

# How Industry Is Approaching the Search for New Diagnostic Markers and Biomarkers\*

J. Werner Zolg†§ and Hanno Langen¶

In the diagnostic and the pharmaceutical industry there is a constant need for new diagnostic markers and biomarkers with improved sensitivity and specificity. During the last 5 years, only a few novel diagnostic markers have been introduced into the market. Proteomics technologies are now offering unique chances to identify new candidate markers. Before a marker can be introduced into the market, three successive developmental phases have to be completed: the discovery phase, in which a variety of proteomics technologies are applied to identify marker candidates; the prototype developmental phase, in which immunological assays are established and validated in defined sample collectives; and finally the product development phase, with assay formats suitable for automated platforms. The hurdles that a potential candidate marker has to pass in each developmental phase before reaching the market are considerable. The costs are increasing from phase to phase, and in industry a number of questions concerning the medical need and the potential return on investment have to be answered before a proteomics discovery project is started. In this review, we will cover aspects of all three developmental phases including the repertoire of discovery tools for protein separation as well as giving an outline of modern principles of mass spectrometry for the identification of proteins. *Molecular & Cellular Proteomics* 3:345–354, 2004.

The proteomics activities in industry aim for two different outcomes: the identification of proteins or pathways that can be used as drug targets represents traditionally a main focus for the pharmaceutical industry. Conceptionally different is the search for proteins (or parts thereof) in the diagnostic industry. Here, analytes that are correlated with disease states (referred to as diagnostic markers) are in the focus of proteomics discovery programs.

A biomarker is a molecule that indicates an alteration of the physiological state of an individual in relation to health or disease state, drug treatment, toxins, and other challenges of the environment. By this definition, a biomarker is not static, it

is changing over time. Therefore, a genetic predisposition, like a single nucleotide polymorphism, is not considered a biomarker.

Biomarkers play an important role in the pharmaceutical industry and are assuming an ever greater role in drug discovery and development (1). The potential benefit of biomarkers is to allow earlier, more robust drug safety and efficacy measurements. Generally, a better understanding of the mechanism of disease progression and therapeutic intervention is needed. The major challenge is the selection and validation of biomarkers including clinical endpoint validation. This requires extended clinical studies to compare the new biomarkers with clinical endpoints determined by other means. Only when this correlation is established will biomarkers identified by proteomics be accepted as surrogate markers.

Many diagnostic markers are also biomarkers. But not all biomarkers meet the criteria set for diagnostic markers in respect to specificity and/or sensitivity. They can still be used for e.g. monitoring inflammation during drug treatment. Inflammation markers generally lack the specificity to identify a given disease, but their reduction during drug treatment can be taken as an early indication of drug efficacy.

Before focusing in the context of “biomarkers” and “diagnostic markers” exclusively on proteomics activities and consequently protein-based markers, one should have in mind that a whole category of markers is actually based on non-protein targets. Although not the subject of this review, diagnostic approaches targeting DNA and specific gene alterations in the DNA (single nucleotide polymorphisms) based on amplification technologies such as PCR are getting more and more attention in the context of predisposition testing, e.g. correlating single nucleotide polymorphisms with the probability of developing a certain disease.

On the mRNA level, expression profiles of given biological material are monitored based on hybridization events using chip-based test formats (see “Genchips”). One of the many applications of such mRNA expression chips is the stratification of patients, e.g. to classify individuals for their ability to metabolize certain drugs.

## POSITIONING OF DIAGNOSTIC MARKERS

Marker discovery by proteomics is based on the assumption that “disease” can be defined as an altered flow of information in a biological system. One group of information carriers are in fact proteins. Consequently, the type and con-

From †Roche Diagnostics GmbH, Centralized Diagnostics, Nonnenwald 2, D-82377 Penzberg, Germany; and ¶Hoffmann-La Roche Ltd., Roche Center for Medical Genomics, CH-4070 Basel, Switzerland

Received, January 21, 2004

Published, MCP Papers in Press, January 28, 2004, DOI 10.1074/mcp.M400007-MCP200

centration of a protein (or a group of proteins) at any given time in a proteome and the correlation of those patterns present in a “disease state” as compared with a “healthy state” can be of high diagnostic value. These disease-specific proteins can be used in different diagnostic applications:

1. Screening markers are markers discriminating the “healthy” state from a beginning “disease” state, preferentially in the asymptomatic phase. Targets for the screening approach can be entire populations or population subgroups being tested regularly at given time intervals. Examples for screening markers would be diagnostic markers for the early detection of cancer events.
2. Prognostic markers are markers that, once the disease status is established, predict the likely course of the disease and thus influence the aggressiveness of the therapy. In rheumatoid arthritis (RA),<sup>1</sup> it is of importance to discriminate “slow” from “fast” progressors, to treat the latter ones with more aggressive drugs to halt joint destruction.
3. Stratification markers, although often DNA-based (see above), can also be proteins (or protein families) and be used to predict the likely response to a drug before starting treatment classifying individuals as “responders” as compared with likely “nonresponders.” This is of major importance in the design of clinical drug trials to define an intended use for the drug under investigation.
4. Efficacy markers are markers that, once the responder status is established, can be used to monitor the efficacy of drug treatment. Again referring to RA, following e.g. the collagen turnover with a set of diagnostic markers can be used to follow the effectiveness of a given drug over time.

### DECISION MAKING IN THE INDUSTRY

Industry is approaching the search for new markers from a different perspective than academic institutions. The scientific value and the gain of knowledge, which are key drivers in an academic research setting, is complemented within the industry by considerations reflecting potential commercial value creation. To this end, a series of questions has to be addressed in industry before any R&D activities are initiated to identify either biomarkers and/or diagnostic markers. Does the marker in question fulfill a true medical need and will its use change medical practice? Will the marker be accepted by

the medical community and how high will be the barriers-to-entry into the market? Will the proposed marker be a “stand-alone marker” or will the marker be part of an existing product line? What will be the competitive situation for the new biomarker (availability of markers with similar claims) or do radically different solutions than protein markers (e.g. imaging technologies) already exist? Can the program be realized with the existing knowledge base and the technical solutions at hand? And, most critical for the ranking of competing proposals, do the incidence and prevalence figures of an indication area under consideration justify the long-term investment of resources, both in manpower and funding? It is mandatory in our industry to answer these and similar business-related questions and compile them into a business plan before submitting any proposal to an internal review board.

