NewsSpots

Summer 2004

HIGHLIGHTS OF 3rd INVESTIGATORS MEETING

The 3rd Investigators Meeting of the NLHBI Proteomics Initiative took place May 18-19 at the Bethesda Marriott, with a total of 90 researchers and trainees in attendance. Key features of the meeting included a poster session and a session on interfaces with developments and priorities in clinical proteomics.

The Clinical Proteomics session included guest speakers Leigh Anderson, PhD, and Chris Granger, MD, who co-chaired the Clinical Proteomics Working Group for NHLBI. The working group report was recently published in Circulation 2004, 109:1697-1703. Dr. Granger is a cardiologist at the Duke University Research Institute, and Dr. Anderson is founder and CEO of the Plasma Proteome Institute in Washington, DC. Joining Drs. Anderson and Granger in a panel discussion and Q&A period were Jennifer Van Eyk, PhD, director of the NHLBI Proteomics Center at Johns Hopkins University, and NLHBI scientist, Pothur Srinivas. Dr. Van Eyk is a member of the Clinical Proteomics Working Group, and Dr. Srinivas is the program officer on the current Clinical Proteomics Request for Applications (RFA-HL-04-019).

Wednesday morning's poster session was a popular event! Meeting participants visited 25 posters, many of them prepared and explicated by postdoctoral fellows working on technology development and applications in the 10 participating centers. There was universal enthusiasm about the information exchange and interactions facilitated by this informal session, which will probably become a regular feature of future meetings in a longer time slot.

MCW ENHANCES PROTEOMICS RESEARCH WITH LATEST BLADE SERVER

The Medical College of Wisconsin (MCW) recently implemented an IBM eServer BladeCenter JS20 as the heart of a new computing infrastructure system for proteomics research. The BladeCenter JS20 is the industry's first blade system based on Power Architecture. MCW is the first customer site to test the Sequest application on the JS20 running SUSE LINUX. Sequest is an application specifically for use in proteomics research.

As key factors in choosing IBM's newest blade server, Dr. Andrew Greene, director of the MCW Proteomics Center, cites the scalable technology platform, good performance/price point, IBM's commitment to support research, and good experience working with IBM previously. Another important reason was the plan of several key software vendors to make programs available for this platform.

Sequest on IBM JS20

- In collaboration with IBM Life Sciences, Thermo Electron
- Installed 8 JS20 Blades
- Port PVM-Sequest to POWER architecture (in progress)

Medical College of Wisconsin
IBM
Thermo Electron Corporation

The BladeCenter JS20 extends blade technology to 64-bit computing, offering customers the choice of POWER processor-based or Intel Xeon processor-based systems. The system supports both SUSE LINUX and Turbolinux, with additional support for Red Hat to be available by mid-2004. In the near future the AIX 5L operating system will be supported as well.

The adaptability of the architecture is a critical feature. It can be configured one way for one project, and another way for the next—in other words, the processors can be selected to suit the needs of the application.

Proteomics Admin. Ctr
171 Ashley Ave., 101 BSB
PO Box 250501
Charleston, SC 29425

Tel: 843.792.0871
Fax: 843.792.1657
proteomics-admin@musc.edu

MUSC CARDIOVASCULAR PROTEOMICS CENTER ADVANCES BIOINFORMATICS PROTOTYPE

The Cardiovascular Proteomics Center at the Medical University of South Carolina (MUSC) is following a research plan centered on integration among the experimental, analytical, and mathematical modeling components of the program. The early version of the integrative bioinformatics prototype enables proteomic screenings by 2D electrophoresis to be submitted and automatically incorporated in data structures in order to automate data warehousing, management, normalization and analysis. All MUSC center participants can access and complement these data structures according to their permission level in order to manipulate particular datasets.

The entirely web-based prototype provides a vertical integration that spans the spectrum from assigning probability levels to observed differential expression of different proteins to going all the way back to visualization of the raw data (image with identified spots). It also includes visualization of the intermediate steps involving data normalization and multivariate exploratory statistical analysis.

