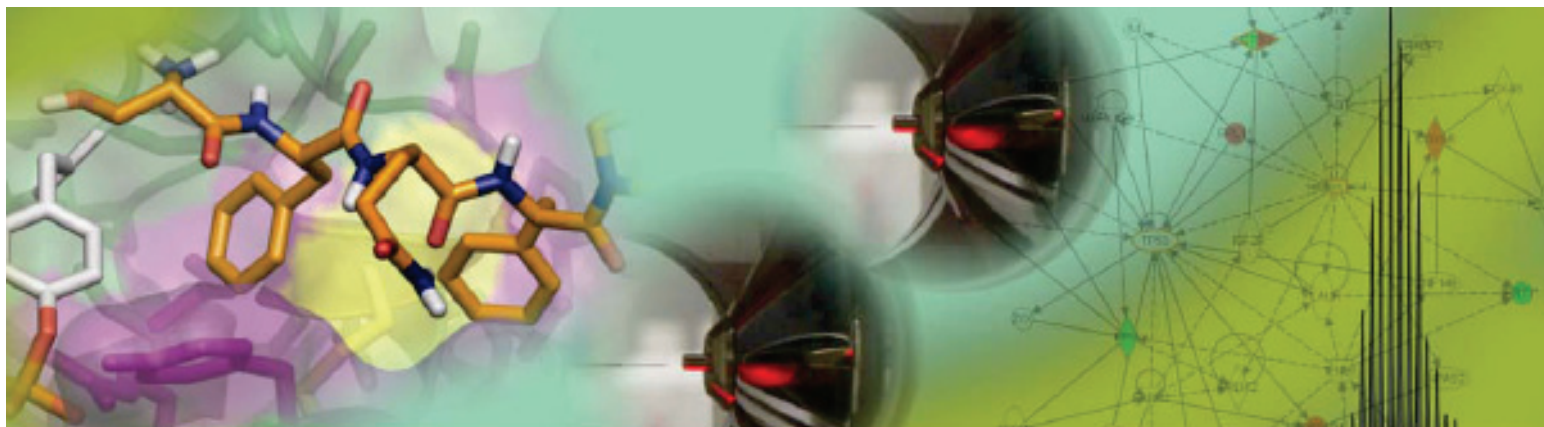


W.M. KECK FOUNDATION BIOTECHNOLOGY RESOURCE LABORATORY



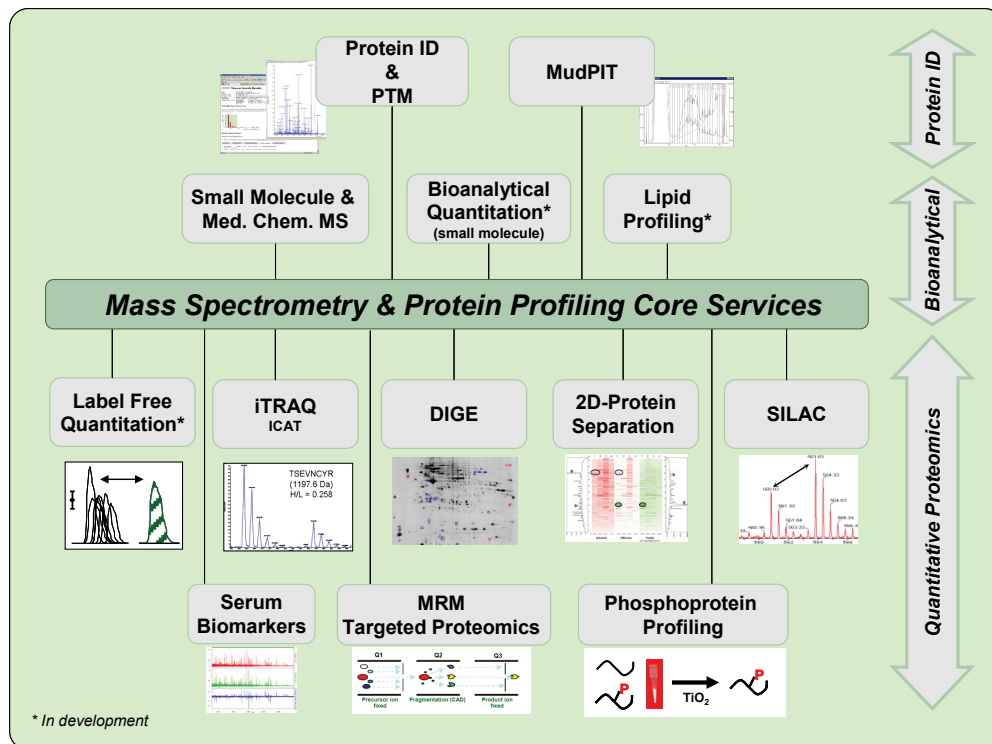
MS & Proteomics Resource

Resource Description

The Mass Spectrometry (MS) and Proteomics Resource of the W.M. Keck Foundation Biotechnology Resource Laboratory strives to provide investigators with access to the most advanced protein profiling technologies. These technologies are implemented with a broad spectrum of mass spectrometry based techniques and protein/peptide chemistries to separate, characterize and quantify analytes from complex biological samples. The resource has 11 staff including 4 with Ph.D. level appointments, with well over 100 years of mass spectrometry and protein chemistry experience. It is equipped with 9 state-of-the-art diverse tandem mass spectrometer systems including TOF-TOF, Quadrupole-TOF, Triple-Quadrupole, FT-ICR, and LTQ-Orbitrap type analyzers with either matrix assisted laser desorption ionization (MALDI) or Electrospray ionization (ESI) sources coupled to HPLC systems.

Given these diverse technology platforms, the resource supports a wide array of scientific research from both Yale and external investigators. Samples for MS analysis includes proteins, oligonucleotides, lipids,

carbohydrates, synthetic peptides, and a wide range of small molecules including chemical intermediates. In FY 2008, for example, the MS and Proteomics Resource, with its dedicated staff and extensive knowledgebase, carried out 11,528 MS



Schematic outline of all the mass spectrometry and proteomics technologies offered by the Keck Biotechnology Resource Laboratory's MS&Proteomics Resource

analyses for investigators from more than 100 institutions in 10 Countries ranging from government laboratories to academic and private research institutions.

