Proteomic Approaches within the NCI Early Detection Research Network for the Discovery and Identification of Cancer Biomarkers

MUKESH VERMA,[a] GEORGE L. WRIGHT, JR.,[b] SAMIR M. HANASH,[c]
RASHMI GOPAL-SRIVASTAVA,[d] AND SUDHIR SRIVASTAVA[a]

[a]Cancer Biomarkers Research Group, Division of Cancer Prevention,
National Cancer Institute, Bethesda, Maryland, USA

[b]Department of Microbiology and Molecular Cell Biology and Virginia Prostate Center,
Eastern Virginia Medical School, Norfolk, Virginia, USA

[c]Department of Pediatrics, University of Michigan, Ann Arbor, Michigan, USA

[d]Division of Extramural Activities, National Cancer Institute, Bethesda, Maryland, USA

ABSTRACT: In the postgenome era, proteomics provides a powerful approach for the analysis of normal and transformed cell functions, for the identification of disease-specific targets, and for uncovering novel endpoints for the evaluation of chemoprevention agents and drug toxicity. Unfortunately, the genomic information that has greatly expounded the genetic basis of cancer does not allow an accurate prediction of what is actually occurring at the protein level within a given cell type at any given time. The gene expression program of a given cell is affected by numerous factors in the *in vivo* environment resulting from tissue complexity and organ system orchestration, with cells acting in concert with each other and responding to changes in their microenvironment. Repositories of genomic information can be considered master "inventory lists" of genes and their maps, which need to be supplemented with protein-derived information. The National Cancer Institute's Early Detection Research Network is employing proteomics, or "protein walking", in the discovery and evaluation of biomarkers for cancer detection and for the identification of high-risk subjects. Armed with microdissection techniques, including the use of Laser Capture Microdissection (LCM) to procure pure populations of cells directly from human tissue, the Network is facilitating the development of technologies that can overcome the problem of tissue heterogeneity and address the need to identify markers in easily accessible biological fluids. Proteomic approaches complement plasma-based assays of circulating DNA for cancer detection and risk assessment. LCM, coupled with downstream proteomics applications, such as two-dimensional polyacrylamide gel electrophoresis and SELDI (surface enhanced laser desorption ionization) separation followed by mass spectrometry (MS) analysis, may greatly facilitate the characterization and identification of protein expression changes that track normal and disease phenotypes. We highlight recent work from Network investigators to demon-

Address for correspondence: Sudhir Srivastava, Ph.D., M.P.H., Chief, Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, 6130 Executive Boulevard, EPN-330F, Rockville, MD 20852-7346. Voice: 301-435-1594; fax: 301-402-8990.

ss1a@nih.gov

strate the potential of proteomics to identify proteins present in cancer tissues and body fluids that are relevant for cancer screening.

KEYWORDS: Biomarkers; EDRN; LCM; MALDI; ProteinChip; Proteomics; SELDI

INTRODUCTION

Genetic aberrations during carcinogenesis disturb the normal function of a cell, primarily through their effect upon the proteins encoded in the altered/mutated genome. Thus, while increasingly sophisticated techniques have rapidly expanded our knowledge of events at the nucleic acid level, protein-based analyses and assays remain desirable as they provide information that is more easily linked to function than nucleic acid–based assays. From a practical point of view, assays of protein tumor markers, due to their ease and robustness, lend themselves to routine clinical practice. Screening, even for high-risk groups, could only be successful when the procedures undertaken are simple, inexpensive, and minimally invasive. A most effective approach is to develop protein-based tests for easily accessible biological fluids such as urine, blood, feces, sputum, bladder, and bronchioalveolar lavage. The development, in particular, of tests to evaluate secretory proteins has proven value in the diagnosis of cancer and in the assessment of precancerous lesions. Proteomics provides the means to achieve the goal of early cancer detection.

An important step in the new postgenome era is to decode the functions of the huge numbers of new open reading frames (ORFs). Various avenues to elucidate what unknown genes do and how they interact within an organism are being utilized, including (1) the simultaneous assessment at the RNA level of all genes that are expressed in a cell population or tissue and (2) the mapping and quantitation of all the proteins expressed in a cell population or tissue.[1,2] Valuable information that cannot be accurately predicted from nucleic acid data, such as protein location and posttranslational modification, may be obtained directly at the proteomic level. The ensemble of proteomic information constitutes the body of knowledge of the "proteome", the total protein output encoded by a genome. Proteome research, or proteomics, is an extension of protein (polypeptide) and gene-product mapping (linking proteins to their encoding genes).[3] The goal of proteomics is a comprehensive, quantitative description of protein expression and of the changes associated with biological perturbations such as disease (cancer) or drug treatment. Proteome analysis can be undertaken independently of genomic or RNA analysis. Its ability to identify and quantify proteins and confirm their expression and structure as presumed from nucleic acid data is highly complementary to genomic science.