These advances in integrative capability have catalyzed a much closer interaction among experimentation, methods development, and modeling research projects. Some recent papers are listed below. Current efforts are directed to the integration of Mass Spectrometry data, automatic model extraction, and biomarker identification by machine learning and translational linkage to clinical research. The director of the NHLBI Proteomics center at MUSC is Dr. Dan Knapp.

RECENT PAPERS

On data management

Stanislaus R, Jiang LH, Swartz M, Arthur J, Almeida JS. (2004) An XML standard for the computation and dissemination of annotated 2D gel electrophoresis complemented with mass spectrometry results. BMC Bioinformatics 5:9.

On data modeling

Almeida JS, Voit EO. (2003) Neural-network-based parameter estimation in S-system models of biological networks. Genome Informatics 14:114-123.

Voit EO, Almeida JS. (2003) Dynamic profiling and canonical modeling: Powerful partners in metabolic pathway identification. In R Goodacre and GG Harrigan (Eds), Metabolite Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Kluwer Academic Publishing. Dordrecht, The Netherlands.

Voit EO, Almeida JS. (2004) Decoupling dynamical systems for pathway identification from metabolic profiles. Bioinformatics, in press.

On computer science and mathematics

Vinga S, Gouveia-Oliveira R, Almeida JS. (2004) Comparative evaluation of word composition distances for the recognition of SCOP relationships. Bioinformatics 20: 206-215.

Almeida JS, Wu S, Voit EO. (2004). XML4MAT: Inter-conversion between MatlabTM structured variables and the markup language MbML. Elsevier Computer Science Preprint Server CSPS: Data/0312002, submitted to Data & Knowledge Engineering.

In the news

BioInform 3/22/2004, page 3, "MUSC develops 2D gel data standard, ontology as part of software suite." (www.bioinform.com)

SEATTLE PROTEOME CENTER STRATEGY RESULTS IN PROTEOMIC IDENTIFICATION OF RARE HUMAN DISEASE

NHLBI-supported investigators at the Seattle Proteome Center have described a generic strategy for determining the specific composition, changes in the composition, and changes in the abundance of protein complexes. This strategy is based on the use of Isotope Coded Affinity Tag (ICAT) reagents and mass spectrometry to compare the relative abundances of tryptic peptides derived from suitable pairs of purified or partially purified protein complexes.[1]

In the first application of the strategy, the genuine protein components of a large RNA Polymerase II (Pol II) preinitiation complex (PIC) were distinguished from a background of non-specifically interacting proteins by comparing the relative abundances of peptides derived from a control sample and the specific complex that was purified from nuclear extracts by a single-step promoter DNA affinity procedure. In this experiment, the Seattle researchers identified a novel 8 kilodalton protein, encoded by an uncharacterized open reading frame as a bona fide component of Pol II PICs.[2] They also showed that the protein is conserved throughout evolution, and mutations in the human ortholog account for a DNA repair-deficient form of the trichothiodystrophy disorder called TTD-A.[3]

The identification of a novel, evolutionarily conserved, core TFIIH subunit is essential for understanding TFIIH function in transcription, DNA repair, and human disease. The study furthermore shows an impressive example where (1) a new technology was developed in the center, (2) its initial application in a model organism led to a new important insight into a transcription, a fundamental biological process, and (3) the extension of the study to human biology could explain the molecular mechanism of a rare human disease.

References

1 Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J and Aebersold R. (2003) The study of macromolecular complexes by quantitative proteomics. Nature Genetics 33(3):349-355.

2 Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, and Aebersold R. Quantitative proteomic analysis of an RNA polymerase II transcription complex identifies TFB5, a new component of general transcription/DNA repair factor IIH (submitted).

3 Giglia Mari G, Coin F, Ranish JA, Hoogstraten D, Theil A, Wijgers N, Jaspers NGJ, Raams A, Argentini A, van der Spek PJ, Botta E, Stefanini M, Egly JM, Aebersold R, Hoeijmakers JHJ and Vermeulen W. A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A and stabilizes TFIIH. (submitted).

