

Protein Profiling Technologies Offered by the Keck Laboratory at Yale University (4/23/03)

Background The remarkable success of genome level DNA sequencing has placed us at a threshold of knowledge that was unimaginable 25 years ago. To enable this watershed of data to be transformed into knowledge that will be of use in diagnosing, staging, understanding, and treating human diseases will require that we not only know the sequences of the estimated >30,000 human proteins but also that we identify key changes in protein expression which portend the impending onset of disease, accurately classify at the molecular level the disease subtype, and that we understand the functions, interactions, and how to modulate the activities of proteins which are intimately involved in disease processes. One of the most fundamental approaches to understanding protein function is to correlate expression level changes as a function of growth conditions, cell cycle stage, disease state, external stimuli, level of expression of other proteins, or other variable. Although DNA microarray analysis offers a massively parallel approach to genome-wide mRNA expression analysis, there often is not a direct relationship between the *in vivo* concentration of an mRNA and its encoded protein. Differential rates of translation of mRNAs into protein and differential rates of protein degradation *in vivo* are two factors which confound the extrapolation of mRNA to protein expression profiles. Gygi *et al*¹ estimate the correlation between protein and mRNA abundance for yeast is only 0.4. They found yeast genes with similar mRNA levels which had protein levels that differed by 20-fold. Conversely, they found invariant, steady-state levels of proteins which had mRNA levels that varied by 30-fold. These and other studies on yeast are summarized in Fig. 1, which shows that protein concentration can vary by >100-fold for a given mRNA concentration². Additionally, microarray analysis is unable to detect, identify or quantify post-translational protein modifications - which often play a key role in modulating protein function. Protein expression analysis offers a potentially large advantage in that it measures the level of the biological effector protein molecule, not just that of its message.

Currently, no protein profiling technology is available that can approach the ability of microarray analysis to simultaneously profile the relative level of mRNA expression of 25,000 or more genes. Fortunately, NIH is very much aware of the need to develop new protein profiling technologies and in October, 2002, Yale University was awarded one of 10 NHLBI contracts nationally (PI: Kenneth Williams, TC: \$18.2 million over 7 years) to establish centers to improve existing and to develop new proteomic biotechnologies. The Yale/NHLBI Proteomics Center (<http://info.med.yale.edu/nhlbi-proteomics/>) will focus on:

- ◆ Implementing and improving existing and developing new protein profiling biotechnologies and
- ◆ Designing new cell permeable, synthetic peptide-based reagents to modulate specific protein:protein interactions *in vivo*.

As new protein profiling methodologies are implemented in the Yale/NHLBI Proteomics Center, they will be made available to the Yale scientific community through the Keck Laboratory. The following sections provide a brief overview of two protein profiling technologies, MALDI-MS based peptide/protein disease biomarker discovery and differential fluorescence 2D gel

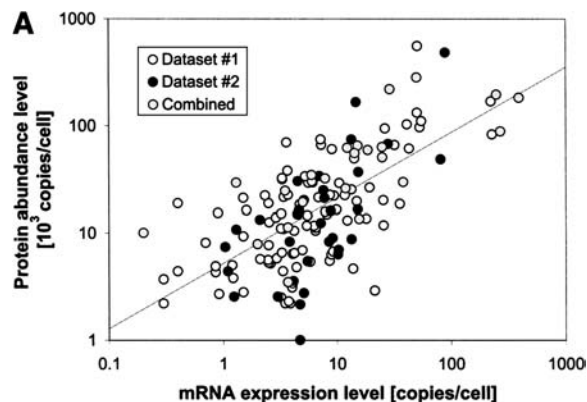


Fig. 1: Comparison of protein abundance and mRNA expression in yeast (from ref.

electrophoresis (DIGE) which are currently available from the Keck Laboratory. Two additional technologies, isotope-coded affinity tag (ICAT)/MS based protein profiling and comparative single and multidimensional LC/MS analysis of tryptic digests of whole cell and partially purified protein extracts, will be available within the next several months. The latter technology will take advantage of the very high sensitivity, mass resolution, and accuracy of a new Fourier Transform Ion Cyclotron Mass Spectrometer (FTICR-MS) funded by a recent NIH High End Instrumentation Grant (PI: Kenneth Williams, TDC: \$1.4 million) to Yale University.

