Advances in analytical biochemistry and systems biology:
Proteomics

Brett Boghigian
Department of Chemical & Biological Engineering
Tufts University
July 29, 2005

Proteomics

- The basics
- History
- Current and developing techniques and technologies
- How we can use proteomics in our research

Genome vs. Proteome

Transcriptional profiles cannot be directly correlated to proteomic profiles

- Genome
  - All of the genetic information and material possessed by an organism
- Proteome
  - The complete profile of proteins expressed in a given tissue, cell, or biological system *at a given time*

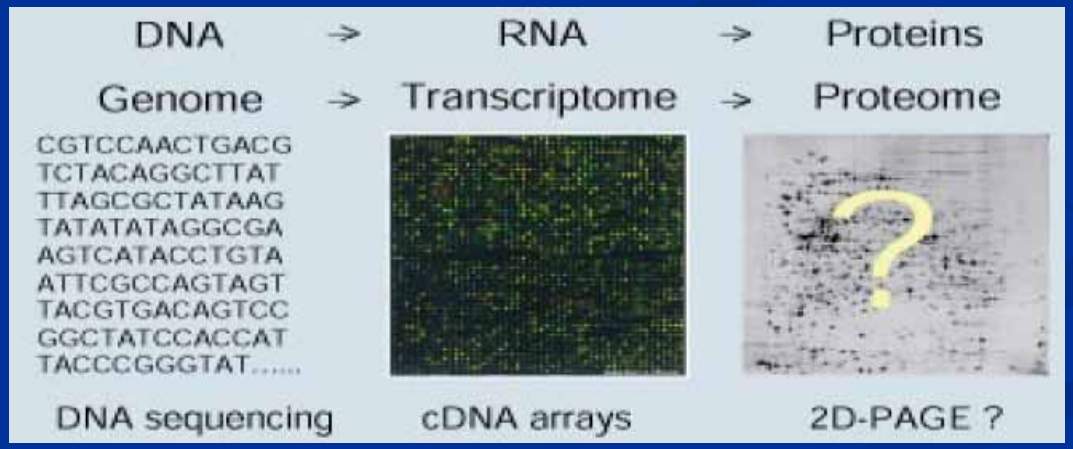


Figure 1 (DNA → RNA → protein)

www.oardc.ohio-state.edu/tomato/Proteomics%20Presentation.pdf

Lee SY, Lee DY, Kim TY. *Trends Biotechnol*. 2005 Jul;23(7):349–58.

TUFTS SCHOOL OF ENGINEERING

Mapping the *genome*

- 1995 – first prokaryote *Haemophilus influence* mapped
- 1996 – first eukaryote *Saccharomyces cervisiae* mapped
- 1998 – first multicellular organism *Caenorhabditis elegans* mapped
- 2000 – first plant *Arabidopsis thaliana* mapped
- 2001 – human mapped

Blomberg A. "Proteomics in Metabolic Engineering." From: *Metabolic Engineering in the Post Genomic Era*. Jan 2004:37–68.

TUFTS SCHOOL OF ENGINEERING

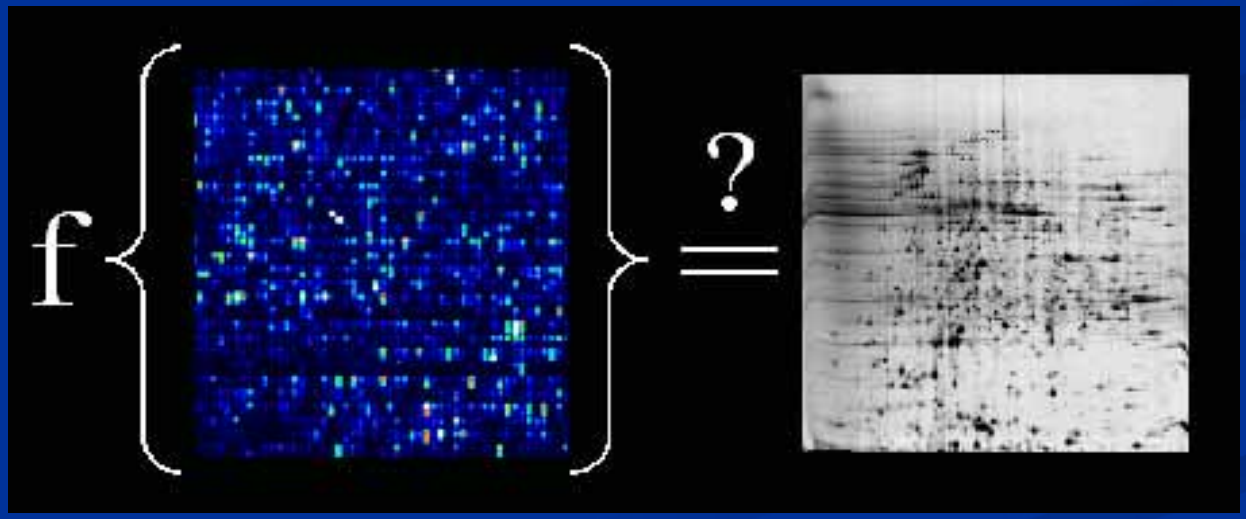
Mapping the *proteome*?

- Genome
  - Essentially static over time
  - Non site-specific
  - Human genome mapped (2001)
    - ~22,000 genes
  - PCR used to amplify DNA
- Proteome
  - Dynamic over time
  - Site-specific
  - Human proteome *not* mapped
    - ~400,000 proteins
  - No equivalent of PCR for proteins

TUFTS SCHOOL OF ENGINEERING

Genome → Proteome

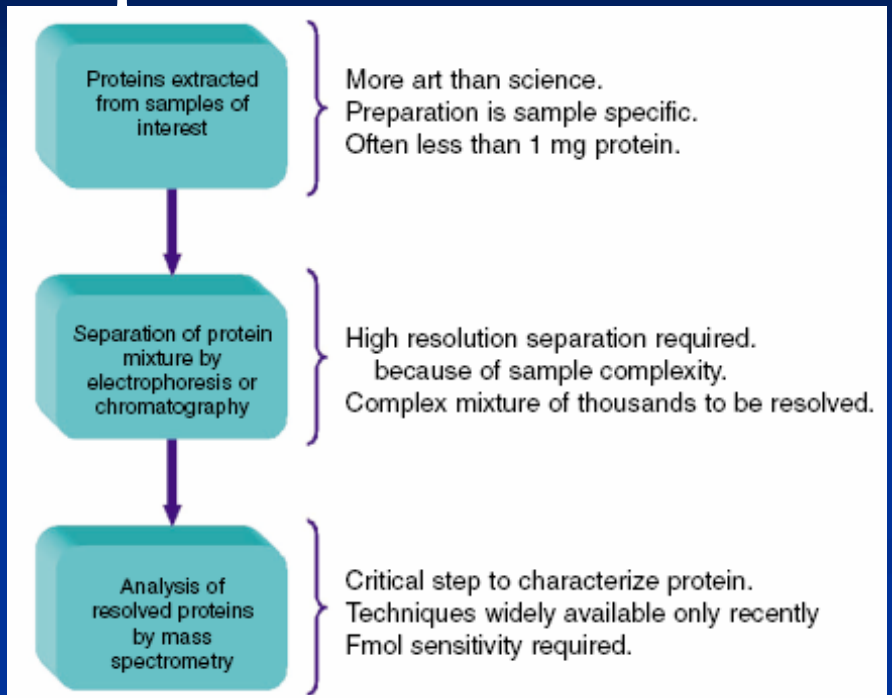
- Large increase in protein diversity due to:
  - Alternative splicing of pre-mRNA (introns and exons)
  - Post-translational modifications (phosphorylation, glycosolation, hydroxylation, etc.) of proteins
  - Cell-age and health



cDNA Array 2D SDS-PAGE

Figure 2 (DNA chip array translates into proteins separated by 2D SDS-PAGE)