Tools from the Seattle Proteome Center

Common format for MS data

Proteomics research utilizes a broad range of mass spectrometers (MS), each with a unique design, data system, performance specifications, strengths and weaknesses for certain types of experiments. Unfortunately, the binary data formats produced by each type of instrument also differ and are usually proprietary. The diverse, non-transparent nature of the data structure complicates integration of new instruments into pre-existing infrastructure, impedes analysis, exchange, comparison and publication of results from different experiments and laboratories, and prevents the bioinformatics community from accessing datasets required for software development. The Seattle Proteome Center recently introduced the mzXML format, an open, generic XML representation of MS data. They also developed an accompanying suite of related programs. The new format should facilitate data management, interpretation and dissemination in proteomics research.

Informatics course

- A quarterly, 6-day course demonstrating a suite of software tools developed at the Center for analysis, validation, storage and interpretation of data from large-scale quantitative proteomics experiments using ICAT reagent labeling, multi-dimensional chromatography, and tandem MS
- Aimed at active proteomics researchers
- >100 nationally and internationally based researchers have taken the course
- See http://www.systemsbiology.org/course/proteomics_index.html for info and application form

Reagents for TAP tagging

The Seattle Proteome Center has generated an extensive suite of expression vectors to express and purify proteins from mammalian cells that are expressed with a tandem affinity tag. The reagents are intended to purify native protein complexes from mammalian cells prior to mass spectrometric analysis. The reagents and protocols for their use are described at http://www.proteomecenter.org/protocols.php. The reagents have been sent to approximately 50 laboratories so far.

BU CENTER FOCUSES ON OXIDATIVE PROTEIN MECHANISMS IN CV DISEASE

The Cardiovascular Proteomics Center at Boston University (BU) focuses on understanding modifications of proteins caused by oxidants that may interfere with cell function in the course of diseases such as atherosclerosis, diabetes, and heart failure. On proteins, a site that is very sensitive to oxidants is the free thiol on the amino acid cysteine. Because free cysteine thiols are known to function importantly in normal cell signaling (for instance, when they react with nitric oxide), identifying those that actively participate in cell signaling and are most sensitive to oxidants is a vital step.

A new method under development at BU uses light and heavy isotope-coded affinity (ICAT) tags to label free cysteines, taking advantage of the fact that the ICAT can bind only to free thiols and cannot bind to oxidized thiols. Mass spectra obtained on ICAT-labeled normal and oxidized samples are analyzed together, and the degree of oxidation of any cysteine can be determined from the ratio of labeling by the ICAT.

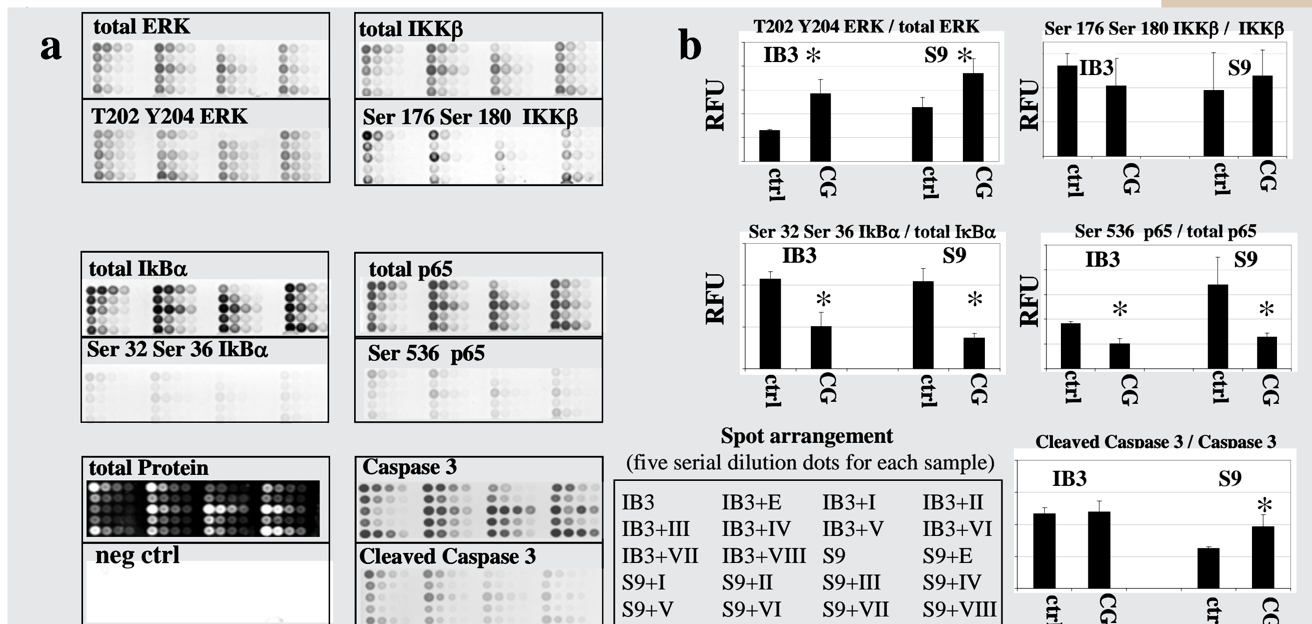
BU investigators recently published a paper showing proof of principle for the method with an important purified cardiac protein, creatine kinase. They also demonstrated that the technique can be applied to complex mixtures of cellular proteins.[1]