These fundamental questions emphasize the different decision trees for the selection of targets between the academic world and the industry. Colorectal cancer (CRC) can serve as an example. CRC is the third most common cancer worldwide, with an incidence of >400,000 new cases a year and a prevalence of >1.6 million in seven major markets (2). The lifetime risk of developing CRC is about 6%, with a sharp increase over the age of 50, making age the most important risk factor for the disease, ahead of genetic and/or familial traits. Of diagnostic relevance is the fact that CRC can be effectively cured through early detection and intervention. Some 37% of all CRC cases are detected in their early localized stages or stages preceding the cancer event (Dukes stage A) or in stages where no invasion of the lymph nodes has yet occurred (Dukes stage B), conditions with a 5-year survival rate of 93% after curative surgery. After infiltration of the lymph nodes (regional stages, Dukes stage C, another 37% of the cases), the survival rate drops to 63%, to fall to only 9% in the distant stages correlated with metastasis (Dukes stage D). This segmentation already underlines the need for early diagnosis in a market segment of important size. The problem of early diagnosis for CRC is compounded by the fact that the available screening methods do have considerable limitations. Still widely used is the test for occult blood in stool. Some nonmalignant conditions (hemorrhoids and ulcers) can lead to false-positive results, as can certain diets and medications (3). Despite the fact that fecal occult blood testing does not reach a satisfactory level of sensitivity and specificity, the test reduces CRC mortality by detecting “true positives” and thus is recommended by many national cancer agencies simply because of the absence of a better test. Colonoscopy is seen as a “gold standard” in CRC testing, but the testing procedure is, by its very nature, highly invasive and thus does not have the wide-spread compliance of the target population (people >55 years of age).

In summary, in CRC screening there is a large diagnostic gap between the established test regimen of stool testing (with all its limitations) on one side and a reliable early detection procedure (which lacks the much needed acceptance) on

<sup>1</sup> The abbreviations used are: RA, rheumatoid arthritis; CRC, colorectal cancer; MS, mass spectrometry; 2D, two dimensional; MudPIT, multidimensional protein identification technology; MALDI, matrix-assisted laser desorption/ionization; ESI, electrospray ionization; LC, liquid chromatography; TOF, time-of-flight; MS/MS, tandem MS; SELDI, surface-enhanced laser desorption/ionization; MRM, multiple reaction monitoring; MTP, microtiter plate; ROC, receiver-operator characteristic; CRP, C-reactive protein.

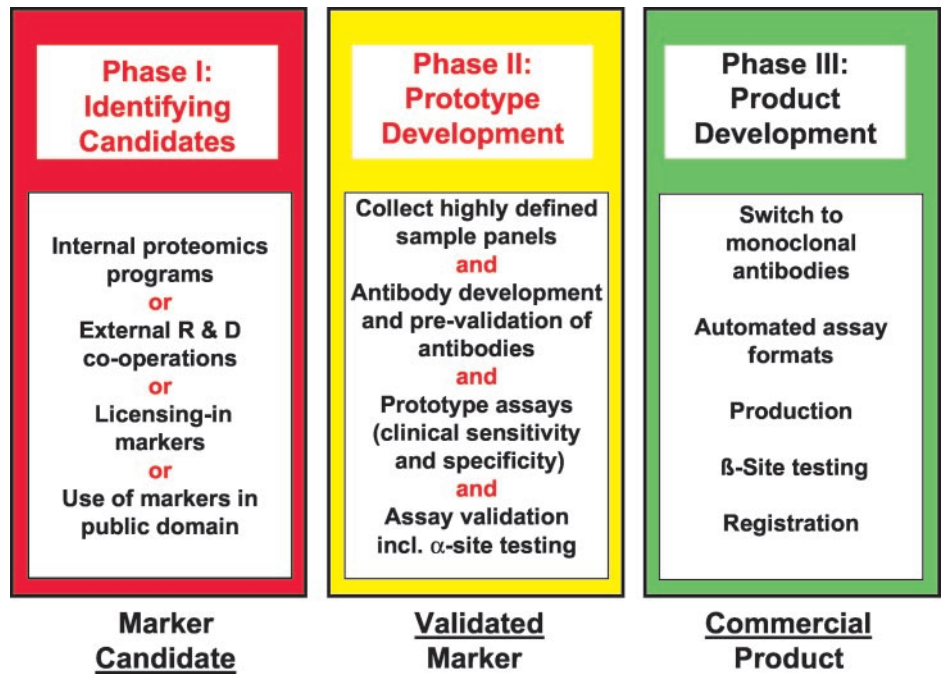


FIG. 1. Example for the workflow in industry from the discovery phase to a commercial product.

the other side. Any new screening marker for the early detection of CRC would be a definite improvement likely be accepted by the patients and the medical community. The example of CRC can therefore be viewed as a model target area of high interest for the diagnostic industry fulfilling most (if not all) industry conditions referred to in the questions above (“Decision Making in the Industry”). That this overlap between key industry considerations and CRC is a close to perfect match is evidenced by the many parallel R&D activities in CRC across the industry to identify suitable diagnostic markers.

SEARCH FOR NEW MARKERS: GENERAL WORKFLOW

Independent of the target, it takes three distinct developmental phases from the identification of a protein to the launch of a diagnostic product based in an immunoassay format (Fig. 1). At the end of the discovery phase, one has an array of marker candidates. The next phase deals with prototype development, resulting in validated markers, followed by the product development, leading to a commercial product. Key elements of the three phases are described below.

PHASE I: DISCOVERY PHASE

Once the decision is made to initiate the search for new diagnostic markers in a particular indication area, industry is generally evaluating four options: 1) the evaluation of markers that are already in the public domain either as stand-alone markers or in combination with markers of the already existing portfolio aiming to improve sensitivity and/or specificity; 2) alternatively, proprietary markers protected by patents can be licensed-in; 3) discovery programs can be conducted with external partners freeing resources for 4) own internal R&D efforts in an indication area considered of higher priority.