The basic proteomics "workflow"



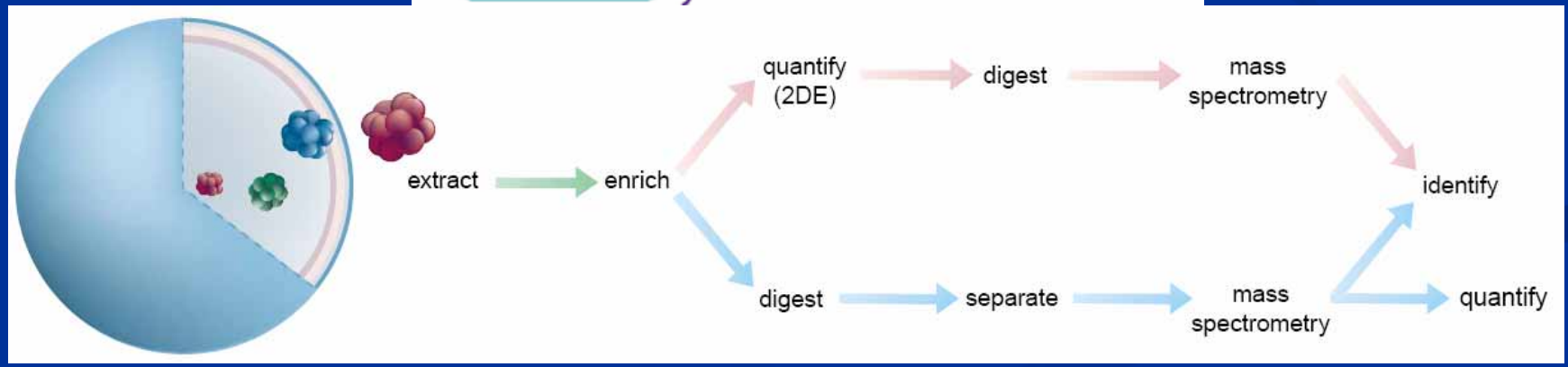


Figure 3 (Basic proteomics workflow)

Lee KH, Reardon KF. *AIChE J*. 2004 Apr;49(11):2682–6.
Patterson SD, Aebersold RH. *Nat Genet*. 2003 Mar;33 Suppl:311─23.

Proteomics

- The basics
- History
- Current and developing techniques and technologies
- How we can use proteomics in our research

History – 2D PAGE

- 1975 – Patrick O'Farrell at CU Boulder
- Two-dimensional polyacrylamide gel electrophoresis
  - Protein separation

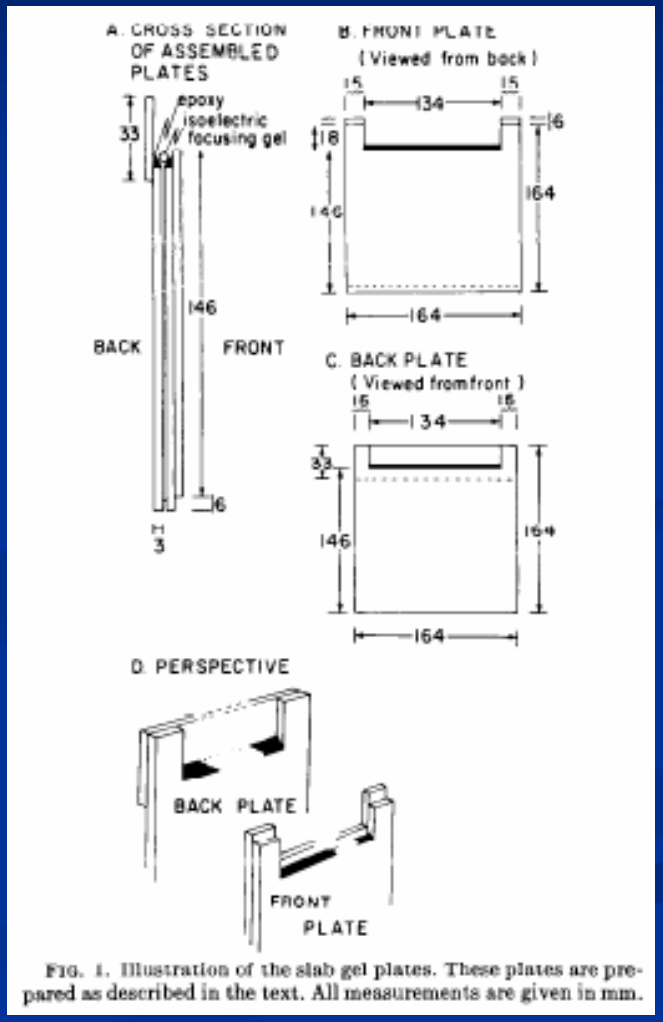
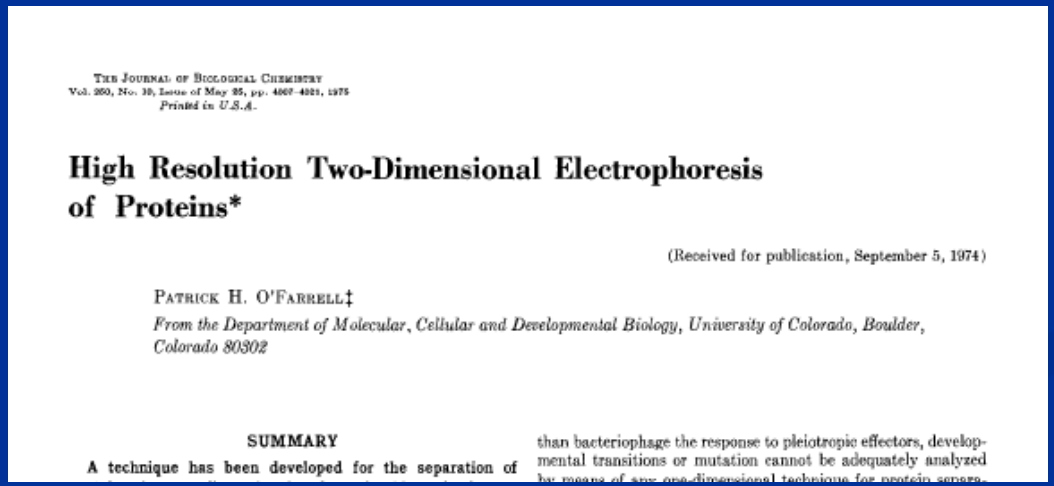


FIG. 1. Illustration of the slab gel plates. These plates are prepared as described in the text. All measurements are given in mm.

Figure 5 (Basic 2-D gel assembly)



THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 250, No. 10, Issue of May 25, pp. 4007–4021, 1975
Printed in U.S.A.

High Resolution Two-Dimensional Electrophoresis of Proteins*

(Received for publication, September 5, 1974)

PATRICK H. O'FARRELL‡
From the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80302

SUMMARY

A technique has been developed for the separation of

than bacteriophage the response to pleiotropic effectors, developmental transitions or mutation cannot be adequately analyzed

Figure 4 (Seminal paper on two-dimensional electrophoresis by O'Farrell)

O'Farrell PH. *J Biol Chem*. 1975 May 25;250(10):4007-21.