Peptide/Protein Disease Biomarker Discovery A publication by Petricoin et al³ is among several recent studies suggesting that naturally occurring peptide disease biomarkers that bind C16 and other supports may be identified by algorithmic analysis of MALDI-MS spectra acquired from comparatively large numbers of disease versus normal serum samples. Since preliminary data obtained in the Keck Laboratory appears to confirm this approach has merit, a new service has been made available that utilizes a similar overall approach as that described by Petricoin et al³ but which is being carried out on a superior MALDI-MS platform and which utilizes an algorithm written by staff in the Keck Biostatistics Resource,

<http://keck.med.yale.edu/biostats/>

which is directed by Dr. Hongyu Zhao. This new Keck Laboratory service, which is depicted in the flow chart in Figure 2, includes robotic desalting of serum and other biological samples followed by automated MALDI-MS data acquisition. Serum, plasma or other liquid samples and extracts (10 µl each) are submitted in Marsh Bioproducts AB0800 PCR plates. A minimum of 8 samples (disease and normal) per plate is required for this service which begins by the robotic addition of 5 µl 0.1% TFA to each sample. After repeatedly (8x) pulling each sample up into a C18 ZipTip and expelling it back into the original sample well, the C18 ZipTip is washed 5x with 20 µl 0.1% TFA. Bound peptides are eluted from the C18 ZipTip with 10 µl of 50% acetonitrile, 0.1% formic acid into a new 96 well plate. Two µl of the eluent are then removed, mixed with 0.5 µl alpha-cyano-4-hydroxycinnamic acid matrix in 50% acetonitrile, 0.05% TFA containing an internal standard of 25 fmol of bradykinin (M+H C12 monoisotopic mass is 1060.569) and subjected to automated MALDI-MS on a Micromass M@LDI-R time-of-flight mass spectrometer. The internal

