



Deshaies 02 - Open twisted Alpha/Beta structure #1, #2, #3 - 53 x 108 in. - 3 -11" x 14" -28 x 36 cm -Acrylic on canvas - www.JacquesDeshaies.com

The Move to Clinical Proteomics

As technologies improve, researchers struggle with the best way to detect biomarkers

By Josh P. Roberts

Clinical proteomics is undergoing a major shift, perhaps even a revolution. What was principally a search for drug targets in the year 2000 is now more a quest for markers of disease. The hunt for a single protein has turned into a pursuit to identify patterns of polypeptides. Buzzwords such as high-throughput, personal medicine, quantitative, and multiplex are bandied about in the same breath as subproteome and bioinformatics.

"What is now emerging is sort of Proteomics II," says Leigh Anderson, CEO of the Washington, DC-based nonprofit Plasma Proteome Institute (PPI). Proteomics I was exploratory, but now there is increased focus on quantitative measurement of proteins, including patterns of quantitative changes, with an emphasis on statistics. "That's a new and important additional level of maturity that's arising in this field," adds Anderson.

Recent advances in instrumentation and computational methodologies help researchers sift through the protein constituents of blood, urine, seminal fluid, and nipple aspirate for new markers of disease. The effort seems to be paying off: Several groups claim to be able to spot pathological changes earlier and with greater accuracy using new technologies. Now researchers and physicians are scrutinizing these claims at meetings, in the published literature, and in current and forthcoming clinical trials.

HOW MANY IS ENOUGH? A biomarker is nothing more than a molecule, such as a protein or metabolite, whose presence or abundance in a biological fluid signals disease. A classic example is the presence of elevated levels of cardiac enzymes in the blood following a heart attack. Finding such indicators is proving problematic, however, and over the past decade the Food and Drug Administration (FDA) has approved only about one new diagnostic marker per year.

Anderson says he believes this bottleneck is due in large part to the genetic heterogeneity among populations; what may indicate disease in one group may be statistically insignificant in another. Others in the field point to the heterogeneity of the dis-

eases themselves. For example, with dozens of histologic types of breast cancer, there is no single marker—one that can be used to "tell people who are normal that they don't have cancer," says Emanuel Petricoin III, FDA's codirector (with Lance Liotta) of the National Cancer Institute (NCI)-FDA Clinical Proteomics Program.

Finding individual proteins that are expressed so disproportionately between the sick and the healthy, such as prostate-specific antigen (PSA) or ovarian cancer antigen 125, is difficult. William Rich, CEO and president of CIPHERGEN Biosystems of Fremont, Calif., points to the "graveyard of diagnostic markers" that have failed to live up to their potential.

The old way of discovering biomarkers, one-by-one, "is wrong, and it's not working," says Petricoin. The future, he says, lies in biomarker patterns—collections of proteins that, in aggregate, illuminate disease—and their enabling tool, mass spectrometry. MS-based proteomics, he says, "allows you to look at hundreds of thousands of things at once, and do this without the illusion of knowledge."

Rich predicts a revolution in the diagnostics market, with Nobel prizes and billions of dollars awaiting the company that can produce multimarker protein tests that have high predictive accuracy.

PROTEINS OR PATTERNS? One highly publicized example concerns ovarian cancer. Notoriously deadly when diagnosed late, ovarian cancer can be treated effectively when caught early. But researchers could find no reliable indicator of early-stage malignancy. In 2002, Petricoin and Liotta's groups showed that proteomic patterns—collections of proteins, none of which is diagnostic on its own—could be used to diagnose early ovarian cancer effectively.¹

The previous year, John Semmes' team at Eastern Virginia Medical School (EVMS) in Norfolk, Va., had shown that differential expression profiles generated by surface-enhanced laser desorption ionization (SELDI)-MS can distinguish diseased, benign, and normal prostate and bladder cell populations.² What distinguished Petricoin's work "is that they took this profiling

that we showed you could do, and hooked it up with an algorithm," recalls Semmes, director of the EVMS Center for Biomedical Proteomics. "Then you had an automated sort of way of detecting disease states." Several other groups, including Semmes, have followed suit with similar results for breast, prostate, and liver cancers.³ In all these cases, the proteomics profiles effectively distinguished disease states without identifying the polypeptides in the profiles.