History – 2D PAGE

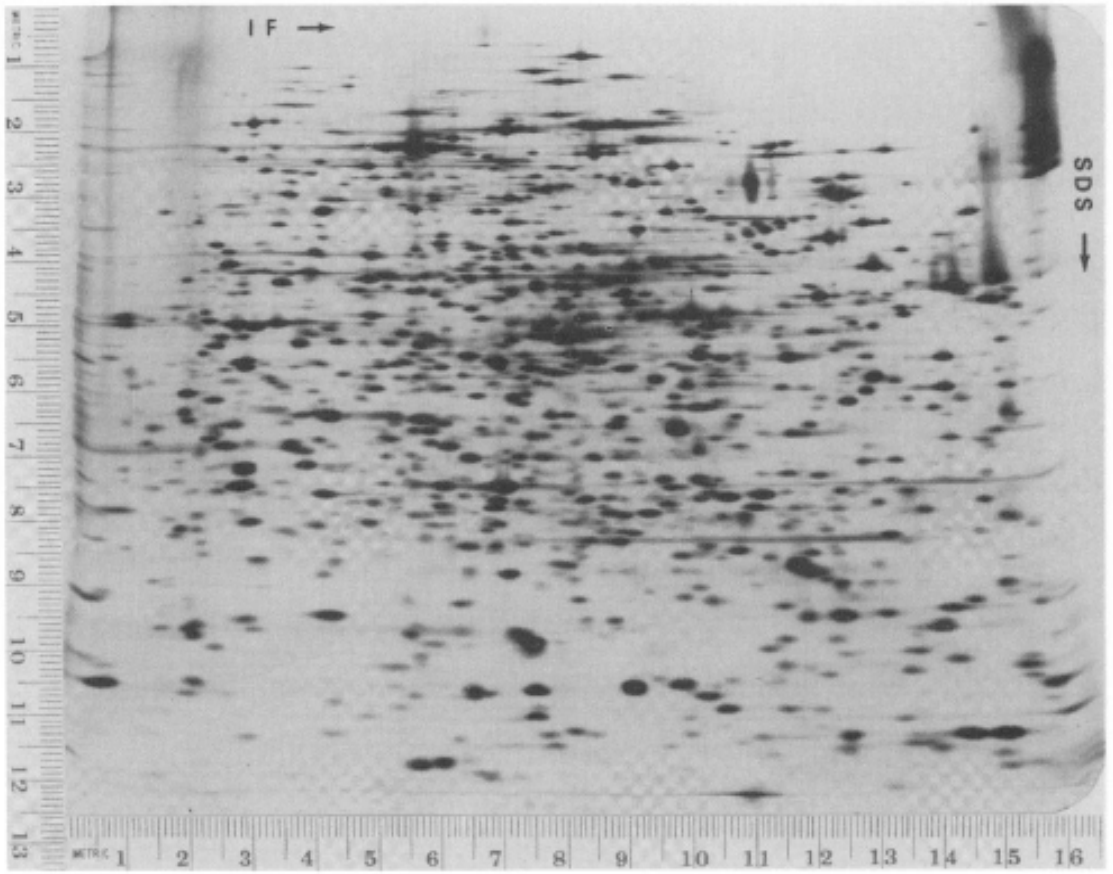


FIG. 2. Separation of *Escherichia coli* proteins. *E. coli* (1100) was labeled with ^{35}C-amino-acids as described under "Materials and Methods." The cells were lysed by sonication, treated with DNase and RNase and dissolved in lysis buffer. Twenty-five microliters of sample containing 180,000 cpm and approximately 10 μg of protein were loaded on the gel. The isoelectric focusing gel was equilibrated for 30 min. The gel in the SDS dimension was a 9.25 to 14.4% exponential acrylamide gradient. A volume of 10 ml of 14.4% acrylamide was used in the front chamber of the gradient mixer. The total volume of the gel was 16 ml. At this exposure, 825 hours, it is possible to count 1000 spots on the original autoradiogram. All autoradiograms of two-dimensional gels were photographed with a metric ruler along two edges of the autoradiogram. These rulers establish a coordinate system which is used to give spot positions. The verticle scale is given in units from top to bottom. The horizontal scale is given in units from left to right. The coordinates are given as horizontal × vertical.

Figure 6 (Separation of *E. coli* proteins)

O'Farrell PH. *J Biol Chem*. 1975 May 25;250(10):4007-21.

History – from protein chemistry to proteomics

- 1980s – advances in molecular biology:
  - Gene cloning
  - Sequencing
  - Expression analysis
- Led to a link between observed activity or function of a single protein and the gene that encoded it → **protein chemistry**

Patterson SD, Aebersold RH. *Nat Genet*. 2003 Mar;33 Suppl:311–23.

History – from protein chemistry to proteomics

TUFTS SCHOOL OF ENGINEERING

- Successful sequencing projects led to the development of "reverse" approaches
  - Focus on phenotype instead of genotype
- Also led to gene sequence databases
- With a wide array of complete gene sequences, why not look at all of an organism's proteins instead of one? → **proteomics**
- Now the only limitation was not correlating genotype to phenotype, but in the ability to sequence proteins and peptides → **mass spectrometry**

History – mass spectrometry

- 1899 – J.J. Thompson publishes first article on mass spectrography
- Principle is that molecules must first be ionized
  - Difficult to do to with larger molecules such as proteins without destroying them
- Late 1980s – advent of electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI) techniques

History – Systems biology technologies

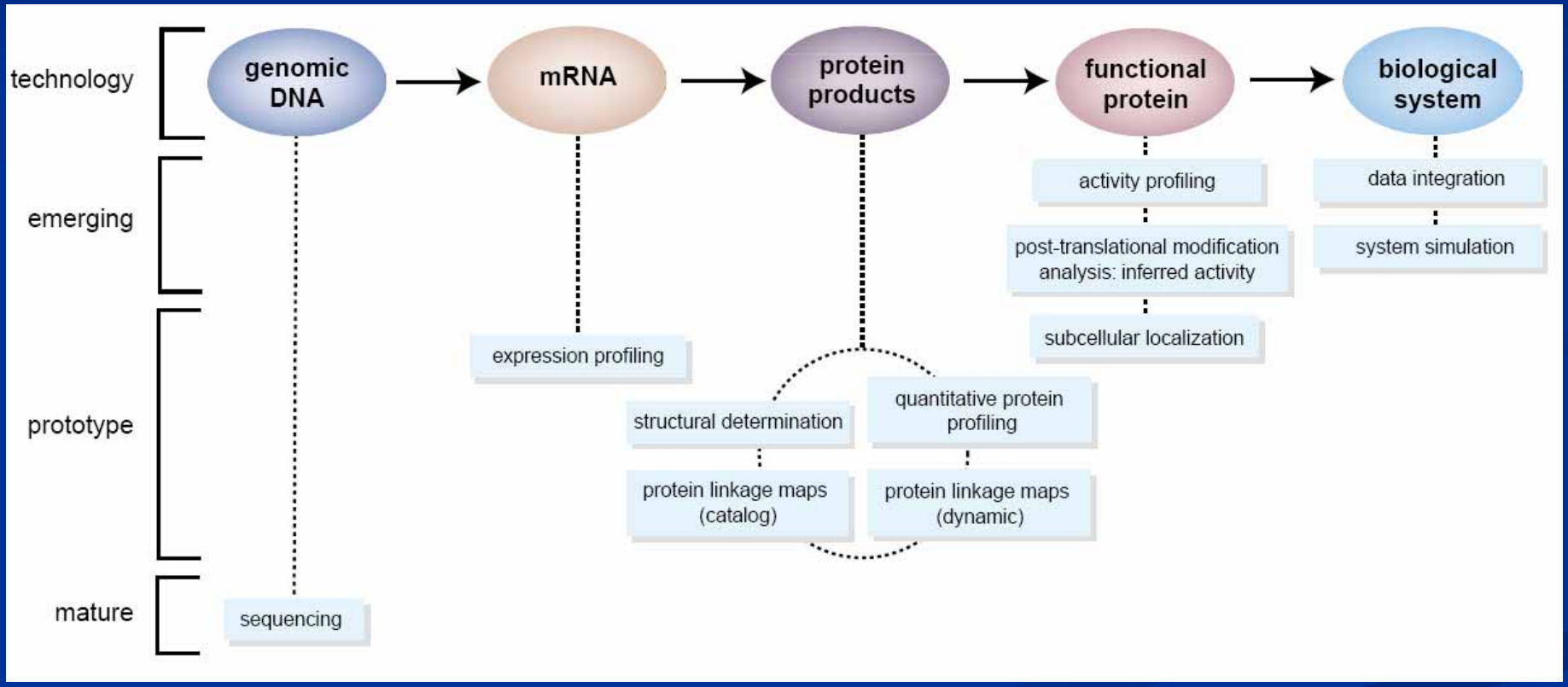


Figure 7 (Systems biology technologies and their relative maturity)

Patterson SD, Aebersold RH. *Nat Genet*. 2003 Mar;33 Suppl:311–23.