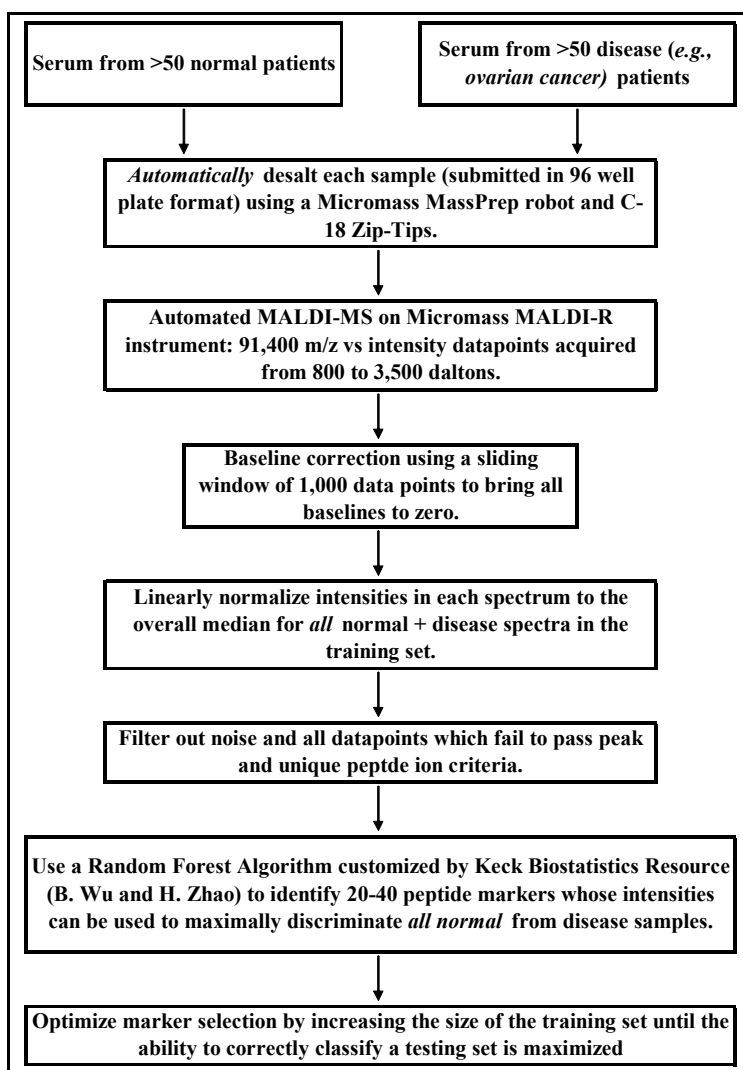


Fig. 2: Flow chart depicting the Keck Laboratory approach to MALDI-MS based peptide disease biomarker discovery.

bradykinin standard is then used as a "lock mass" to calibrate the spectrum.

The MALDI-R mass spectrometer automatically acquires data in positive ion detection over the mass range extending from 800-3,500 Da. Although the latter mass range is adjustable, it is difficult to acquire meaningful data below about 800 Da because of interference from the matrix and on the current reflectron (only) equipped instrument detection drops off substantially as the mass range is increased above about 2,500. The expected mass resolution is 14,000 at M+H 2465 and the mass accuracy is better than approximately ± 70 ppm. We recommend that reasonably large numbers (e.g., >25 each) of disease and normal samples from different patients be analyzed to obviate biological diversity so that disease-specific peptide biomarkers may be identified with higher confidence. Although not yet tested, we believe this service may be equally applicable to other biological fluids (e.g., amniotic fluid, urine, and tissue extracts, etc). This service will soon be further expanded by upgrading the current M@LDI-R to a M@LDI-L/R instrument which will enable the use of both linear and reflectron analyzers. Using an alpha-cyano matrix the linear analyzer on the new instrument will permit the mass range to be extended to about 25,000 Da. It is important to note that with this expanded mass range, the disease biomarker discovery technology will nicely complement the DIGE technology described below. Since (even with a linear analyzer) MALDI-MS response decreases with increasing MW (e.g., a 10^3 -fold larger amount of a 50,000 Da protein may be needed to generate the same MALDI-MS response as that generated by a 2,500 Da peptide), the disease biomarker service will optimally detect the *naturally* occurring forms of small proteins and peptides that span the 800-10,000 Da range while DIGE will optimally detect differential protein expression above this range.

To take optimum advantage of both the high resolution of the reflectron analyzer and the higher mass range of the linear MALDI-MS analyzer, the current plan is to acquire the higher resolution data permitted by the reflectron analyzer below about 3,000 daltons and to then acquire data on the same samples with the linear analyzer from about 3,000 - 25,000 daltons. The resulting spectra will be joined together and then analyzed by the Random Forest-based algorithm now being used to identify peptide disease biomarkers in MALDI-MS spectra acquired with the reflectron analyzer (only) over the 800-3500 Da mass range. Currently, approximately 91,400 m/z versus intensity data points are acquired over the latter mass range. As described in Wu et al⁴, the customized Random Forest disease biomarker discovery algorithm written by the Keck Biostatistics Resource is designed to identify a relatively small number (e.g., 20-40) of biomarkers whose relative intensities can be used to best discriminate all disease from normal serum samples in a training set that often is composed initially of approximately 48 sera from disease and 48 sera from normal patients. The validity of the resulting biomarkers are then assessed by using them to classify a testing set which also is composed of approximately equal numbers of sera from disease and normal patients. A straightforward approach to optimize the size of the training set is to gradually increase its size. Following each incremental increase of perhaps 25 samples, the ability of the resulting biomarkers to correctly classify a testing set (of perhaps 50 sera from approximately equal numbers of disease and normal patients) is determined as a function of the size of the training set.