Because mass spectrometric analysis provides the degree of oxidation of any cysteine and also the sequence of the protein in which it resides, BU investigators hope to identify many new proteins that are affected by oxidants, how affected cysteines are involved in cell signaling, and how interference in their normal status by oxidants affects cell function. They then intend to study these proteins in tissues and cells affected by cardiovascular disease.

This Proteomics Center is also developing a novel technology—cryogenic Fourier transform mass spectrometry—that will enable identification of oxidant-induced protein modifications with dramatically increased precision and speed and facilitate new approaches to analyze the increasingly complex nature of the cellular systems in which the proteins participate.

Dr. Catherine Costello, director of the BU Mass Spectrometry Resource and Cardiovascular Proteomics Center, presented a web-broadcast lecture at the National Institutes of Health as part of the NIH Proteomics Seminar Series on May 14, 2004. This lecture, entitled New Mass Spectral Strategies for Proteomics and Glycomics, is available for viewing as a download from the NIH videocast archives at http://videocast.nih.gov/PastEvents.asp?c=36.

1 Sethuraman M, McComb ME, Heibeck T, Costello CE, Cohen RA. (2004) Isotope-coded affinity tag (ICAT) approach to identify and quantify oxidant-sensitive protein thiols. Mol Cell Proteomics 3: 273-278.



Cardiac glycosides block activation of IκBα and NFκB, p65 in CF lung epithelial IB3-1 cells.
[Figure is from Srivastava M, Eidelman O, Zhang J, Paweletz C, Caohuy H, Yang Q-F, Jacobson KA, Heldman E, Jozwi, C, Pollard BS, Pollard HB. Digitoxin mimics CFTR-gene therapy, and suppresses hypersecretion of proinflammatory interleukin-8 (IL-8) from cystic fibrosis lung epithelial cells. (2004) Proc Nat Acad Sci (USA) 101:7693-7698.]

RESEARCH AT USU POINTS TO NEW ROLE FOR DIGOXIN IN CYSTIC FIBROSIS

Cystic fibrosis (CF) is a common lethal genetic disease in the US, affecting more than 30,000 children, and responsible for the death of nearly one afflicted child each day. The principal site of CF disease is the lung, where massive inflammation leads to infection and lung destruction. Lung transplantation is the only long-term solution, although the 5-year survival for transplant recipients is only 50%.

Researchers at the NHLBI Proteomics Center of the Henry M. Jackson Foundation, based at the Uniformed Services University (USU), have approached the problem of developing a therapy for CF children by searching for small molecules that could cause the proteomic signatures of CF cells to closely resemble that of normal cells.

Their newest work shows that digitoxin, a drug with a two-century history of use for heart failure, can also cause genomic and proteomic changes in CF lung cells that mimic gene therapy. Digitoxin also suppresses the hypersecretion of the inflammation signaling molecule interleukin-8 (IL-8). These findings suggest that "off-label" use of digitoxin may prove to be therapeutically useful for the treatment of cystic fibrosis.

UTSW SCORES BREAKTHROUGH WITH SMALL MOLECULES THAT BIND LIKE ANTIBODIES

Antibodies are critical reagents in biomedical research. For proteomics work, the ideal situation would be to have antibodies against all proteins in humans, mice and other model organisms. However, antibody preparation on this scale would be a massive endeavor. Therefore, investigators have considerable interest in identifying simpler molecules that exhibit antibody-like binding properties (ie, high affinity and specificity for a target protein).