History – Proteomics technologies

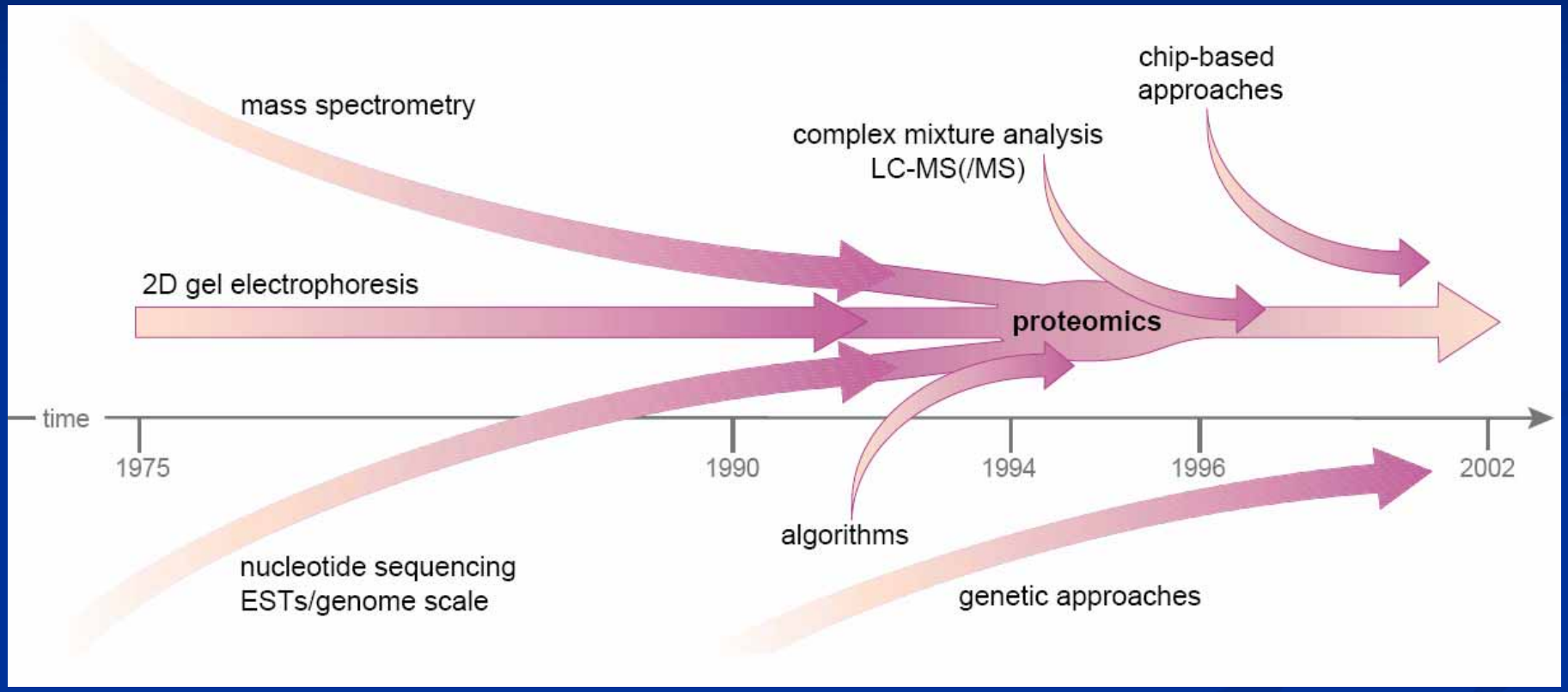


Figure 8 (Timeline of proteomics technologies)

Patterson SD, Aebersold RH. *Nat Genet*. 2003 Mar;33 Suppl:311–23.

Proteomics

- The basics
- History
- Current and developing techniques and technologies
- How we can use proteomics in our research

Gel vs. gel-free methods

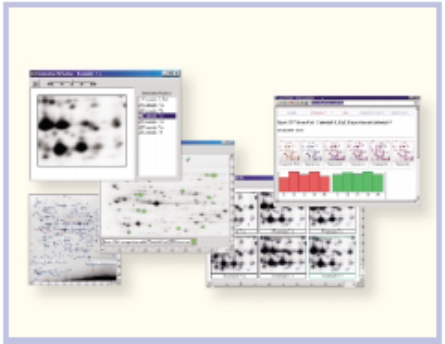
- 2D SDS-PAGE
  - Gold standard in proteomic analysis
  - Very low and very high MW, and very basic and very acidic proteins have limited sensitivity
  - Can detect post-translational modifications with various protein stains
  - Multiple proteins in one spot
- HPLC
  - Developing technology typically used in tandem with ESI-MS/MS

Gel-based proteomic workflow

IMAGING

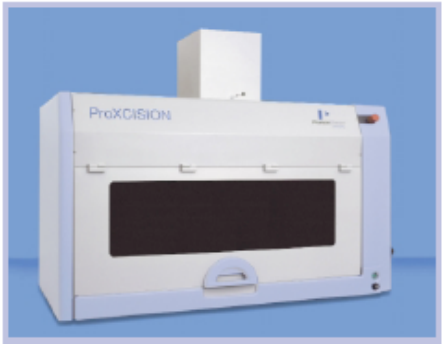
ProXPRESS™ Proteomics Imaging System

ANALYSIS



Progenesis™ 2-D Image Analysis Software

SPOT EXCISION



ProXCISION™ Proteomics Gel Cutting Robot

SAMPLE PREP and TARGET SPOTTING

MultiPROBE® II Proteomics Workstation

TARGET SPOTTING

TOFprep™ MALDI Spotting System

PROTEIN IDENTIFICATION

prOTOF™ 2000 MALDI O-TOF Mass Spectrometer

Figure 9 (PerkinElmer's proteomic workflow and series of instruments)

Step 1: Protein separation

- 2D SDS-PAGE
  - First dimension – separate by charge
  - Second dimension – separate by size

New technologies in gel-based protein separation

- Increased first-dimensional focusing power

Figure 10 (Bio-Rad ReadyStrip IPG Strips)

Figure 11 (Bio-Rad PROTEAN IEF Cell)