Petricoin does not worry about what constitutes the pattern, as long as it can be reproduced reliably in a clinical setting. Many others, however, are concerned. Eleftherios Diamandis of Toronto's Mount Sinai Hospital, for example, writes in a dialog with Petricoin and Liotta that "the identity of these molecules is not absolutely necessary for their use as biomarkers, but without this knowledge, the method will remain empirical and probably difficult to validate, reproduce, standardize, and quality control."⁴

Diamandis suspects that the molecules constituting the distinguishing pattern are "epiphenomena of cancer and that they are produced by other organs in response either to the presence of cancer or to a generalized condition of the cancer patient" He remains doubtful as to whether they are able to distinguish among the varieties of maladies that could potentially generate them.

The ability to replicate the pattern reliably is frequently mentioned as a principle objection to using patterns as biomarkers. "If you haven't identified the analytes, the technology has to be perfect to give you reproducible answers," says Anderson. Otherwise it's "just pixels." He sees the option of not identifying the analytes as "purely a stop-gap, temporary situation."

"Scientists want to know not only that you're looking at real proteins, they want to know what they are," says Rich, whose company manufactures a SELDI-based instrument. "They want to know how they relate back to the biology."

Making such identifications and relationships is the role of bioinformatics. Each MS instrument has its own embedded software for protein identification and database searching, says Gilbert Omenn, a physician and researcher at the University of Michigan in Ann Arbor. But the information contained in each database varies considerably and even between subsequent versions of a given database.

SETTING THE STANDARDS The Human Proteome Organization (HUPO), an international collection of organizations, scientists, and institutions, is putting forth an overall effort to define community standards for data representation in proteomics, which will facilitate data comparison, exchange, and verification. This Proteomics Standards Initiative (PSI), in turn, is a precursor to the larger goal of mapping the components of at least three distinct sets of proteins: those found in plasma (or serum), the brain, and the liver.

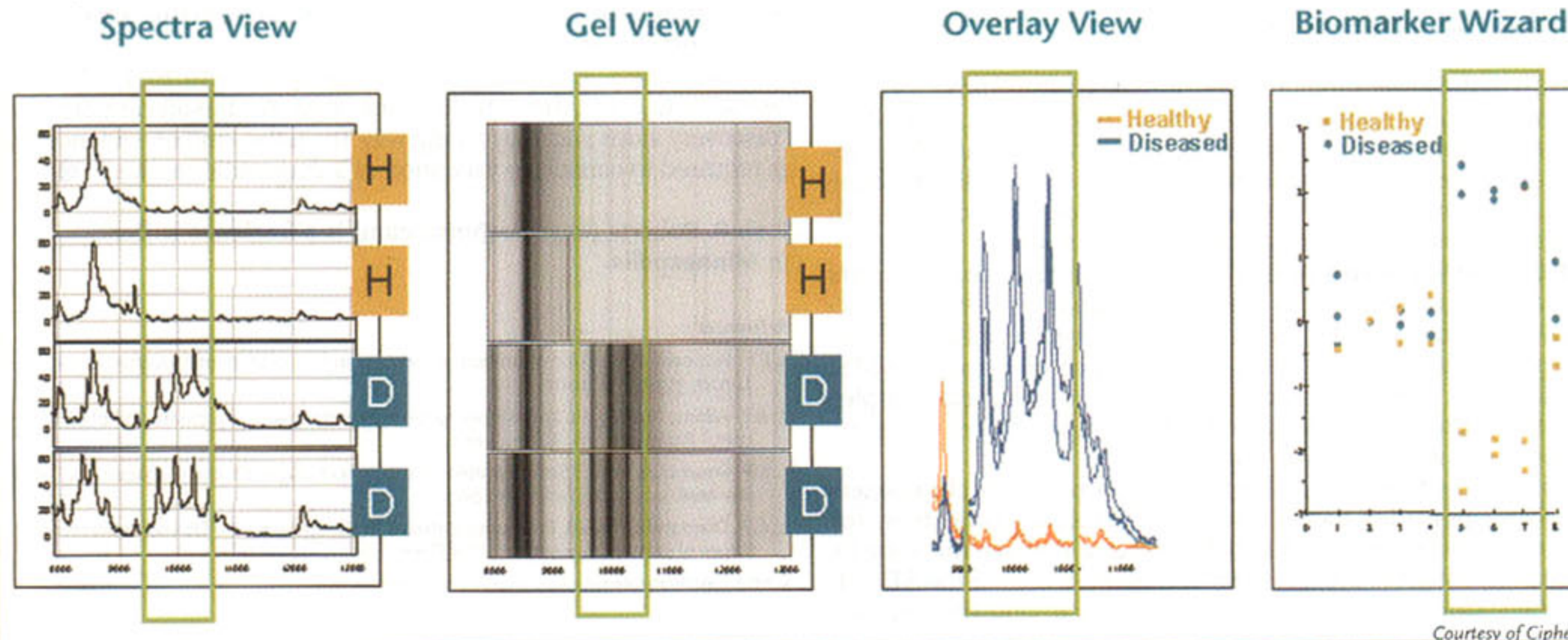
The PSI currently is contemplating specimen collection and handling, and considering such questions as whether protease inhibitors should be added during collection, whether plasma or serum should be used, how to deal with highly abundant proteins, and the advantages and limitations of a variety of technology platforms (including bioinformatics). This pilot phase has "very specific, feasible goals that will set the foundation for many kinds of longer-term, population-based, and clinical studies for public health and epidemiology and drug trials," notes Omenn, project leader of HUPO's Plasma Proteome Project initiative.