To address the broad range of complexity inherent in many proteomics projects, the Keck Laboratory offers a comprehensive array of instrumentation and services, and if necessary can develop complementary methods to achieve successful sample analysis. The methods generally follow two mainstream workflows. One workflow uses a gel based Quantitative or Differential 2D gel electrophoresis (DIGE) combined with tandem mass spectral identification of proteins, and another, uses an automated “bottom-up” or “shot-gun” tandem mass spectral peptide identification approach. Both of these approaches can be quantitative with chemical and stable-isotope tagging of proteins or peptides (e.g., ICAT, iTRAQ, SILAC). In addition, two new cutting edge MS based technology services involving phosphoproteome profiling, and absolute quantitative (“targeted”) proteomic analysis of multiple pre-selected biomarker proteins have been recently developed and being implemented.

Finally, a web based platform called Yale Protein Expression Database (YPED) is utilized for users to integrate, analyze, and visualize their results. The brief descriptions of all the MS and proteomics services provided by the MS and Proteomics Resource can be found below.

List of services:

A. Protein Identification:

- 1. Peptide sequence** based Protein Identification by tandem mass spectrometry (LC-MS/MS) from Immunoprecipitated, or SDS gel purified protein sample preparations.
- 2. MudPIT:** Multi-Dimensional Protein Identification Technology
Peptide based Shotgun or bottom-up tandem mass spectrometric analysis approach to identifying hundreds of proteins isolated from complex samples like cells, tissues, etc.
- 3. Top-down/bottom-up approach for protein identification:** Highly purified proteins (either intact or enzymatically digested) are identified (without LC separation) and post translational modifications are characterized on an ultra-high mass resolving power ($m/\Delta m_{50\%} > 400,000$) and ultra-high mass accuracy (1 ppm or better for internally calibrated spectra) FT-ICR mass spectrometer platform using specialized protein fragmentation techniques like Electron Capture Dissociation (ECD) and InfraRed Multi-Photon Dissociation (IRMPD).