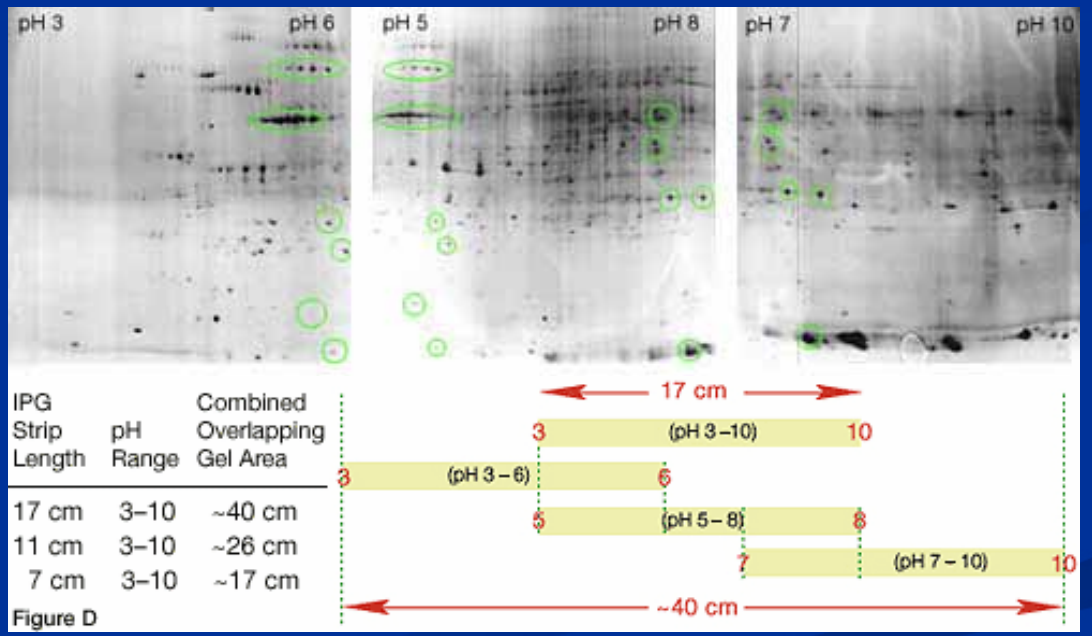


Figure 12 (Bio-Rad ReadyStrip IPG Micro-Scale Strips)

Step 2: Gel dying

- Coomassie Brilliant Blue (CBB) stain

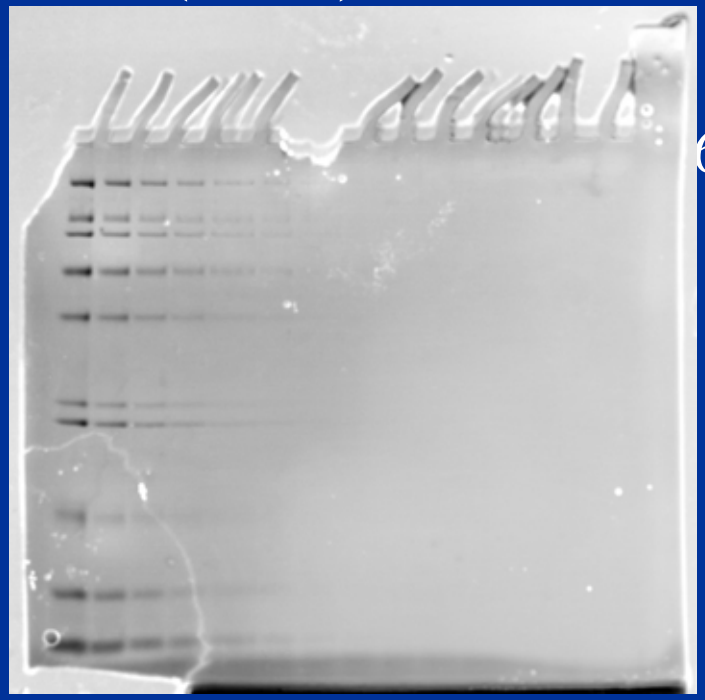


Figure 13 (Pierce GelCode Blue stain)

- Silver stain
  - Complex staining

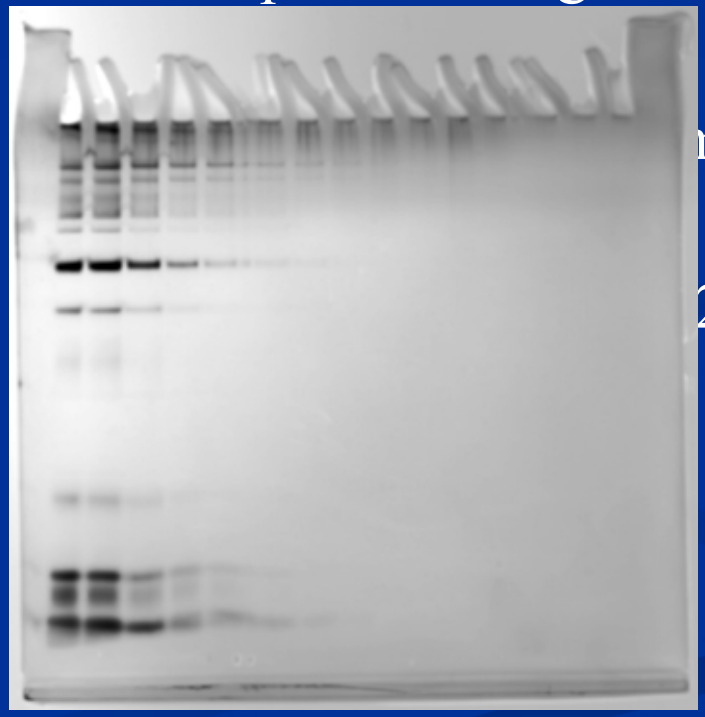


Figure 14 (Bio-Rad Silver stain)

Stehmann C, *et al*. *Proteomics*. 2005. In press.
Berggren K, *et al*. *Electrophoresis*. 2000 Jul;21(12):2509-21.

New technologies in protein gel staining

- Fluorescent stains and multiplexing proteomics
  - SYPRO Ruby (all proteins)
  - Pro-Q Diamond (phosphoproteins)
  - Pro-Q Emerald (glycoproteins)

Figure 15 (1D gels stained with Pro-Q Diamond (left) and SYPRO Ruby (right))

Step 3: Gel imaging and analysis

- Charge-coupled device (CCD) cameras
  - Filters define wavelength diversity
  - Higher resolution but relatively slow

Figure 16 (PerkinElmer ProXPRESS 2D Proteomic Imaging System)

- Laser imaging devices
  - Fixed wavelengths
  - Lower resolution but relatively fast

Figure 17 (Bio-Rad Molecular Imager FX Pro Plus System)

