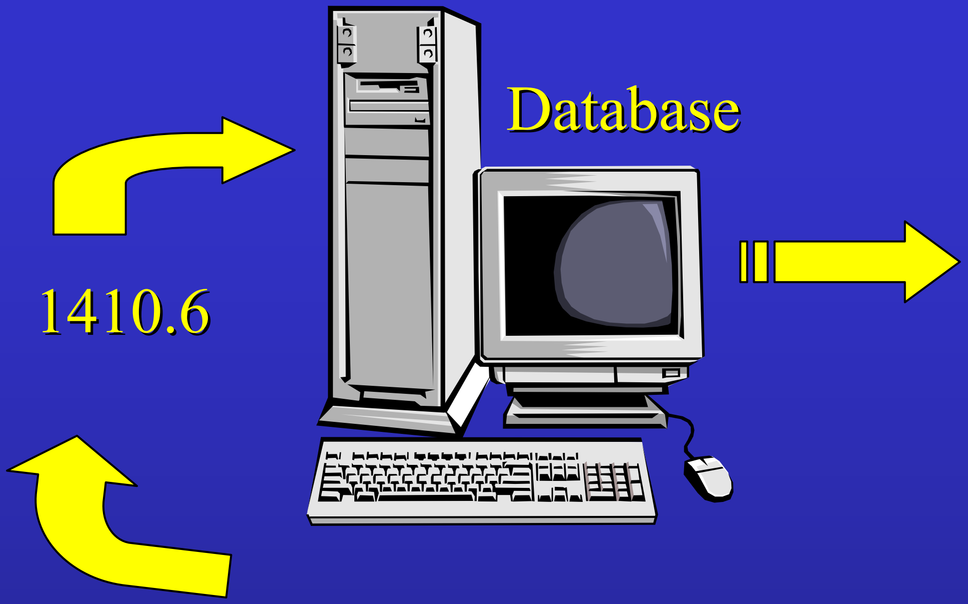
→ Alternate Splicing
Unantic. Mod.
Unknown ORFs

Tabb et al (submitted)

DTASelect

Post Analysis Review Software

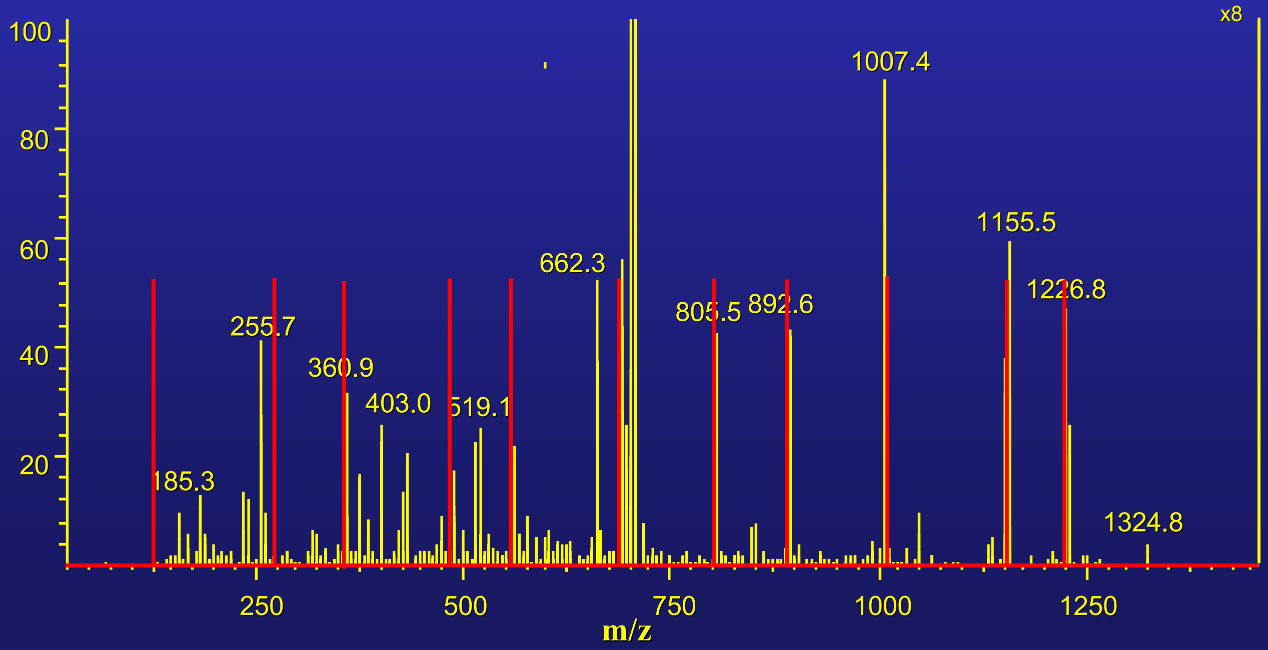
Link et al. *Nature Biotech.* **17**, 676-682 (1999)
Tabb et. al. *J. Proteome Res.* 1, 26, (2002)



1410.6

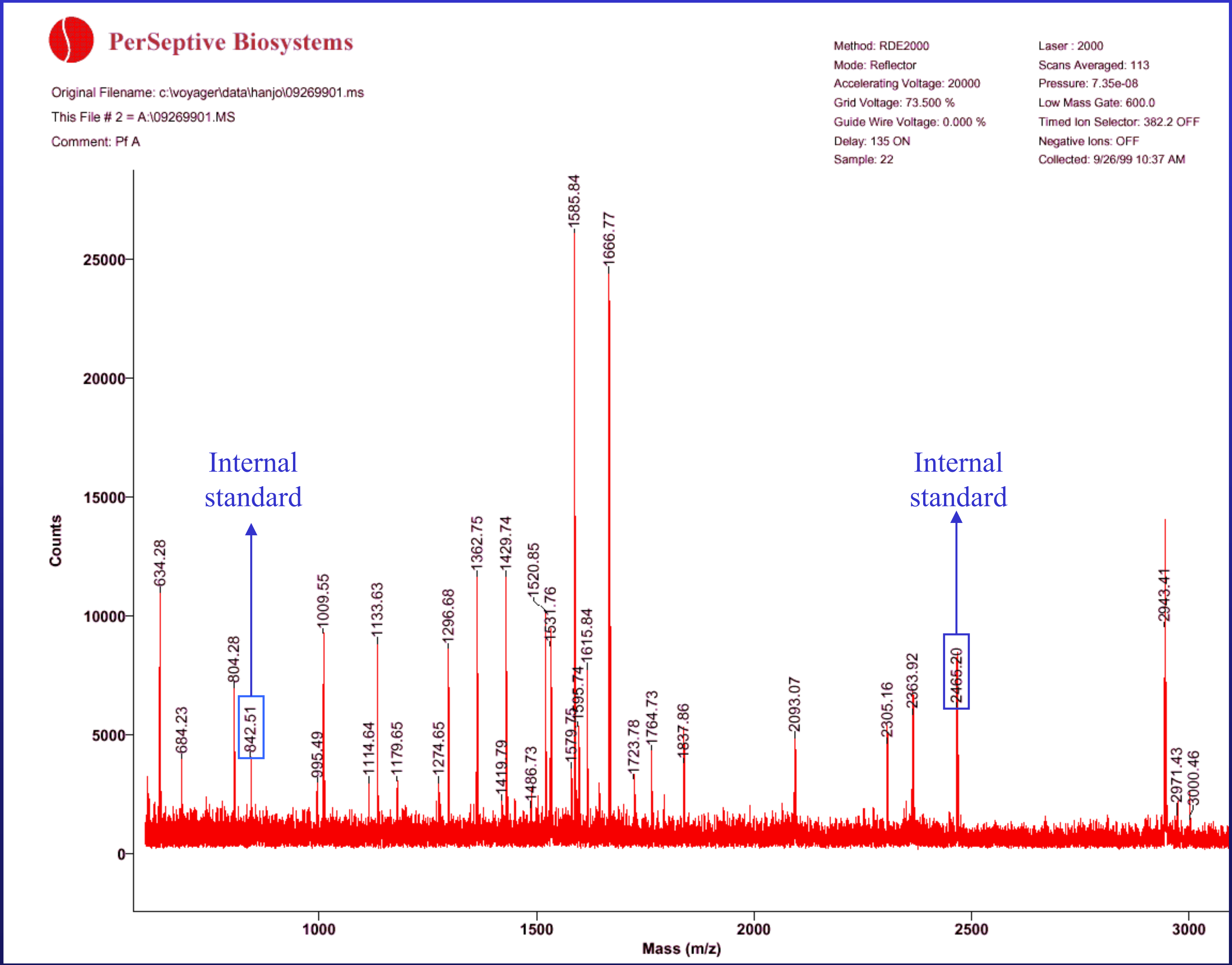
Database

GGILAQSPFLIIK
IIGHFYDDWCPLK
AFDSLPDDIHEK
SPAFDSIMAETLK



real spectrum Cross-Correlated with model spectrum

MALDI-TOF Mass Spectrum



PerSeptive Biosystems

Original Filename: c:\voyager\data\hanjo\09269901.ms
This File # 2 = A:\09269901.MS
Comment: Pf A

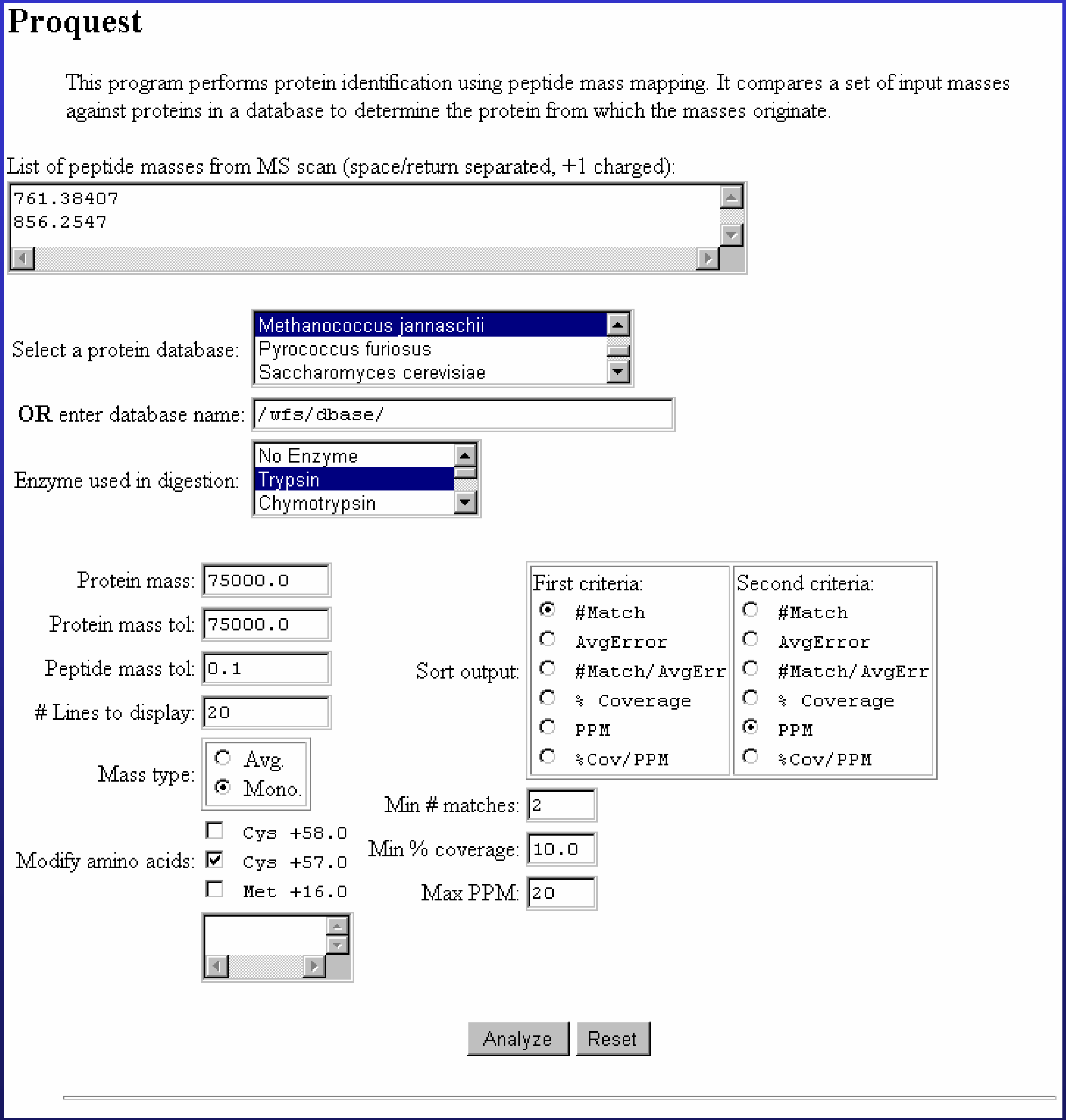
Method: RDE2000
Mode: Reflector
Accelerating Voltage: 20000
Grid Voltage: 73.500 %
Guide Wire Voltage: 0.000 %
Delay: 135 ON
Sample: 22

Laser : 2000
Scans Averaged: 113
Pressure: 7.35e-08
Low Mass Gate: 600.0
Timed Ion Selector: 382.2 OFF
Negative Ions: OFF
Collected: 9/26/99 10:37 AM

Internal standard

Internal standard

PROQUEST



Proquest

This program performs protein identification using peptide mass mapping. It compares a set of input masses against proteins in a database to determine the protein from which the masses originate.

List of peptide masses from MS scan (space/return separated, +1 charged):

```
761.38407
856.2547
```

Select a protein database: Methanococcus jannaschii (selected) / Pyrococcus furiosus / Saccharomyces cerevisiae

OR enter database name: /wfs/dbase/

Enzyme used in digestion: No Enzyme / Trypsin (selected) / Chymotrypsin

Protein mass: 75000.0

Protein mass tol: 75000.0

Peptide mass tol: 0.1

Lines to display: 20

Mass type: ○ Avg. ● Mono.

Modify amino acids:
- ☐ Cys +58.0
- ☑ Cys +57.0
- ☐ Met +16.0

Sort output:

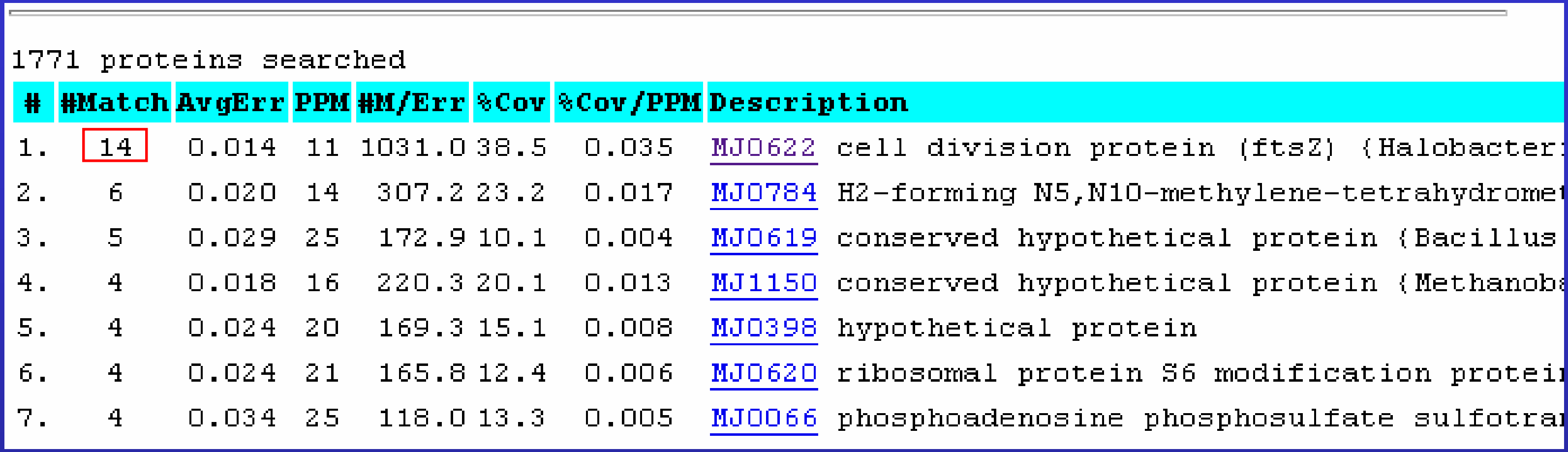
First criteria:	Second criteria:
● #Match	○ #Match
○ AvgError	○ AvgError
○ #Match/AvgErr	○ #Match/AvgErr
○ % Coverage	○ % Coverage
○ PPM	● PPM
○ %Cov/PPM	○ %Cov/PPM

Min # matches: 2

Min % coverage: 10.0

Max PPM: 20

Proteins are identified by number of peptide matches.