B. Protein Profiling: Relative Quantitation of Protein Expression

- 1. DIGE:** Differential Fluorescence (2D) Gel Electrophoresis
 - a. Up to 3 samples can be compared in the same gel, and any number of samples may be compared across gels in a multiplexed experiment. Typically, differential expression ratios are obtained on 1,000 or more protein spots.
 - b. New AB Model 4800 MALDI ToF-ToF identifies one or more proteins in >85% of spots that are automatically selected via user-defined criterion (e.g., differential expression between a control versus cancer sample of >1.3-fold) and subjected to robotic spot excision, trypsin digestion, and MALDI-MS/MS analysis
 - c. Protein identification is based on highly accurate masses *and* MS/MS sequence data from *multiple* tryptic peptides (e.g., 5-10 peptides/protein is common)
- 2. iTRAQ:** Isobaric tagging technology for relative and absolute quantitation (NEW)
 - a. Quantify differences in protein expression for up to 8 different samples in parallel.
 - b. All protein samples are digested and peptides are tagged with up to 8 different iTRAQ labels before mixing. Peptides are both identified and quantified with the same experiment by multi dimensional LC and MS/MS
 - c. Quantitation is done at the MS/MS spectral level, and typically 400 or more identified proteins from a crude cellular or tissue extract are quantified.
- 3. SILAC:** Stable isotope labeling by amino acids in cell culture (NEW)
 - a. Quantify differences in protein expression from up to 2 or 3 samples from cells. Cells are grown with amino acids (i.e. Arg and Lys) that are incorporated with heavy stable isotopes.
 - b. Ideal for quantitation of over expressed proteins, especially phosphoproteins IP'd from cells. Larger scale SILAC experiments can also be done.
 - c. Samples can be mixed immediately after the cells are grown. As a result, downstream processes would all be subjected to the same magnitude of experimental variations, reducing the error on the final intensity ratios of the samples analyzed.
 - d. Tissue samples cannot be compared
- 4. ICAT:** Isotope Coded Affinity Tags
 - a. A pioneering gel-free MS based technique that is used to compare up to 2 samples.

- b. Quantitation is based on LC-MS peak areas of stable isotope profiles of cysteine-containing peptides derived from tryptic digests of protein extracts.
 - c. Not as widely used technique in our laboratory since the introduction of other profiling methods.
- 5. Label free Quantitation** using FT-ICR MS platform (NEW)
- a. A newly developed nanoLC MS based profiling/quantitation method without the need for labeling. Quantitation can be performed for large number of samples; however, statistically significant numbers of technical repeats are necessary.
 - b. Wide range of applicability to any comparative samples; especially ideal for human clinical samples
 - c. Can be utilized for biomarker discovery.
 - d. Simpler and less expensive sample preparation
 - e. Data analysis can be statistically demanding, and requires the use of our Biostatistics Core
- 6. PF2D: 2 Dimensional Chromatofocusing/Reverse Phase HPLC Protein Profiling**
- a. Quantitation is based on UV absorption of fractionated samples after 2D chromatofocusing (pI) and non-porous Reversed Phase LC separation of proteins in the Beckman-Coulter Proteomelab PF2D System.
 - b. Proteins from the collected fractions can also be identified after tryptic digestion and LC-MS/MS analysis.
- C. Protein Profiling: Absolute Quantitation of Protein Expression (Targeted Proteomics):**
1. a method also called absolute quantitation (AQUA) of peptides based on the quantitatively calibrated measurements of their synthetic stable isotope incorporated peptides
 2. Carried out on a newly purchased triple Quadrupole mass spectrometer (ABI QTRAP 4000) and based on an acquisition technique called Multiple Reaction Monitoring (MRM) of 40-50 *pre-selected* tryptic peptides from potential protein disease biomarkers.
 3. The dynamic range for protein quantitation can be 5 orders of magnitude with limits of detection 10 times higher than currently available LC-MS/MS techniques.
- D. MS based Phosphoproteomics**
1. Phosphopeptide Enrichment: The first step requirement to most MS based phosphoproteomic analysis. Currently our core uses several approaches. One approach is based on the use of strong cation exchange chromatography (SCX) to enrich and fractionate for early eluting phosphopeptides and another very robust approach uses TiO₂ as the affinity support.
 2. Identification of sites of phosphorylation: Allows LC-MS/MS analysis of enriched phosphopeptides from affinity purified phosphoproteins for the discovery or verification of sites of phosphorylation.
 3. Large Scale phosphoprotein Profiling: Uses both SCX fractionation and phosphopeptide enrichment methodology to identify large number of phosphopeptides from complex protein extracts using our newly acquired Thermo Fisher LTQ-Orbitrap mass spectrometer.
 4. Relative quantitation of phosphorylation: Using already established quantitative protein profiling techniques like SILAC, iTRAQ and Label free Quantitation, relative abundance of phosphorylation between different samples can be compared
- E. Bioanalytical Quantitation of Small Molecules**
1. Determination of concentration of small molecules in biological matrices using triple quadrupole mass spectrometer.
 2. The molecules of interest could be drugs, metabolites, etc.
 3. For each compound of interest sample extraction and LC-MS/MS methods are developed.
- F. MS based small molecule profiling using direct infusion and LC FT-ICR MS**
1. Relative quantitation can be performed for large number of clinically relevant samples via LC MS approach. It's similar to label free quantitation for proteins, except metabolites/small molecules instead of the proteins are extracted from the samples.
 2. Data analysis can be statistically demanding
 3. Exact mass determination provides potential elemental composition of small molecular species. Subsequent MS/MS analysis utilizing FT-ICR mass spectrometry can reveal structural information for the small molecules.
- G. Other Mass Spectrometry Based Services:**
1. Exact mass determination of small molecular weight compounds using FT-ICR mass spectrometry for determining elemental composition.
 2. Use of multi-fragmentation techniques of FT-ICR for small molecule to determine structural information.
 3. Ultra high resolution FT-ICR for mass spectral separation of compounds with mass difference of <1mDa in a complex mixture.
 4. Routine molecular weight determination of synthesis products, oligos & other biomolecules