Over the next few months several approaches will be taken to increase the dynamic range and likely biological importance of the resulting peptide/protein disease biomarkers. These approaches will include methodologies to remove very abundant albumin and immunoglobulin proteins prior to analysis and reverse phase/LC fractionation and automated spotting of sera onto MALDI-MS targets. The latter will permit each sample to be separated into 10 or more fractions which will each be subjected to MALDI-MS. This should help to increase

the dynamic range of the analysis. Finally, it is anticipated that the disease biomarker technology eventually will be moved to the very high performance MALDI-FTICR-MS platform and this overall approach may be extended to the analysis of complex tryptic digests of partially purified and/or whole cell protein extracts from normal versus disease sera and tissue extracts.

2D Fluorescence difference gel electrophoresis (DIGE) utilizes mass- and charge-matched, spectrally resolvable fluorescent dyes (Cy3 and Cy5) to label two different protein samples *in vitro* prior to 2D electrophoresis. Compared to conventional 2D electrophoresis, DIGE has the major advantage that both the control and experimental sample are run in the *same* polyacrylamide gel. These samples are then imaged separately but because they were run in the same gel, the images can be perfectly overlaid without "warping". This reduces the number of gels that must be run to make statistically valid comparisons and raises the confidence with which protein changes between samples can be detected and quantified. Hence, changes in relative level of protein expression may be detected that are as little as 1.3-fold above background. Use of a third dye (Cy2) permits an internal standard to be created by pooling an equal aliquot of all biological samples in the experiment. The internal standard is then run on every gel in the experiment. This means that every protein spot from all samples will be represented in the internal standard. This in turn allows more accurate quantification and spot statistics between gels. Based on the literature⁵, it is possible to profile up to 1,000 protein spots on properly prepared samples that provide well resolved 2D gels. The limit of detection for quantifying protein expression ratios is between 0.25 - 0.95 ng protein which is similar to that for silver staining^{6,7}. Because detection is based on fluorescence, the DIGE approach has a large dynamic range of about 10^4 , which permits differential expression analysis of relatively low copy number proteins⁶. An additional advantage of this system is the ability to detect many protein post-translational modifications, such as phosphorylation, which often play a key role in modulating protein function and which cannot generally be detected by ICAT-MS based protein profiling (see below). The overall DIGE approach involves the following major steps:

1. *In vitro* label 50 µg of the control and 50 µg of the experimental protein extracts with Amersham Biosciences Cy-3 and Cy-5 N-hydroxysuccinimidyl ester dyes. It is recommended that a third dye (Cy-2) be used as an internal (pooled 25 µg control + 25 µg experimental) standard to permit normalization of multiple gels and for internal normalization.
2. Mix control, experimental, and internal standard samples together (*i.e.*, 150 µg total protein) and subject to isoelectric focusing using Immobiline (IPG) Drystrips.
3. Carry out the SDS polyacrylamide gel electrophoresis (second) dimension on a 10 inch wide by 7.5 inch tall by 1 mM thick 12.5% polyacrylamide gel with one glass plate coated with Bind-Silane.
4. Immediately after SDS PAGE, the gel (which is still held between two glass plates) is scanned at all 3 wavelengths simultaneously on an Amersham Typhoon 9400 Imager. After scanning, 16 bit tiff files of each color channel are exported for image analysis using the differential in-gel analysis module of the Amersham DeCyder software package. After spot detection (which includes automatic background correction, spot volume normalization and volume ratio calculation), a user defined "dust filter" may be applied to each gel. This has the effect of automatically removing non-protein spot features from the gel and is followed by recalculation of experimental parameters.
5. The front glass plate is removed and the gel is then fixed and stained with Sypro Ruby, which is the fluorescent stain that will be used as a guide to excise spots of interest from the gel. The reason for using Sypro Ruby, which stains all protein in the gel, is that the Cy-dye

labeling is carried out such that the extent of incorporation will be <5% in terms of mole Cy-dye/mole protein. Since the Cy-dye has a MW of about 500 Da when coupled to a protein, low MW proteins (e.g., 10 Kd) labeled with Cy-dyes will not exactly co-migrate in the SDS PAGE dimension with their non-labeled counterparts.