The human genome encodes for some 30,000 genes that are scattered among 3.2 billion nucleotides. Added complexity stems from the findings of alternative splicing and due to the occurrence of posttranslational modifications that give rise to several protein products from a given gene.[2] The profiles of expressed proteins in normal and cancer states are different as some genes are up- or downregulated, whereas others are expressed constitutively.[4] To generate these profiles, a variety of technologies are employed for sample collection (cells, tissue, or fluid), processing and analysis of samples, and storage of data. Cytosolic proteins are most readily amenable to proteomic analysis. Whereas secreted proteins are easy to solubilize,

their low abundance is problematic. On the other hand, membrane-bound proteins both occur in low abundance and require additional steps for their solubilization prior to analysis.[5]

TECHNOLOGIES INVOLVED

The most widely used techniques for the characterization of proteins are two-dimensional gel electrophoresis (2-DGE), mass spectrometry (MS), amino acid composition analysis, peptide sequence tagging, and high-throughput microchip arrays. Matrix assisted laser desorption ionization (MALDI), surface enhanced laser desorption ionization (SELDI), laser capture microdissection (LCM), and capillary electrophoresis (CE) have been added to the proteomics tool set.[6] We summarize these technologies below.

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) remains as the standard high-resolution technique for protein separation and is the method of choice when complex samples need to be characterized.[7] Quantitative 2-D PAGE can be used for protein-based gene expression analysis and allows large-scale screening of the protein component of normal or cancer cells or tissues. The separation range of 2-D gels can extend from the most acidic to the most basic proteins as needed for specific applications.

MS has been used for protein identification and sequencing and has given a major impetus to proteomics.[8–10] For sequence-based determinations by MS, proteins are cleaved with trypsin or cyanogen bromide into peptide fragments and separated by high-pressure liquid chromatography (HPLC) followed by tandem mass spectrometry. For protein identification, the mass spectra of peptides from digested proteins are utilized to search sequence databases. Modified peptides, such as those containing an oxidized methionine, can be identified by accounting for modifications. If the peptide under investigation can be matched correctly to the mass value predicted by genomic sequencing, it can be deemed unmodified. However, if a particular peptide does not have a match in the sequence of a particular candidate protein, specific modifications can be inferred that may account for mass discrepancy. The confidence achieved in matches obtained from peptide-mass fingerprinting can be significantly increased by additional sequence information from within the proteolytically derived peptide fragments. This approach is termed "peptide sequence tagging". The short (2–4 amino acids) sequence can be derived by fragmentation of the parent ion into three complementary segments: the mass of an N-terminal fragment, the mass of a C-terminal fragment, and the partial amino acid sequence between them.

Amino acid composition analysis is an efficient method for protein identification.[11] The method is complementary to MS and represents a useful analytical tool for the mapping of proteins of interest.

The handling of protein samples to be separated by 2-D PAGE is still very much a manual process. High-throughput screening (HTS) in proteome analysis needs improvement through automation.[12] Although the current methodologies to process samples for proteome analysis are feasible, a manual approach is incompatible with the industrial scale needed to conduct genomic style analyses. Studies such as those of Aicher *et al.*,[13] who identified an association between decreased calcium-binding

protein, calbindin-D 28-kDa, urinary calcium wasting, and intratubular cortico-medullary calcification in rat kidney using proteomics, can be substantially facilitated through automation.

MS-based methodologies span the vast expanse of proteomics and drug discovery. Both electrospray ionization (ESI) and MALDI support proteomics-based research projects by identifying proteins separated and isolated by PAGE. MALDI/MS-based surface scanning of one-dimensional isoelectric focusing gels, "virtual 2-D gel electrophoresis", represents a potentially high-throughput means to map proteins and to determine protein profiles. MS can also be used to directly study the covalent and noncovalent interactions of drug molecules and biomolecular targets.

Capillary electrophoresis/time-of-flight mass spectrometry (CE/TOFMS) has been used for analysis of in-gel digests of protein spots excised from 2-D PAGE gels. An off-line purification and preconcentration procedure with a Zip Tip is used before CE/TOFMS analysis, which allows for detection of protein spots with <1 picomole of material from 2-D gels. The off-line procedure provides sufficient purification for analysis while maintaining the quality of the CE separation. Using this procedure, proteins from Coomassie Blue and zinc negatively stained gels can be identified by database searching based on the peptide maps generated. CE/TOF tandem mass spectrometry is used to confirm database query results and to provide structural analysis of peptides that do not match the expected peptide maps derived from the database search in order to identify structural modifications.

