

Proteomics

Ian Musselman
latex: Dan Simola

January 2005

Overview

- Proteome: The entire protein complement expressed by a genome of a cell or organism.
- Genomes are static, proteomes are highly dynamic
- Proteomics: Any large scale or systematic characterization of the proteome of a cell, tissue or simple organism.
- Global Proteomics: Attempt to detect (composition) or quantitatively compare (protein profiling) all proteins present in a cell, tissue or organism
 - Feasible for simple organisms
 - Only get a small portion of total proteins in higher organisms
 - Challenge is to interpret significance and determine changes that correlate with biology/disease
- Target Proteomics Reproducibly isolate and characterize a subproteome, usually within range of protein profiling capacities for mammalian cells and tissues
 - Organelles, complexes and protein groups
 - Keys: Reproducibility and distinguishing non-specific interactions vs moderate affinity interactions
 - Simpler data sets, question is more focused.
- Types of protein studies
 - Protein Profiling: quan. Comparisons of 2+ samples
 - Protein Composition: id of all components of a complex/organelle
 - Protein interactions: define interacting proteins
- Critical advances: Complete genomes, MS, Bioinformatics, Protein separation techs, improved robotics, advances in antibody arrays
- Applications
 - Discover genes implicated in disease mechanisms and therapeutics
 - Discover biomarkers: diagnostics, stratify patients, monitor therapies/toxicity
 - Id new drug targets: since most are proteins
 - Systems biology

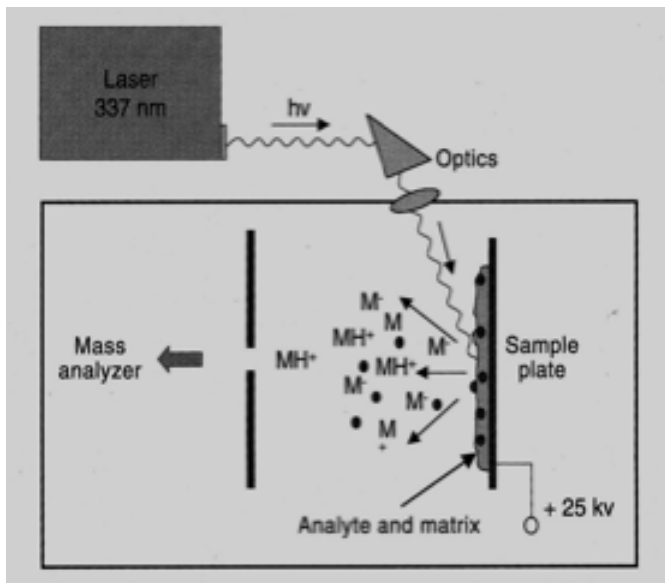
Protein Profiling

- Tools: 2D gels, LC/LC-MSMS, antibody arrays, nucleotide microarrays
- 2D gel
 1. run sample on gel/strip separating by isoelectric point (P_I)
 2. run on SDS gel – this must be second because SDS is denaturing and gives sample a negative charge
- Advantages over nucleic acid microarrays:
 - Protein determines phenotype
 - mRNA levels do not frequently correlate with protein levels
 - Post translational modifications
 - Functions of many proteins are affected by changes in subcellular location
- Disadvantages
 - No method to amplify proteins
 - Less sensitive
 - Existing methods can screen only a small portion of mammalian proteomes
 - Low throughput, minimal automation
- Similar challenges
 - Mine databases for knowledge of identified proteins/genes
 - Determine which changes are interesting
 - Validate discovery by independent methods
 - Integration of knowledge (systems biology)
- - Scope of human proteome comparisons problem
 - 35 000 genes → alt splicing of mRNA 200 000 → PTMS ~10 unique species/gene → total unique protein components = 2 million
 - Dynamic range is 100 to 10^6 copies per cell
 - ~20 000 components per cell, each cell expresses a subset of the full complement
 - Physiological variation
 - Secondary effects include inflammation, oxidative stress, proteolysis
- 2D gels → IEF followed by SDS-PAGE (MW)
 - Technically difficult, variable, not able to automate
 - Dynamic Range 20X to 1000X
 - Resolves around 2000 spots per gel (MAX)
- To improve resolution: (prefractionation)
 - Chromatography, organelle, isolation, extraction/solubilization
 - IEF fractionation (ZoomIEF)
 - Follow this with 2D or 1D gels and MS
- Proteomics of Biological fluids
 - Cells leak proteins into fluids
 - Large dynamic range, a small number of uninteresting proteins overwhelm interesting ones
 - Can deplete these proteins to see low copy proteins (Ab columns)

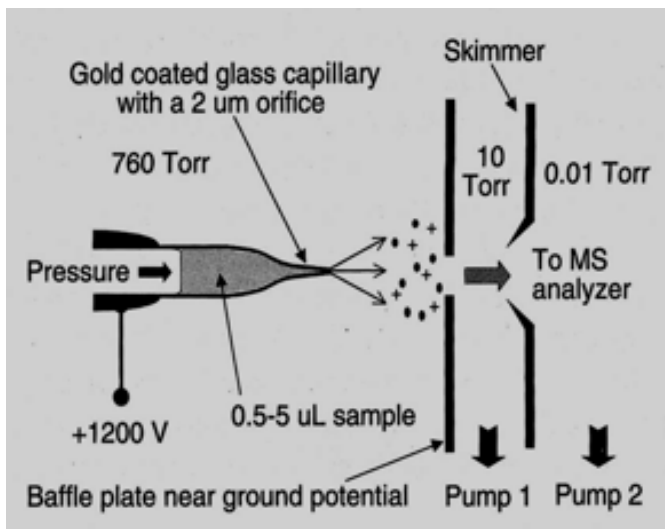
- * top 6 filtering: albumen, IgG, apolipoprotein, ...
- Or add additional levels of separation
- Cancer biomarkers
 - Would greatly enhance disease discovery (early stage diagnosis key)
 - Current markers all were found by indirect methods, > 35 kDa
 - High throughput options needed for discovery based methods
- SELDI: Solution profiling
 - Chemically enhanced surfaces with a low powered MALDI device
 - Surfaces allow more complex mixtures, direct sample application
 - Challenges: pattern recognition, quantitative comparisons
 - * Comparing complex patterns
 - * Data integration: proteomic and genomic data

Mass Spec

- MALDI Matrix Assisted Laser Desorption/Ionization
 - Tolerant of sample solution, fast (sec-min/sample), reasonable mass accuracy, $M_n < 15\text{kDa}$



- Electrospray Interfaces (ESI)
 - Low tolerance of salts, buffers, etc., only very clean liquids or direct interface with Reverse phase column, uses quadrupoles



- Performance
 - * Resolution
 - * Mass Accuracy
 - * Sensitivity

MS vs. MS/MS vs. MS_n

- MS = measure intact masses only, proteins, peptides, etc. as introduced into instrument
- MS/MS or MS₂ = mass of intact component → fragment in mass spectrometer → measure masses of fragments (daughter ions)
- MS/MS_n = continued isolation and subsequent fragmentation of daughter ions
- Proteins are typically identified using a mass fingerprint of tryptically digested proteins
- Tandem MS: use tryptic peptide masses followed by fragmentation with inert gas and measure masses of daughter ions
- Steps in protein ID
 1. Excise protein band from gel
 2. Digest with trypsin
 3. Use liquid chromatography (LC) to separate peptides
 4. Analyze with MS
 5. Search database for fingerprint of peptide
 6. THIS IS NOT SEQUENCING