*Proteomics Methods for the Identification of Diagnostic Markers*

Proteomics research combines high-resolution separation techniques applied to complex protein mixtures with state-of-the-art identification methods such as mass spectrometry (MS). It is generally agreed that none of the existing separation and identification methodologies on its own can give a full account of the protein composition or the protein expression in complex mixtures, e.g. biological fluids such as serum, synovial fluid, cerebrospinal fluid, urine, whole cells, cell fractions, or tissue extracts. This limitation, however, has not prevented the use of existing methods (or the combination of several existing technologies) to provide valuable information on a wide range of proteins, especially when either their absence or presence or their level of expression can be correlated to a disease state. There are two strategies for the search of protein biomarkers in body compartments. One is the direct search in the peripheral fluids where the marker concentration is expected to be relatively low and today’s detection methods will not allow their identification. The other approach is the search for new markers in diseased tissue where the marker protein is presumably present at a much higher concentration, facilitating the protein identification by MS as discussed below. An inherent risk in the tissue approach is the fact that the candidate marker identified in e.g. tissue cannot later be detected in peripheral fluid such of serum.

*Removal of Abundant Proteins from Serum*—The biggest hurdle to overcome in the discovery phase is the fact that the analytical tools used at the end of the process chain such as MS have a definite detection limit for finite amounts of pro-

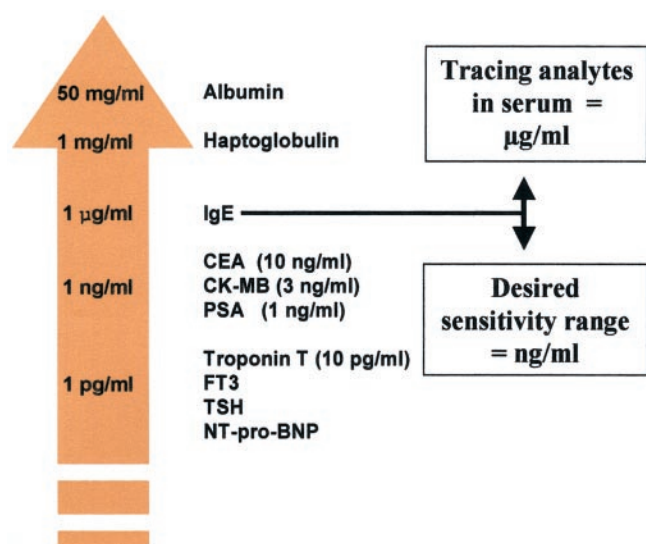


FIG. 2. **Dynamic range of serum proteins (approx. concentrations).** Abbreviations: CEA, carcinoembryonic antigen; CK-MB, creatine kinase muscle-brain isomer; PSA, prostate-specific antigen; FT 3, free triiodothyronine; TSH, thyrotropin; NT-pro-BNP, N-terminal pro-brain natriuretic peptide.

teins (or peptides derived thereof). To fully exploit the sensitivity limits for peptide identification by MS, it is necessary to enrich the remaining protein mixture for the potential marker candidate. To this end, the concentration of abundant proteins present in complex peripheral fluids such as serum has to be reduced as much as possible. The seven most abundant proteins in serum (albumin, immunoglobulins both large and small chains, transferrin,  $\alpha$ -macroglobulin, antitrypsin, and haptoglobulin) already amount to some 97% of the total serum proteins (4). If one adds up the 30 most abundant proteins present in a concentration of  $>100 \mu\text{g/ml}$  serum and assuming an average total protein content of some 80 mg/ml serum, then these “nontarget” proteins amount to about 99% of the total proteins, leaving just 1% of all proteins present as prime targets for the identification of novel markers. Assuming that the markers of interest are not bound or complexed to one of those major proteins, they in fact have to be removed early in the process in order to reduce the complexity. Size-exclusion chromatography under both denaturing and non-denaturing conditions is one of the options at hand to separate abundant proteins of a larger size from those in the 20- to 50-kDa range. For albumin removal, common affinity adsorbers based on dyes such as Sepharose Blue® and other modifications are used (5). Immunoglobulins are commonly removed using Protein G affinity chromatography (6). In our proteomics depletion strategy, we are using a series of immuno-adsorbers with specific monoclonal antibodies immobilized to magnetic beads on a preparative scale. On an analytical scale, similar immuno-adsorbers (Multiple Affinity Removal System) are commercially available, removing six abundant proteins such as albumin, IgG, IgA, transferrin, an-

titrypsin, and haptoglobulin (Agilent Technologies, Waldbronn, Germany). This attempted reduction of complexity becomes even more challenging if one looks at the dynamic range of the protein distribution (Fig. 2). Taking albumin with about 50 mg/ml as an example, if any purification scheme were able to remove 99.9% of albumin from the serum, the remaining (contaminating) concentration of albumin would still be 50  $\mu\text{g/ml}$  or a factor 50,000 higher than well-known tumor markers such as the prostate-specific antigen present in about 1 ng/ml.

Other diagnostically relevant analytes are present in even lower concentrations (Fig. 2). These two key factors, the overall concentration of abundant proteins (masking the potential marker candidates) and their dynamic range, have to be addressed in any purification strategy. In addition, the ratio of potential marker candidates to known proteins (which are carried over after the removal step or are difficult to remove in the first place) has to be increased further in favor of the unknown candidates by applying common protein separation steps such as ion exchange chromatography or columns based on hydrophobic interactions before attempting MS analysis. Independent of the overall analytical strategy followed (e.g. two-dimensional (2D) gels or multidimensional protein identification technology (MudPIT), see following section), the comparison of the peptide maps or peptide sequences between “healthy” and “diseased” specimen after MS are used to identify the respective proteins in databanks. Those annotated proteins are at the end of phase I, the proteomics discovery phase (Fig. 1).