Researchers in the NHLBI Proteomics Center at the University of Texas – Southwestern Medical Center (UTSW) have recently made a breakthrough in this area with development of a simple method for isolating high affinity protein-binding peptides or peptoids from combinatorial synthetic libraries. This work has just been accepted for publication[1] in Chemistry & Biology, the chemistry journal of Cell press.

The protocol involves screening a library for a modest affinity ligand for a target protein. Once this is obtained, a new library is synthesized and all of the molecules in this library are linked to the modest affinity lead molecule. The new library is then re-screened under conditions too demanding for the lead compound to provide a clean 'hit.' The idea is that the lead compound and a library-derived species must collaborate to bind the target protein in a bivalent fashion to provide very high affinity.

Figures 1 and 2 show assays that measure the binding affinity of specificity of a simple 15 amino acid peptide that was isolated to bind ubiquitin via this procedure. The equilibrium dissociation constant was found to be 6 nM, and the peptide exhibited a 10,000-fold preference for ubiquitin relative to a mixture of non-specific proteins derived from an *E coli* extract. These parameters are comparable to what one would find using a good antibody.

Members of the NHLBI Proteomics Initiative believe that this is the first report of the successful production of small, easy-to-synthesize molecules that evince the binding properties of monoclonal antibodies. Current research at UTSW is focused on increasing the throughput of the current protocol and isolating binding molecules against a variety of proteins of interest.

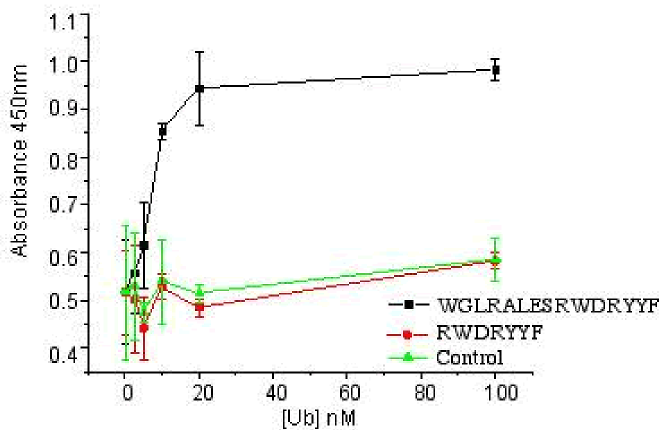


Fig 1. Measurement of the equilibrium dissociation constant of the ubiquitin and ubiquitin-binding peptide complexes (the lead peptide and the high affinity bivalent ligand). The data were obtained in an ELISA-like assay using immobilized peptide and an HRP-conjugated anti-ubiquitin antibody. The apparent Kd of the 15mer peptide-Ub complex is 6 nM.

1 Reddy MM, Bachhawat-Sikder K, Kodadek T. Transformation of low affinity lead compounds into high affinity protein capture agents. Chemistry & Biology, in press.