Affinity mass spectrometry offers a novel proteomic approach for the identification and measurement of cancer-associated biomarkers. Based on the work of Hutchens and Yip,[14] Ciphergen Biosystems (Fremont, CA) has developed the SELDI ProteinChip® MS technology that brings to the field of proteomics a user-friendly methodology.[15,16] SELDI has several advantages over other existing technologies, such as liquid chromatography–mass spectrometry (LC-MS), 2-D gel electrophoresis–coupled MS (2DGE-MS), ELISA, and fluorescent-based binding assays for high-throughput screening, due to its versatility, ease of use, speed, and cost. It is rapid, reproducible, and highly sensitive (i.e., detection limit in the femto-molar range) and is readily adaptable to a diagnostic format.[10] Its major limitations are the difficulty in identifying proteins detected in mass spectra and the limited ability to detect most proteins in complex samples.[17]

The most tedious job in proteomics is sample preparation, as some proteins are insoluble in a heterogeneous mixture of cells.[5] An important advance in sample preparation is the development of LCM.[18] Prior to LCM, obtaining "pure" cell populations from tissues was performed manually. Although manual microdissection can achieve good precision, it is time-consuming and labor-intensive, and requires a high degree of manual dexterity. The LCM system yields a one-step procurement of selected human cell populations from a section of complex, heterogeneous tissue. LCM has been successfully used to obtain pure populations of cancer cells from both frozen and paraffin-embedded tissues, stained or unstained, for molecular analysis of macromolecules. LCM is capable of isolating single cells, making it possible to procure pure populations of neoplastic cells from lesions less than 1 mm in diameter without encroachment of adjacent nonneoplastic cells. Using this technique, matched normal epithelium, stroma, benign, preneoplastic neoplasia, and cancer cells from the same specimen have been successfully isolated.
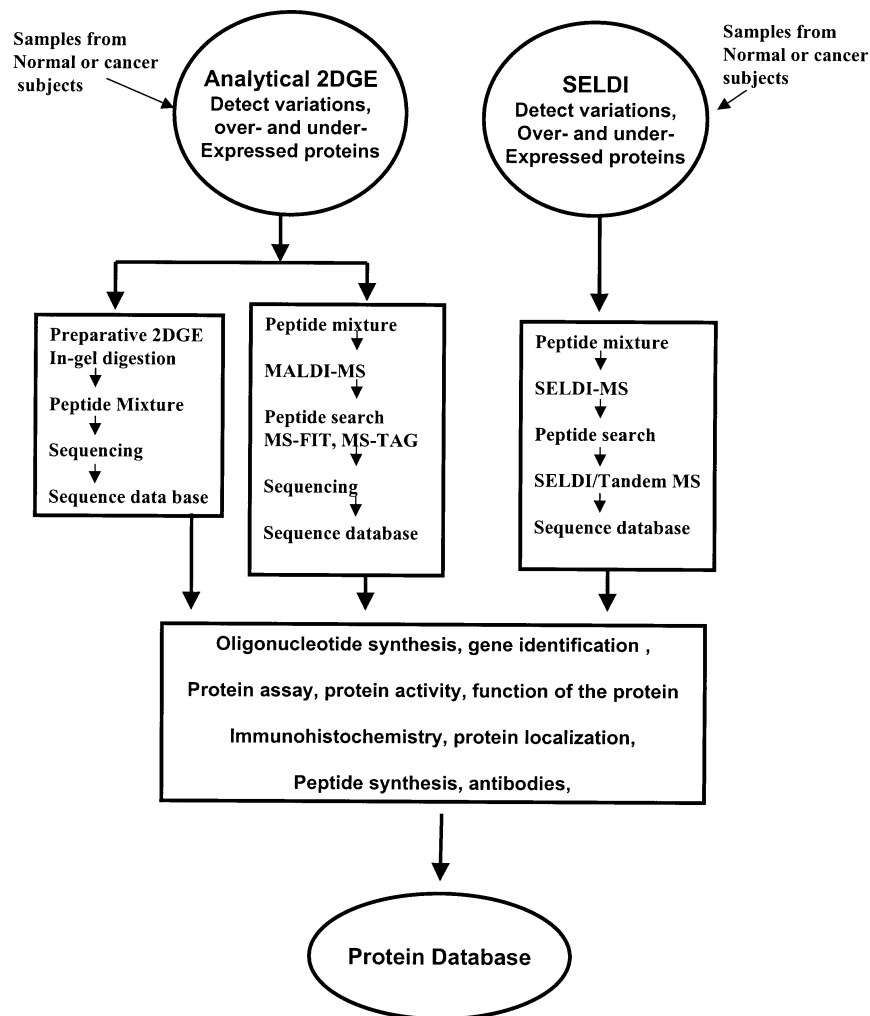
FIGURE 1. General strategies for detection and characterization of cancer-associated proteins.