*Tissue Approach*—In the tissue-based discovery of protein markers, the large difference in protein expression has to be overcome to a certain extent. The following methods can be used for fractionation:

1. Specific enrichment of the cell types involved in the disease process. Laser capture microdissection (LCM) is one of the most specific ways to obtain cells out of tissue that are related to the disease process. The retrieval of selected cells is achieved by capturing them on a transfer film placed in contact with a tissue section after directing a laser beam to the area of interest with the help of an inverted microscope. The use of the LCM system in combination with 2D electrophoresis has been demonstrated in several recent studies (7, 8).
2. Subcellular fractionation by differential centrifugation is one of the classical biochemical methods for protein purification. Intact organelles are purified by the different sedimentation coefficients. Plasma membrane proteins and secreted proteins are the main sources of proteins leaking into plasma. In addition, proteins from other organelles like mitochondria or the nucleus have been identified as diagnostic markers. Many proteins that are present at a low concentration in a crude

extract can be successfully traced in the subcellular fractions (9–12).

3. Chromatography-based protein separation with or without the combination of subcellular fractionation is a further step to increase the proteomics coverage of differences in protein expression. Many different types of chromatography methods can be applied like ion exchange chromatography, hydroxyapatite (13), chromatofocusing (14), hydrophobic interaction chromatography (15), and heparin chromatography (16), to list a few options.
4. Affinity-based protein purification can selectively enrich certain subclasses of proteins. Antibodies against specific proteins can reveal not only the target protein of the corresponding antibody but also the functional interaction partners. The specificity of this approach is poor because many proteins are sticky and can interact nonspecifically with other proteins. This method was improved by the introduction of the so-called tandem affinity purification protocol (17). Because in the search for biomarkers the nature of the protein of interest is usually unknown, the affinity-based methods are not widely used in the early stages of the purification schemes.

#### *Protein Identification Methods*

Mass spectroscopy is the method of choice for the analysis of proteins and peptides. One modification of MS, the matrix-assisted laser desorption/ionization technology (MALDI) (18), is ideally suited for analyzing the low-complexity mixtures of proteins derived from protein spots after separation by 2D electrophoresis (see “2D Gel Electrophoresis”).

Another modification of MS is electrospray ionization (ESI), which is a solution-based analytical technology (19) optimized for coupling liquid chromatography (LC) technology directly to the MS detection units. We are using the ESI-MS modification preferentially for proteins and peptides of low molecular mass that cannot be resolved by 2D gels. Another class of proteins that is not amenable for the separation by 2D gels and lends itself to the analysis by ESI are the hydrophobic membrane proteins. These proteins are analyzed by preparing the corresponding subcellular fractions (see “Tissue Approach”) followed by one-dimensional SDS electrophoresis. The bands are cut out, the proteins in the gel pieces are digested by trypsin, and the resulting peptides are injected onto a capillary reversed-phase column coupled to MS (LC-MS). The separated peptides are directly sprayed into the mass spectrometer, and the peptide masses are recorded (20). In MS modifications such as ion trap instruments (21) or quadrupole instruments (22), the peptides can be further fragmented and the fragment information can be used for further database searches.

One of the most important recent developments in MS is

the increased quality of the mass spectral results. This is made possible by the introduction of a new generation of mass spectrometers, the “time-of-flight/time-of-flight” (TOF/TOF) instruments (23). With MALDI-TOF/TOF instruments, MS fragmentation spectra can be obtained within a few seconds.

To optimally use of the analytical capability of these instruments, prefractionation schemes have to be developed to overcome the differences in expression rates of proteins (see “Removal of Abundant Proteins from Serum”). Those fractionation methods need to be applied before the proteins are separated by 2D gels or, in the case of shotgun approaches, before the proteins are digested with trypsin and separated by high-performance LC.

**2D Gel Electrophoresis**—After nearly 30 years (24), 2D gel electrophoresis is still the method of choice for high-resolution protein separation (25). On a 2D gel, up to 1,000 protein species can be separated followed by the identification with MS. Why is there still a resistance to the widespread use of 2D gels? One reason is the substantial workload involved in performing this separation technology. The hands-on time has somewhat been reduced by the availability of prefabricated strips for the first dimension (isoelectric focusing) and with pre-cast gels used in the second dimension. Another hurdle in the acceptance of 2D gels are the in-built limitations of the technology. Several classes of proteins are difficult to resolve, among them the very basic proteins, small proteins, and the hydrophobic proteins like G-coupled protein receptors with several transmembrane regions. Despite all these limitations, 2D gels were successfully used for differential protein display many years even before the term “proteomics” was coined.

In most of the proteomics-based work, a 2D gel-matching procedure (gel imaging) is used to compare two sets of protein mixes run under highly standardized conditions. A minimum of two parallel gels (preferentially three gels) have to be run to obtain a sound basis for applying the image technology and to directly compare preparations of protein mixtures from “disease” and “healthy” states. The imaging software recognizes the spot differences, and only those different spots are further analyzed. But even the most advanced imaging software packages still need a manual intervention to assist in the correct matching. Even under the most standardized conditions, not all protein spots on two given gels can be correctly matched, and thus ambiguities will arise. In addition, with the imaging approach misleading protein identifications can occur because proteins are identified irrespective of their origin including protein isoforms. Assuming that a given protein is present in three isoforms in the “disease” state and only in two isoforms in the “healthy” state, the expression of an additional protein in the “disease” state would be scored by the imaging technology as a “new protein,” whereas in fact the same protein (in an additional isoform) has been expressed.

To avoid some of these problems related to the image analysis of 2D gels, we have implemented in our proteomics

strategy the picking of all spots on the gels and determine the differences in corresponding gel sets after the peptide analysis by MS. The resulting protein list is more extensive and complete than the lists obtained by the imaging procedure. Moreover, we have observed in many occasions that two different proteins can be hidden in one single protein spot, a fact only revealed after comparing the peptide files. These “hidden” proteins would have been missed by applying the “imaging technology.” We have automated our approach to analyze each and every spot after 2D gel electrophoresis to a very high degree. We use appropriate software and “spot-picking robots” that transfer the gel pieces to microtiter plates, which in turn are processed in automated washing stations followed by automated trypsin digestion and subsequent spotting on MS target plates. With this automated workflow, we are able to process as many as 15,000 spots per day.