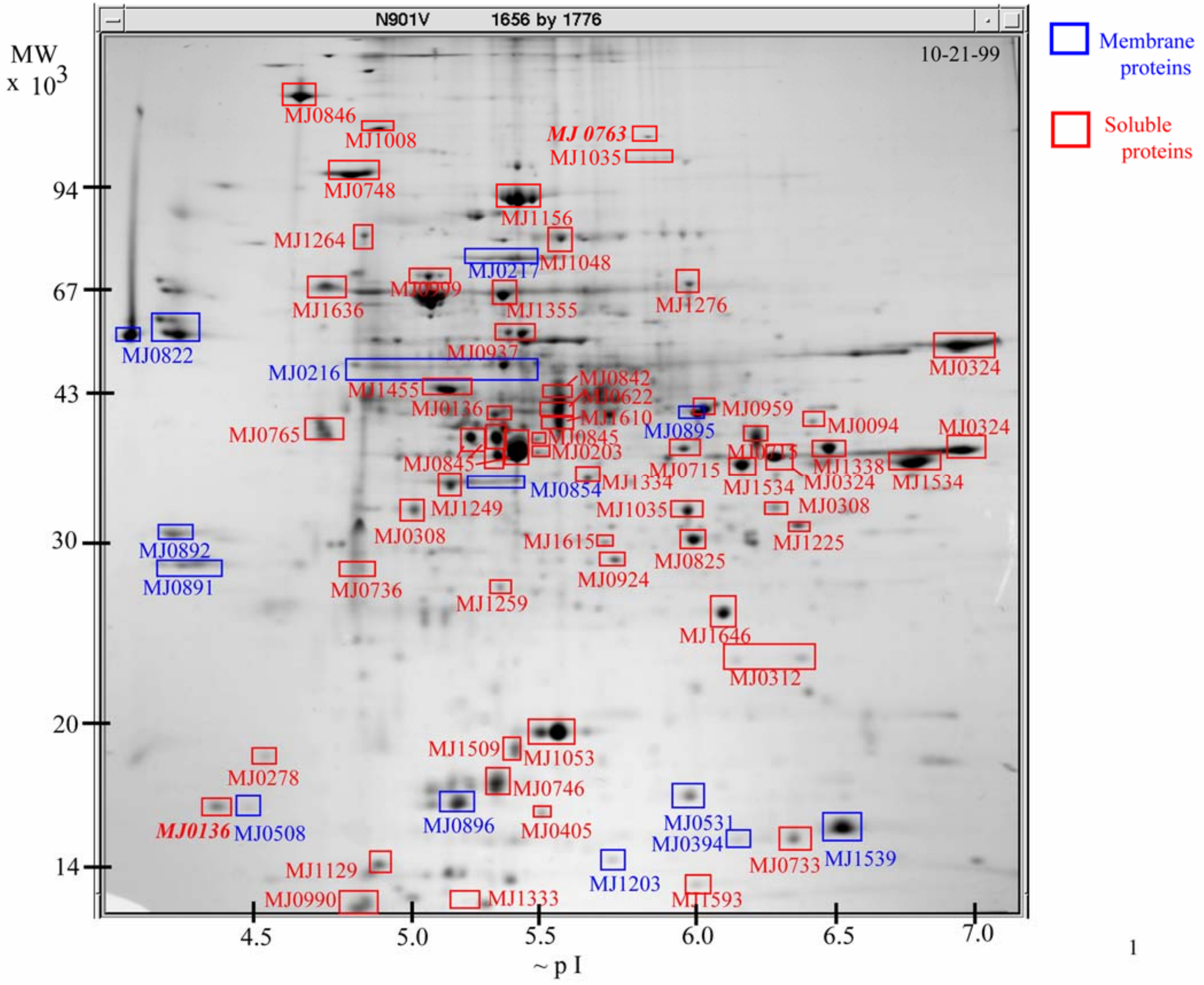


1771 proteins searched

#	#Match	AvgErr	PPM	#M/Err	%Cov	%Cov/PPM	Description
1.	14	0.014	11	1031.0	38.5	0.035	MJ0622 cell division protein (ftsZ) {Halobacter
2.	6	0.020	14	307.2	23.2	0.017	MJ0784 H2-forming N5,N10-methylene-tetrahydrome
3.	5	0.029	25	172.9	10.1	0.004	MJ0619 conserved hypothetical protein {Bacillus
4.	4	0.018	16	220.3	20.1	0.013	MJ1150 conserved hypothetical protein {Methanob
5.	4	0.024	20	169.3	15.1	0.008	MJ0398 hypothetical protein
6.	4	0.024	21	165.8	12.4	0.006	MJ0620 ribosomal protein S6 modification protei
7.	4	0.034	25	118.0	13.3	0.005	MJ0066 phosphoadenosine phosphosulfate sulfotra

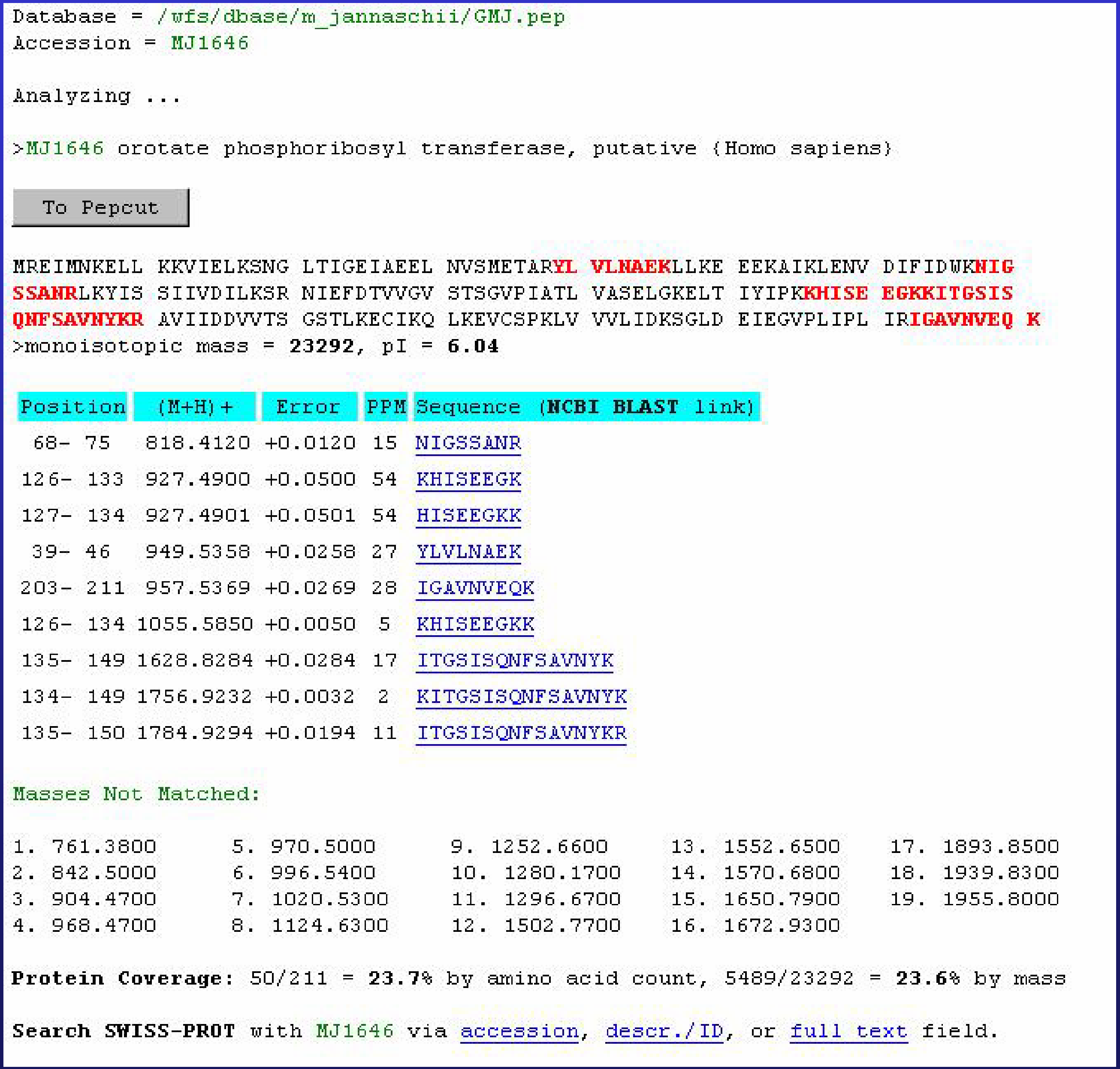
Identified *M. jannaschii* ORFs

Methanococcus jannaschii (Composite Gel Picture)



* The numbers indicate the open reading frames (ORF) containing the greatest number of matched peptides in the spots from MS/MS.

PROQUEST Sequence Output Screen



Database = /wfs/dbase/m_jannaschii/GMJ.pep
Accession = MJ1646

Analyzing ...

>MJ1646 orotate phosphoribosyl transferase, putative (Homo sapiens)

MREIMNKELL KKVIELKSNG LTIGEIAEEL NVSMETAR**YL VLNAEK**LLKE EEKAIKLENV DIFIDWK**NIG**
SSANRLKYIS SIIVDILKSR NIEFDTVVGV STSGVPIATL VASELGKELT IYIPK**KHISE EGKKITGSIS**
QNFSAVNYKR AVIIDDVVTS GSTLKECIKQ LKEVCSPKLV VVLIDKSGLD EIEGVPLIPL IR**IGAVNVEQ K**
>monoisotopic mass = **23292**, pI = **6.04**

Position	(M+H)+	Error	PPM	Sequence (**NCBI BLAST** link)
68- 75	818.4120	+0.0120	15	NIGSSANR
126- 133	927.4900	+0.0500	54	KHISEEGK
127- 134	927.4901	+0.0501	54	HISEEGKK
39- 46	949.5358	+0.0258	27	YLVLNAEK
203- 211	957.5369	+0.0269	28	IGAVNVEQK
126- 134	1055.5850	+0.0050	5	KHISEEGKK
135- 149	1628.8284	+0.0284	17	ITGSISQNFSAVNYK
134- 149	1756.9232	+0.0032	2	KITGSISQNFSAVNYK
135- 150	1784.9294	+0.0194	11	ITGSISQNFSAVNYKR

Masses Not Matched:

1. 761.3800
2. 842.5000
3. 904.4700
4. 968.4700
5. 970.5000
6. 996.5400
7. 1020.5300
8. 1124.6300
9. 1252.6600
10. 1280.1700
11. 1296.6700
12. 1502.7700
13. 1552.6500
14. 1570.6800
15. 1650.7900
16. 1672.9300
17. 1893.8500
18. 1939.8300
19. 1955.8000

Protein Coverage: 50/211 = **23.7%** by amino acid count, 5489/23292 = **23.6%** by mass

Search SWISS-PROT with MJ1646 via accession, descr./ID, or full text field.

SAGA (Spt-Ada-Gcn5-Acetyltransferase) Complex

A subset of TAF$_{II}$s in the SAGA complex are required for nucleosome acetylation and transcriptional stimulation

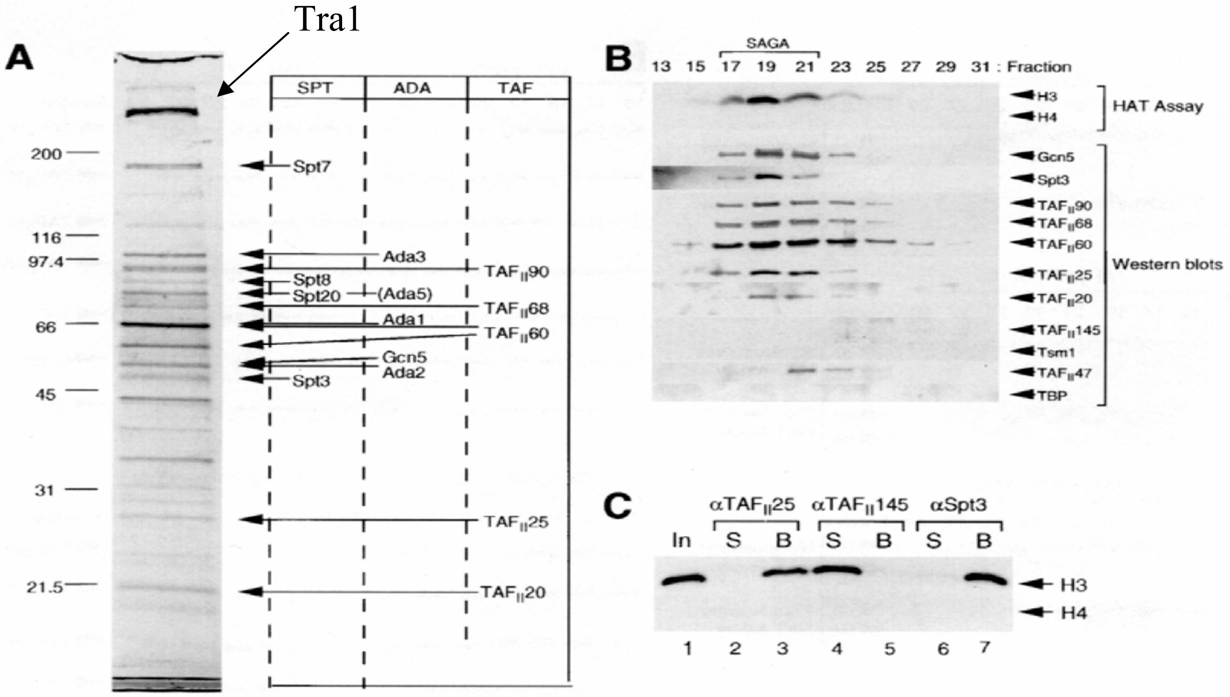


Figure 2. Identification of TAF$_{II}$s as Components of SAGA

(A) Silver stain gel of the SAGA complex. The single peak SAGA fraction from the final Mini Q column (Figure 1) was analyzed by silver staining of proteins separated on a 4%–20% SDS polyacrylamide gel. The components of SAGA identified by mass spectrometry are grouped according to their three different classes of transcriptional regulators, namely Adas, Spts, and TAF$_{II}$s. The position of a protein, which may represent TAF$_{II}$20/17, is also indicated.

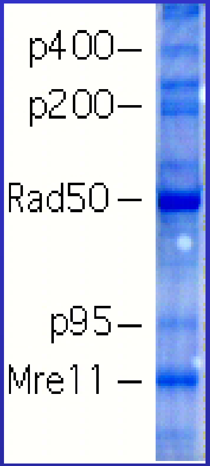
(B) Fluorogram of HAT assays and Western blots of purified SAGA. Fractions from the final Superose 6 column were incubated with nucleosomes and assayed for HAT activity or assayed by Western blotting with the indicated antisera. The position of histones H3 and H4 are indicated, as determined by Coomassie staining of the same gel.

(C) Immunoprecipitation of the SAGA complex. Purified SAGA complex was incubated with anti-TAF$_{II}$25, TAF$_{II}$145, or Spt3 antisera immobilized on protein A–Sepharose beads. A fluorogram of HAT reactions performed with free histones and either 20% of input material (In), 25% of supernatants (S), or 20% of beads (B).

P. Grant, D. Schieltz, M. Pray-Grant, D. Steger, J. Reese, J. Yates, J. Workman, *Cell* **94**, 45 (1998)

P. Grant, D. Schieltz, M. Pray-Grant, J. Yates, J. Workman, *Molecular Cell* **2**, 863 (1998)

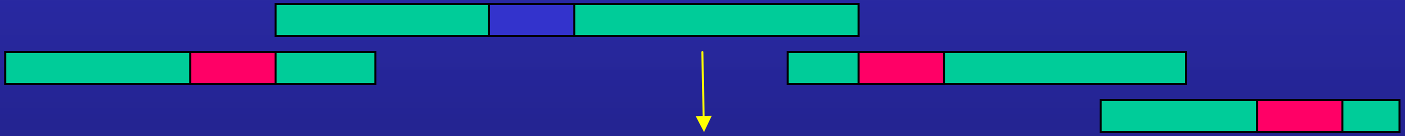
Nijmegen Breakage Syndrome: Disruption of Double Strand DNA Damage Repair



Immunoaffinity Isolation hRad50/hMre11 Protein Complex

Tandem Mass Spectrometry

EST Database



DNA Sequence Full Length cDNA

Map to Chromosome Location

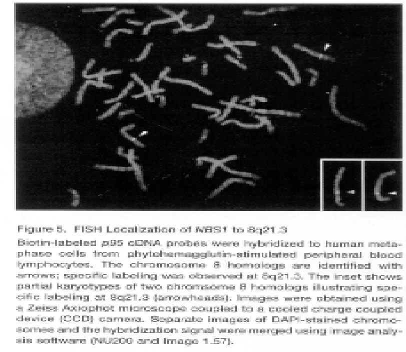


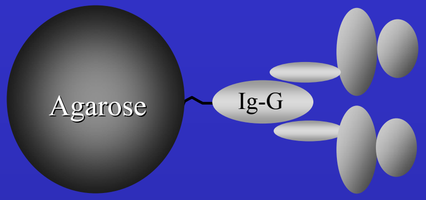
Figure 5. FISH Localization of NBS1 to 8q21.3

Biotin-labeled *p95* cDNA probes were hybridized to human metaphase cells from phytohemagglutin-stimulated peripheral blood lymphocytes. The chromosome 8 homologs are identified with arrows; specific labeling was observed at 8q21.3. The inset shows partial karyotypes of two chromosome 8 homologs illustrating specific labeling at 8q21.3 (arrowheads). Images were obtained using a Zeiss Axiophot microscope coupled to a cooled charge coupled device (CCD) camera. Separate images of DAPI-stained chromosomes and the hybridization signal were merged using image analysis software (NU200 and Image 1.57).

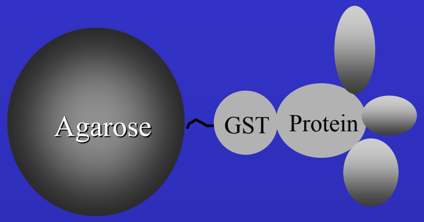
Nijmegen Breakage Syndrome

J.H.J Petrini, et al. *Cell* **93**, 477 (1998)

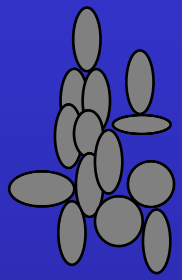
Protein-Protein Interactions



Co-immunoprecipitation



Protein Interaction
Chromatography



Multiprotein Complex

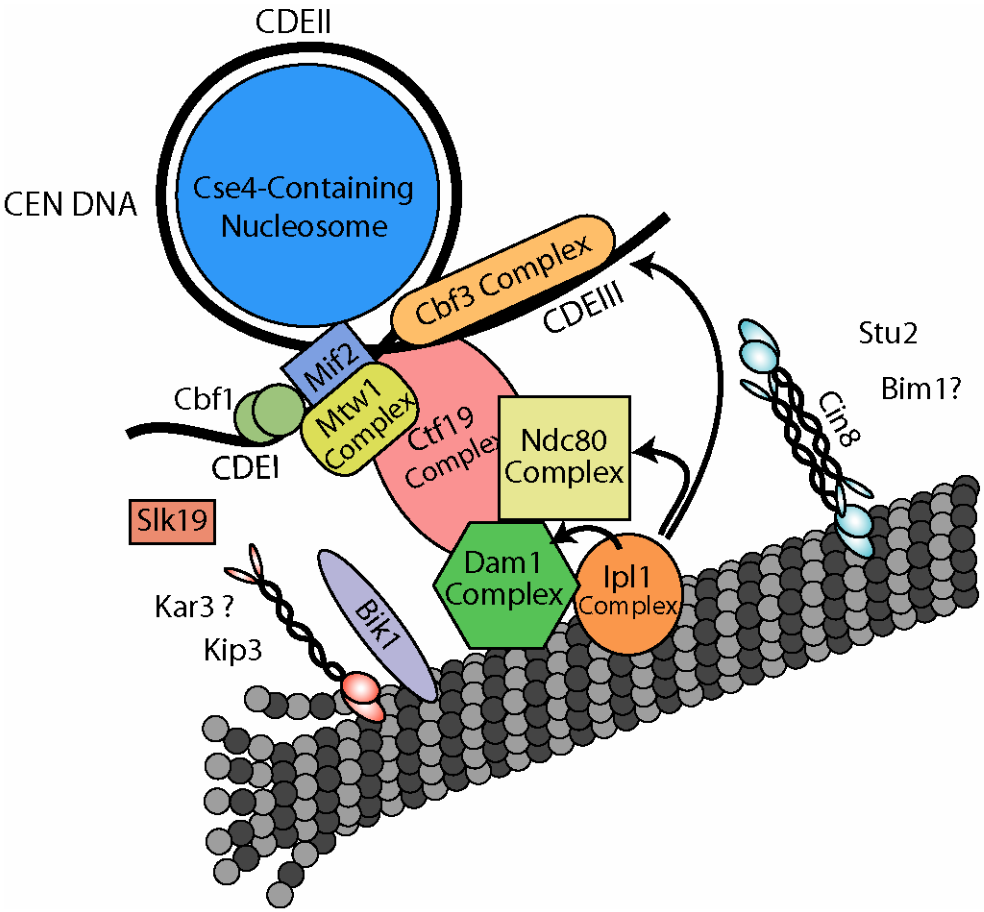
Coronin-linking actin and
microtubule cytoskeletons