PROTEOMICS APPROACHES IN THE EARLY DETECTION RESEARCH NETWORK (EDRN)

General strategies commonly used to detect and analyze cancer-associated proteins by 2-DGE and SELDI are shown in FIGURE 1. EDRN investigators have applied proteomic approaches to identify and characterize potential markers for the early detection of cancer from samples such as serum, nipple aspirate, urine, seminal
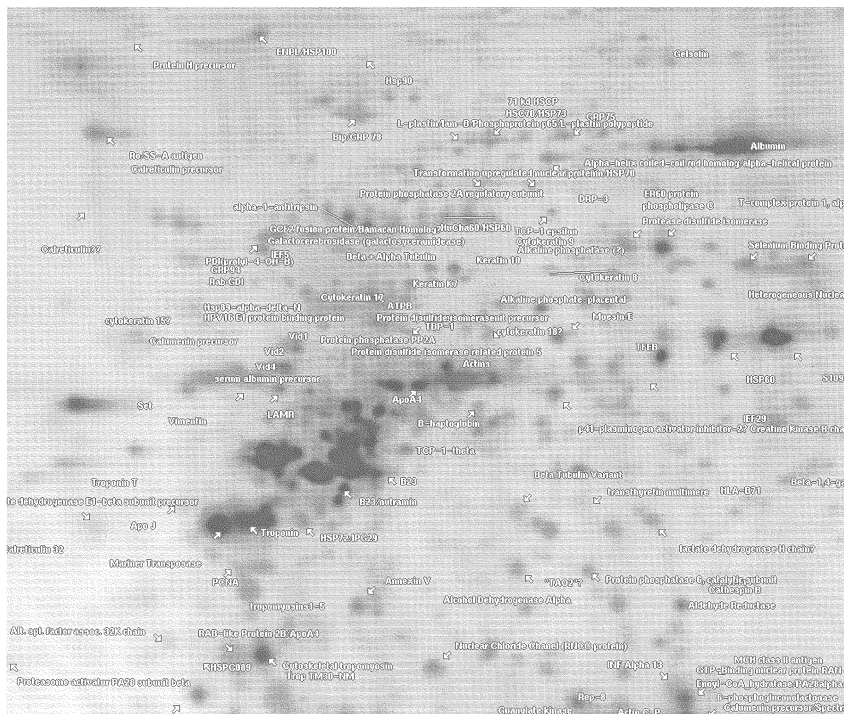
FIGURE 2. A 2-D PAGE master image showing a subset of proteins identified in lung tumors.

plasma, and microdissected cells. At the University of Michigan, the focus of the research is on the detection and identification of protein antigens that induce humoral response and on proteins secreted by tumor cells in lung and colon cancer. Tumor proteins may induce a humoral response as a result of their overexpression, increased turnover, posttranslational modification, or other unique processing in neoplastic cells. There is evidence that a humoral response to such protein antigens may predate the diagnosis of cancer, indicating the utility of assays for specific tumor protein antigens or antibodies in serum or biological fluids for the early detection of cancer. A large number of proteins of interest have been characterized and their derived information has been entered into a queryable database. Several lung cancer–specific protein markers have been identified and are being characterized (FIG. 2).

The focus of the Eastern Virginia Medical School (EVMS) Biomarker Developmental Laboratory is the application of the SELDI ProteinChip®–Surface MS system to the discovery of cancer-associated biomarkers, with a major emphasis on prostate cancer (PCA). A previous report from this laboratory described SELDI methodologies for detecting known prostate cancer biomarkers, such as prostate specific antigen (PSA), prostate acid phosphatase (PAP), and prostate specific mem-

brane antigen (PSMA) in cell lysates from LCM microdissected cells and body fluids (i.e., seminal plasma and serum).[19]

Ongoing studies are directed at biomarker discovery and assay development. For the discovery phase, efforts have focused on SELDI analyses of cell lysates from microdissected prostate cells and seminal plasma. Pure populations of cells, that is, cancer, benign, and normal epithelium, were obtained from the same patients' surgical tissue specimens using the Pixcell-II®–Laser Capture Microdissection microscope (Arcturus Engineering, Mountain View, CA). Approximately 2000–3000 cells were procured for each cell type, cell lysates were prepared, and the lysates were subjected to SELDI analysis. Three proteins (33 kDa, 8.5 kDa, and 9.5 kDa) were found to be overexpressed in PCA cells compared to benign and normal epithelial cells. An interesting finding was the observation that, when the 33-kDa protein was overexpressed, the free form of PSA appeared to be underexpressed. Using a ratio of peak intensity of the 33-kDa with the 28-kDa free PSA protein, the cancer population could be clearly differentiated from the benign and normal cell populations ($p < 0.002$). Overall, PSA appears to be downregulated in cancer compared to normal epithelium in the same tissue specimen, confirming a recent report.[20] Perhaps most notable has been the recent discovery of four proteins under 10 kDa that are overexpressed in preneoplastic lesions, that is, prostate intraepithelial neoplasia (PIN), and that continue to be upregulated in cancer. Efforts are in progress to validate whether these proteins represent an early protein signature of neoplasia, to isolate and identify these potential diagnostic biomarkers, and to determine if they are also present in body fluids. Other reports coupling LCM with SELDI describe similar findings with the generation of distinct protein profiles for each cancer type evaluated.[21]