**Multidimensional Chromatography**—While the use of one-dimensional and 2D gels is considered essential to many proteomics approaches, the emergence of shotgun sequencing based on high-performance LC and tandem MS (MS/MS) is a powerful alternative (26). Both single-dimensional high-pressure LC and multidimensional LC (LC/LC) can be directly interfaced on-line with MS to allow for automated collection of large datasets (20). This approach is known as MudPIT (27). In the shotgun proteomics approach, one does not analyze the intact proteins but peptides generated by specific enzymes such as trypsin. This fragmentation can offer definite advantages because even very large proteins, very hydrophobic, or very basic proteins (protein classes that are difficult to handle) will give rise to peptides of sufficient size. However, because several peptides are generated from each protein, the complexity of the mixture to be analyzed is increased. Consequently, more instrument time and computing power are needed for the shotgun approach as compared with the 2D gel approach. In addition, in the shotgun approach, the information about the integrity of the original protein or its potential posttranslational modifications are lost to a large extent. The shotgun approach is therefore better suited for less complex mixtures. With the increased number of different peptides, the matching of the mass spectral fingerprints with the protein databases becomes more demanding and time consuming and leads, without careful manual inspection of the spectra, easily to false-positive identification of proteins, e.g. the tentative identification based on one peptide only (“one-hit wonders”). We have extensively studied the information gain achieved with the MudPIT approach in comparison to the 2D gel approach. We consistently find that both technologies are highly complementary to each other, e.g. proteins identified by the 2D gel/MALDI approach were not found with MudPIT. On the other hand, additional proteins were in fact found with MudPIT that were never seen with the 2D gel/MS method. In consequence, we routinely analyze all protein mixtures by both strategies in parallel.

**Peptidomics**—Peptides play a central role in many physiological processes. In order to analyze comprehensively all “natural” peptides and small proteins of a whole organism (peptidome), an approach described as peptidomics is used (28). Neither the 2D gel approach (which is limited to proteins >10 kDa (29)) nor the shotgun approach can therefore be used to analyze the naturally occurring peptides. The workflow of peptidomics usually starts with the removal of the proteins over 20 kDa by size-exclusion filtration or size fractionation with membranes of suitable pore size. Alternatively, precipitation techniques using trichloroacetic acid can be employed. Next, the peptides can be identified by fragmentation in modern MS/MS instruments like the TOF/TOF or quadrupole/TOF instruments.

**SELDI**—The surface-enhanced laser desorption/ionization (SELDI)-TOF-MS technology (30) uses chromatographic surfaces coupled to the target plate. The plates are then extensively washed and the bound material is directly analyzed by MALDI-MS. SELDI covers peptides and proteins predominantly in the low molecular mass range. This technology is limited to the major abundant peptides and proteins as long as a suitable upfront purification scheme is not integrated (31). The SELDI technology leads only to a pattern and not to the identification of peptides as is the case in the peptidomics approach (28). In the published examples using SELDI, only about 20–50 well-resolved peaks are detected (32), compared with several hundred peptides species detected in the peptidomics strategy. For the best possible reproducibility and mass accuracy using SELDI, high-resolution MS is mandatory.

**MALDI Imaging**—MALDI-MS has been used to generate protein and peptide mass lists of tissues directly mounted to the target plate of the mass spectrometer. The plate is scanned in the corresponding xy coordinates to give a list of masses that correspond to the molecular components present (33). The high sensitivity of the technique (low femtomole to attomole levels for proteins and peptides) allows the study of biomarkers for example in tissue sections from cancer patients to generate a molecular fingerprint of each tumor type. Using this technology, Yanagisawa *et al.* were able to classify several subtypes of lung cancer tumors (34).

**Genchips**—In several recent studies, cDNA microarray analysis has revealed significantly elevated expression of the prostaticin and osteopontin genes in ovarian cancers, and correspondingly high levels of these proteins have subsequently been detected in the serum of ovarian cancer patients (35). This proves that transcript profiling to identify genes encoding secreted proteins in human carcinomas can be achieved using commercially available microarrays. Genchips led to the identification of 2,300 genes and classified them as being extracellular. Seventy-four of these genes were overexpressed in one or more carcinoma types relative to healthy control tissue. The overexpression of several of these genes was confirmed at the protein level in the serum of cancer patients.

*Selective Ion Monitoring*—Before starting the development of immunological tools for a given marker candidate (phase II in Fig. 1), especially if the discovery phase was conducted using tissue as sample material, we verify the presence of the marker candidate in question in serum by applying the multiple reaction monitoring technology (MRM) (36, 37). MRM combines the quantitative bioanalytical LC-MS/MS with stable isotope labeling of peptides. MRM has been shown to be a powerful technology for the quantitative analysis of peptides in complex mixtures like plasma and is well established in pharmacokinetic studies of small molecules but has not yet found a widespread use in proteomics studies. Recent publications describe the use of  $^{13}\text{C}$ -labeled peptides as internal standards for an accurate quantification of proteins in plasma (35, 38). Two or three tryptic peptides of the protein to be analyzed are selected for peptide synthesis based on their ionization properties. During synthesis,  $^{13}\text{C}$  is incorporated into the leucine moieties of the peptides. The plasma proteins to be analyzed are digested with trypsin and the labeled peptides are spiked in. The ionization properties for the labeled synthetic peptides are the same as for the naturally occurring nonlabeled peptides, the only difference being the mass difference due to the number of  $^{13}\text{C}$  atoms incorporated in the amino acid leucine of the synthetic peptides. Knowing the concentration of the labeled peptide, a direct quantification between the synthetic and the corresponding “natural” peptides can be made. In a recent study, candidate markers for RA identified in synovial fluid could be quantified directly in serum.<sup>2</sup>