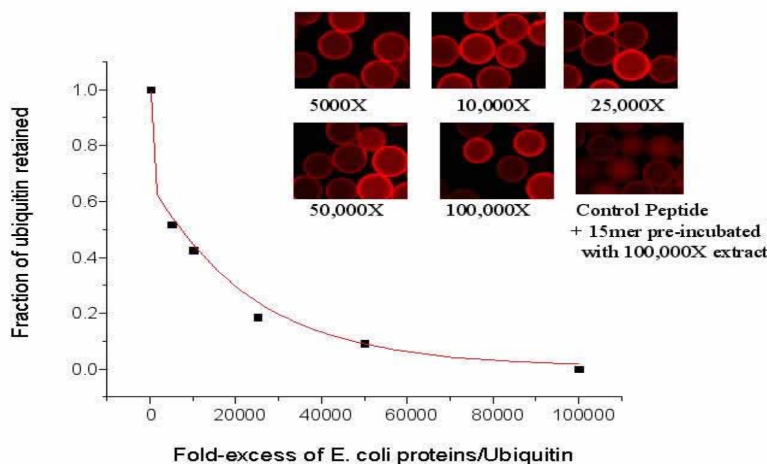


Fig 2. A competition experiment to measure the specificity of the 15 amino acid ubiqutin-binding peptide. The peptide was displayed on TentaGel beads. The beads were incubated with either Texas Red-labeled ubiquitin alone or Texas Red-labeled ubiquitin + the indicated molar excess of unlabeled proteins derived from an E coli extract (ie, a source of competitor proteins). In each micrograph, some beads from the ubiquitin-only incubation were mixed with those from the competition experiment, and the beads were photographed together to provide a direct comparison. Note that significant differences between the two bead populations appear visually only at 25,000-fold excess bacterial proteins. The final panel is a micrograph of a control bead population (displaying a different peptide) mixed with the beads displaying the ubiqutin-binding peptide that had also been exposed to 100,000-fold excess of competitor protein. At this point, specific binding of ubiqutin is almost lost. The fluorescence intensities of the beads were quantified and the graph shown was derived. The data indicate that half of the labeled ubiquitin is lost when approximately a 10,000-fold excess of bacterial proteins are present.

HIGHLIGHTS FROM THE NHLBI PROTEOMICS CENTER AT JHU

The focus of the NHLBI Proteomics Center at Johns Hopkins University (JHU) is the proteomics of adapatation to ischemia and hypoxia in heart, lung, and blood. A number of highlights from the Technical Implementation and Coordination Core (TICC) characterize recent progress at this site. Dr. Bob Cole oversees this service core for the JHU proteomics group. To accommodate expanding capabilities and meet increasing demands, the core recently moved into 2,000 sq. ft. of new laboratory space in Hopkins' new Broadway Research Building.

Two LTQ linear ion trap mass spectrometers (ThermoFinnigan) will arrive at TICC this summer. One of them will be interfaced with a vMALDI (a vacuum MALDI instrument) for obtaining sequence tags from proteins separated by offline fractionation or 2D gel electrophoresis. The other LTQ will be interfaced with an HPLC system for online fractionation and analysis of peptides. The speed and sensitivity of the LTQs permit digging more deeply into the proteome.

In addition, updating the Q-Star software with Nanomate takes advantage of Q-STAR's precursor ion scanning capabilities to identify and sequence phosphopeptides. The setup will also be useful for obtaining sequence tags from proteins separated by offline fractionation or 2D gel electrophoresis.

The DB database—a web-based, open source informatics storage database aimed at integrating in a single database all types of proteomic data obtained by the Hopkins group—is now up and running for 2D gel electrophoresis (2DE). It will soon be running for various mass spectrometers.

TICC offers a variety of training workshops. The MALDI-TOF workshop is currently offered every Thursday for 3 to 4 attendees. More than 250 graduate students, postdoctoral fellows, and faculty members have been trained since November 2000 to operate the MALDI-TOF spectrometers.

The DIGE workshop is 4-day lecture and laboratory workshop on differential protein expression analysis using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and difference gel electrophoresis (DIGE) technology. 2D-PAGE DIGE analysis allows direct comparison of two different samples on the same 2D gel by labeling each protein sample with a green or red fluorescent Cy dye. This approach solves the reproducibility problem for 2D gel analysis. This short course was first offered in December 2003 and is currently offered every quarter.

A new Sample Prep workshop will be initiated this year as a 4-day lecture and laboratory workshop focusing on sample preparation techniques to complement 2D-PAGE and DIGE technology. Attendees will work with a variety of sample conditions to illustrate effect of sample preparation on spot number and resolution. This short course will be offered in August 2004. ■

WORK AT YALE PROTEOMICS CENTER COMPARES PROTEIN ABUNDANCE AND mRNA EXPRESSION LEVELS

While DNA microarray analysis has the ability to interrogate the relative level of mRNA expression of 25,000 or more genes, the true causative forces in the cell are the concentration of proteins, their post-translational modifications (eg, phosphorylation), and their interactions. A major focus of the NHLBI Proteomics Center at Yale University, co-directed by Drs. Ken Williams and Bill Sessa, is to study protein concentrations that correspond to observable mRNA concentrations as well as changes in these concentrations.