Proteolysis
LC/MS/MS
SEQUEST

Identification of Protein Components

McCormack et al. *Analytical Chemistry* **69**, 767 (1997)
Lane-Goode et al. *J. Cell Biology* 144, 83 (1999), Meeusen et al. *J. Cell Biology* **145**, 291 (1999)

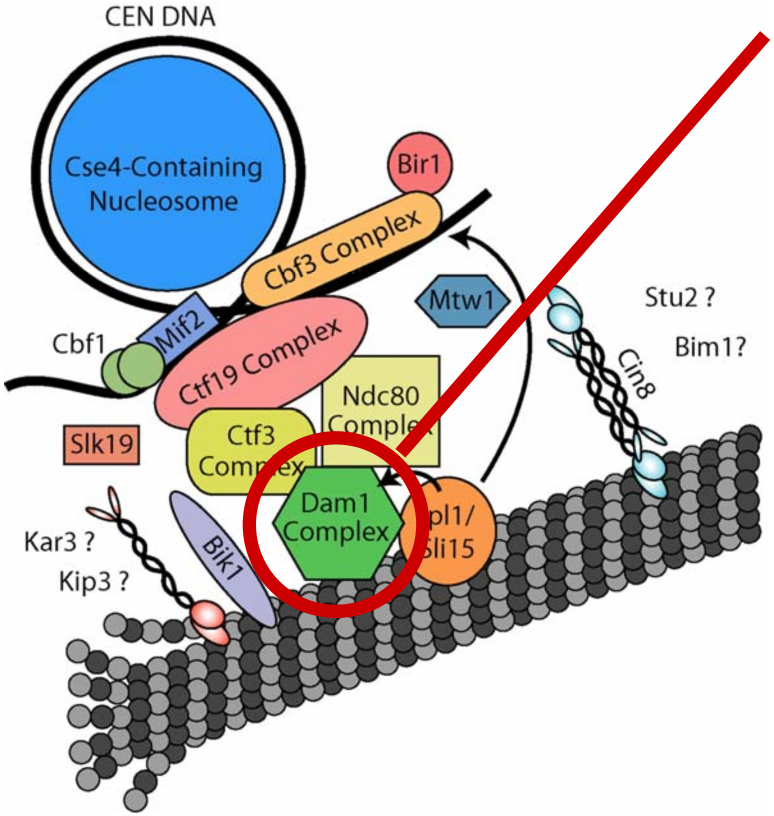
Molecular Analysis of Kinetochore Composition and Organization

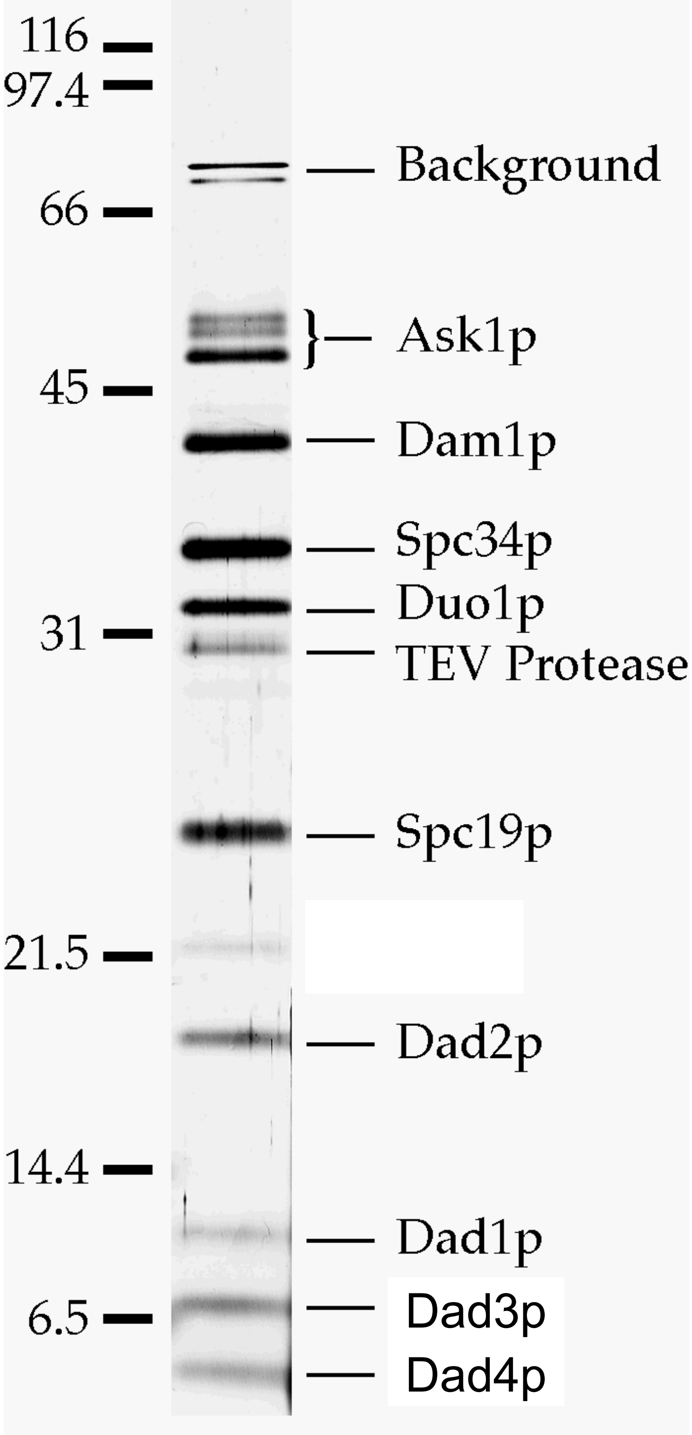


Cheeseman, Drubin, Barnes- UCB

The Dam1p Complex

Cheeseman et al. 2001
Janke et al. 2002
Li et al. 2002





Mapped Phosphorylation Sites

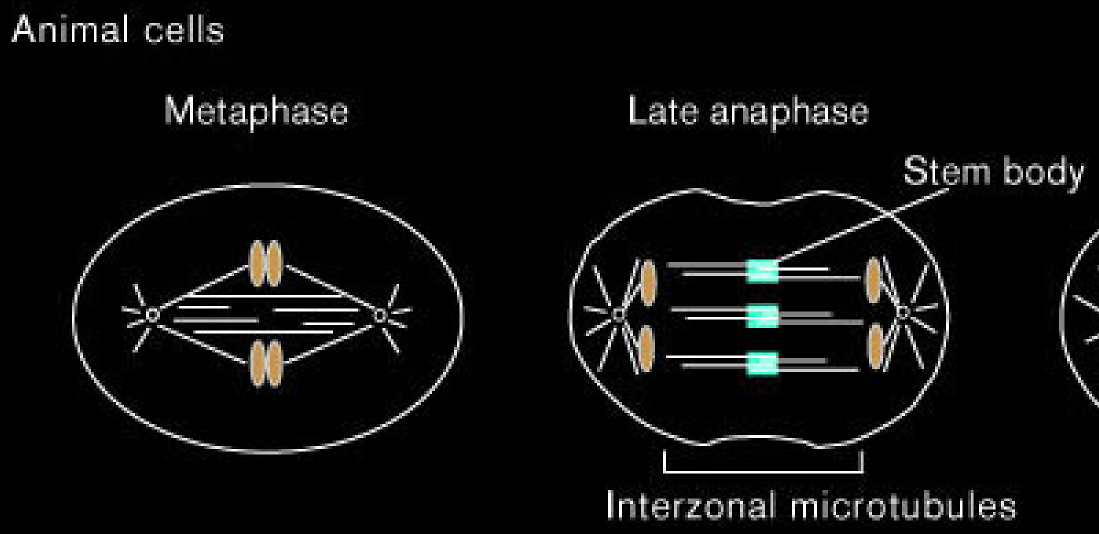
Complexes	*in vivo* Phosphorylation Sites	Ipl1p Targets
Dam1p Complex -	13	6
Ndc80p Complex -	1	1
Ctf19p Complex -	0	0
Ipl1p Complex -	4	3
Bim1p Complex -	2	1?
Mif2p Complex -	3	1?

Validation or Credentialing of Findings

- Simple model organisms such as yeast classical and molecular genetics is straightforward and high-throughput
- Animal systems: classical and molecular genetics is often possible but not high throughput

The Midbody

What is it?



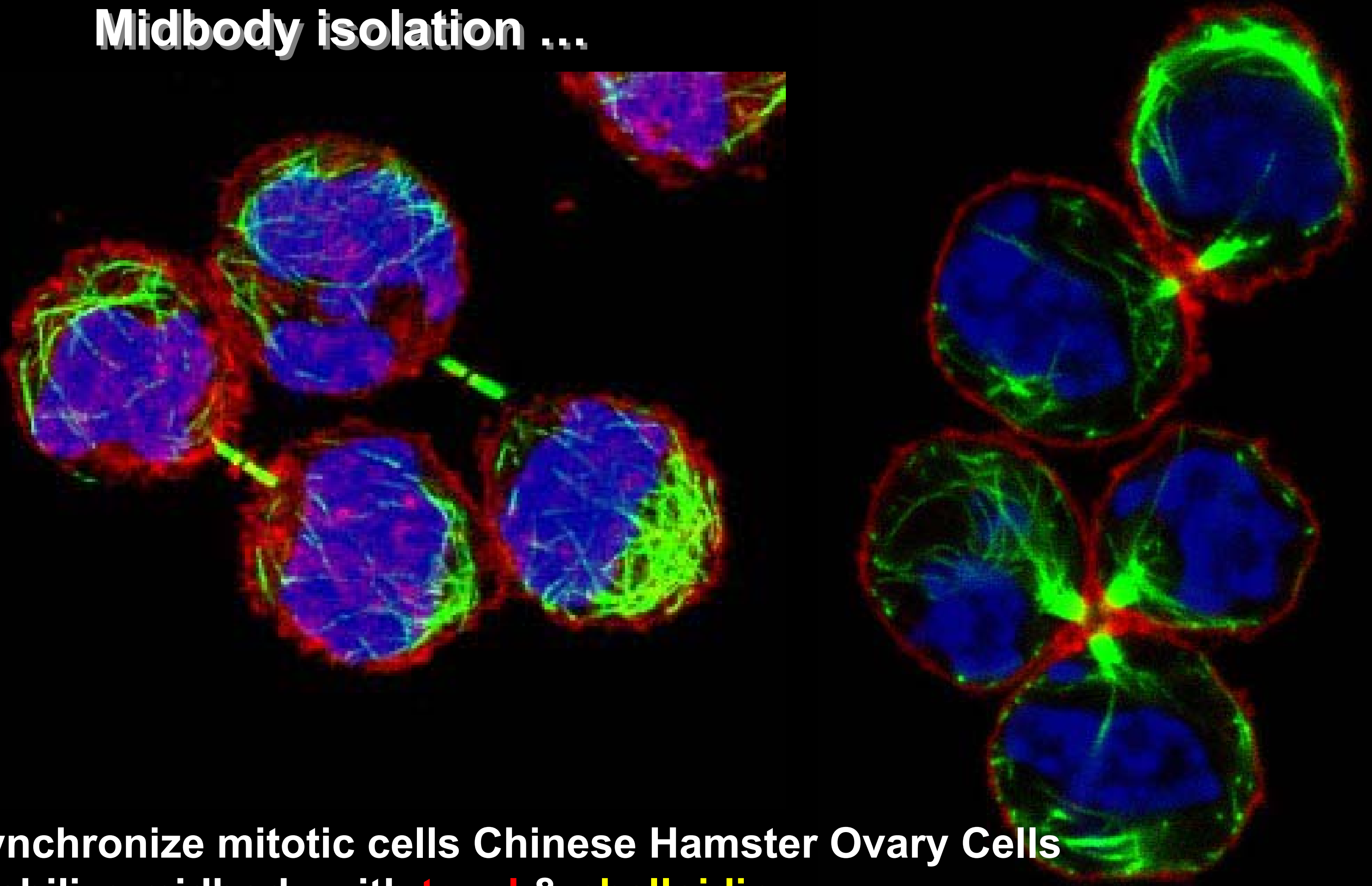
-formed from microtubules that span the spindle midzone in late anaphase that are bundled by the cleavage furrow

-function of midbody is unknown (except in spindle alignment)

Length(early) ~3-5um
Length (late) ~1um

A. Skop, R. Heald, B. Meyers- UC Berkeley

Midbody isolation ...



-synchronize mitotic cells Chinese Hamster Ovary Cells

-stabilize midbody with taxol & phalloidin

-lyse cells in a hypotonic medium with detergent (PIPES & Triton)

-separation of the midbody from other cellular components by centrif.

(adapted from Mullins & McIntosh, 1982 & Kuriyama etal, 1984)

Midbody LC/LC Tandem Mass Spectrometry Data

160 Mammalian Proteins-91% have homologs:

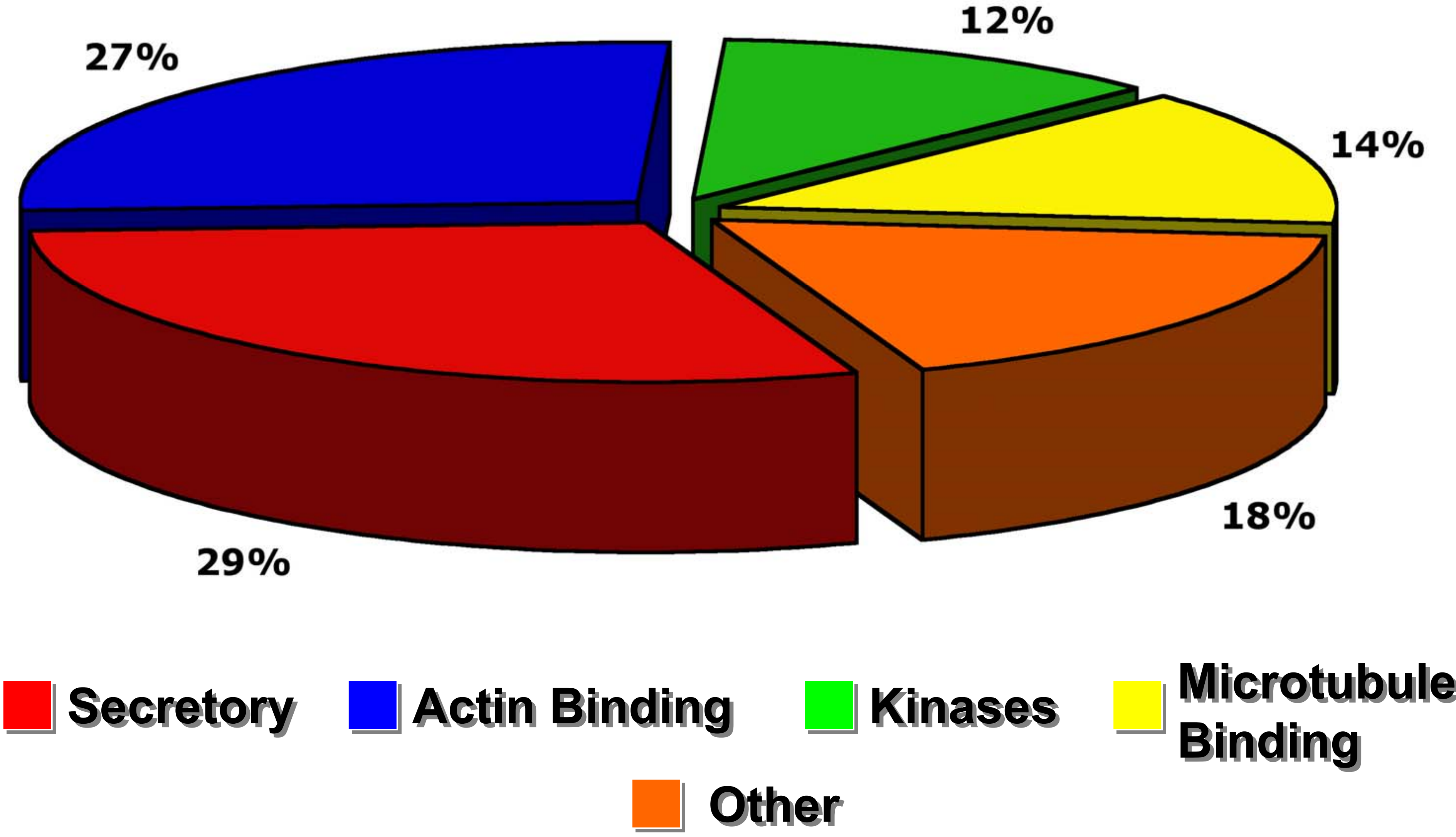
◆52 proteins were known cytokinesis proteins in mammalian, yeast, amoebae, plant and *Drosophila* systems…33%

◆21 proteins were well characterized cytokinesis proteins in *C.elegans*

◆9 novel mammalian proteins & corresponding *C. elegans* genes

106 proteins have an uncharacterized role in cytokinesis

Functional Classes of Mammalian Midbody Proteins



Functional Assay
- RNAi Screen in *C. elegans*

Do these proteins produce defects in cytokinesis after RNAi?