6. Amersham Biosciences DeCyder software is used to quantify the gel image and to identify a "pick list" of differentially expressed protein spots to be excised and subjected to automated trypsin digestion followed by MS-based protein identification. The DeCyder software can analyze any two Cy-dyed gel images, either on the same gel or on different gels, match the spots between the two images, and then identify differentially expressed protein spots. The DeCyder software automatically outputs a listing of statistically significant differences in protein expression including t-test values, using the Cy-2 internal standard. Differentially expressed spots may be identified using a number of criteria including area, volume, 3D peak slope, 3D peak height, and/or statistical variation. Protein spots that show different degrees of intensity between the two samples are highlighted by the software so they can be manually confirmed. The DeCyder software can also analyze Sypro Ruby images, match the spots found with Sypro staining to those identified with the Cy-dye stains, and then choose a 'pick list' from the Sypro stained gel image. DeCyder data can be read by labs without the DeCyder software using an HTML format.
7. The protein spot pick list is transferred to the Amersham Biosciences Ettan Spot Picker instrument which automatically excises the selected protein spots from the gel and transfers them into a 96-well microtiter plate.
8. The excised protein spots are then subjected to automated in-gel trypsin digestion on the Amersham Biosciences Ettan TA Digester.
9. The 96 well plate is then transferred to a Micromass MassPrep Sample Handler where an aliquot of each digest is spotted (along with matrix) onto a MALDI-MS target.
10. High mass accuracy, automated MALDI-MS spectra are acquired on each target and the resulting peptide masses are subjected to database searching using ProFound and Mascot algorithms as described at: http://keck.med.yale.edu/prochem/procmald.htm#auto_ms_id
11. The remaining aliquots of digests of protein spots that are not identified by this approach may be subjected to nanospray or LC/MS/MS analysis (on the Keck Laboratory's Micromass Q-ToF instrument) with the resulting MS/MS spectra then being subjected to Sequest database searches to identify proteins present in the sample.

Isotope coded affinity tag (ICAT)-based protein profiling. While both MALDI-MS based peptide/protein disease biomarker discovery and DIGE analyses comparatively profile the *naturally* occurring forms of peptides and proteins, ICAT analysis profiles the relative amounts of cysteine-containing peptides derived from tryptic digests of protein extracts. Recognizing that only a single tryptic peptide is needed to quantify the expression of the corresponding parent protein, the ICAT reagent was designed to affinity isolate and quantify via the use of a stable isotope the relative concentrations of cysteine-containing tryptic peptides obtained from digests of control versus experimental samples. Hence, the newest ICAT reagent from Applied Biosystems has a thiol-specific reactive group adjacent to an alkyl linker which contains either nine [¹²C] or nine [¹³C] atoms. This results in a mass difference of 9 daltons between the control and experimental version of the same tryptic peptide. In addition to incorporating a cleavable biotin group (see below), another advantage of this new ICAT reagent is that unlike the previous deuterated ICAT reagent, [¹²C] and [¹³C]-ICAT derivatized forms of the same tryptic peptide co-elute on RP-HPLC. This greatly simplifies their relative quantitation by mass spectrometry. A very nice feature of the ICAT approach is that the *in vitro* incorporation of a stable isotope into

one of the two samples being compared obviates the need to analyze by mass spectrometry the control and experimental samples separately. The alkyl linker in the ICAT reagent is connected to a (cleavable) biotin group which allows rapid affinity isolation of cysteine-containing tryptic peptides. While a tryptic digest of a whole cell human protein extract might produce 550,000 peptides, less than 100,000 of these might be expected to contain cysteine. Based on a search of the Swiss Database, <5% of human proteins lack cysteine and would be missed. As depicted in Fig. 3, following derivatization of the control protein extract with the [¹²C]-ICAT reagent and of the experimental protein extract with the [¹³C]-ICAT reagent, the pooled samples are subjected to trypsin digestion followed by cation exchange chromatography. Typically, a whole cell or tissue protein extract would be divided into 30 cation exchange fractions with each of them being subjected to avidin chromatography isolation of cysteine-containing tryptic peptides followed by LC/MS/MS analysis to identify ICAT peptide pairs and quantify the relative [¹²C]/[¹³C] ratios. The resulting ICAT data, which is analogous to that obtained via the use of two different fluorescent dyes in DNA microarray analysis of mRNA or DIGE analysis of protein expression, provides the corresponding ratio for the level of expression of the parent protein in the control versus experimental sample. Currently, the largest number of proteins profiled by this approach from a single sample were 491 proteins contained in microsomal fractions of naïve and *in vitro* differentiated human myeloid leukemia cells⁸. Based on reports on Applied Biosystems web pages and a personal communication from Dr. David Han (3/19/03) from the University of Connecticut, we believe it may be possible to routinely identify and profile approximately 300 proteins/sample that has been separated into 30 cation exchange HPLC fractions prior to LC/MS/MS.