The difficulty in trying to predict protein from mRNA concentrations is illustrated in Figure 1. Although there is a general trend for protein concentration to rise with mRNA levels, the actual correlation is weak and protein concentrations can vary by more than two orders of magnitude for a given mRNA level. Clearly, Figure 1 illustrates a powerful impetus to develop improved proteomic biotechnologies.

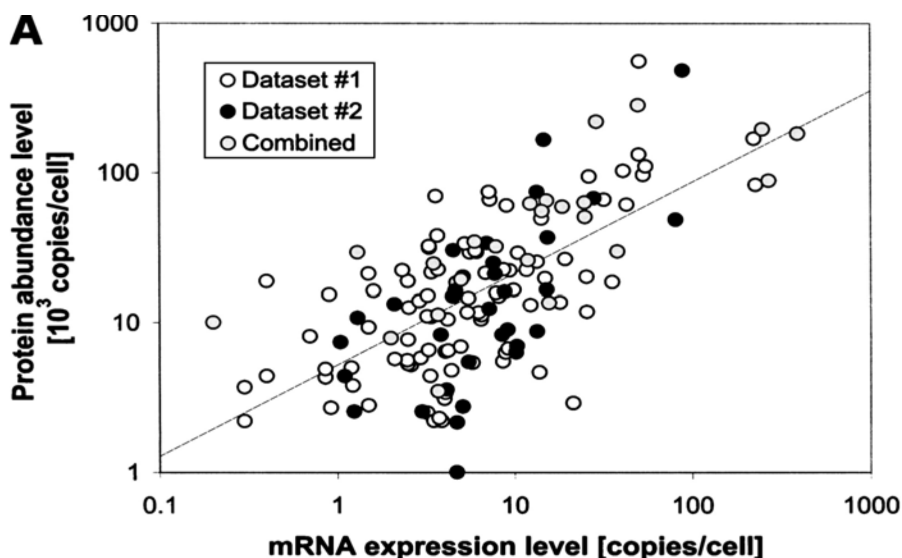


Figure 1. A direct comparison of protein abundance and mRNA expression in yeast (Greenbaum et al, 2001).

One approach to this challenge is to find correlations between mRNA and protein expression data that might allow more accurate extrapolation of protein from the more easily obtained mRNA expression data. Only a few such studies have as yet been done, most notably in human cancers and yeast cells, for the most part reporting only minimal or limited correlations.

Training Workshops offered by the TICC at the Hopkins Proteomics Center

MALDI-TOF

- 3 hours of hands-on instruction in sample preparation and analysis using a Voyager DE-STR MALDI-TOF mass spectrometer
- Covers acquiring spectra for proteins, peptides, oligonucleotides and small molecules
- Offered every Thursday for 3 to 4 attendees
- Open to outside participants

2D-PAGE DIGE

- 4-day lecture and laboratory workshop on differential protein expression analysis using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and Difference Gel Electrophoresis (DIGE) technology
- Provides detailed instruction and hands-on experience with Cy dye labeling, 2D PAGE, gel staining, gel scanning and image analysis using two different software packages
- Restricted to JHU personnel
- Offered quarterly

Day 1: *Theory and practices of 2D-PAGE, DIGE analysis, sample preparation and protein stain techniques. Attendees label two different protein samples with Cy dyes, evaluate efficiency of labeling, learn how to scan fluorescent gels and set up first dimension (IEF) separation of their samples.*

Day 2: *Attendees prepare their IpG strip for the second dimension, run 2D-PAGE gels, and stain them for total protein. They obtain all the gels images necessary for image analysis using either DeCyder or Progenesis software, which requires multiple scans of their gels for the three different Cy dye images and total protein images.*

Day 3: *Attendees receive instruction then perform image analysis of their 2D-PAGE DIGE gels using DeCyder software. They generate statistics and pick list for proteins of interest from their gels.*

Day 4: *Attendees receive instruction then perform image analysis of their 2D-PAGE DIGE gels using Progenesis software. They generate statistics and pick list for proteins of interest from their gels.*

Sample Prep

- New 4-day lecture and laboratory workshop on sample preparation techniques that complement 2D-PAGE and DIGE technology.
- Attendees will work with a variety of sample conditions to illustrate effect of sample preparation on spot number and resolution.
- Limited to JHU personnel.