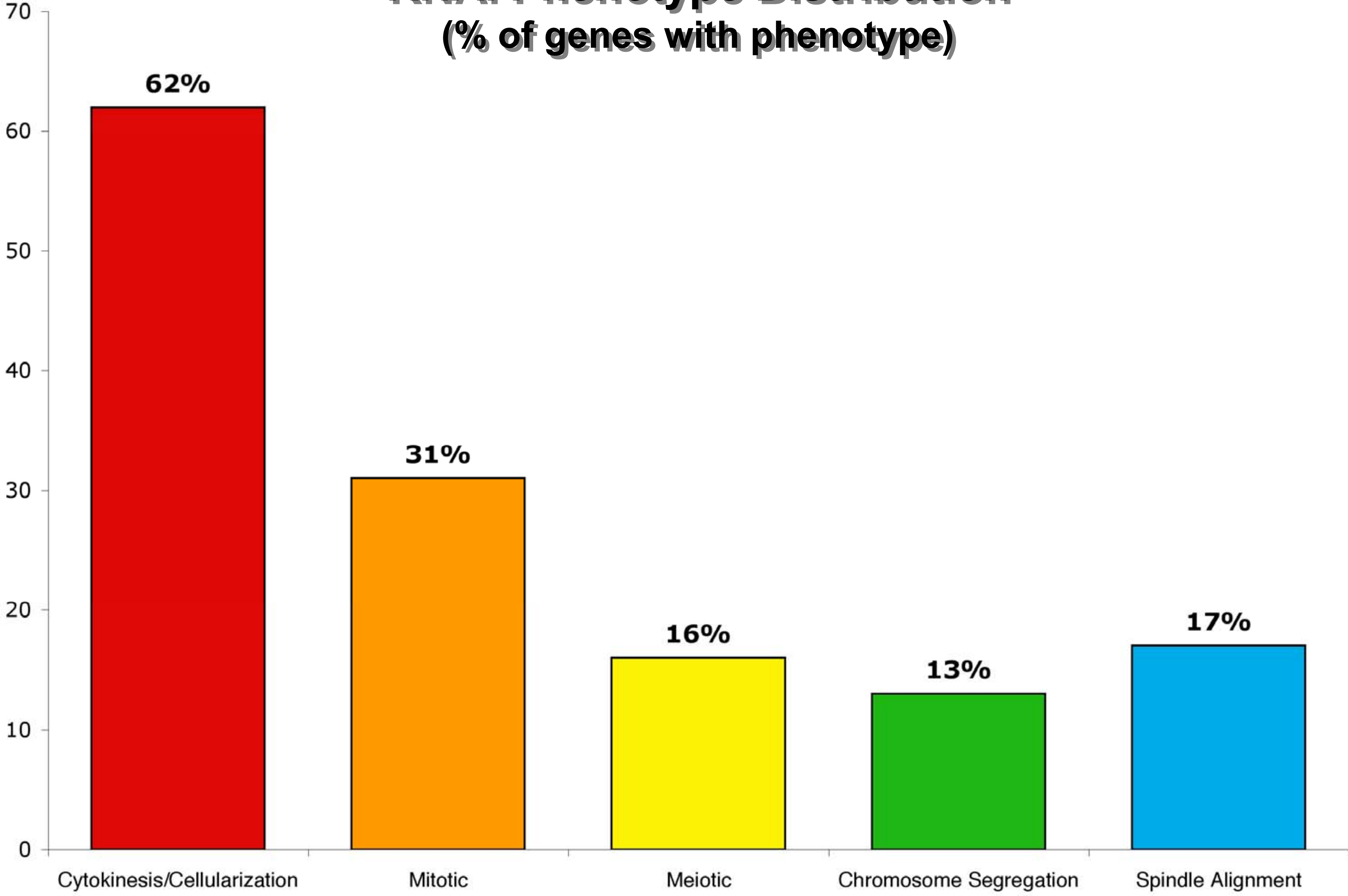
Find *C. elegans* homologs

- --~94% have homologs in *C. elegans*
- --suggests that cytokinesis is well conserved

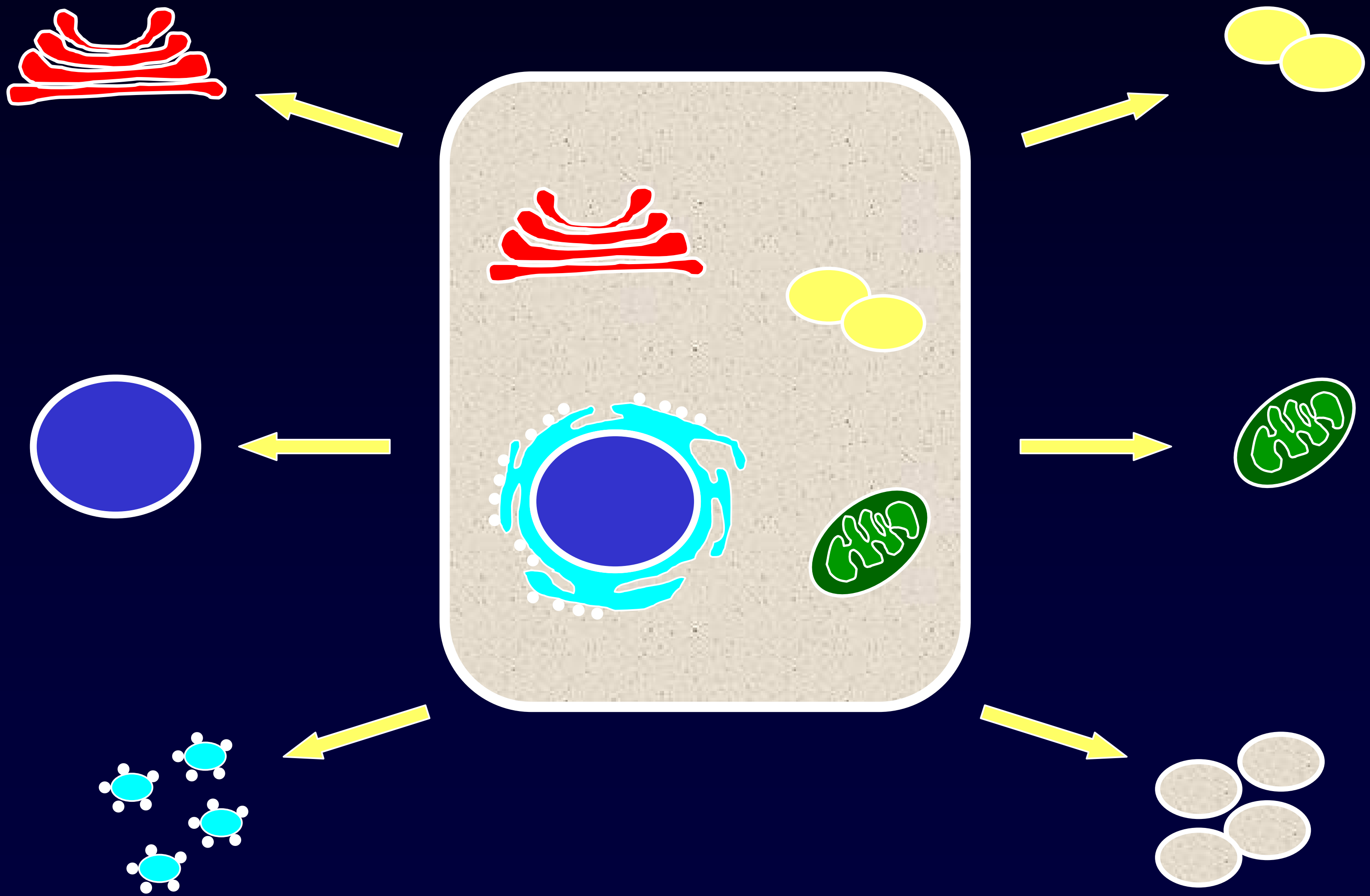
RNAi

- -Make dsRNA
- -Inject histone & tubulin::GFP strain
- -Wait 24-36hr
- -Film & Look for defects

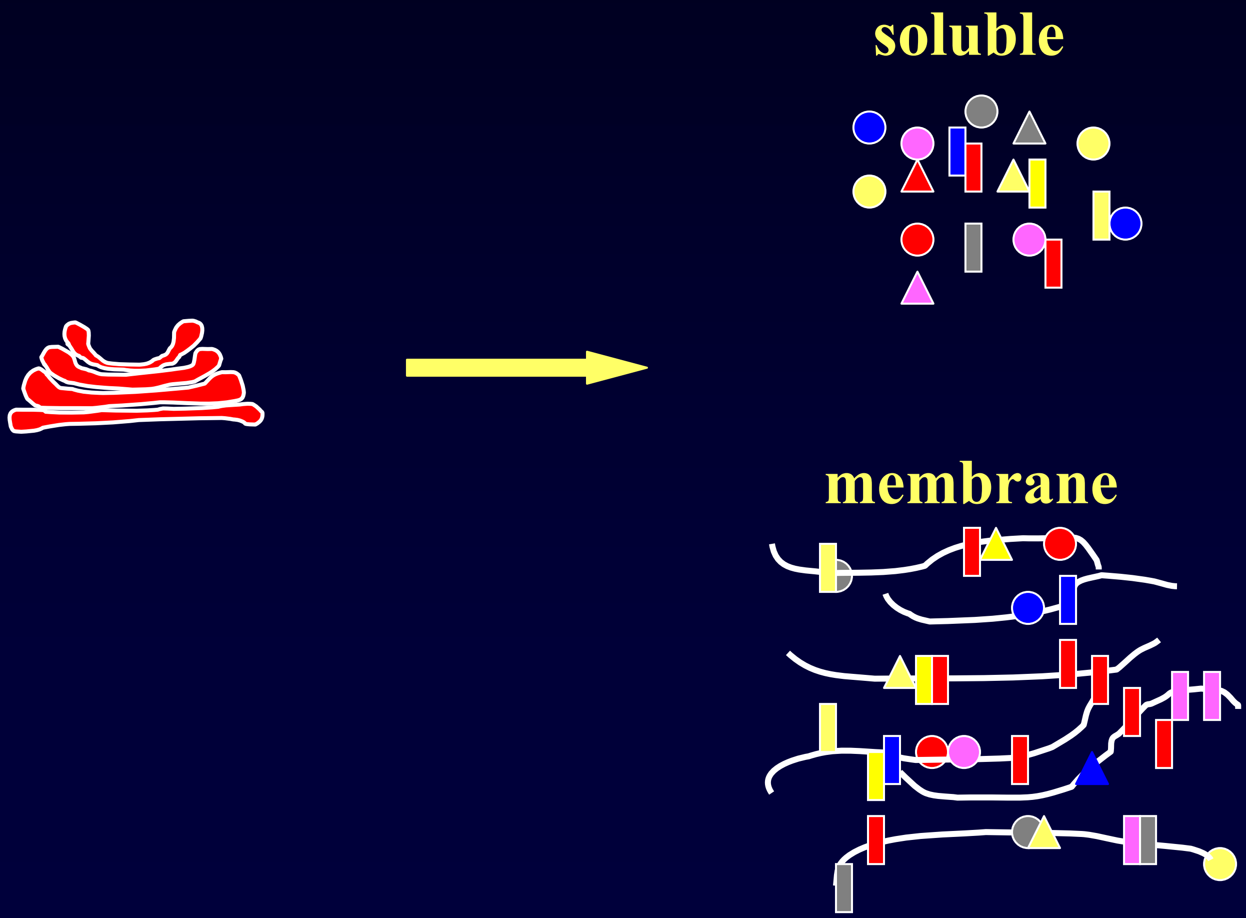
RNAi Phenotype Distribution
(% of genes with phenotype)



Cell Biology - Organelle Fractionation
Lumenal and Membrane Proteins



Biochemical fractionation to enrich for specific populations of proteins

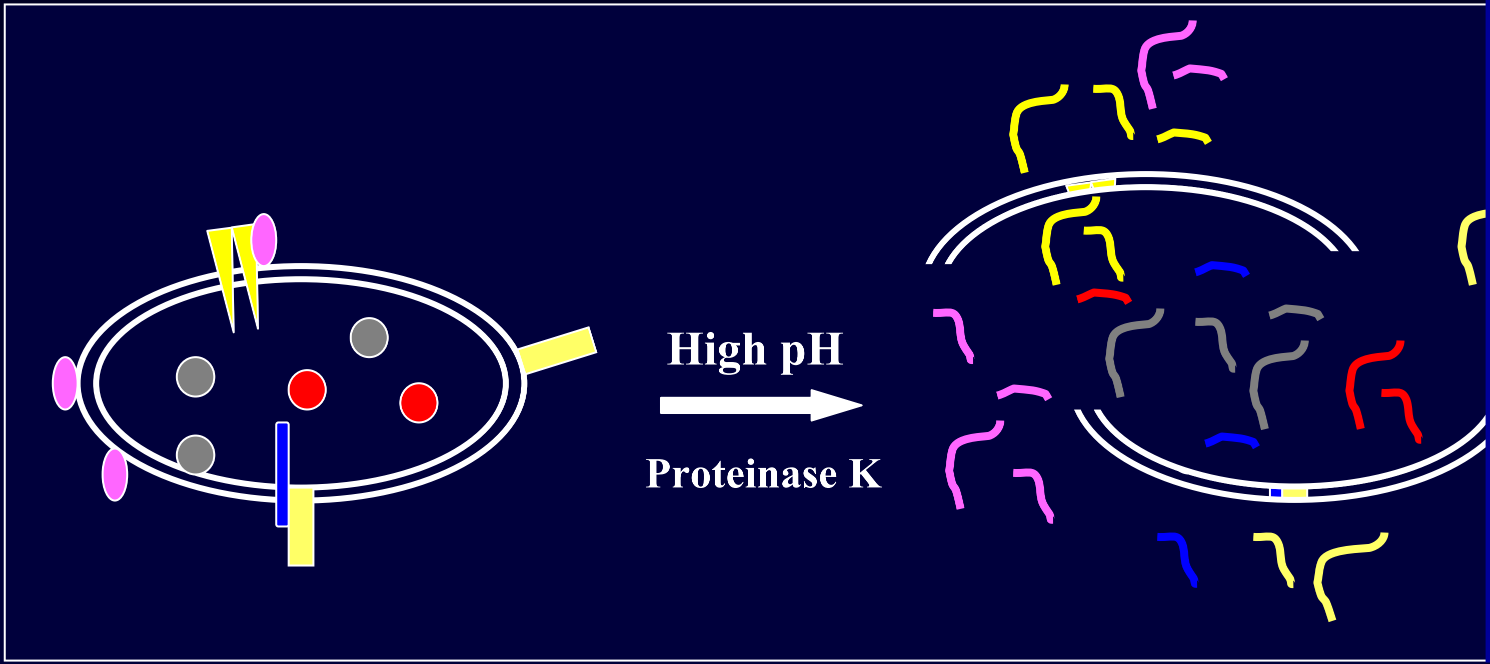


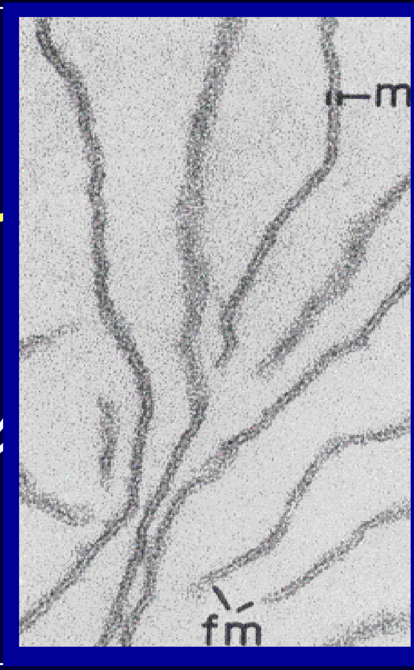
Taylor et al., Electrophoresis 21:3441-3459 (2000)

Technical Challenges For Proteomics

- Complex mixtures (whole cell lysates)
- Membrane proteins
- Quantitative comparisons

High pH/Proteinase K Method
(hpPK Method)





Wu et al., Nat. Biotech. 21:532-538 (2003)

Howell and Palade, J. Cell Biol. 92:822-832 (1982)

Overlapping Peptide Coverage

(TM7) gi|13794265|ref|NP_056312.1| DKFZP564G2022 protein

MAAAAWLQVL PVILLLLGAH PSPLSFFSAG PATVAAADRS
KWHIPIPSGK NYFSFGKILF RNTTIFLKFD GEPCDLSLNI
TWYLKSADCY NEIYNFKAEE VELYLEKLKE KRGLSGNIQT
SSKLFQNCSE LFKTQTFSGD FMHRLPLLGE KQEAKENGTN
LTFIGDKTAM HEPLQTWQDA PYIFIVHIGI SSSKESSKEN
SLSNLFTMTV EVKGPYEYLT LEDYPLMIFF MVMCIVYVLF
GVLWLAWSAC YWRDLLRIQF WIGAVIFLGM LEKAVFYAEF
QNIRYKGESV QGALILAELL SAVKRSLART LVSIVSLGYG
IVKPRLGVTL HKVVVAGALY LLFSGMEGVL RVTGAQTDLA
SLAFIPLAFL DTALCWWIFI SLTQTMKLLK LRRNIVKLSL
YRHFTNTLIL AVAASIVFII WTTMKFRIVT CQSDWRELWV
DDAIWRLLFS MILFVIMVLW RPSANNQRFA FSPLSEEEEE
DEQKEPMLKE SFEGMKMRST KQEPNGNSKV NKAQEDDLK**W**
VEENVPSSVT **DVALPALLDS** **DEER**MITHFE RSKME

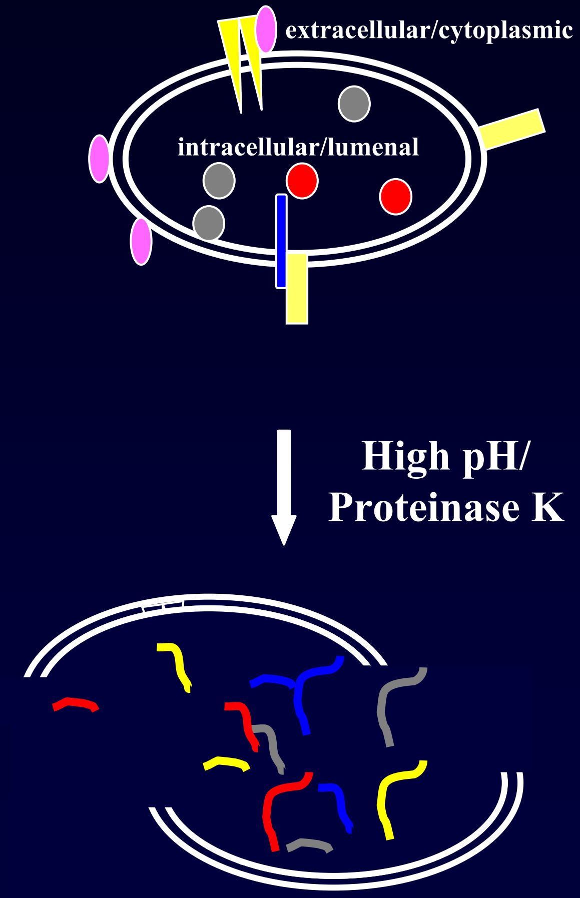
WVEENVPSSVTDVALPALLDS*DEER
VEENVPSSVTDVALPALLDS*DEER
EENVPSSVTDVALPALLDS*DEER
ENVPSSVTDVALPALLDS*DEER
VPSSVTDVALPALLDS*DEER
PSSVTDVALPALLDS*DEER
LPALLDS*DEER
PALLDS*DEER

(TM6) gi|14249524|ref|NP_116213.1| hypothetical protein FLJ14681

MVAACRSVAG LLPRRRRCFP ARAPLLRVAL CLLCWTPAAV
RAVPELGLWL ETVNDKSGPL IFRKTMFNST DIKLSVKSFH
CSGPVKFTIV WHLKYHTCHN EHSNLEELFQ KHKLSVDEDF
CHYLKNDNCW TTKNENLDCN SDSQVFPSLN NKELINIRNV
SNQERSMDVV ARTQKDGFHI FIVSIKTENT DASWNLNVSL
SMIGPHGYIS ASDWPLMIFY MVMCIVYILY GILWLTWSAC
YWKDILRIQF WIAAVIFLGM LEKAVFYSEY QNISNTGLST
QGLLIFAELI SAIKRTLARL LVIIVSLGYG IVKPRLGTVM
HRVIGLGLLY LIFAAVEGVM RVIGGSNHLA VVLDDIILAV
IDSIFVWFIF ISLAQTMKTL RLRKNTVKFS LYRHFKNTLI
FAVLASIVFM GWTTKTFRIA KCQSDWMERW VDDAFWSFLF
SLILIVIMFL WRPSANNQRY AFMPLIDDSD DEIEEFMVTS
ENLTEGIKLR ASKSVSNGTA KPATSENFDE DLK**WVEENIP**
SSFTDVALPV **LVDSDEEIMT** **RS**EMAEKMFS SEKIM