#### PHASE II: PROTOTYPE DEVELOPMENT

##### *Sample Banks*

Before the transfer of a marker candidate (after verification by MRM) into the prototype development phase (Fig. 1), a complex infrastructure regarding samples has to be in place in an industrial setting. These indication area-specific sample banks used in a developmental program in industry have to satisfy different needs (including regulatory ones with detailed patient consent forms and approval by ethic commissions) than selected samples used in an academic environment. Most likely different between the two settings are both the sample bank size and the degree of sample characterization before a sample can be included into the respective collection. We are opting for 500–1,000 samples per medical condition with up to (and sometimes exceeding) 100 individual data points per patient. These large collectives allow the formation of statistically valid patient subgroups according to disease progression or previous medical treatment. In the case of serum and/or plasma, sufficiently large volumes per patient (several ml) are another prerequisite facilitating the direct comparison of dozens of markers with the very same

biological specimen over extended periods of time. We also place great emphasis on characterization of samples to be included in the “healthy control” group. The absence of the disease under investigation in “healthy” individuals has to be specifically confirmed, e.g. in the case of RA, all individuals included in the “healthy” control group were seen by a rheumatologist who confirmed the absence of any rheumatological symptoms based on anamnesis and extensive check lists as well as laboratory parameters. It is prudent, apart from having age- and sex-matched samples in the “healthy” collective, to also take the potential co-morbidities into consideration as well as the duration and type of any medication taken. It is often underestimated that the time frame for building up suitable sample collections can take years, especially when chronic progressive diseases such as RA or Alzheimer’s disease have to be monitored by prospective follow-up sampling. It takes a highly computerized bioinformatics infrastructure to design versatile case report forms and appropriate retrieval systems to allow the selection of any desired combination of clinical manifestation in an individual sample before the prototype developmental phase can be sensibly stated in an industrial setting.

##### *Immunological Tools*

Many of the strategies to set up immunological prototype assays are explained in the respective tutorials and are not the subject of this review (39). For reasons of throughput, we are working in the prototype developmental phase exclusively with polyclonal antibodies raised in rabbits using the 105-day bleeds throughout. For immunization, we follow a dual strategy. Several peptides are selected per protein using appropriate software such as compiled on the Expert Protein Analysis System (ExPASy; at [us.expasy.org](http://us.expasy.org)) to identify nonhomologues regions of high immunogenicity. In parallel, we express the full-length protein by means of recombinant expression systems starting with but not restricted to *Escherichia coli*. Although each particular assay requires a tailored strategy and many options have to be considered, one of the many modifications we use is as follows. Anti-rabbit antibodies are immobilized on the walls of microtiter plates (MTP), and the raw serum (containing the antibodies in question) is added followed by the biotinylated peptide used to immunize. The antibody-captured biotinylated peptide is then detected by streptavidin coupled to horseradish peroxidase. This robust entry-screening system allows a first estimation of the combined antibody concentration and antibody affinity and thus a selection of the serum charge to be followed-up. Next, the IgG fraction of the selected polyclonal antibody pool is purified using conventional purification schemes. The recombinant full-length protein is used to establish a standard curve to obtain more detailed information as to binding characteristics of the antibodies intended for the set-up of prototype assays.

<sup>2</sup> B. Guild, personal communication.

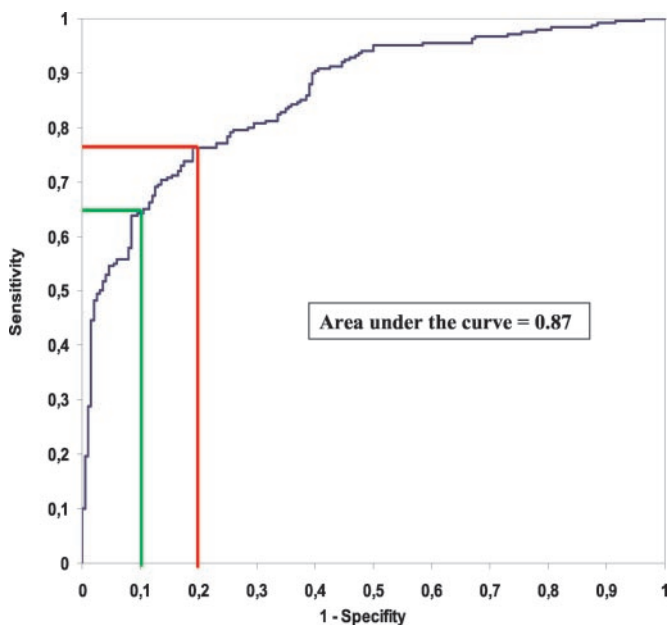


FIG. 3. ROC for CRP in RA versus controls (blue plot). Corresponding specificity and sensitivity is shown for two examples. Setting the specificity at 80%, the resulting sensitivity is 76% (red line). If the specificity is raised to 90%, the corresponding sensitivity will be 64% (green line).

In certain applications such as cancer, it is helpful to test the specificity of the antibodies before setting up the prototype assay format. To this effect, we use e.g. lysates of different cancer tissues and probe them in the Western blot format with the antibodies in question, often enriching the antigens by immunoprecipitating the lysates beforehand. Alternatively, immunohistological screens with immobilized cells obtained by LCM from various tumors (see “Tissue Approach”) are used to get a first impression about the specificity of the antibodies.

### ROC Curves

After deciding on a suitable assay format and optimizing the prototype assay for minimal interference (eliminating matrix effects), a receiver-operator characteristics (ROC) plot is established (40) using a panel of a minimum of 250 control samples and 250 highly characterized samples of diseased individuals. The clinical performance of a laboratory test depends on its diagnostic accuracy or the ability to correctly classify subjects into clinically relevant subgroups (e.g. “healthy” versus “diseased”). The ROC plot depicts the overlap between the two populations by plotting the sensitivity versus  $(1 - \text{specificity})$  (Fig. 3). The true-positive fraction (sensitivity) is plotted on the y axis and is defined as the  $(\text{number of true-positive test results}) / [(\text{number of true-positive test results}) + (\text{number of false-negative test results})]$ . This is referred to as “positivity” in the presence of a disease and is calculated solely from the affected subgroup. On the x axis, the false-positive fraction is plotted or  $(1 - \text{specificity})$  defined

as the  $(\text{number of false-positive results}) / [(\text{number of true-negative results}) + (\text{number of false-positive results})]$ . It is an index of specificity and is calculated entirely from the unaffected subgroup. Because the true- and false-positive fractions are calculated from two different subgroups, the ROC plot is independent of the prevalence of disease in the sample cohort.