In research supported by NHLBI at Yale, investigators have found that smaller homogenous subpopulations of yeast genes have significantly higher (and lower) correlations than occur in a global 'all against all' comparison. In particular, some sub-cellular localization categories—for example, the nucleolus —have significantly higher correlations (r=0.80) than the global correlation (r=0.66). Other localizations present significantly less correlation between mRNA and protein data, for example, the mitochondria (r=0.42), possibly reflecting the heterogeneous nature and function of this organelle.

In terms of functional categories, they found that some categories, such as cell rescue, show a lower than average correlation (r=0.45), while others, such as cell cycle, show a significant increase in correlation (r=0.71). Logically, this increased correlation reflects the co-regulated nature of the proteins in this functional category.

The Yale group postulates at least two probable reasons for the poor correlation between the levels of mRNA and protein. Firstly, mRNAs differ in their rates of translation into protein and secondly, proteins differ in their *in vivo* half-lives.

As one means of examining the first option, they looked at correlations between mRNA and protein abundance for those genes that had varied versus steady levels of mRNA expression over the course of the cell cycle. Broadly speaking, the cell can control the levels of protein at the transcriptional level and/or at the translational level.

The Yale group logically assumed that those genes that show a large degree of variation in their expression are controlled at the transcriptional level. The rationale is that the cell has already put significant energy into dictating the final level of protein through tightly controlling the mRNA expression, thus there would be minimal control at the protein level. They predicted and found a very high degree of correlation (r=0.89) between the mRNA and protein levels for these particular genes.

They also reasoned that those genes showing minimal variation in their mRNA expression throughout the cell cycle are more likely to have little or no correlation with the final protein level, as the cell would be controlling these genes at the translational and/or post-translational level, with the mRNA levels being somewhat independent of the final protein concentration. Indeed, the investigators found strikingly less correlation between protein and mRNA expression for these genes (r=0.2).

Furthermore, they found that genes with higher than average levels of ribosomal occupancy—that is, a large percentage of their cellular mRNA concentration is associated with ribosomes (being translated)—have well correlated mRNA and protein expression levels (r=0.78). These cases probably represent a situation wherein the cell, having significantly controlled the mRNA expression to produce a specific level of protein, will probably not also employ mechanisms to control the translation. Alternatively, those proteins with very low occupancy rates have uncorrelated mRNA and protein expression (r=0.30). Thus, given that the cell has not tightly controlled the mRNA expression for these genes, it will dictate the resulting protein levels through rigorous controls of its translation (that is, through tight limits on occupancy).

Future studies of computational correlation of mRNA and protein expression levels are needed. Further examination of the protein classes that allow for very high correlations, along with incorporation of data associated with mRNA and protein turnover rates, will allow the Yale proteomics group to create a more rigorous methodology that should enable more accurate extrapolation of protein from mRNA abundance levels. In addition, the Yale group plans to extend these studies, which are described in Greenbaum et al. (2003), by using the same approaches to examine relative changes rather than just the absolute concentrations of mRNA molecules and their corresponding proteins.

Because of the paucity of published human protein and mRNA expression data on heart, lung and blood cells, initial work has been carried out on yeast. Hence, a major goal is to extend a similar type of analysis to the protein expression data that will be obtained during the course of profiling studies now underway in the Yale Proteomics Center. An important objective of these studies is to reach the point where extrapolation from mRNA to protein expression data is a reliable research tool, for example, to guide the selection of antibodies to be spotted onto microarrays or the identification of proteins to be targeted by directed MS-based protein browsing and other technologies, in order to facilitate the independent measurement of selected proteins of highest potential interest.

References

Greenbaum D, Luscombe NM, Jansen R, Qian J, Gerstein M. (2001) Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. Genome Res 11: 1463-1468.

Greenbaum D, Colangelo C, Williams K, Gerstein M. (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biology 117: 1-118.

Proteomics Admin. Ctr
171 Ashley Ave., 101 BSB
PO Box 250501
Charleston, SC 29425

Tel: 843.792.0871
Fax: 843.792.1657
proteomics-admin
@musc.edu