Journal of proteome research reviews

Proteomics: From Basic Research to Diagnostic Application. A Review of Requirements & Needs†

Frank Vitzthum,*,‡ Fritz Behrens,‡ N. Leigh Anderson,‖ and John H. Shaw§

Dade Behring Marburg GmbH, Emil-von-Behring-Str. 76, 35041 Marburg, Germany, Dade Behring Inc., 1717 Deerfield Road, Deerfield, Illinois 60015-0778, and The Plasma Proteome Institute, P.O. Box 53450, Washington, DC 20009-3450

Received March 28, 2005

For several years proteomics research has been expected to lead to the finding of new markers that will translate into clinical tests applicable to samples such as serum, plasma and urine: so-called in vitro diagnostics (IVDs). Attempts to implement technologies applied in proteomics, in particular protein arrays and surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF MS), as IVD instruments have initiated constructive discussions on opportunities and challenges inherent in such a translation process also with respect to the use of multi-marker profiling approaches and pattern signatures in IVD. Taking into account the role that IVD plays in health care, we describe IVD requirements and needs. Subject to stringent costs versus benefit analyses, IVD has to provide reliable information about a person's condition, prognosis or risk to suffer a disease, thus supporting decisions on treatment or prevention. It is mandatory to fulfill requirements in routine IVD, including disease prevention, diagnosis, prognosis, and treatment monitoring or follow up among others. To fulfill IVD requirements, it is essential to (1) provide diagnostic tests that allow for definite and reliable diagnosis tied to a decision on interventions (prevention, treatment, or nontreatment), (2) meet stringent performance characteristics for each analyte (in particular *test accuracy*, including both *precision of the measurement* and *trueness of the measurement*), and (3) provide adequate *diagnostic accuracy*, i.e., *diagnostic sensitivity* and *diagnostic specificity*, determined by the desired positive and negative predictive values which depend on disease frequency. The fulfillment of essential IVD requirements is mandatory in the regulated environment of modern diagnostics. Addressing IVD needs at an early stage can support a timely and effective transition of findings and developments into routine diagnosis. IVD needs reflect features that are useful in clinical practice. This helps to generate acceptance and assists the implementation process. On the basis of IVD requirements and needs, we outline potential implications for clinical proteomics focused on applied research activities.

Keywords: in vitro diagnostic • clinical chemistry • clinical proteomics

Introduction

Proteomics opens new horizons in many research areas of life sciences. This is particularly true for research efforts in the field of medicine. Clinical proteomics may be defined as a subset of proteomics activities in the field of medicine, which promises to accelerate the discovery of new drug targets and protein disease markers useful for in vitro diagnostics (IVD). IVD is based on the extracorporeal analysis of tissues and body fluids. Thus, it is expected that new pharmaceutical treatment opportunities will emerge and that the number and value of protein diagnostics will increase. The latter is of special interest, because reliable diagnostic information, in particular IVD data, is essential for choosing the appropriate intervention.[1]

Technologies applied in proteomics research, in particular SELDI-TOF MS and protein array techniques, are thought to be moving from research-focused applications to clinical laboratories as routine instruments for protein analysis.[2] In conjunction with the routine implementation of such technologies in the clinical laboratory it has been argued that multimarker profiling approaches or pattern signatures will be the next generation of protein IVD's and shift paradigms in IVD.[3]

These goals and expectations for clinical proteomics should be assessed critically in view of the role of IVD in health care. We believe that the success of clinical proteomic marker searches and of the technologies applied in proteomics and their ability to enter routine IVD testing depends on the ability

† Part of the Biomarkers special issue.

* To whom correspondence should be addressed. Dade Behring Marburg GmbH, New Marker Research & Development, P.O. Box 1149, 35001 Marburg, Germany. Tel: +49 (6421) 39-4473. Fax: +49 (6421) 39-5347. E-mail: frank_vitzthum@dadebehring.com.

‡ Dade Behring Marburg GmbH.

§ Dade Behring Inc.

‖ The Plasma Proteome Institute.

10.1021/pr050080b CCC: $30.25
Published on Web 06/22/2005

to fulfill IVD requirements and to adequately address IVD needs.[4] Thus, we set out to provide an industry perspective on the role of IVD in health care, a view of the requirements and needs of IVD, and the potential implications for clinical proteomics that should be considered, if a timely and effective transfer of research results to laboratory diagnostics is sought.