The Keck Laboratory's ICAT-based, quantitative protein profiling technology is expected to become available during the summer of 2003 and will be carried out on a new Applied Biosystems API QStar XL mass spectrometer which has recently been ordered and which will be dedicated to carrying out ICAT-based protein profiling. The Keck Laboratory expects to follow the Applied Biosystems protocol for reduction and trypsin digestion of 100 µg amounts of extracted protein which then will be fractionated into 30 pools by cation exchange HPLC⁸. Avidin chromatography will be used to isolate Cys-containing tryptic peptides from each pool which then will be individually subjected to LC-MS/MS on the QStar mass spectrometer at a flow rate of 300 nl/min on a 75 micron x 15 cm Vydac C-18 column equilibrated with 0.5% acetic acid, 5% acetonitrile and eluted with a 60 min acetonitrile gradient.

ICAT derivatized peptide pairs that differ by exactly 9 Da will be identified and quantified by the Applied Biosystems ProICAT software. ProICAT can perform modified database searches by extracting and using only data obtained on cysteine-containing

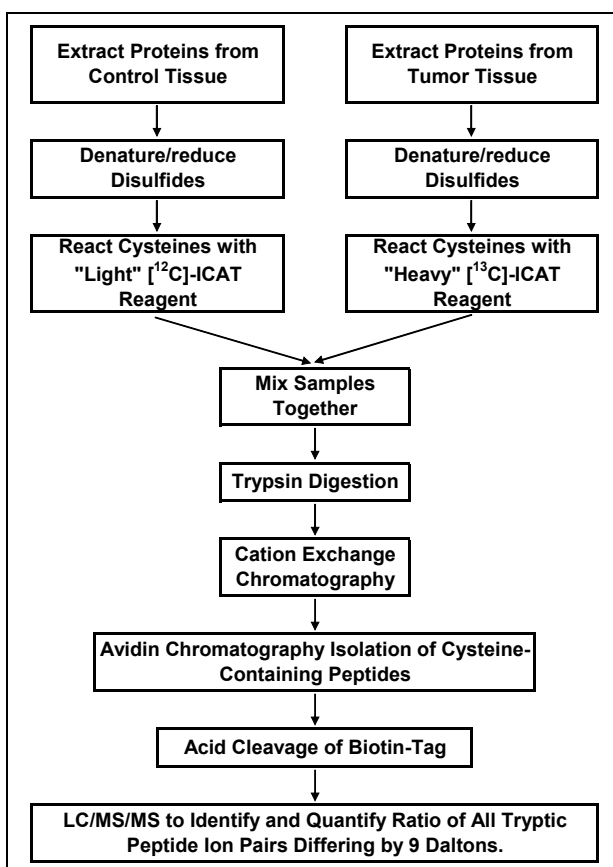


Fig. 3: Flow chart depicting ICAT/MS-based protein profiling.

peptides, thus significantly reducing search and data analysis times. ProICAT uses a 3-dimensional LC/MS reconstruct algorithm to locate and accurately determine experimental:control (heavy:light) peak ratios in complex proteomic samples. Applied Biosystems has been working closely with Spotfire to link/export the ProICAT quantitation data directly into DecisionSite which would then enable dynamic visualization and use of analytic tools similar to those currently used for analysis of gene expression data. The QStar mass spectrometer will be interfaced with a LC Packings Ultimate Capillary/Nanoflow HPLC System which consists of a UltiMate™ Micro Pump and Detection Module for accurate and reproducible micro- and nanoflow delivery, a FAMOS™ Micro Autosampler for automated injections of small volume samples with zero sample loss, and a Switchos™ Micro Column Switching Module that allows for automated sample preparation and multidimensional (e.g., 2D, 3D) LC.