An application of SELDI is the simultaneous analysis of multiple proteins to establish "fingerprint" profiles that discriminate cancer from noncancer. The premise for this approach is that no single protein by itself will improve the early detection/diagnosis of cancer. Seminal plasma and serum from normal age-matched men and from patients diagnosed with benign prostate hyperplasia (BPH) or PCA were subjected to SELDI protein profiling. A representative example of SELDI analysis of serum is shown in FIGURE 3. The raw time-of-flight data were analyzed by two different learning algorithms developed by investigators at the EDRN Data Management Coordinating Center (i.e., Binary Combinations and Wavelet algorithms). For seminal plasma, using the Binary Combinations algorithm, a sensitivity of 88% and a specificity of 89% were obtained. This protein profiling algorithm, which analyzes combinations of 8 proteins, correctly "diagnosed" 21/24 cancers (88%), 21/23 BPH (91%), and 25/28 normals (89%). For serum, 167 sera from PCA patients and 96 sera from normal age-matched men were used to train the Wavelet algorithm. The analyses of 11 protein masses resulted in a specificity of 100% and a sensitivity of 97% for the training set. A test set of samples consisting of 30 normals and 15 cancers, all coded, yielded correct diagnosis for all 30 normals (100%) and 14/15 cancers (93%). A SELDI protein profiling study of urine from bladder cancer patients resulted in an improved detection rate of low-grade bladder cancer by 44% (to 77%) compared to voided urine or bladder washing cytology (33%). Five proteins and 7 protein clusters between 2000 and 20,000 daltons were used in this analysis. One of the proteins (3.3 kDa) was identified as defensin.[22] These preliminary studies are very encouraging and demonstrate that SELDI fingerprint protein
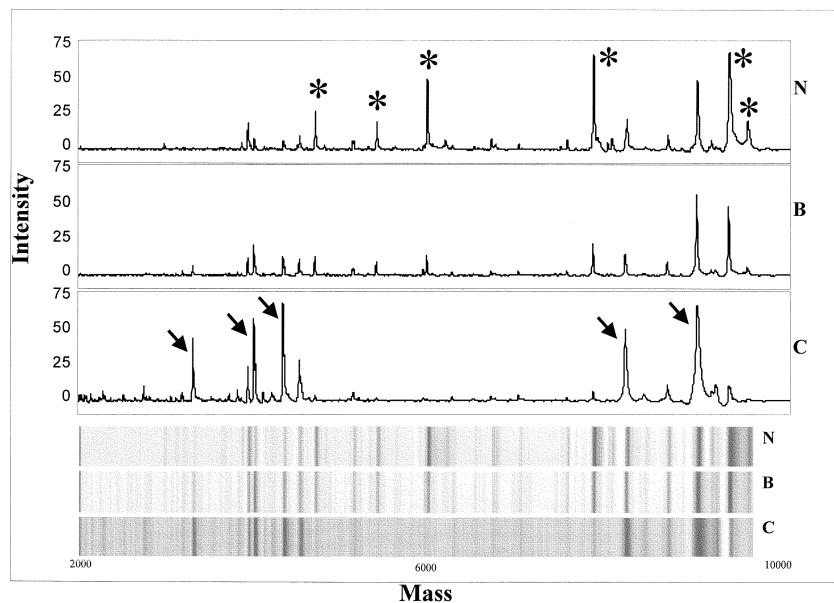
FIGURE 3. Representative SELDI retentate map of serum from a patient diagnosed with prostate cancer (C), compared to the retentate maps of sera from a patient with benign prostate hyperplasia (B) and a normal age-matched male donor (N). The raw mass spectra are shown in the top three panels; gel views are shown on the bottom. Only the low mass region (2–10 kDa) of the protein profiles is shown. The asterisks (*) identify those proteins underexpressed, and the arrows identify those proteins overexpressed in the prostate cancer specimen.

profiling coupled with a learning algorithm may lead to innovative clinical assay for the early detection/diagnosis of prostate and bladder cancers. With the SELDI approach, protein identification may not be necessary for development of clinical assays based on the SELDI protein profile.