A test with perfect discrimination (no overlap in the distribution of the two subgroups) has an ROC plot that would pass through the upper left corner, where the true-positive fraction is 1.0, or 100% (perfect sensitivity), and the false-positive fraction is 0 (perfect specificity) (Fig. 3). The theoretical plot for a test with no discrimination at all (identical distributions of results for the two groups) is a 45° diagonal line from the lower left corner to the upper right corner. Most plots fall in between these two extremes. The most common measure is the area under the ROC plot. This is a quantitative, descriptive expression of how close the ROC plot of an assay is to the perfect diagnostic assay (ROC area = 1.0).

Fig. 3 shows an example of a ROC plot measuring the C-reactive protein (CRP) in a cohort of “healthy controls” versus patients with confirmed RA. The ROC area is 87%, indicating a rather high discrimination power of CRP in the two reference populations and confirming the validity of CRP measurements in RA. Depending on the diagnostic question at hand, one can set either the specificity or the sensitivity of the assay under investigation and can directly read off the corresponding sensitivity and specificity. In the example shown in Fig. 3, the specificity was set arbitrarily at 80%, resulting in a corresponding sensitivity of 76% (Fig. 3, red line). If one wishes to adjust the assay to a higher specificity, e.g. 90%, one has to take a reduced sensitivity of 64% into account (Fig. 3, blue line).

This balancing of specificity versus sensitivity is of great importance in diagnostics if one considers the different requirements for different diagnostic applications (see “Positioning of Diagnostic Markers”). Taking a potential screening marker in oncology as an example, in this particular application one does not want to miss any asymptomatic individuals that are about to develop cancer. Consequently, the sensitivity of an ideal screening marker has to be in the >95–99% range. Inevitably, based on the corresponding lower specificity of the marker (see ROC plot), few individuals will be included that in fact do not develop the disease (“false positives”). Apart from the emotional stress caused by such a misclassification, undue therapeutic intervention has to be avoided by the immediate confirmation of an initial positive test result by other diagnostic measures. The opposite requirements exist for e.g. differentiation markers used after a condition is diagnosed with some degree of certainty. In such a diagnostic situation, related disease complexes (requiring different treatment) have to be ruled out and consequently the specificity of the assay has to be very high in favor of a lower sensitivity. Such a scenario is given if one tries e.g. to dis-



criminate RA from rheumatologic conditions “other than RA” showing in fact similar or overlapping symptoms.

After the completion of performance characteristics with samples out of the highly defined collectives, the prototype assays are then evaluated for the first time in a clinical setting ( $\alpha$ -site testing, Fig. 1). This is to confirm the intended use of the assay (verification of specificity) and to conduct a technical feasibility for robustness under “non-R&D” conditions with samples as they will be available in the commercial setting once the developmental phase is concluded.

#### PHASE III: PRODUCT DEVELOPMENT

Marker candidates identified in the discovery phase and having been put through the prototype development phase and fulfilling the requirements regarding clinical sensitivity and specificity (determined by the ROC analysis) will enter the product developmental phase (Fig. 1). In this phase, extensive guidelines structure the entire product developmental process. On one hand, the mandatory requirements have to be meticulously followed and implemented in order to be in accordance with the various national and international regulations and laws governing the development of diagnostic devices. On the other hand, there is a demanding operational need to coordinate the many interactions of the different departments involved in the launch of an industrial product such as R&D, Clinical Trials, Production, Business Development, Marketing, and Regulatory Divisions. These interactions are usually structured by a milestone process (which might vary to some degree within the industry) but basically ensures that the right deliverables from each department involved are at hand in the required completeness to advance on the timetable to launch without delay. Each assay has to pass each and every of the successive milestones within the overall product developmental phase to advance to the next one.

For reasons of product safety, reproducibility, and constant supply of the raw materials, one of the first steps in product development will usually be the replacement of the polyclonal antibodies in the prototype assay configuration by monoclonal antibodies for the envisioned commercial product (Fig. 1). This replacement will have to be verified by extensive test series to show the equivalency of the two antibody classes. Next, the manual prototype assays in the MTP format used in phase II will have to be converted into an automated assay format such as Elecsys® format, the automated immunochemistry analyzer in place in our affiliation. This complex undertaking requires the (re)design of all assay reagents to make them compatible with the magnetic beads replacing the MTP walls as the solid phase and electrochemoluminescence used as the detection system in the Elecsys® test format. The automated assay system differs from the MTP formats by a much faster time-to-result (18 min *versus* hours for MTP assays) and covers a much larger dynamic range (five orders of magnitude *versus* two orders of magnitude in the corresponding MTP format). Again, test series with selected samples are

used to establish the technical performance of the assay. This includes, among other parameters, precision data, recovery rates of the analytes, reagent stability, and lot-to-lot variation. Upon confirmation of all those performance data, the standard-operating procedures are finalized and the assay is transferred to production. Different production lots are again assayed for their performance and reproducibility after different times of storage under different conditions. Upon meeting those specifications, the production lots are released for  $\beta$ -site testing planned according to rigid criteria by the Clinical Trials Department (Fig. 1). In these test series with the final product (no alterations of the design or of the entire assay handling are allowed anymore), large patient collectives are tested and datasets collected that are used to formulate the final specification claims as will be included in the package inserts. Furthermore, these data are critical to file for approval both with the national and international registration agencies. In parallel to these product developmental activities, the pre-launch activities have to be set in motion to inform the medical community and the world-wide net of affiliates that a new marker is about to be launched and commercially available.

*Acknowledgments*—We thank Norbert Wild for his helpful discussions and for providing the data on the ROC plots and Johann Karl and Marie-Louise Hagmann for their critical reading of the manuscript.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Roche Diagnostics GmbH, Nonnenwald 2, D-82377 Penzberg, Germany. Tel.: 0049-8856-60-4145; E-mail: werner.zolg@roche.com.