HUPO has an important task, observes John N. Weinstein, senior investigator with NCI's Center for Cancer Research. But he predicts that systematizing proteomics will be more difficult than was systematizing genomics, because proteins vastly outnumber genes and are also much more heterogeneous.

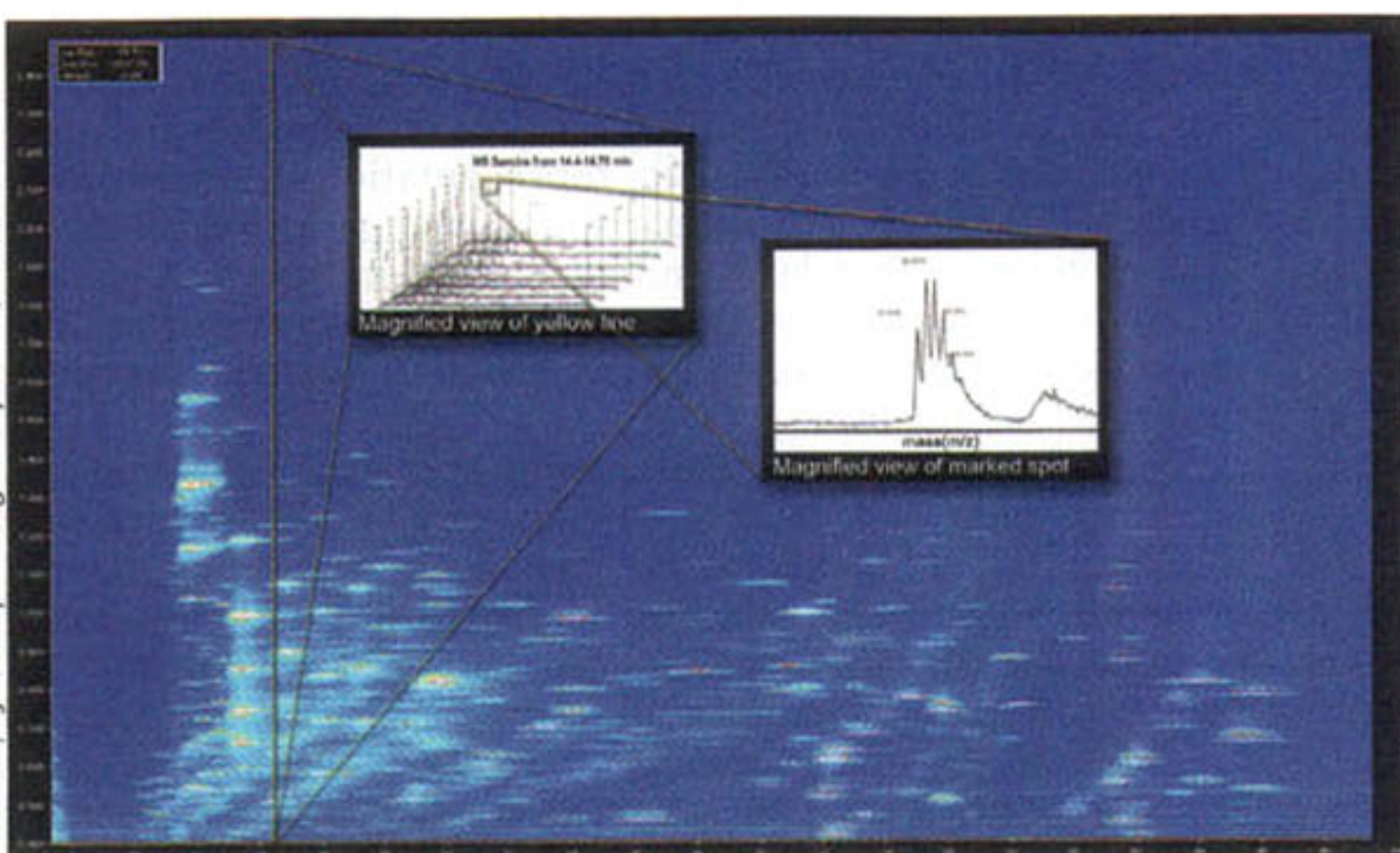
A STEP AHEAD On the other hand, proteomics may ultimately be a more informative exercise. Two steps removed from genomics (through mRNA), proteomics more directly relates to phenotype, says Omenn. Many downstream effects are the result of multiple gene products interacting with each other, as well as interactions of genes with nongenetic variables such as metabolism, behaviors, exposure to infectious agents or chemicals, or social and psychological interactions. The mediators for most of those interactions, Omenn explains, turn out to be proteins.