www.bio-rad.com
www.perkinelmer.com

New technologies in gel imaging and analysis

- Three-dimensional gel analysis

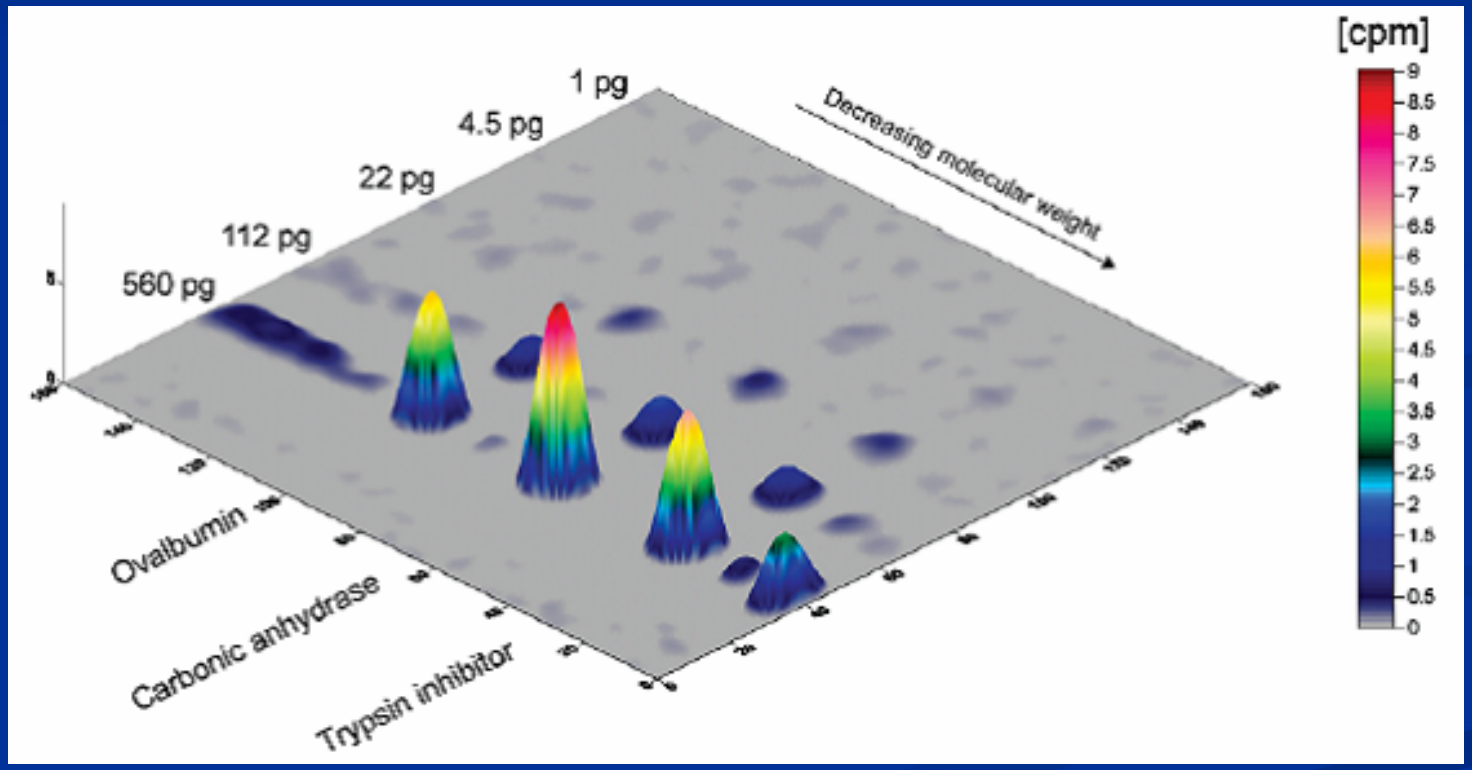


Figure 18 (3-D image analysis of a 1-D gel)

Kleiner O, *et al*. *Proteomics*. 2005 Jun;5(9):2322-30.

New technologies in gel imaging and analysis

- Three-dimensional gel analysis

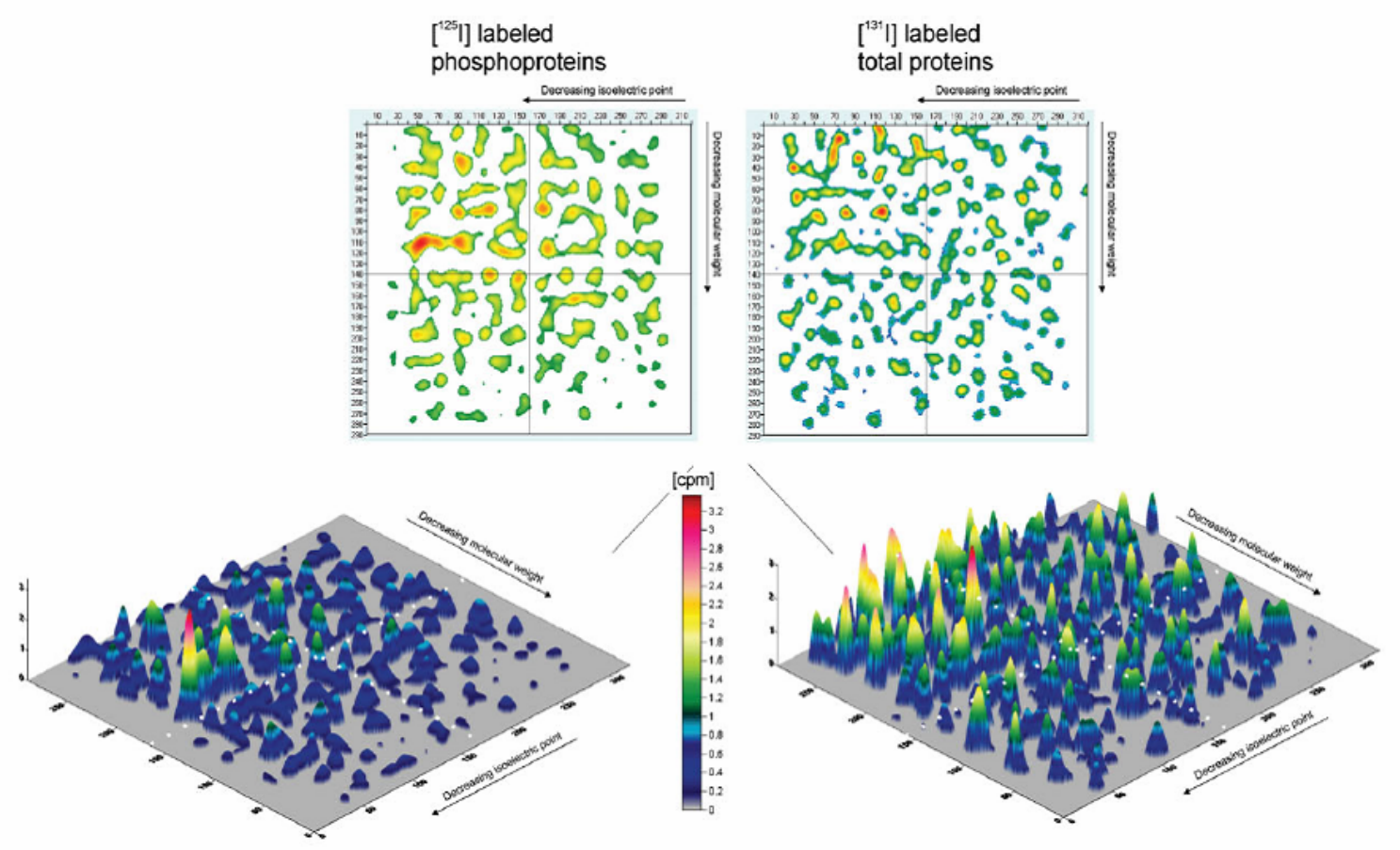


Figure 19 (3-D image analysis of multiplexed 2-D gels)

Kleiner O, *et al*. *Proteomics*. 2005 Jun;5(9):2322-30.

TUFTS SCHOOL OF ENGINEERING

Step 4: Gel-spot excision

- Cut protein spots from 1-D and 2-D gels for subsequent mass spectrometry analysis
- Typically done manually
  - Extremely labor intensive and time-consuming
  - May miss some of a spot and therefore miss protein
    - Disrupt quantitative-intensive experiments

New technologies in gel-spot excision

- Automation
  - Some systems have a CCD camera on board
  - Use in tandem with a standalone imager

Figure 20 (Genomic Solutions Investigator ProPic II)