H. **Data Analysis Services**

1. Yale Protein Expressions Database (YPED) addresses the storage, retrieval, and integrated analysis of proteomics data generated by the Proteomics Core Facility. YPED is Web-accessible and currently handles sample requisition, result reporting and sample comparison for ICAT, iTRAQ, DIGE and MUDPIT samples. Researchers can view, subset and download their data through a secure Web interface.
2. X!Tandem is an improved method for running X!tandem protein identification algorithm in a parallel computing environment. Subsequent results can be analyzed through the Trans-Proteomic pipeline and imported in YPED.

- I. **Protein identification tools**, such as MASCOT (cluster version), Sequest (cluster version), and ProteinPilot are all available to users for database searching. The results from any of these programs can be imported into YPED.

Sample preparation requirements:

Protein extracts as dried pellets, in solution, or in gel format are accepted. For buffer restrictions that interfere with MS based analyses please consult our staff listed in the contact information section.

Turnaround time:

Turn around times vary from 2 days to 6 weeks depending on the nature of analysis and the scope of the project. Analysis times for small molecule molecular weight determination may be as little as 2 days, whereas some of the large scale quantitative proteomics services may take up to 6 weeks. Most routine protein analysis like identification of proteins from gel slices may take 2 to 3 weeks.

Contact Information:

Erol E. Gulcicek, Ph.D.
Co-Director, Protein Profiling
Director/PI, NBC Proteomics Core
300 George Street, Room G007
New Haven, CT 06511
Office: (203) 785-6296
erol.gulcicek@yale.edu

Kathryn L. Stone
Director, Keck MS and Proteomics Resource
300 George Street, Room G009
New Haven, CT 06511
Office: (203) 737-2204
kathryn.stone@yale.edu

Christopher M. Colangelo, PhD
Co-Director of Protein Profiling
Director of Protein Microarrays
300 George Street Room G006
New Haven, CT 06511
(203) 737-2636 (W)
christopher.colangelo@yale.edu

TuKiet (Tu) T. Lam, PhD
Director, FT-ICR MS Resource
300 George Street, Room G008
New Haven, CT 06511
Email tukiet.lam@yale.edu
Office (203) 785-5086

Terence Wu,
Co-Director, DIGE technology
300 George Street, Room G001
New Haven, CT 06511
Email terence.wu@yale.edu
Office (203) 785-6481