Proteins present three major kinds of complexity. First, Omenn notes, their concentrations vary enormously, perhaps up to 10 orders of magnitude in cells and the circulation, so finding methods to detect and then identify these proteins represents a significant challenge. Second, protein concentrations are dynamic, sometimes markedly changing with diurnal variation, stress, or disease. Third, proteins can be modified by cleavage or

FOUR VIEWS OF BIOMARKER DISCOVERY: Using Ciphergen's ProteinChip software, spectra can be visualized in a variety of different graphical formats, combined in synthetic maps, and screened for group-specific differences by the Biomarker Wizard, a software tool enabling the rapid discovery of single biomarkers.



Courtesy of Mosaïque Diagnostics/DiaPat, Harald Mischak



DELVING INTO DATA: *DiaPat* uses a capillary electrophoresis/mass spectrometry approach to screen for biomarkers. This plot (of migration time versus mass-to-charge ratio), obtained after one 50-minute run, illustrates the information challenge the company faces. Signal intensity is color-coded, ranging from blue to white. The spectra contained within the yellow line are shown above, and one of the peaks (circled) is further enlarged to illustrate the data's resolution.

by addition of new functional groups, changes that may affect both folding and function.

Many scientists are interested in specifically identifying the "sub-proteomes" of these modified proteins. Studying phosphorylated proteins and peptides (the "phosphorylome"), for example, may allow them to dissect "a snapshot of the state of the circuitry, the wiring diagram of what's happening in that cell," says Petricoin. This is important, he says, because "what you're looking at are the drug targets themselves." HUPO's Plasma Proteome Project is looking into other subproteomes, such as the collection of proteins with attached carbohydrate moieties (the "glycoproteome").

Not all polypeptides of interest to the medical community are whole proteins. "The vast majority of diagnostic information [is] contained within protein fragments and peptides that all reside in this low-molecular-weight region of the proteome," Petricoin observes. "It's an archive that's entirely unexplored, because all past biomarker-based efforts utilized technology such as 2DE [two-dimensional gel electrophoresis] ... and all those previous technologies couldn't resolve the region of the [lower] proteome below, say, 10,000 Daltons."