WVEENIPSSFTDVALPVLVDS*DEEIMTR
IPSSFTDVALPVLVDS*DEEIMTR
PSSFTDVALPVLVDS*DEEIMTR
SFTDVALPVLVDS*DEEIMTR
TDVALPVLVDS*DEEIMTR
TDVALPVLVDS*DEEIMTRS
DVALPVLVDS*DEEIMTR
VALPVLVDS*DEEIMTR
ALPVLVDS*DEEIMTR
ALPVLVDS*DEEIMTRS
LPVLVDS*DEEIMTR
PVLVDS*DEEIMTR
PVLVDS*DEEIMTRS

hpPK Method

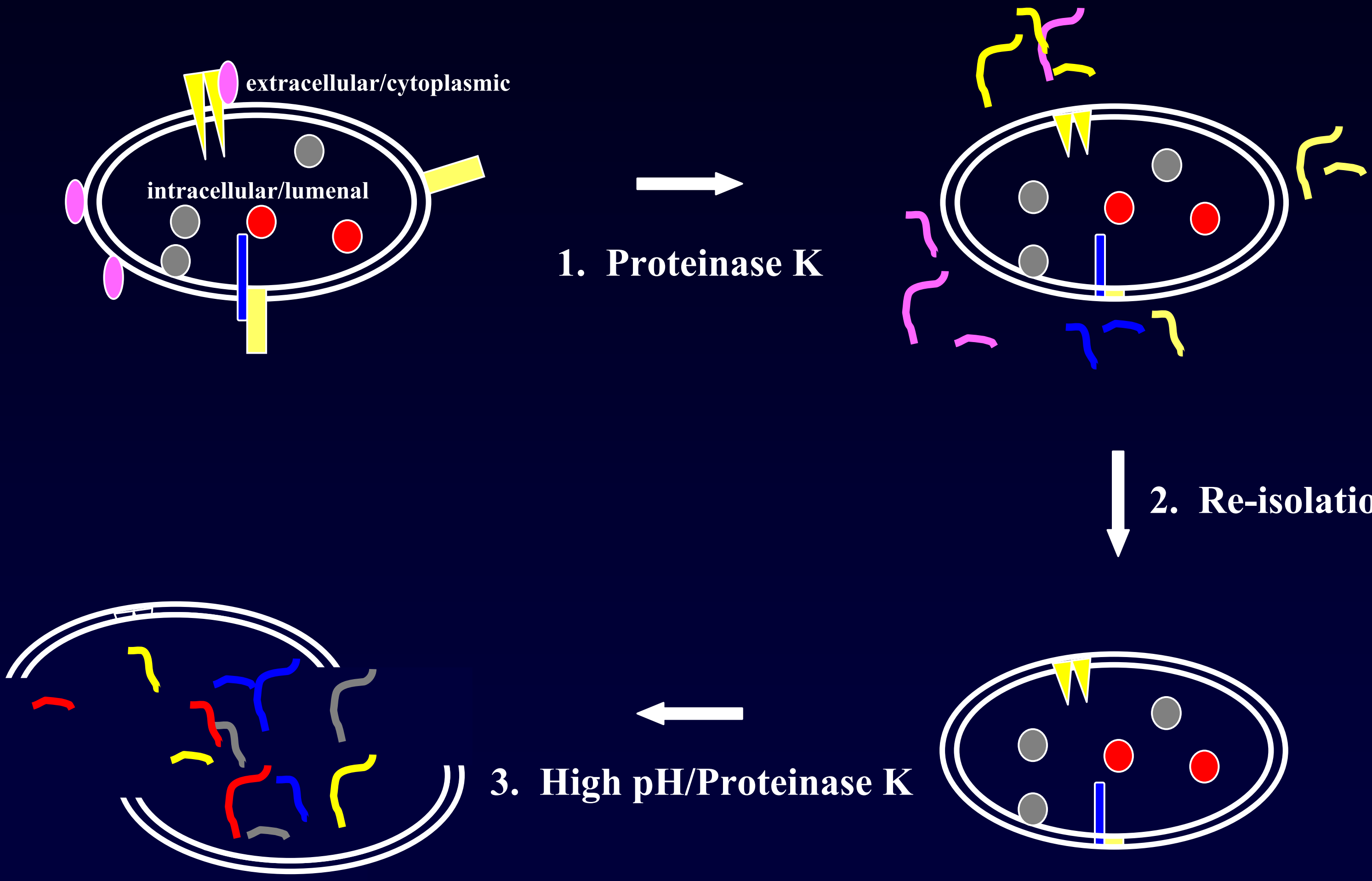


extracellular/cytoplasmic

intracellular/lumenal

High pH/
Proteinase K

hpPK Method Coupled with Protease Protection



1. **Proteinase K**

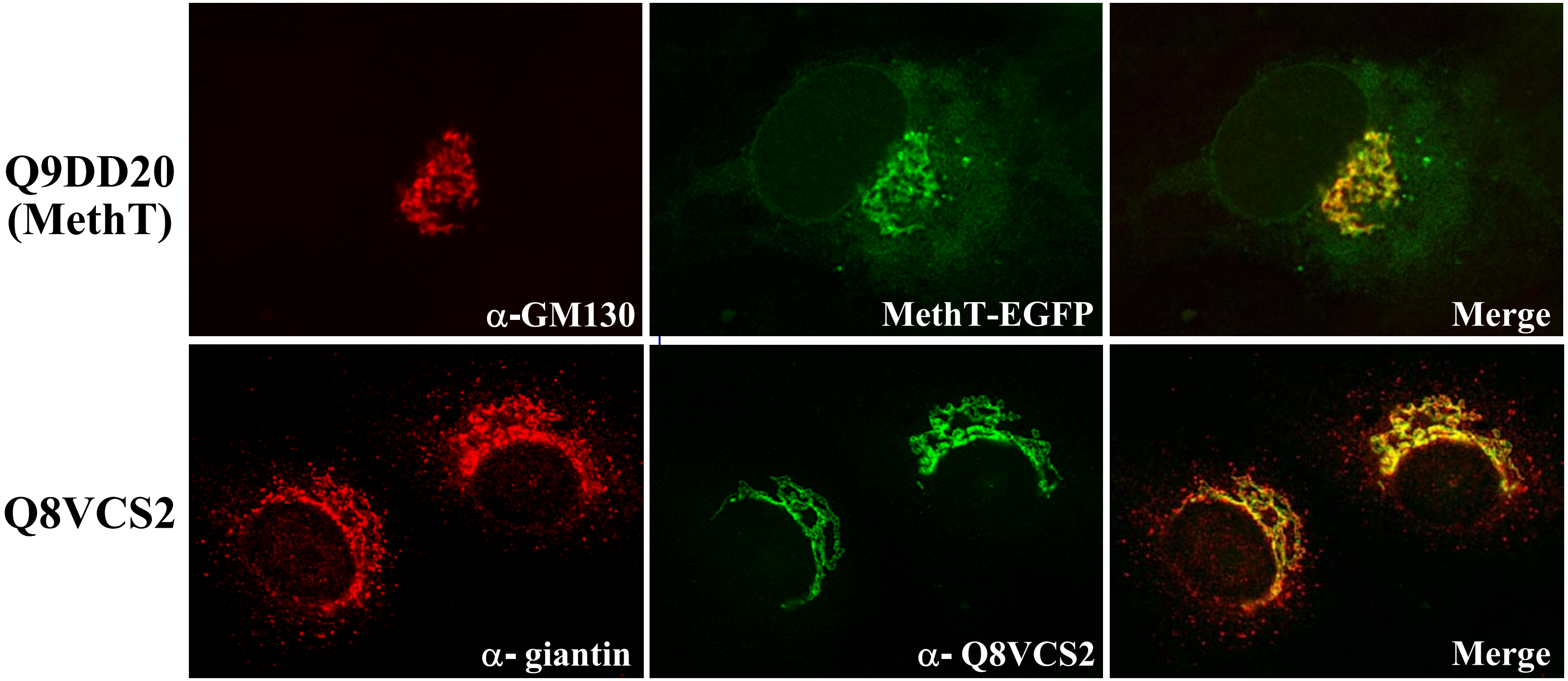
2. **Re-isolation**

3. **High pH/Proteinase K**

Topology

Rat Liver Golgi Proteome

- 421 unique identifications (5 peptide minimum for identification)
- 220 predicted membrane proteins
- 52.3% membrane proteins
- >100 PTMs
- >20% proteins with partial topology / localization data


Q9DD20
(MethT)
α-GM130
MethT-EGFP
Merge
Q8VCS2
α- giantin
α- Q8VCS2
Merge

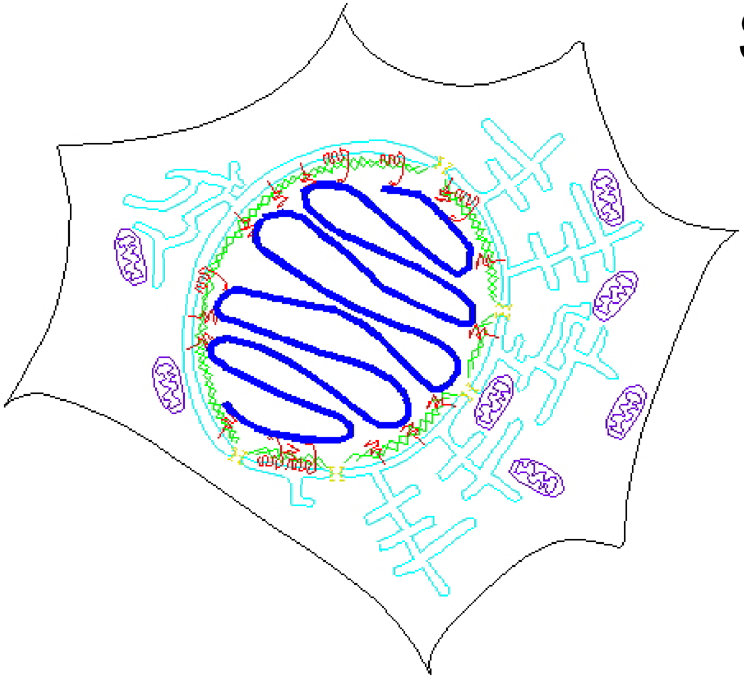
Integral Membrane Proteins

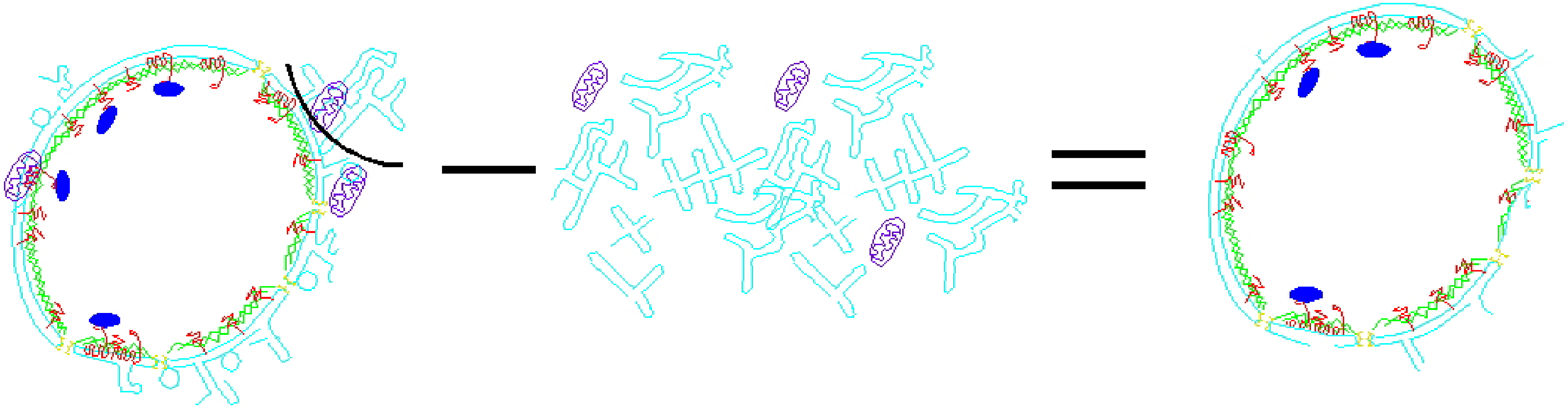
Novel Nuclear Membrane Proteins Identified by Proteomics