In order to perform high-throughput proteomic profiling using ICAT technology, it is essential to automate as many steps as possible. The standard ICAT procedure from Applied Biosystems requires *manual* syringe-based purification steps on both cation-exchange and avidin cartridges. These steps are extremely time-consuming (2-3 hours/sample), labor intensive, and are prone to errors. Due to these limitations one person can only process 3-4 ICAT samples in an 8-hour time-period. To address these challenges, the Keck Laboratory has also recently ordered and installed an Applied Biosystems Vision workstation which automates both these steps. By using the Vision workstation we will be able to automate *both* the cation exchange HPLC and avidin cartridge chromatography steps. This will greatly enhance our ability to process samples, to maximize sample throughput on the QSTAR, and to substantially reduce the possibility of errors associated with manual syringe-based purification. The Vision workstation is a computer controlled biocompatible PEEK™ based HPLC system that enables unattended analysis and fraction collection for protein purification. It is equipped with two positive displacement piston pumps, a robotic sample handling system, eight column switching valve, and UV-Visible, pH, and conductivity monitors. The robotic sample handling system acts as both a fraction collector and autoamplifier, thus allowing *automated* re-injection of cation-exchange fractions onto the avidin cartridge. Besides automating the cation-exchange HPLC and avidin cartridge steps, the Vision workstation will in the near future (*i.e.*, Applied Biosystems is currently completing the development of the programming needed for this new application) also be able to be used to automate on-line cleavage/removal of the biotin group from the Cleavable ICAT reagent which again will be a major advantage in terms of more fully automating the overall ICAT-based protein profiling strategy.

Quantification based on LC-MS peak areas of stable isotope, internal standard analogs of an analyte has been used extensively and thus ICAT technology rests on a very firm foundation in this regard. The resulting peak area reports will be analyzed by the Keck Biostatistics Resource using statistical criteria that will be developed in the Yale/NHLBI Proteomics Center. Protein identification will be based on database searches of the resulting MS/MS spectra using Sequest and other algorithms. The interpreted datasets will be returned to individual investigators via a password-protected Web interface that launches an FTP file transfer and that will be similar to the interface used for the Yale Microarray Database (<http://info.med.yale.edu/microarray/>).

Single and multidimensional LC/FTICR MS/MS analysis can be utilized to compare tryptic digests of isolated proteins, protein complexes, partially purified and whole cell/tissue extracts. An example where a single dimension LC/FTICR MS analysis might be carried out would be to compare the profile of a tryptic digest of the *naturally* occurring form of a protein isolated from HeLa cells versus the same protein that had been cloned and expressed in *E. coli*. Since relatively few protein post-translational modifications occur in *E. coli*, this type of

comparative analysis can quickly identify tryptic peptides containing post-translational modifications. These peptides can then be isolated in the mass spectrometer and subjected to MS/MS structural analysis using either collision induced dissociation or other techniques. In this regard, a unique feature of FTICR-MS is the capability to also fragment ions using both Infra Red Multiphoton Dissociation (IRMPD) and/or Electron Capture Dissociation (ECD) techniques. These three MS/MS approaches often provide complementary structural information which can be particularly helpful when analyzing peptides containing post-translational modifications. Hence, detailed phosphopeptide sequencing has been carried out using ECD FTICR-MS/MS⁹ and combined ECD/IRMPD FTICR-MS/MS has been used to provide detailed structural information regarding glycoproteins¹⁰.