Much of the lower proteome seems to be cleavage products, and "what we're likely looking at is the product of proteases... and other biological processes that seem to be tied to the disease development," says Semmes. He adds that, for many diseases like prostate cancer, "there's nothing left to see" in the larger proteome, which researchers have been mining with 2DE for 20 years.

PROTEOMIC HARDWARE Whichever specific components researchers choose to focus on, the first analytical step is fractionation, as raw serum is too complex for matrix-assisted laser desorption/ionization-MS (MALDI-MS) or electrospray MS, Semmes says. Some researchers fractionate biological samples using multidimensional liquid chromatography. Other use 2DE, and some use a bio-panning approach.

Ciphergen's SELDI-enabling ProteinChip affinity chips selectively bind peptides based on their chemical properties, for instance, resulting in simplified MS spectra that are more easily scoured for potential biomarkers. Yet, many critics of the SELDI-generated pattern approach echo Diamandis in pointing out that

the peaks identified as markers by different laboratories, and by the same laboratory at different times, tend to differ from each other, "making validation difficult."⁴

The FDA/NCI admits on its Web site that the original Ciphergen systems have "too low mass resolution and too high mass drift for our specific needs. We could not run the same sample on the same machine at a later date and have the spectra align correctly."⁵ Ciphergen's CEO concedes the point as well: "I think they're really good research machines at this point, but they're not at a stage where they're going to be robust enough to be put into a diagnostic setting easily," says Rich. He adds that the next generation of systems, due out around the end of the year, "will really be robust enough to be used in a reference laboratory setting and in an esoteric testing setting."

The current FDA/NCI system, from Applied Biosystems in Foster City, Calif., is the Hybrid Pulsar QqTOF instrument (Q-Star) fitted with a Ciphergen SELDI source. It "has given us the opportunity to bin data such that the same intensities fall into the same 'bucket' every day, every week, and every machine," the Web site says. "The goal is reproducibility."

THE ALTERNATIVES Other pre-MS fractionation systems are in the works. For example, South San Francisco-based startup Biospect has a microfluidics-based sample preparation system in the pipeline. It is the task of HUPO's Plasma Proteome Project to compare these technologies.

"We don't expect one way to perform best for every need," says Omenn. The choice will depend on budget, the required level of detail and throughput, and the biological and clinical question that's being investigated, he says.

If the proteins to be investigated are known ahead of time, investigators have a wide variety of protein microchip and other techniques at their disposal to examine them. For example, Milagen in Richmond, Calif., uses antibodies to pan for biomarkers.

At NCI, Weinstein has taken the opposite approach, binding serum to chips and probing it with antibodies. NCI's research center has established a committee to evaluate the appropriateness of using these "reverse-phase arrays" in each of its clinical trials, relates director J. Carl Barrett.

Companies such as Diapat in Hannover, Germany, have begun capitalizing on the multiple-marker approach to diagnostics, using its capillary electrophoresis MS technology to screen for kidney and other disorders. In the meantime, several clinical trials are underway or pending to examine whether diagnoses based on SELDI-generated patterns will prove viable alternatives to the current single-marker tests.

Proteomics' move to the clinic could be imminent. But lest researchers count their chickens too soon, Anderson cautions that, "discovery stops short, by a long way, from the kind of evidence that is required to constitute validation of a diagnostic marker." ☉

Josh P. Roberts (tcwriter@msn.com) is a freelance writer in Minneapolis.

References

1. E.F. Petricoin et al., "Use of proteomic patterns in serum to identify ovarian cancer," *Lancet*, 359:572-7, 2002.
2. B.L. Adam, "Proteomic approaches to biomarker discovery in prostate and bladder cancers," *Proteomics*, 1:1264-70, 2001.
3. K.P. Rosenblatt et al., "Serum proteomics in cancer diagnosis and management," *Ann Rev Med*, 55:97-112, February 2004.
4. E.P. Diamandis, "Point: Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics?" *Clin Chem*, 49:1272-5, 2003.
5. <http://ncifdaproteomics.com/ppatterns.php>