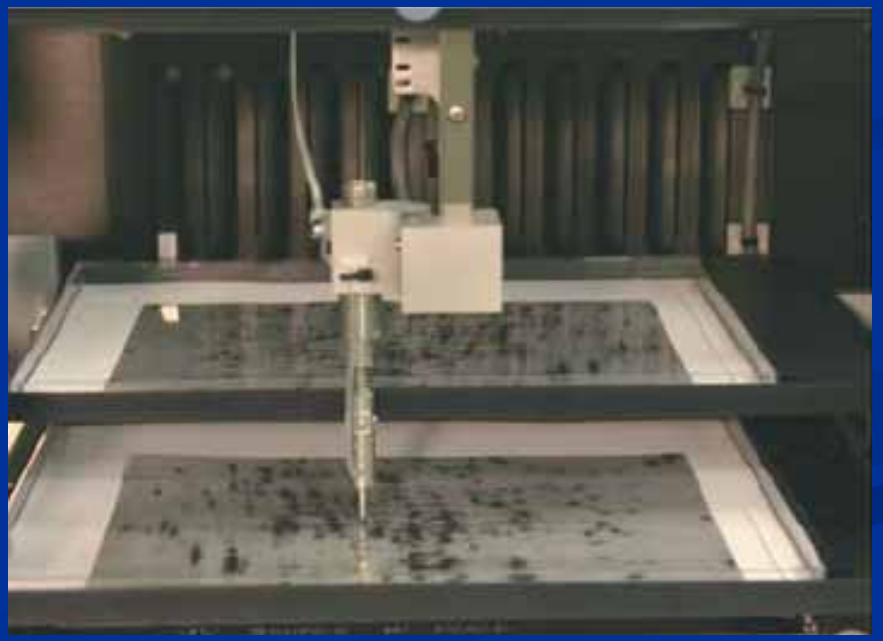


Figure 21 (Robot cutting two, 2-D gels)

Step 5: Sample digestion and plating

- Typically done manually
  - Extremely labor intensive and time-consuming
  - Large problem with contamination

New technologies in sample digestion and plating

- Automation
  - Saves time and manpower

Figure 22 (Genomic Solutions Investigator ProPrep)

Step 6: Mass spectrometry

- Identify and quantify proteins and peptides

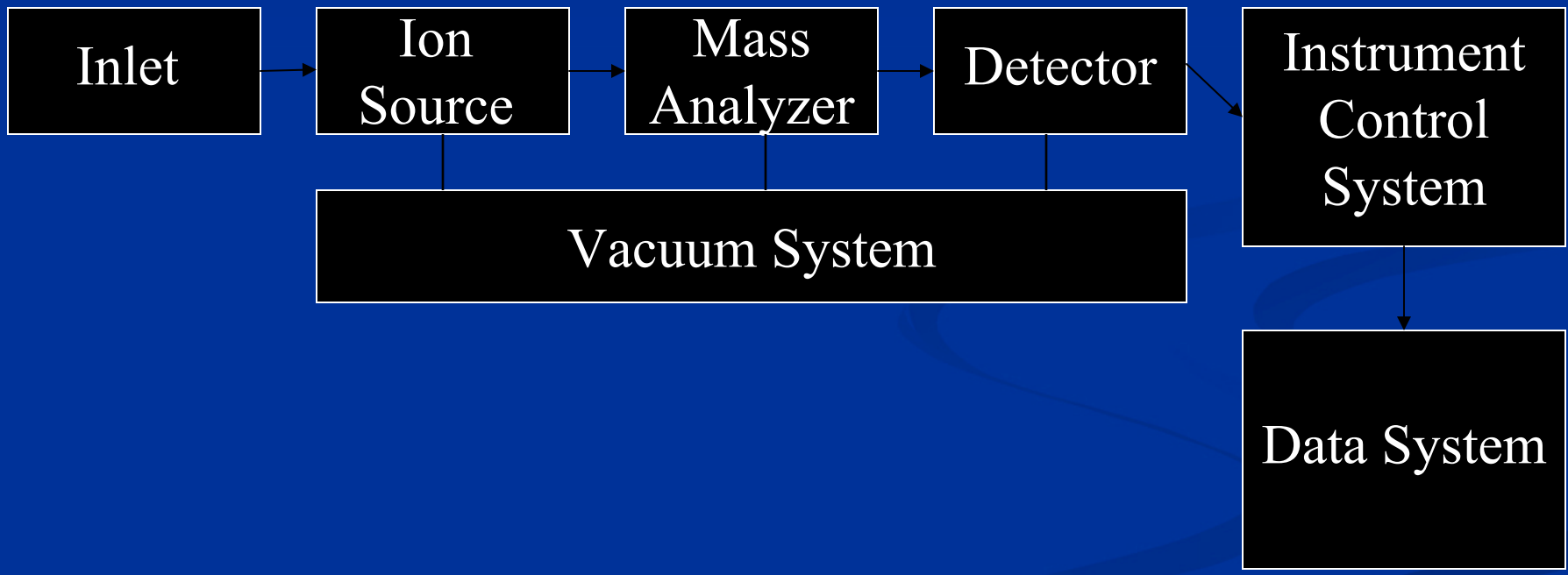


Figure 23 (General flow diagram of a mass spectrometer)

Step 6: Mass spectrometry

- Ion sources (ESI and MALDI) are coupled with mass analyzers
  - MALDI is commonly coupled with a time-of-flight (TOF) analyzer
  - ESI is commonly coupled with a ion-trap or quadrupole analyzer

MALDI-TOF MS

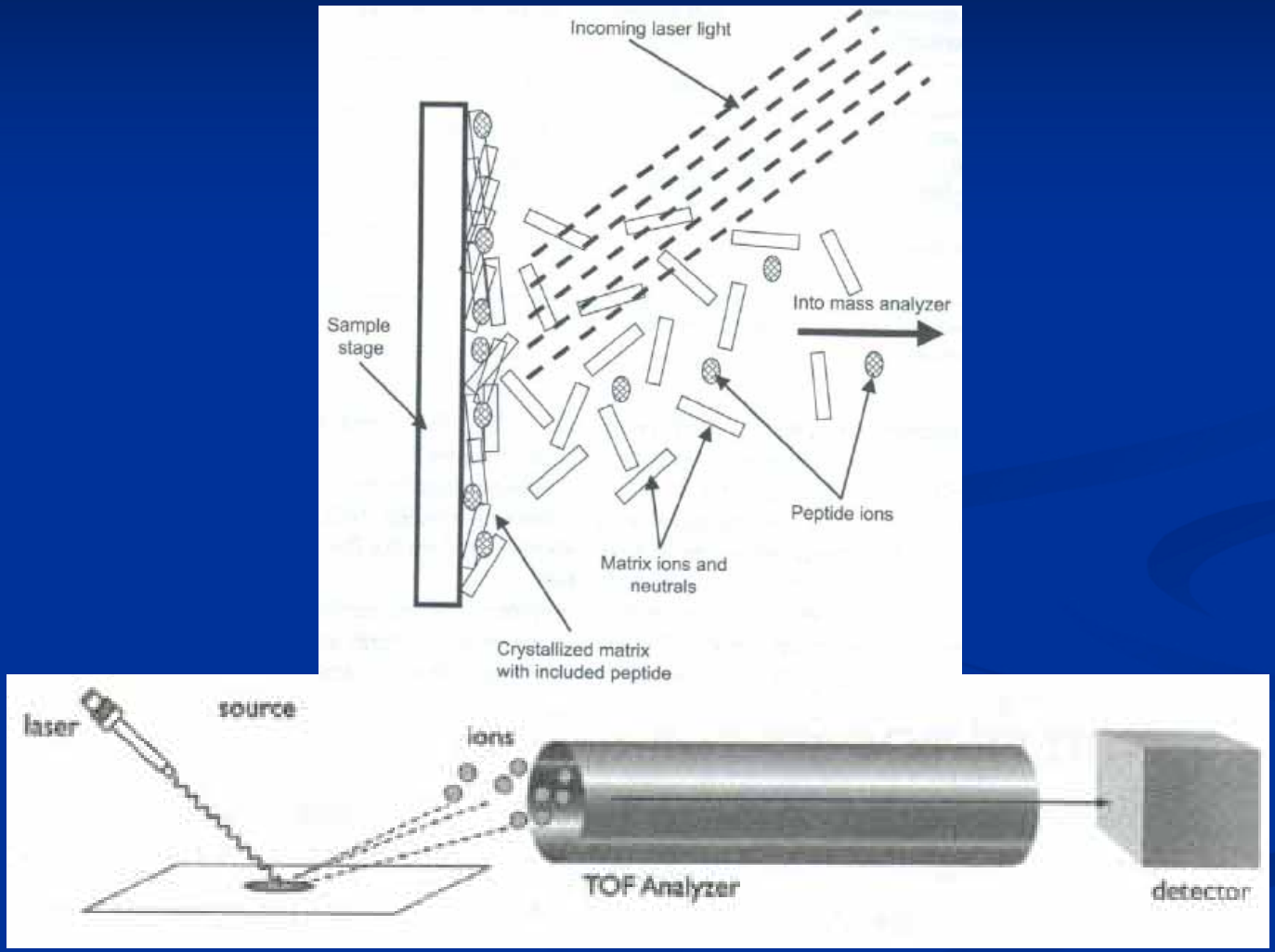


Figure 24 (MALDI ion source and TOF mass analyzer)