When working with partially purified or whole cell/tissue protein extracts one of the more difficult challenges is dealing effectively with the typically low extent of protein post-translational modifications. For example, only about 30% of the expected 30,000 human proteins are thought to be phosphorylated¹¹ and those proteins that are phosphorylated are usually modified at much less than equimolar ratios. Often, less than <5% of a given protein is phosphorylated at a certain position. This typically necessitates enriching the sample *prior* to MS analysis. Hence, following tryptic digestion of whole cell/tissue extracts, immobilized metal affinity chromatography (IMAC) can be used to enrich for phosphopeptides which can then be analyzed by LC FTICR-MS¹¹. Using this type of approach, Ficarro *et al*¹¹ detected more than 1,000 phosphopeptides in a whole cell lysate from *Saccharomyces cerevisiae*.

For very complex samples (*e.g.*, whole cell/tissue protein extracts) the Multidimensional Protein Identification Technology or "MudPit" approach, which utilizes a microcapillary column filled with a strong cation exchange (SCX) packing followed by a reverse phase (RP) column, provides a facile means to enhance both resolution and dynamic range¹². A sample of tryptically digested proteins is loaded onto the columns and a specific sub-set of peptides (related to overall charge) is eluted from the SCX column using a step gradient of increasing salt concentration onto the front of the RP section. Then using an RP gradient, peptides elute from the RP column according to their relative hydrophobicity and enter the mass spectrometer for analysis. After the RP gradient is complete, the next step of the salt gradient releases another sub-set of peptides from the SCX column onto the RP column and the process repeats itself. Comparing the MudPit tandem LC approach to ordinary single stage HPLC, it has been demonstrated that the tandem approach greatly increases the number of peptides identified in a single run and, most importantly, greatly facilitates the identification of peptides from low-abundance proteins. Using this approach on the yeast proteome, Wolters *et al*¹² identified 5,540 unique peptides from 1,484 proteins and demonstrated a dynamic range of detection of 10,000. This method may be extended even further to include (LC)ⁿ separations. In addition to identifying as many proteins as possible in very complex whole proteome samples, the MudPit would be ideal for identifying large numbers of proteins present in sub-cellular organelles and fractions as well as in large protein complexes brought down by immunological and "tag" approaches. It may also be possible to use comparative MudPit analysis to qualitatively identify differentially expressed proteins and their post-translational modifications.

Literature Cited

- ¹ Gygi, S.P., Rochon, Y, Franza, B.P., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720-1730.
- ² Greenbaum, D., Luscombe, N.M., Jansen, R., Qian, J., and Gerstein, M. (2001) Interrelating different types of genomic data, from proteome to secretome: homing in on function. *Genome Research* 11, 1463-1468.
- ³ Petricoin, E.F., Ardekani, A.M., Hitt, B.A, Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simine, C., Fishman, D.A., Kohn, E.C., and Liotta, L.A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 359, 572-77.
- ⁴ Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., Ward, D., Williams, K., and Zhao, H. (2003) Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, in press.
- ⁵ Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J., Hu, N., Taylor, P., Emmert-Buck, M., Liotta, L.A., Petricoin, E.F., Zhao, Y.. (2002) "2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers." *Molecular & Cellular Proteomics*. 1(2), 117-24.
- ⁶ Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins E., Currie, I., Davison, M. (2001) "Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. " *Proteomics* 1, 377-96.
- ⁷ Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M., Cramer, R., Waterfield, M., and Timms, J., (2002) "Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system." *Molecular and Cellular Proteomics* 1, 91-98.
- ⁸ Han, D., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotechnology* 19, 945-951.
- ⁹ Stensballe, A., Jensen, O., Olsen, J., Haselmann, K., and Zubarev, R. (2000) Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom.* 14, 1793-1800.
- ¹⁰ K. Hakansson, H.J. Cooper, M.R. Emmett, C.E. Costello, A.G. Marshall, and C.L. Nilsson (2001) Electron Capture Dissociation and Infrared Multiphoton Dissociation MS/MS of an N-Glycosylated Tryptic Peptide To Yield Complementary Sequence Information. *J. Anal. Chem.* 73, 4530-4536.
- ¹¹ Ficarro, S., McClelland, M., Stukenberg, P., Burke, D., Ross, M., Shabinowitz, J., Hunt, D., and White, F. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotechnology* 20, 301-305.
- ¹² Wolters, D., Washburn, M., and Yates, J. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem.* 73, 5683-5690.