1. Analysis of Complex Fractions
2. Subtractive Proteomics
3. Integral Membrane Proteins
4. Use of Complex Tissues

Larry Gerace and Eric Schirmer

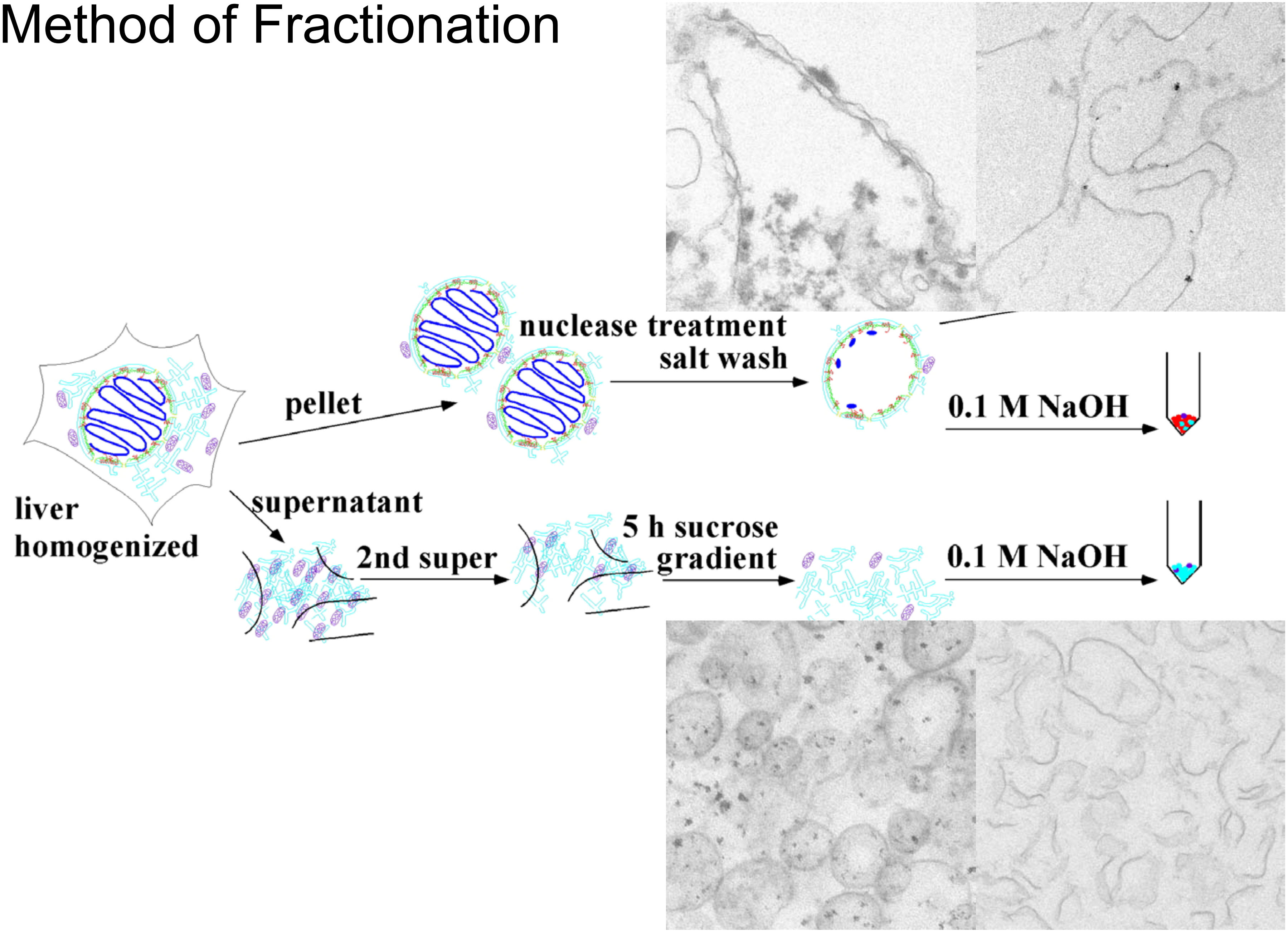
Subtractive Proteomics



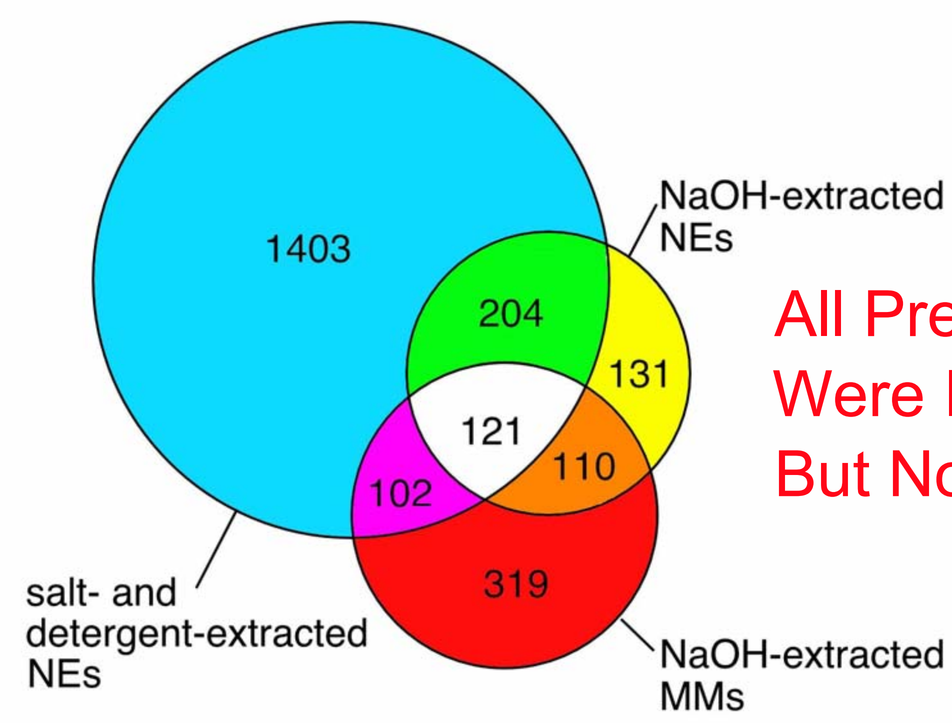


Nuclear Envelopes — Microsomal Membranes = Focus

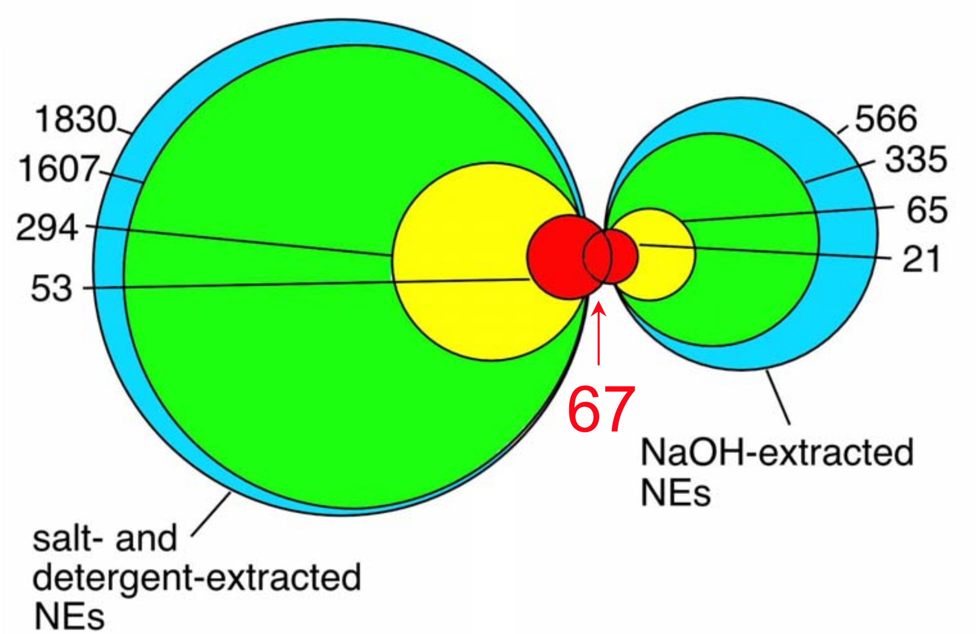
Method of Fractionation



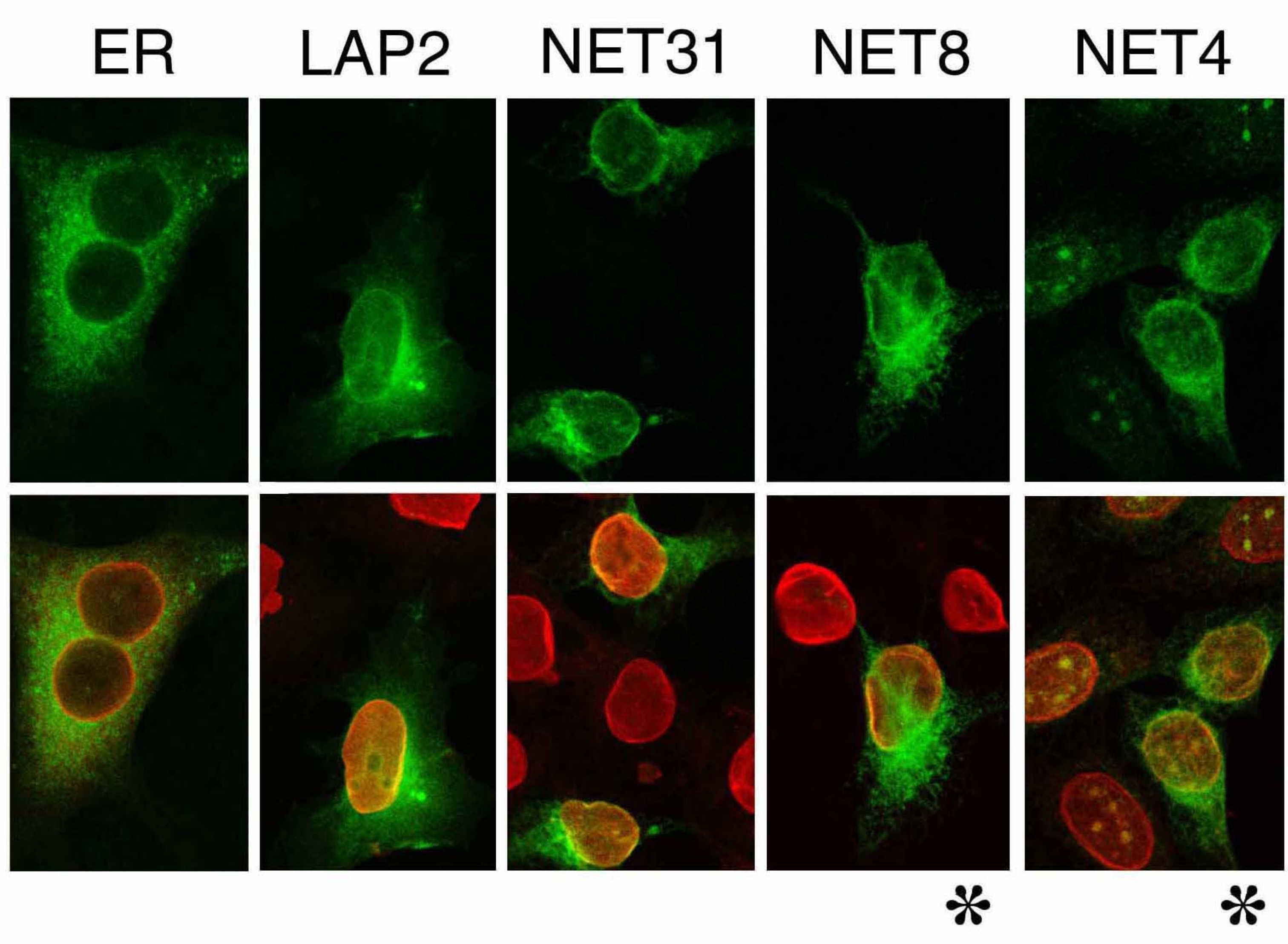
All Previously Known INM Proteins Were In the NE Chaotrope Fraction But Not In the MM Fraction



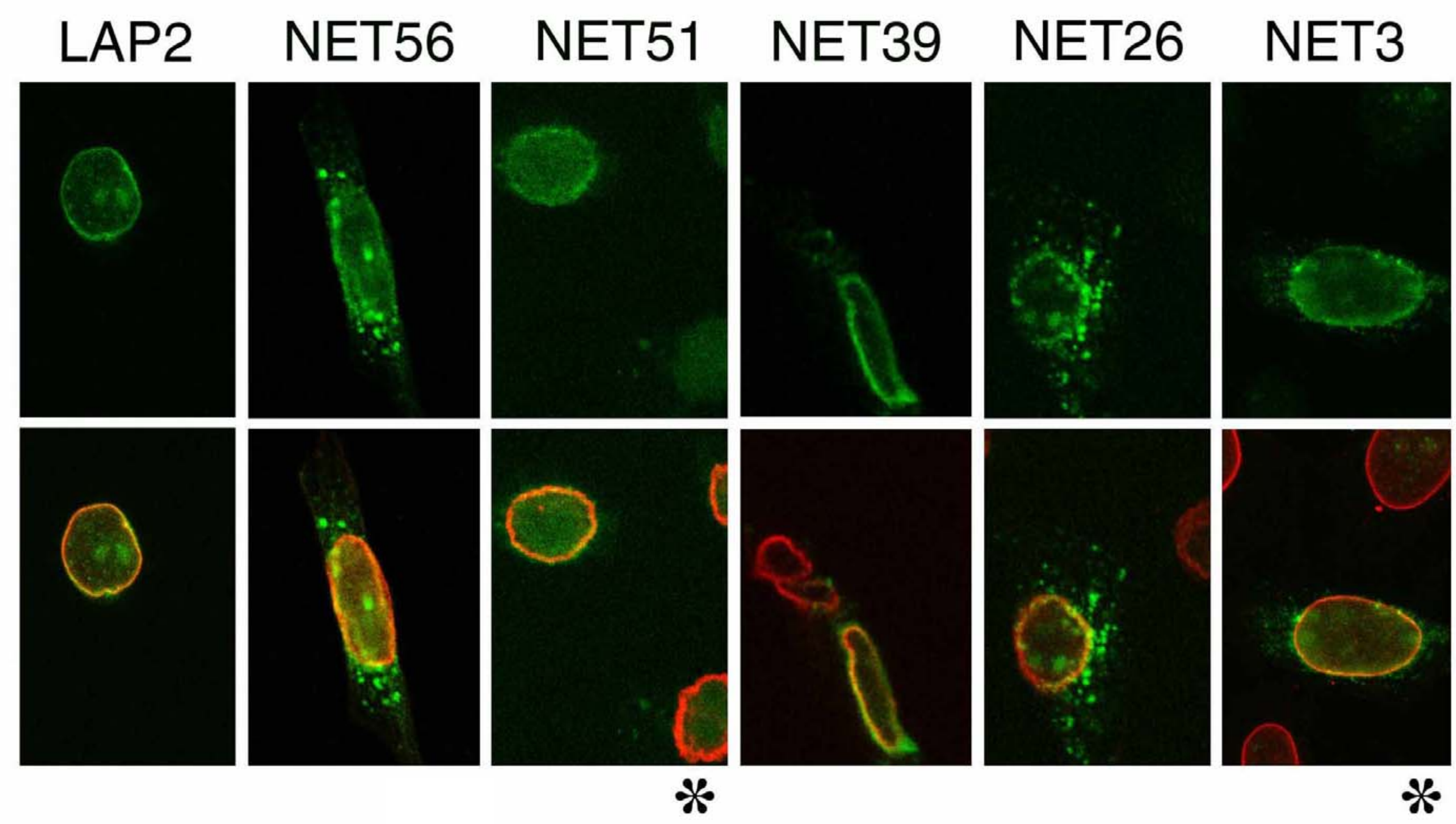
Total # Proteins
Subtracted MM
Hypothetical ORFs
Membrane Proteins



All Proteins Cloned
Targeted to the Nuclear Envelope



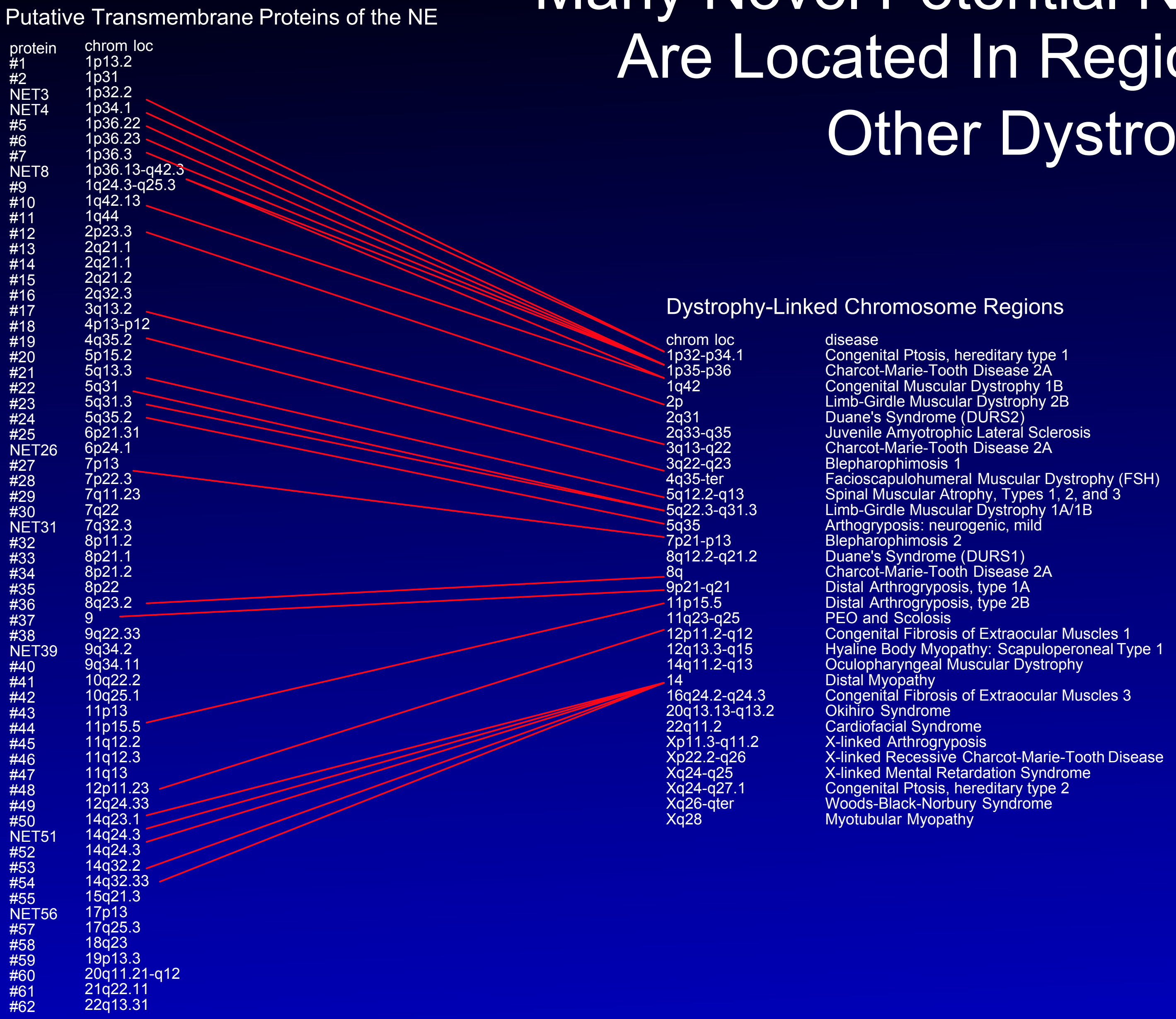
Several After Pre-Extraction With Detergent, Indicating a Tight Association With the Lamina



Nuclear Envelope Protein Mutations That Cause Human Disease Were Identified Rapidly

- Lamin A/C
  - Muscular Dystrophies
    - Limb girdle muscular dystrophy type 1B (2000)
    - Autosomal dominant Emery-Dreifuss muscular dystrophy (1999)
    - Autosomal recessive Emery-Dreifuss muscular dystrophy (2000)
    - Dilated cardiomyopathy with conduction defect 1 (1999)
    - Hauptmann-Thannhauser muscular dystrophy (2002)
  - Partial Lipodystrophy Syndromes
    - Dunnigan-type familial partial lipodystrophy (2000)
    - Mandibuloacral dysplasia (2002)
  - Neuropathy
    - Charcot-Marie-Tooth disorder type 2B1 (2002)
- Emerin
  - Muscular Dystrophies
    - X-linked Emery-Dreifuss Muscular Dystrophy (1994)
- LBR
  - Immune Development Anomaly
    - Pelger-Huet Anomaly (2002)

Many Novel Potential NE Proteins Are Located In Regions Where Other Dystrophies Map

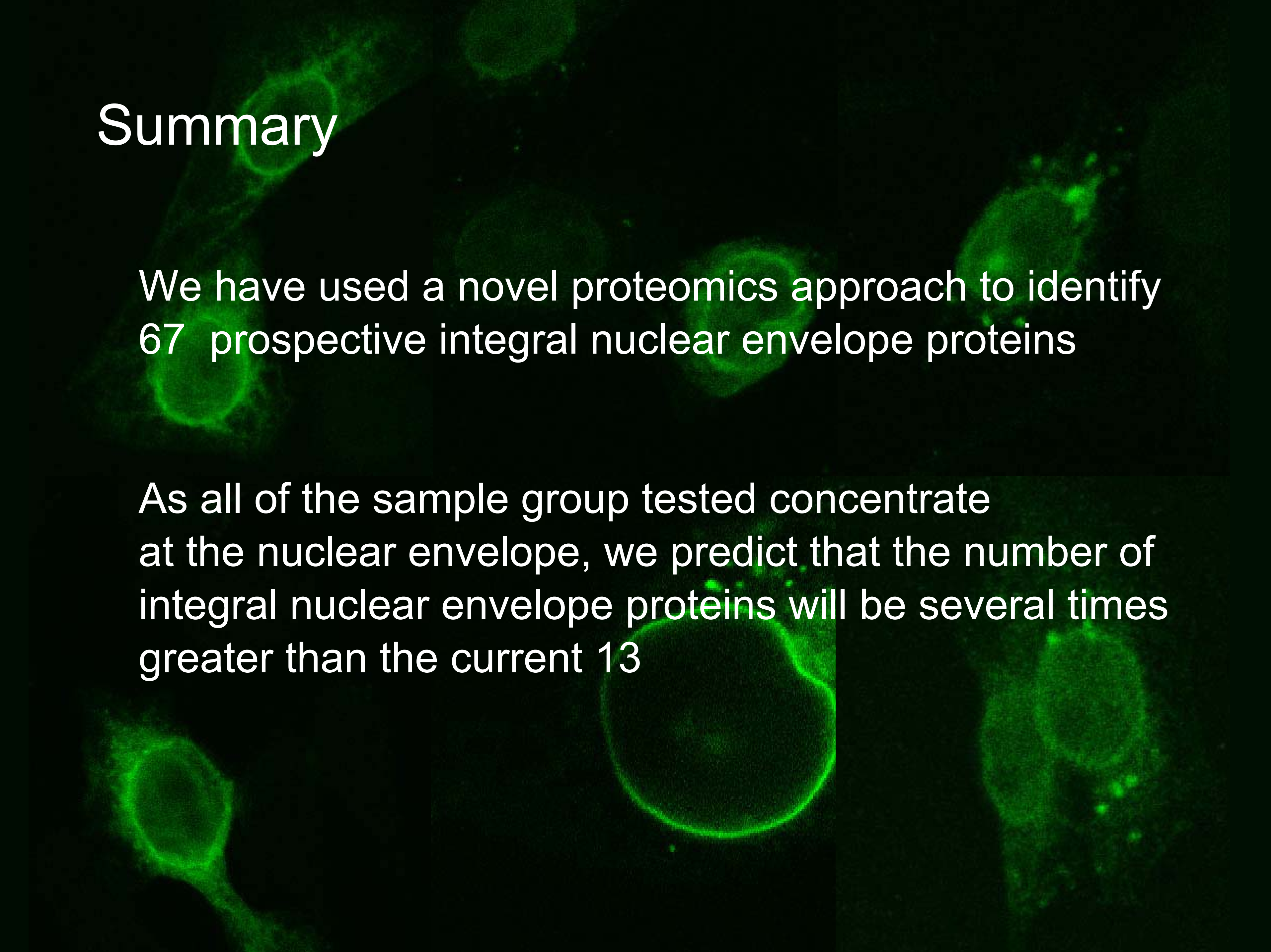


Putative Transmembrane Proteins of the NE

protein	chrom loc
#1	1p13.2
#2	1p31
NET3	1p32.2
NET4	1p34.1
#5	1p36.22
#6	1p36.23
#7	1p36.3
NET8	1p36.13-q42.3
#9	1q24.3-q25.3
#10	1q42.13
#11	1q44
#12	2p23.3
#13	2q21.1
#14	2q21.1
#15	2q21.2
#16	2q32.3
#17	3q13.2
#18	4p13-p12
#19	4q35.2
#20	5p15.2
#21	5q13.3
#22	5q31
#23	5q31.3
#24	5q35.2
#25	6p21.31
NET26	6p24.1
#27	7p13
#28	7p22.3
#29	7q11.23
#30	7q22
NET31	7q32.3
#32	8p11.2
#33	8p21.1
#34	8p21.2
#35	8p22
#36	8q23.2
#37	9
#38	9q22.33
NET39	9q34.2
#40	9q34.11
#41	10q22.2
#42	10q25.1
#43	11p13
#44	11p15.5
#45	11q12.2
#46	11q12.3
#47	11q13
#48	12p11.23
#49	12q24.33
#50	14q23.1
NET51	14q24.3
#52	14q24.3
#53	14q32.2
#54	14q32.33
#55	15q21.3
NET56	17p13
#57	17q25.3
#58	18q23
#59	19p13.3
#60	20q11.21-q12
#61	21q22.11
#62	22q13.31