Besides discovery and protein profiling applications, SELDI can also be used in an immunoassay platform. For this application, antibody rather than a chemical matrix is bound to the chip array to capture the protein (antigen). The EVMS lab has successfully used the SELDI immunoassay to measure PSA and PSMA in body fluids, in both single and multiplex formats.[19,23] Using the SELDI/PSMA immunoassay, it was possible to discriminate PCA patients from patients with BPH and normal age-matched men ($p < 0.001$). Importantly, all patients with serum PSA values between 4 and 10 ng/mL were correctly identified as either BPH or PCA. This is a significant observation since the specificity of PSA in the 4–10 ng/mL range is only 25–30%. If validated, PSMA may be a clinically useful biomarker to improve the accuracy for differential diagnosis of BPH and PCA. The SELDI ProteinChip system, as shown in these studies, is a powerful and innovative proteomic technology for biomarker discovery and diagnostic/prognostic assay development. A major advantage of SELDI over other proteomic technologies is its versatility. Not only can SELDI be used for biomarker discovery, but the same platform can be utilized to

develop clinical assays, such as protein profiling and immunoassays for simultaneous analyses of multiple biomarkers. The promising results obtained with prostate and bladder cancer suggest that SELDI will be applicable to all human solid cancers.

Research by EDRN investigators at the Georgetown University Lombardi Cancer Center aims to characterize differences in protein signatures between foci of normal breast epithelium, ductal carcinoma *in situ* (DCIS), and invasive cancer from fresh surgical specimens; to determine whether differences identified above can be detected in nipple aspirate fluid (NAF) samples from women without breast abnormalities, with DCIS, or with early invasive breast cancer; and to determine whether differentially expressed proteins can be detected in serum samples from women with different stages of breast cancer development. A number of potential candidates have been identified using LCM- and SELDI-based approaches.

Like the EVMS team, investigators at the M. D. Anderson Cancer Center have also discovered differentially expressed proteins in urine from patients with bladder cancer using SELDI. Most of these proteins are of low molecular mass. Analysis of the urine from patients with bladder squamous cell carcinoma (SSC) has found a number of significant differentiation markers, such as the presence of keratins 10, 13, 15, and 19, and loss of a number of proteins, including glutathione-S-transferase. Another putative urinary SCC marker, psoriasin, a calcium-binding protein that is widely expressed in tissues containing stratified squamous epithelium, has also been studied.

Thus, a number of potential biomarkers have been identified by EDRN investigators, which will be validated in different laboratories using a large number of samples. If validated, these markers would be extremely valuable for screening populations at high risk for cancer and perhaps developing prevention strategies for a variety of cancers.

GENOMIC AND PROTEOMIC APPROACHES ARE COMPLEMENTARY

Studying individual proteins in the context of other cellular proteins is complementary to the information gathered from a genomic-based approach. The human genome sequence provides a framework for proteomics.[2] Proteome analysis already represents a means of efficiently complementing differential display (DD), high-density expression arrays (HDEA), expressed sequence tags (ESTs), direct or suppressive subtraction hybridization (SSH), chromosomal linkage studies, and nucleic acid sequencing as a problem-solving tool in molecular biology. DNA sequence information alone cannot predict the efficiency of translation of gene products, their relative concentrations, the extent of posttranslational modifications, and the biological consequences of altered gene expression.

Proteomics complements, but certainly does not replace, other technologies employed in the study of molecular processes within cells. Cellular and molecular biology dictate that information pertaining to DNA and RNA must be supplemented by corresponding information at the level of protein translation. Currently, the throughput of proteomic analysis is lagging behind that using other approaches based on nucleic acids. However, proteomics is still in its early phase of development and the field is poised for major expansion. The mapping of various proteomes including identification of posttranslational modifications such as phosphorylation

promises to provide a very powerful means for investigating a wide variety of signaling phenomena, including cross talk between different signaling pathways, activation or inactivation of particular pathways by endogenous or exogenous cofactors, and even uncovering consequences of genetic defects in signaling cascades.

One limitation of proteome studies is its inability to efficiently resolve low-abundance proteins on 2-D gels. The sensitivity of the staining techniques will allow only those proteins present at more than 0.1 ng to be visualized (by silver staining). Furthermore, electrotransfer to PVDF or nitrocellulose membrane as well as the extensive staining/destaining/washing procedures involved contribute to a significant amount of protein loss before any molecular characterization can be attempted. Since peptide-mass fingerprinting and protein sequencing via MS are capable of characterizing protein at the femtomole level, the limiting factor is the visualization of proteins on 2-D gels and transfer membranes. However, the advancement made by Ciphergen Biosystems employing SELDI has improved visualizing of proteins significantly in the range of 2–10 kDa.