www.oardc.ohio-state.edu/tomato/Proteomics%20Presentation.pdf

ESI Quadrupole MS

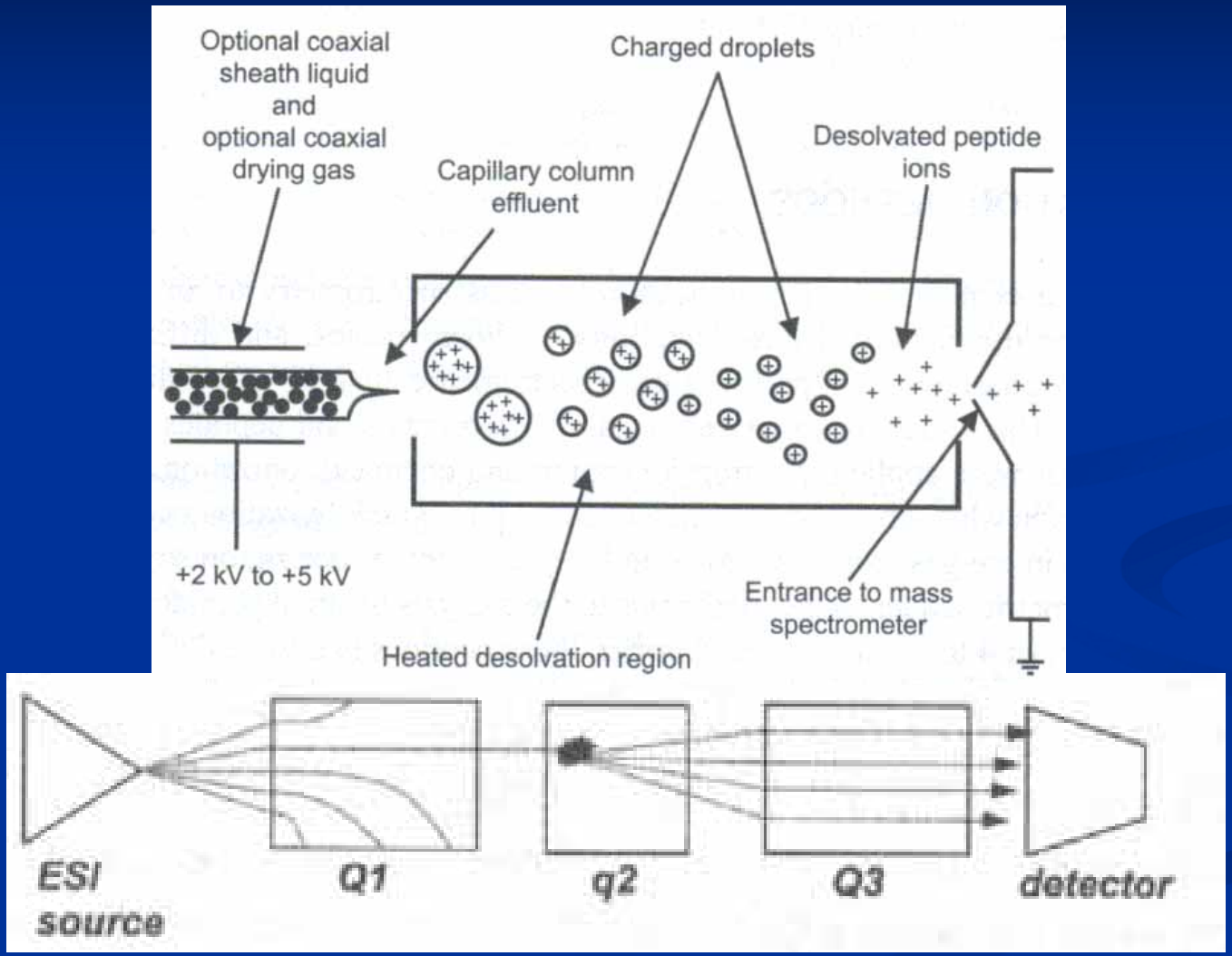


Figure 25 (ESI ion source and quadrupole mass analyzer)

www.oardc.ohio-state.edu/tomato/Proteomics%20Presentation.pdf

MALDI-TOF vs. ESI-quadrupole

- MALDI-TOF
  - Sample on a slide
  - Spectra generate masses of peptide ions
  - Protein identification by peptide mass fingerprinting
  - Expensive
  - Good for sequenced genomes
- ESI-quadrupole
  - Sample in solution
  - MS-MS spectra reveal fragmentation patterns – amino acid sequence data possible
  - Protein identification by cross-correlation algorithms
  - Very Expensive
  - Good for unsequenced genomes

www.oardc.ohio-state.edu/tomato/Proteomics%20Presentation.pdf

TUFTS SCHOOL OF ENGINEERING

Step 7: Protein databases

- **SWISS-PROT** – database of annotated protein sequences. Contains additional information on protein function, protein domains, known post-translational modifications, etc. (http://us.expasy.org/sprot)
- **TrEMBL** – computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.
- **PIR-International** –annotated database of protein sequences. (http://www-nbrf.georgetown.edu/)
- **NCBInr** – translated GenBank DNA sequences, Swiss-Prot, PIR.
- **ESTdb** – expressed sequence tag database (NIH/NSF)
- **UniProt** – proposed new database. Will join Swiss-Prot, TrEMBL, PIR. http://pir.georgetown.edu/uniprot/

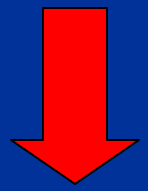
www.oardc.ohio-state.edu/tomato/Proteomics%20Presentation.pdf

Proteomics

- The basics
- History
- Current and developing techniques and technologies
- How we can use proteomics in our research

Applying proteomic techniques to metabolic engineering

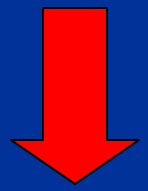
- **Advantage** – 2DE provides excellent quantitative results on protein expression
- Better resolution than running 1D gels
  - DEBS1 (370 kDa), DEBS2 (380 kDa), and DEBS3 (332 kDa) are large proteins; therefore, difficult to view on 1-D gels



- **Specific Aim 1** – correlate DEBS protein expression under standard culture conditions to 6dEB production

Applying proteomic techniques to metabolic engineering

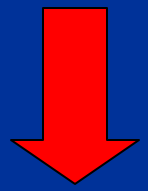
- **Advantage** – MS provides extremely sensitive measurements based on *m/z* ratio
- Detect proteins (DEBS) and organics (6dEB) on very low levels
  - Much more sensitive than HPLC



- **Specific Aim 2** – use MALDI-TOF MS or LC-MS/MS to test viability of MS instruments on our system

Applying proteomic techniques to metabolic engineering

- **Advantage** – fluorescent multiplexing techniques provide information on post-translationally modified proteins
- Using specific stains, view phosphorylated and glycosolated proteins



- **Specific Aim 3** – after 2-D protein maps have been generated, run again with different stains to multiplex images

TUFTS
SCHOOL OF ENGINEERING

Questions?