Dystrophy-Linked Chromosome Regions

chrom loc	disease
1p32-p34.1	Congenital Ptosis, hereditary type 1
1p35-p36	Charcot-Marie-Tooth Disease 2A
1q42	Congenital Muscular Dystrophy 1B
2p	Limb-Girdle Muscular Dystrophy 2B
2q31	Duane's Syndrome (DURS2)
2q33-q35	Juvenile Amyotrophic Lateral Sclerosis
3q13-q22	Charcot-Marie-Tooth Disease 2A
3q22-q23	Blepharophimosis 1
4q35-ter	Facioscapulohumeral Muscular Dystrophy (FSH)
5q12.2-q13	Spinal Muscular Atrophy, Types 1, 2, and 3
5q22.3-q31.3	Limb-Girdle Muscular Dystrophy 1A/1B
5q35	Arthogryposis: neurogenic, mild
7p21-p13	Blepharophimosis 2
8q12.2-q21.2	Duane's Syndrome (DURS1)
8q	Charcot-Marie-Tooth Disease 2A
9p21-q21	Distal Arthrogryposis, type 1A
11p15.5	Distal Arthrogryposis, type 2B
11q23-q25	PEO and Scolosis
12p11.2-q12	Congenital Fibrosis of Extraocular Muscles 1
12q13.3-q15	Hyaline Body Myopathy: Scapuloperoneal Type 1
14q11.2-q13	Oculopharyngeal Muscular Dystrophy
14	Distal Myopathy
16q24.2-q24.3	Congenital Fibrosis of Extraocular Muscles 3
20q13.13-q13.2	Okihiro Syndrome
22q11.2	Cardiofacial Syndrome
Xp11.3-q11.2	X-linked Arthrogryposis
Xp22.2-q26	X-linked Recessive Charcot-Marie-Tooth Disease
Xq24-q25	X-linked Mental Retardation Syndrome
Xq24-q27.1	Congenital Ptosis, hereditary type 2
Xq26-qter	Woods-Black-Norbury Syndrome
Xq28	Myotubular Myopathy



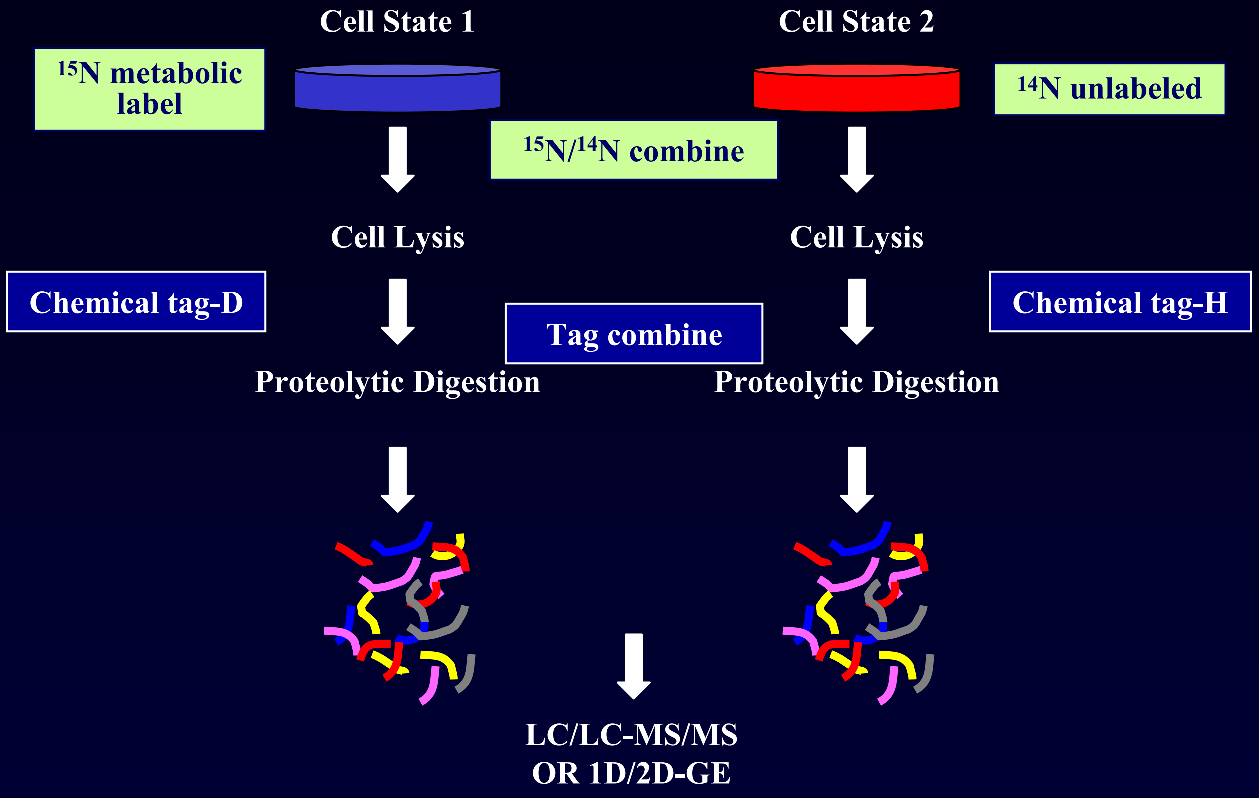
Summary

We have used a novel proteomics approach to identify 67 prospective integral nuclear envelope proteins

As all of the sample group tested concentrate
at the nuclear envelope, we predict that the number of integral nuclear envelope proteins will be several times greater than the current 13

Proteomics

- PROTEOMES ARE DYNAMIC
- QUANTITATION MEASURES DIFFERENCES



Cell State 1

Cell State 2

^{15}N metabolic label

^{14}N unlabeled

^{15}N/^{14}N combine

Cell Lysis

Cell Lysis

Chemical tag-D

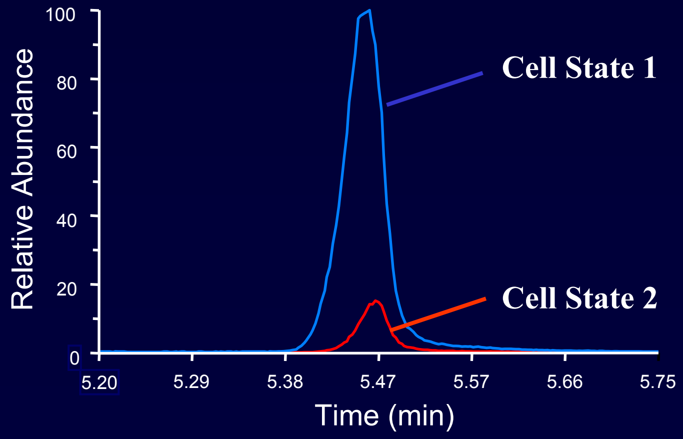
Tag combine

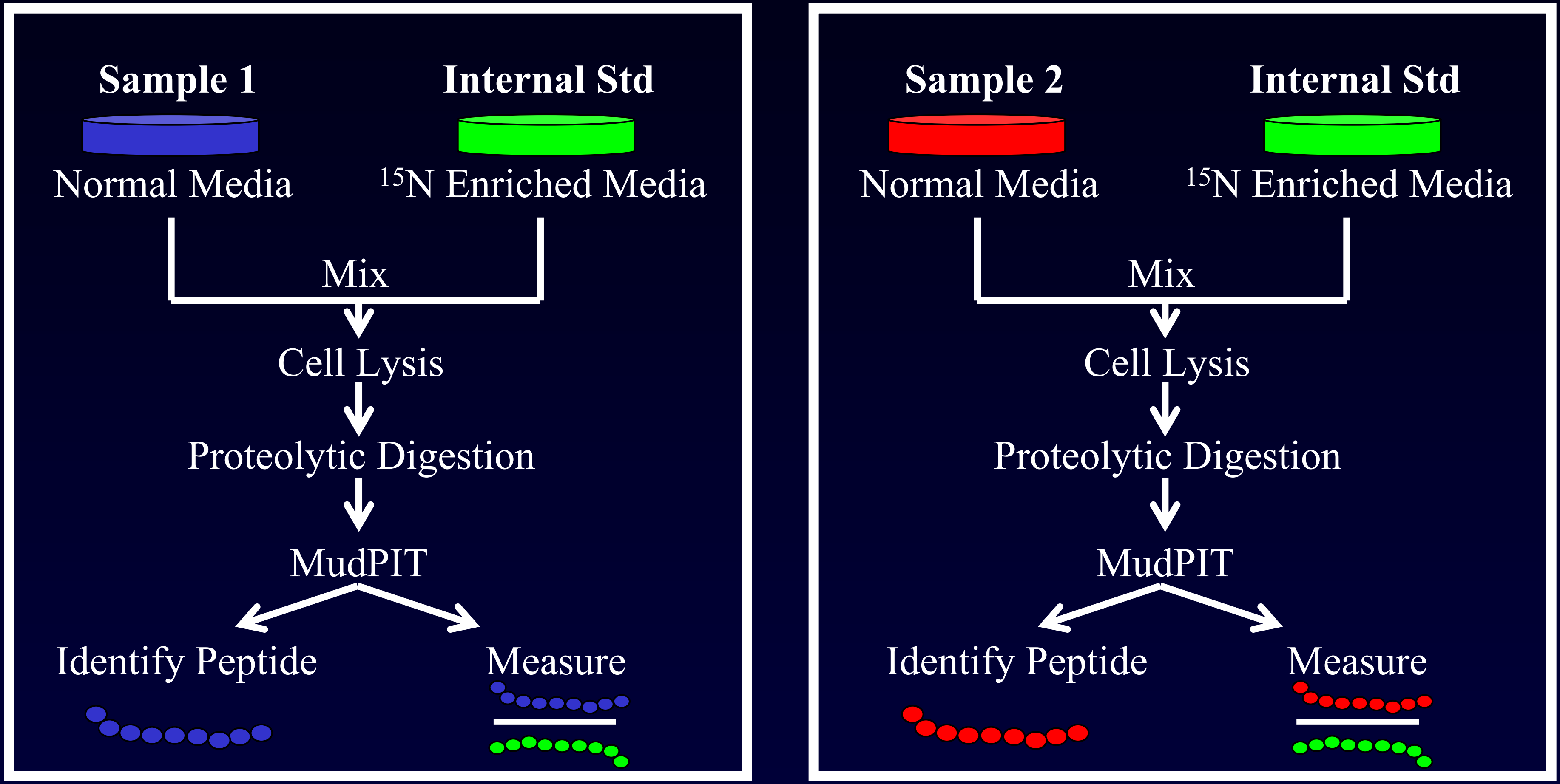
Chemical tag-H

Proteolytic Digestion

Proteolytic Digestion

LC/LC-MS/MS
OR 1D/2D-GE





Sample 1

Internal Std

Normal Media

^{15}N Enriched Media

Mix

Cell Lysis

Proteolytic Digestion

MudPIT

Identify Peptide

Measure

Sample 2

Internal Std

Normal Media

^{15}N Enriched Media

Mix

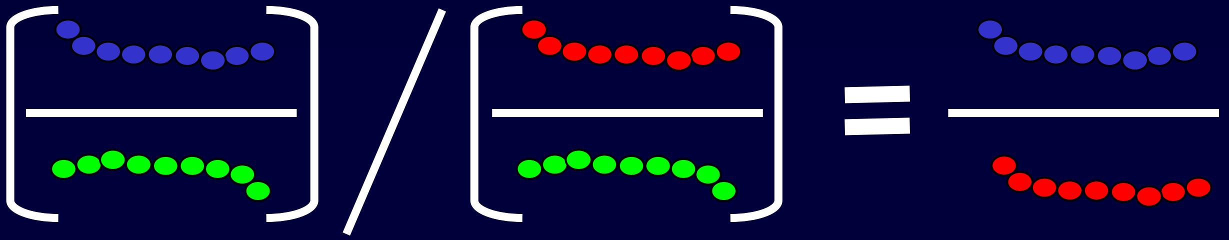
Cell Lysis

Proteolytic Digestion

MudPIT

Identify Peptide

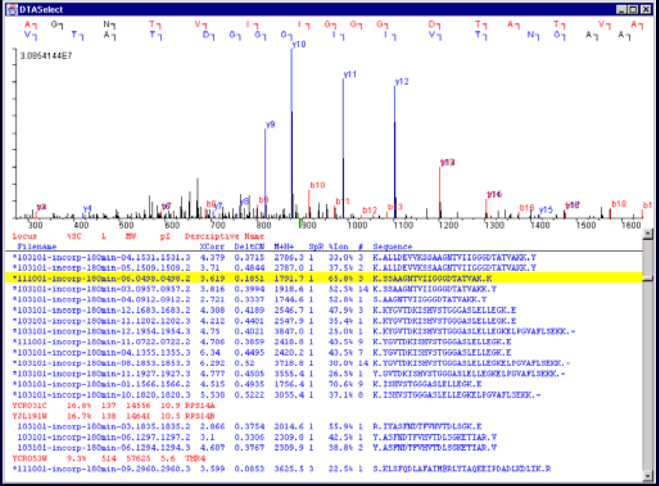
Measure



Fractional Change

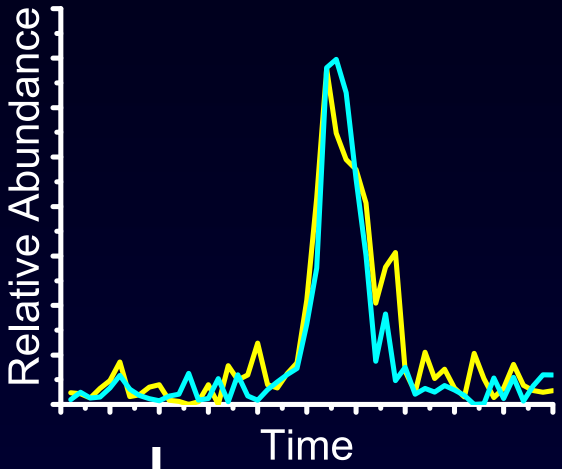
Measuring Changes in Protein Levels Using RelEx

Filtering SEQUEST Results

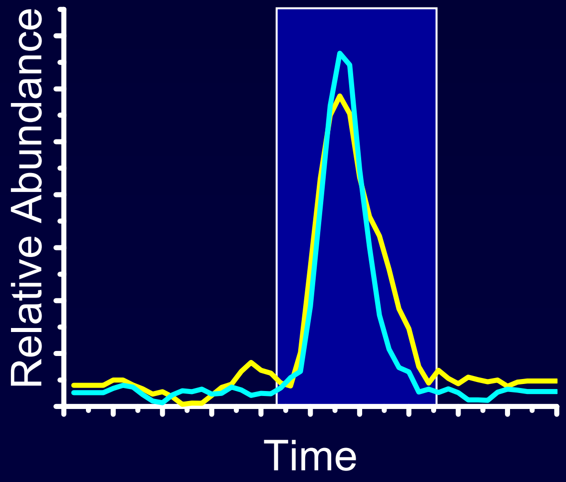


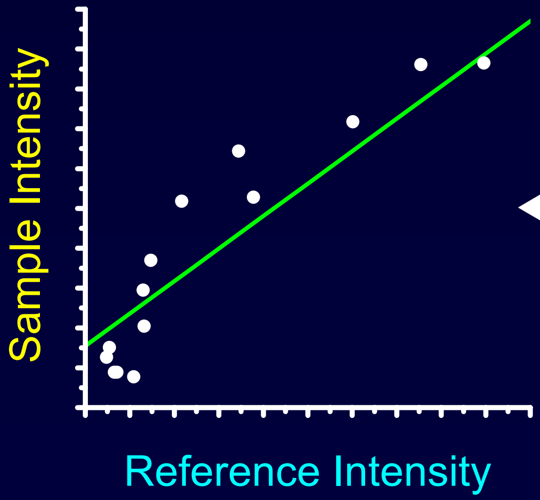
```
Locus     %C    L    MW     pI   Descriptive Name
 Filename                          XCorr  DeltCN  M+H+   SpR %Ion  #  Sequence
*103101-incorp-180min-04.1531.1531.3  4.379  0.3715  2706.3 1   33.0% 3  K.ALLDEVVKSSAAGNTVIIGGGDTATVAKK.Y
*103101-incorp-180min-05.1509.1509.2  3.71   0.4844  2787.0 1   37.5% 2  K.ALLDEVVKSSAAGNTVIIGGGDTATVAKK.Y
*111001-incorp-180min-06.0498.0498.2  3.619  0.1851  1791.7 1   65.8% 3  K.SSAAGNTVIIGGGDTATVAK.K
*103101-incorp-180min-03.0957.0957.2  3.816  0.3994  1918.6 1   52.5% 14 K.SSAAGNTVIIGGGDTATVAKK.Y
*103101-incorp-180min-04.0912.0912.2  2.721  0.3337  1744.6 1   52.8% 1  S.AAGNTVIIGGGDTATVAKK.Y
*103101-incorp-180min-12.1683.1683.2  4.308  0.4189  2546.7 1   47.9% 3  K.KYGVTDKISHVSTGGGASLELLEGK.E
*103101-incorp-180min-11.1202.1202.3  4.212  0.4401  2547.9 1   35.4% 1  K.KYGVTDKISHVSTGGGASLELLEGK.E
*103101-incorp-180min-12.1954.1954.3  4.75   0.4021  3847.0 1   25.0% 1  K.KYGVTDKISHVSTGGGASLELLEGKELPGVAFLSEKK.-
*111001-incorp-180min-11.0722.0722.2  4.706  0.3859  2418.8 1   43.5% 9  K.YGVTDKISHVSTGGGASLELLEGK.E
*103101-incorp-180min-04.1355.1355.3  6.34   0.4495  2420.2 1   43.5% 7  K.YGVTDKISHVSTGGGASLELLEGK.E
*103101-incorp-180min-08.1853.1853.3  6.292  0.52    3718.8 1   30.0% 14 K.YGVTDKISHVSTGGGASLELLEGKELPGVAFLSEKK.-
*103101-incorp-180min-11.1927.1927.3  4.777  0.4505  3555.4 1   26.5% 1  Y.GVTDKISHVSTGGGASLELLEGKELPGVAFLSEKK.-
*103101-incorp-180min-01.1566.1566.2  4.515  0.4935  1756.4 1   70.6% 9  K.ISHVSTGGGASLELLEGK.E
*103101-incorp-180min-10.1820.1820.3  5.538  0.5222  3055.4 1   37.1% 8  K.ISHVSTGGGASLELLEGKELPGVAFLSEKK.-
YCR031C    16.8%  137  14556  10.9 RPS14A
YJL191W    16.7%  138  14641  10.5 RPS14B
 103101-incorp-180min-03.1835.1835.2  2.066  0.3754  2014.6 1   55.9% 1  R.IYASFNDTFVHVTDLSGK.E
 103101-incorp-180min-06.1297.1297.2  3.1    0.3306  2309.8 1   42.5% 1  Y.ASFNDTFVHVTDLSGKETIAR.V
 103101-incorp-180min-06.1294.1294.3  4.607  0.3767  2309.9 1   38.8% 2  Y.ASFNDTFVHVTDLSGKETIAR.V
YCR053W     9.3%  514  57625  5.6  THR4
*111001-incorp-180min-09.2960.2960.3  3.599  0.0853  3625.5 3   22.5% 1  S.KLSFQDLAFAIM@RLYTAQEEIPDADLKDLIK.R
```

Extract Intensity vs Scan Data for Unlabeled and ^{15}N-Labeled Peptide m/z



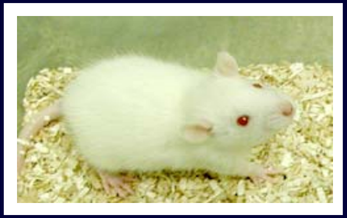
Smoothing and Peak Detection





Metabolic labeling works well with cells in culture and microorganisms because the growth media can be defined.