Proteome analysis has additional limitations. Proteome analysis is unlikely to detect point mutations or minor insertions, deletions, or recombination events. There is, however, at least one example where a point mutation in lactate dehydrogenase (LDH) was detected. The mutated version of LDH had reduced substrate binding specificity.[24] Proteome analysis is also not able to detect all possible gene products associated with every gene/ORF within a genome due to the phenomenon of temporarily restricted gene expression as all genes are not expressed constitutively. This restriction is linked to suppression (or induction) of gene activity associated with a variety of stimuli occurring as a result of disease state, environmental changes, stress, embryological development, or housekeeping activities. The influence of the relationship between the length of the cell cycle or replication time and the half-life of both protein and mRNA gene products is critical to the relative abundance of intracellular proteins.

The principal technical challenge in proteomics is achieving a level of comprehensiveness that matches that of genomics, that is, complete coverage. This is clearly a much harder task with proteins than with nucleic acids. In the first place, genes are approximately equimolar in genomic DNA, whereas proteins may span seven or eight orders of magnitude in functional abundance in a cell type such as the hepatocyte and potentially wider ranges in distributed media such as serum. Very few detection methods exist that are usable over seven orders of magnitude and, if one could be found, problems remain such as that of abundant spots obscuring minor ones in 2-D gels. The relative chemical homogeneity of DNA and the availability of technologies and resources such as reverse transcriptases, restriction enzymes, PCR, and sequence complementarity have all contributed to the ease of genomics as compared to proteomics. There may be a temptation to assert that complete proteomes are just around the corner, an assertion that represents a major leap of technological faith at this point.

BIOINFORMATICS AND PROTEOMICS

Like the genome project, more sophisticated data mining and learning algorithm software programs will be required to reproducibly and efficiently collect and ana-

TABLE 1. Protein databases on the Internet

Web address	Description
http://expasy.hcuge.ch	Access to several protein databases using ExPASy server
http://www.mdli.com/chemscape/chime	To view structure in 3-D (MDL or PBD format)
http://www.biochem.ucl.ac.uk/bsm/ dbbrowser/PRINTS/PRINTS.html	To align protein fingerprints
http://www.ncbi.nlm.nih.gov/BLAST	To search for homologues
http://www.pdb.bln.gov	PDB home-page (3-D structures)
http://www.ncbi.nlm.nih.gov/Structure/ cn3d.html	To view structure in 3-D (ASN.1 format)
ftp://ftp.pdb.bnl.gov/pub/kinemage	To view structures published in *Protein Science* and other journals
ftp://ftp.pdb.bnl.gov/pub/other-software/ Rasmol	To download by anonymous ftp Rasmol
http://www.prosci.uci.edu/scop	To compare structures
http://www.embl-heidelberg.de/Service	To see secondary structure
http://motif.stanford.edu/emotif	To align proteins
http://biobase.dk/cgi-bin/celis	To search disease-related proteome database

lyze the tens of thousands of bits of information being generated by the proteomic technologies. Furthermore, current proteomic projects are generating a wealth of data for a wide range of tissues, and it is critical for the success of any study that the results produced are organized and presented in a form that makes analysis as simple and accessible as possible. To this end, a number of databases have been constructed (TABLE 1, FIG. 2), containing both visual and textual data with extensive links to both protein and nucleic acid databases. The ability to effectively search such databases for the identification of protein spots from 2-DGE has become an essential step in the study of proteomes. The ability of analytical techniques used in protein characterization and their associated database query programs to determine identity at the functional group level has been examined for proteins with low levels of homology at the gene/protein sequence level. Such theoretical data manipulations may provide the means to predict the utility of data acquired experimentally with nonsequence-dependent software for proteome analysis. Eight organisms have now been fully sequenced and distributed within publicly accessible gene and protein databases. These databases of model organisms have been proposed as a starting point for studies of either the total proteome or the "functional" proteome (defined as gene-product expression under specific environmental or laboratory conditions). The proteome of the yeast, *Saccharomyces cerevisiae*, has been studied extensively by 2-DPAGE of cells grown under a variety of conditions. Approximately 600 visible protein spots have been characterized from whole-cell extract. These protein spots have been identified by amino acid analysis, peptide mass fingerprinting, and associated sequence tags in conjunction with nanoelectrospray tandem mass spectrometry and conventional N-terminal microsequencing.

CONCLUSIONS

The inside of a cell is a crowded and dynamic compartment, where proteins are perpetually being produced and degraded. Understanding the structures, interactions, and functions of all of a cell's proteins is one of the grand goals of the postgenome era. Prospects for introducing proteomics technologies into the clinical laboratory will depend on their robustness, their user-friendliness, and the clinical relevance of the added information they provide that cannot be captured through other technologies in use in the clinical laboratory. The primary focus of proteome analysis is the rapid characterization of gene products as either "previously studied" or "novel". The overall goal is to apply this information for cancer screening and risk assessment. Efforts made in the EDRN are directed towards identification and characterization of biomarkers that can be used for early cancer detection and risk assessment.