#### REFERENCES

- Colburn, W. A. (2003) Biomarkers in drug discovery and development: From target identification through drug marketing. *J. Clin. Pharmacol.* **43**, 329–341
- Globocan (2000) Cancer incidence, mortality and prevalence worldwide. [www.dep.iarc.fr/globocan.html](http://www.dep.iarc.fr/globocan.html).
- Mandel, J. S., Bond, J. H., Church, T. R., Snover, D. C., Bradley, G. M., Schuman, L. M., and Ederer, F. (1993) Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N. Engl. J. Med.* **328**, 1365–1371
- Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
- Travis, J., Bowen, J., Tewksbury, D., Johnson, D., and Pannell, R. (1976) Isolation of albumin from whole human plasma and fractionation of albumin-depleted plasma. *Biochem. J.* **157**, 301–306
- Haney, P. J., Draveling, C., Durski, W., Romanowich, K., and Qoronfle, M. W. (2003) SwellGel: A sample preparation affinity chromatography technology for high throughput proteomic applications. *Protein Expr. Purif.* **28**, 270–279
- Banks, R. E., Dunn, M. J., Forbes, M. A., Stanley, A., Pappin, D., Naven, T., Gough, M., Harden, P., and Selby, P. J. (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—Preliminary findings. *Electrophoresis* **20**, 689–700
- Craven, R. A., and Banks, R. E. (2001) Laser capture microdissection and proteomics: Possibilities and limitation. *Proteomics* **1**, 1200–1204
- Dreger, M. (2003) Subcellular proteomics. *Mass Spectrom. Rev.* **22**, 27–56

10. Huber, L. A., Pfaller, K., and Vietor, I. (2003) Organelle proteomics: Implications for subcellular fractionation in proteomics. *Circ. Res.* **92**, 962–968
11. Navarre, C., Degand, H., Bennett, K. L., Crawford, J. S., Mortz, E., and Boutry, M. (2002) Subproteomics: Identification of plasma membrane proteins from the yeast *Saccharomyces cerevisiae*. *Proteomics* **2**, 1706–1714
12. Rout, M. P., and Field, M. C. (2001) Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. *J. Biol. Chem.* **276**, 38261–38271
13. Fountoulakis, M., Langen, H., Evers, S., Gray, C., and Takacs, B. (1997) Two-dimensional map of *Haemophilus influenzae* following protein enrichment by heparin chromatography. *Electrophoresis* **18**, 1193–1202
14. Fountoulakis, M., Langen, H., Gray, C., and Takacs, B. (1998) Enrichment and purification of proteins of *Haemophilus influenzae* by chromatofocusing. *J. Chromatogr. A* **806**, 279–291
15. Fountoulakis, M., Takacs, M. F., and Takacs, B. (1999) Enrichment of low-copy-number gene products by hydrophobic interaction chromatography. *J. Chromatogr. A* **833**, 157–168
16. Karlsson, K., Cairns, N., Lubec, G., and Fountoulakis, M. (1999) Enrichment of human brain proteins by heparin chromatography. *Electrophoresis* **20**, 2970–2976
17. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: A general procedure of protein complex purification. *Methods* **24**, 218–229
18. Hillenkamp, F., and Karas, M. (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol.* **193**, 280–295
19. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71
20. McDonald, W. H., and Yates, J. R., III (2002) Shotgun proteomics and biomarker discovery. *Dis. Markers* **18**, 99–105
21. Wu, S. L., Amato, H., Biringer, R., Choudhary, G., Shieh, P., and Hancock, W. S. (2002) Targeted proteomics of low-level proteins in human plasma by LC/MSn: Using human growth hormone as a model system. *J. Proteome Res.* **1**, 459–465
22. Devreese, B., Vanrobaeys, F., and Van Beeumen, J. (2001) Automated nanoflow liquid chromatography/tandem mass spectrometric identification of proteins from *Shewanella putrefaciens* separated by two-dimensional polyacrylamide gel electrophoresis. *Rapid Commun. Mass Spectrom.* **15**, 50–56
23. Yergey, A. L., Coorssen, J. R., Backlund, P. S., Jr., Blank, P. S., Humphrey, G. A., Zimmerberg, J., Campbell, J. M., and Vestal, M. L. (2002) De novo sequencing of peptides using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **13**, 784–791
24. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021
25. Hamdan, M., and Righetti, P. G. (2003) Assessment of protein expression by means of 2-D gel electrophoresis with and without mass spectrometry. *Mass Spectrom. Rev.* **22**, 272–284
26. Hancock, W. S., Wu, S. L., and Shieh, P. (2002) The challenges of developing a sound proteomics strategy. *Proteomics* **2**, 352–359
27. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
28. Schrader, M., and Schulz-Knappe, P. (2001) Peptidomics technologies for human body fluids. *Trends Biotechnol.* **19**, S55–60
29. Fountoulakis, M., Juranville, J. F., Roder, D., Evers, S., Berndt, P., and Langen, H. (1998) Reference map of the low molecular mass proteins of *Haemophilus influenzae*. *Electrophoresis* **19**, 1819–1827
30. Merchant, M., and Weinberger, S. R. (2000) Recent advancements in surface-enhanced laser desorption/ionization-time-of flight-mass spectrometry. *Electrophoresis* **21**, 1164–1177
31. Tang, N., Tornatore, P., and Weinberger, S. R. (2004) Current developments in SELDI affinity technology. *Mass Spectrom. Rev.* **23**, 34–44
32. Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., and Liotta, L. A. (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577
33. Caprioli, R. M., Farmer, T. B., and Gile, J. (1997) Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760
34. Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M., and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **362**, 433–439
35. Welsh, J. B., Sapinoso, L. M., Kern, S. G., Brown, D. A., Liu, T., Bauskin, A. R., Ward, R. L., Hawkins, N. J., Quinn, D. I., Russell, P. J., Sutherland, R. L., Breit, S. N., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2003) Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3410–3415
36. Kusmierz, J. J., Sumrada, R., and Desiderio, D. M. (1990) Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. *Anal. Chem.* **62**, 2395–2400
37. Dass, C., Kusmierz, J. J., and Desiderio, D. M. (1991) Mass spectrometric quantification of endogenous  $\beta$ -endorphin. *Biol. Mass Spectrom.* **20**, 130–138
38. Kuhn, E., Wu, J., Karl, J., Liao, H., Zolg, W., and Guild, B. (2004). Quantification of C-reactive protein of serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and  $^{13}\text{C}$ -labeled peptide standards. *Proteomics*, in press
39. Diamandis, E. P., and Christopoulos, E. P. (1996) *Immunoassay*. Academic Press, San Diego
40. Zweig, M. H., and Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **39**, 561–577