What about tissues?



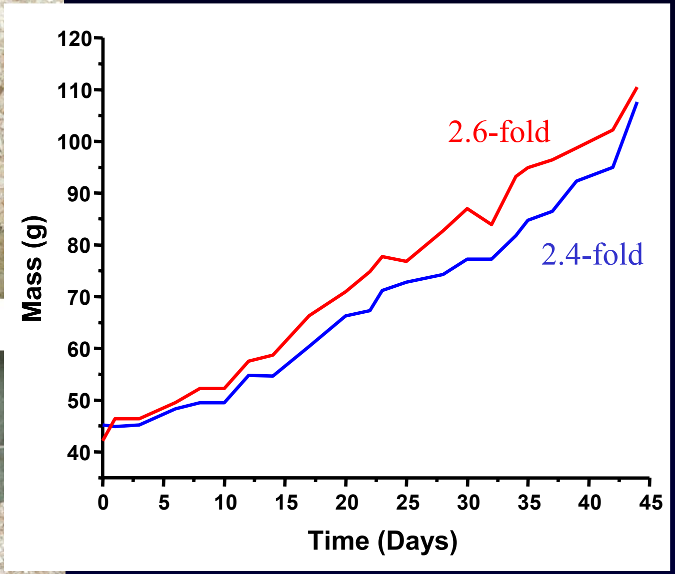
Can a rat be labeled with ^{15}N to provide internal standards for proteomics?

^{15}N metabolic labeling

^{15}N-labeled *protein* provided as only source of protein in rat chow

liver **plasma** **gut** **kidney** **lung** **heart** **muscle** **brain**

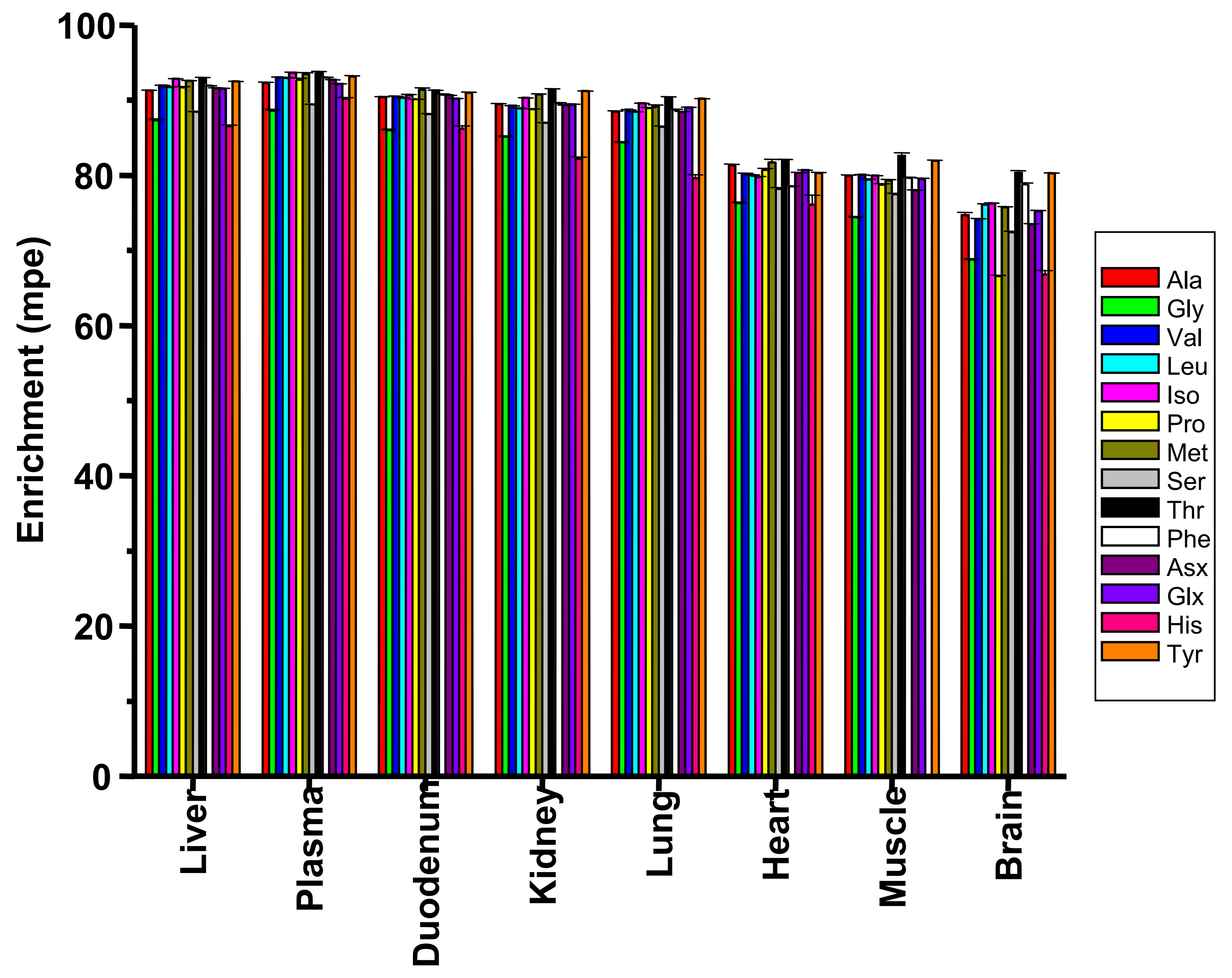
^{15}N Metabolic Labeling of a Rat



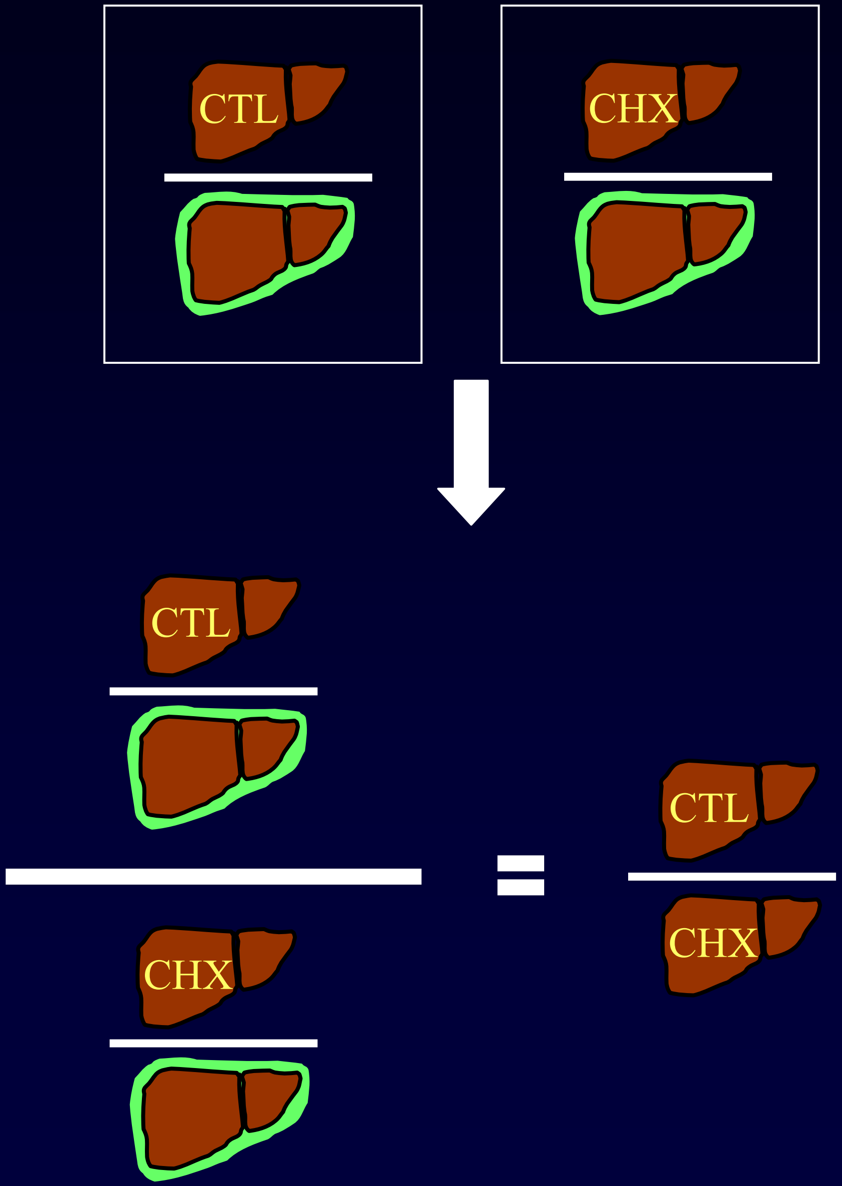
^{15}N-labeled

unlabeled

^{15}N amino acid enrichment in rat tissues



Cycloheximide-treated and untreated liver comparison



Quantitative Results

- Relative ratios were determined for 369 proteins with high signal/noise from both samples
- 62 of these had ratios that changed relative to actin ($p<0.1$)
- Automated analysis using RelEx

Cytochrome P450:

Xenobiotic and steroid metabolism

3A decreased

2C increased

	Swiss-Prot AC	Gene	Enzyme	TMDs	Description	Control			Chx				
						Avg Ratio	Std Dev	n	Avg Ratio	Std Dev	n	Factor Change	p-value
				3 IN	Cytochrome P450 3A2	6.05	0.030	2	4.39	0.000	1	-1.51	0.03
				4 IN	Cytochrome P450 3A18	1.18	0.205	15	0.68	0.085	17	-1.90	0.00
				3 IN	Cytochrome P450 2C11	1.79	0.315	31	3.51	0.979	46	1.78	0.00
				0 OUT	Cytochrome P450 2C23	1.09	0.233	12	1.60	0.309	12	1.34	0.05
5	P09875	Ugt2b1							0.71	0.105	10	1.32	0.09
6	P08542	Ugt2b3							1.39	0.458	16	1.33	0.08
7	Q8K154 mouse	LOC305264							1.04	0.297	4	1.73	0.08
8	P20070	Dia1							1.02	0.205	13	1.32	0.09
9	P04166	Omb5							1.06	0.038	3	1.56	0.09
10	P07687	Ephx1							0.92	0.159	16	1.37	0.03
11	P55159	Pon1							190.78	0.000	1	30.78	0.01
12	Q9EQ76	Fmo3							3.65	1.447	15	1.35	0.03
	ER												
13	P07153	Rpn1							1.16	0.157	10	1.70	0.00
14	P25235	Rpn2							1.20	0.248	7	1.39	0.09
15	O54734 mouse	LOC313648							1.31	0.341	7	1.71	0.05
16	P35565	Canx							1.58	0.095	6	1.59	0.00
17	P18418	Calr							1.86	0.237	6	1.68	0.00
18	P04785	P4hb							1.39	0.250	14	1.66	0.01
19	P06761	Hspa5							1.27	0.148	12	1.31	0.07
20	P55072 human	Vcp							2.71	0.650	27	1.46	0.00
21	P16303	Ces10							1.81	0.620	28	1.20	0.10
22	Q8K3R0	LOC246252							1.37	0.012	2	1.73	0.04
23	O09175	Rnpep	EC 3.4.11.6	1 IN	Aminopeptidase B	0.56	0.065	9	0.87	0.026	3	1.42	0.09
24	Q63617	Orp150	NA	0 IN	150 kDa oxygen-regulated protein	0.93	0.132	10	1.49	0.234	15	1.46	0.01
25	P08011	Mgst1	EC 2.5.1.18	3 OUT	Microsomal glutathione S-transferase 1	0.72	0.069	12	1.27	0.115	20	1.60	0.00
	Lipid/Cholesterol/Steroid Metabolism												
26	P18163	Facl2	EC 6.2.1.3	1 IN	Fatty acid Coenzyme A ligase, long chain 2	0.67	0.120	3	1.17	0.155	13	1.58	0.07
27	O88813	Facl5	EC 6.2.1.3	1 IN	Fatty acid Coenzyme A ligase, long chain 5	0.44	0.097	8	0.72	0.099	27	1.48	0.05
28	P49889	Ste	EC 2.8.2.4	0 OUT	Estrogen sulfotransferase, isoform 3	4.20	0.527	6	6.69	0.367	3	1.45	0.00
29	P15589	Sts	EC 3.1.6.2	2 OUT	Steryl-sulfatase	0.94	0.146	21	1.53	0.260	30	1.48	0.00
30	Q8VCH6 mouse	LOC298298	NA	0 OUT	24-dehydrocholesterol reductase	0.57	0.044	16	0.85	0.174	6	1.35	0.10
31	Q63448	Acox3	EC 1.3.3.6	3 IN	Acyl-coenzyme A oxidase 3, peroxisomal	0.57	0.065	7	0.93	0.104	5	1.50	0.06
32	Q8CHN6	Spl	EC 4.1.2.27	0 OUT	Sphingosine-1-phosphate lyase	0.64	0.059	12	1.09	0.247	9	1.55	0.03
33	Q02769	Fdft1	EC 2.5.1.21	2 OUT	Farnesyl-diphosphate farnesyltransferase	0.42	0.080	2	1.94	0.000	1	4.16	0.08
34	Q64591	Decr1	EC 1.3.1.34	0 OUT	2,4-dienoyl CoA reductase 1, mitochondrial	0.33	0.097	14	0.22	0.064	29	-1.69	0.07
	Metabolic Enzymes												
35	Q8CE68 mouse	LOC295399	NA	0 OUT	Similar to amylo-1,6-glucosidase, 4-alpha-glucanotransferase isoform 1	1.15	0.239	7	0.58	0.024	4	-2.18	0.03
36	P09811	Pygl	EC 2.4.1.1	0 OUT	Glycogen phosphorylase, liver form	0.67	0.168	13	0.40	0.168	15	-1.82	0.03
37	P07824	Arg1	EC 3.5.3.1	1 IN	Arginase 1, liver	0.63	0.074	12	0.92	0.064	4	1.34	0.10
38	P09034	Ass	EC 6.3.4.5	1 IN	Arginosuccinate synthase	0.46	0.060	13	0.75	0.085	10	1.51	0.03
39	P25093	Fah	EC 3.7.1.2	0 OUT	Fumarylacetoacetase	0.67	0.091	20	1.02	0.092	17	1.38	0.01
40	O88655	Hppd	NA	0 OUT	4-hydroxyphenylpyruvate dioxygenase	0.69	0.052	10	1.06	0.155	21	1.39	0.03
41	Q9D8C9 mouse	LOC297713	NA	0 OUT	Similar to Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)	0.78	0.071	11	1.20	0.279	6	1.41	0.08
42	O09171	Bhmt	EC 2.1.1.5	1 IN	Betaine-homocysteine S-methyltransferase	0.81	0.153	21	1.41	0.226	33	1.58	0.00
43	P14141	Ca3	EC 4.2.1.1	0 OUT	Carbonic anhydrase III	1.97	0.679	33	3.31	1.761	14	1.53	0.00
	Translation												
44	Q9QX72	Sbp2	NA	0 IN	SECIS-binding protein 2	16.58	12.828	3	31.98	0.000	1	1.76	0.05
45	P05197	Eef2	NA	0 OUT	Elongation factor 2	0.35	0.078	2	0.85	0.156	17	2.19	0.08
46	Q07984	Ssr4	NA	2 IN	Translocon-associated protein, delta subunit	0.51	0.071	5	0.89	0.122	9	1.59	0.06

Next Steps in Animal Labeling

- Fully label animal starting with mother pre-pregnancy
- Compare quantitative distribution of membrane proteins in cerebellum, cerebrum, hypothalamus/striatum between fetal and adult brain
- All other organ systems will be available for additional studies

Summary

- Mass spectrometry provides a sensitive and facile method to identify proteins and their modifications.
- Protein Complexes and Large Protein Structures can be readily identified and their organization studied
- Mammalian organisms can be metabolically labeled to near completion with ^{15}N using an isotope-enriched diet
- Metabolic labeling provides a non-biased method to create internal standards for proteomics at the whole animal/tissue level. Enables the use of animal models of disease states

Professor John R. Yates, III
The Scripps Research Institute

Christine Wu, Ph.D.
Scott Anderson
Claire Delahunty, Ph.D.
Laurence Florens, Ph.D.
Johannes Hewel, Ph.D.
Jeff Johnson
Hongbin Liu, Ph.D.
Michael MacCoss, Ph.D.
Hayes McDonald, Ph.D.
Zee-Yong Park, Ph.D.
Rovshan Sadygov, Ph.D.
David Tabb
John Venable, Ph.D.
James Wohlschlegel, Ph.D.
Wenhong Zhu, Ph.D.

National Institutes of Health
R33 CA81665

NIH NCRR
RR11823

American Cancer Society
PF-03-065-01-MGO

The Scripps Research Institute