REFERENCES

1. COLLINS, F.S. & V.A. MCKUSICK. 2001. Implications of the human genome project for medical science. JAMA **285:** 540–544.
2. VENTER, J.C., M.D. ADAMS, E.W. MYERS *et al.* 2001. The sequence of the human genome. Science **291:** 1304–1351.
3. WASINGER, V.C., S.J. CORDWELL, A. CERPA-POLJAK *et al.* 1995. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. Electrophoresis **16:** 1090–1094.
4. JUNGBLUT, P.R., U. ZIMNY-ARNDT, E. ZEINDL-EBERHART *et al.* 1999. Proteomics in human disease: cancer, heart, and infectious diseases. Electrophoresis **20:** 2100–2110.
5. HERBERT, B. 1999. Advances in protein solubilization for two-dimensional electrophoresis. Electrophoresis **20:** 660–663.
6. GODOVAC-ZIMMERMANN, J., V. SOSKIC, S. POZNANOVIC *et al.* 1999. Functional proteomics of signal transduction by membrane receptors. Electrophoresis **20:** 952–961.
7. HATZIMANIKATIS, V., L.H. CHOE & K.H. LEE. 1999. Proteomics: theoretical and experimental considerations. Biotechnol. Prog. **5:** 312–318.
8. FABRIS, D., M. VESTLING, M. CORDERO *et al.* 1995. Sequencing electroblotted proteins by tandem mass spectrometry. Rapid Commun. Mass Spectrom. **9:** 1051–1055.
9. YATES, J.R., J.K. ENG & A.L. MCCORMACK. 1995. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal. Chem. **67:** 3202–3210.
10. FERAS, E.R., D.J. STEPHENS, C.E. WALTERS *et al.* 1999. The role of cholesterol in the biosynthesis of beta-amyloid. Neuroreport **10:** 1699–1705.
11. HUMPHERY-SMITH, I., S.J. CORDWELL & W.P. BLACKSTOCK. 1997. Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis **18:** 1217–1242.
12. DOVE, A. 1999. Proteomics: translating genomics into products? Nat. Biotechnol. **17:** 233–236.
13. AICHER, L., D. WAHL, A. ARCE *et al*. 1998. New insights into cyclosporine A nephrotoxicity by proteome analysis. Electrophoresis **19:** 1998–2003.
14. HUTCHENS, T.W. & T.T. YIP. 1993. New desorption strategies for the mass analysis of macromolecules. Rapid Commun. Mass Spectrom. **7:** 576–580.
15. MERCHANT, M. & S.R. WEINBERGER. 2000. Recent advancements in surface-enhanced laser desorption/ionization–time of flight mass spectrometry. Electrophoresis **21:** 1164–1177.
16. FUNG, E.T., G.L. WRIGHT, JR. & E.A. DALMASSO. 2000. Proteomic strategies for biomarker identification: progress and challenges. Curr. Opin. Mol. Ther. **2:** 643–650.

17. YIP, T.T., J. VAN DE WATER, M.E. GERSWIN *et al.* 1996. Cryptic antigenic determinants on the extracellular pyruvate dehydrogenase complex/mimeotope found in primary biliary cirrhosis. J. Biol. Chem. **271:** 32825–32833.
18. BANKS, R.E., M.J. DUNN, M.A. FORBES *et al.* 1999. The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—preliminary findings. Electrophoresis **20:** 689–700.
19. WRIGHT, G.L., JR., L.H. CAZARES, S.M. LEUNG *et al*. 1999. ProteinChip surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. Prostate Cancer Prostatic Dis. **2:** 264–276.
20. ORNSTEIN, D.K., C. ENGLERT, J.W. GILLESPIE *et al*. 2000. Characterization of intracellular prostate-specific antigen from laser capture microdissected benign and malignant prostatic epithelium. Clin. Cancer Res. **6:** 353–356.
21. PAWELETZ, C.P., J.W. GILLESPIE, D.K. ORNSTEIN *et al.* 2000. Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip. Drug Dev. Res. **49:** 34–42.
22. VLAHOU, A., P.F. SCHELLHAMMER, S. MENDRINOS *et al.* 2001. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. Am. J. Pathol. In press.
23. XIAO, Z., X. JIANG, M.L. BECKETT *et al.* 2000. Generation of a baculovirus recombinant prostate-specific membrane antigen and its use in the development of a novel protein biochip quantitative immunoassay. Protein Expression Purif. **19:** 12–21.
24. LISZESKI, K. 1999. New twists in gene therapy. Genet. Eng. News **19:** 8–34.