Role of in Vitro Diagnostics in Health Care

IVD plays an essential role in health care by (1) providing reliable information about a person's condition and (2) supporting treatment decision making,[5] both occurring in conjunction with a physicians' examination, which may include anamnesis, physical examination, and in vivo diagnostics (e.g., computed tomography and nuclear magnetic resonance imaging). A third purpose, which has been proposed by McNeil and Adelstein,[6] is to better understand disease mechanisms and evolution. This last purpose is outside the realm of routine IVD, but is nevertheless important, since it should provide better IVDs that will hopefully enter that realm.

IVD tests are estimated to contribute up to 94% of the objective data in clinical records and may influence 60−70% of critical decision making.[1,7] In addition, major cost reductions within the worldwide health expenditures could be achieved through a more efficient use of IVDs or better IVDs (www.VDGH.de).

Despite the importance of IVD and the essential role of the clinical laboratory, it usually accounts for less than 5% of the expenditures in clinics (www.VDGH.de).[1,7] The proportional share of laboratory costs in worldwide health expenditures is even lower, accounting for around 1% to 2% of the health care costs (www.VDGH.de). Even so, the expenditures necessary for laboratory diagnosis are often deemed too high, and the value of laboratory diagnosis is very often underestimated.

Together with the financial constraints in health care, this misperception generates enormous cost pressure on clinical laboratories and consequently on any equipment or test provider that serves this market directly or indirectly. For this reason, stringent cost and benefit considerations are a permanent challenge in IVD and in the translation of research applications generated by clinical proteomics into useful tests for the clinical laboratory.

Nevertheless, IVD testing represents a significant value proposition for improving outcomes, shortening hospital stays, etc. and substantially reducing the costs associated with healthcare. Though there will continue to be a need to justify the costs associated with IVD tests, the primary responsibility of IVD is to help reduce mortality, morbidity and improve the quality of life. It is evident that IVD can help to save lives and costs (www.VDGH.de).[8] There will continue to be great potential for improving IVDs and this dynamic offers clinical proteomics various opportunities.

Considerations Underlying Industry Perspectives and Decision Making Processes

A decision making process in the diagnostics industry has been outlined comprehensively by Zolg and Langen.[9] From the perspective of an IVD company, new markers and technologies must lead to products that improve diagnosis and also give a reasonable return on investment. Each company has to consider numerous factors before investing the limited human and financial resources at its disposal (List 1). A detailed discussion of these aspects is outside the scope of this article. Thus, we will focus on the crucial necessity to fulfill the requirements of the IVD market and to address unmet diagnostic needs.

List 1: Overview on critical IVD requirements, additional needs, and IVD industry considerations.

Critical IVD requirements:

1. Tests have to provide definite and reliable diagnostic information, that support decisions on interventions (prevention, treatment or non-treatment)
2. Test accuracy should meet objective specifications, e.g. imprecision of a measurement < 0.5 CV_I and bias < $0.250 \cdot (CV_I^2 + CV_G^2)^{1/2}$
3. Based on disease frequency and determined by the desired positive and negative predictive values, tests have to exhibit a required level of diagnostic accuracy, including both diagnostic sensitivity and diagnostic specificity.

Additional IVD needs that reflect considerations of clinical laboratories and physicians upon the implementation of tests or instruments:

- Costs
- Ease of use
- Automation
- Throughput
- Turn-around-time
- Down time
- Reliability
- Service quality and continuity
- Menu breadth
- Open channels (for non-proprietary tests)
- Instrument footprint
- Connectivity to the laboratory information management system (LIMS)
- Compatibility with existing laboratory processes (e.g. application of serum or plasma specimens and pre- and post-analytics)

IVD industry considerations

- Address diagnostic needs
- Return on investment
- Strategic fit
- Portfolio compatibility of new technologies
- Portfolio and instrument compatibility of new tests and markers (technical feasibility)
- Requirements of research and development phases and processes
- Manufacturability
- Marketing and sales
- Post-sales customer support
- Challenges and risks of entering the market (e.g. competitive situation, market acceptance, regulatory requirements, reimbursement, third party patent rights)

Addressing Diagnostic Needs Justifies the Significant Efforts for Product Development

The significant efforts and risks associated with product development are justified, if a new test addresses major diagnostic needs.

Examples of major diagnostic needs are cardiovascular diseases, infectious diseases, malignant neoplasms, and chronic diseases of the central nervous system, among others. It is encouraging to see that proteomics strives to address these diseases,[10] in particular cardiovascular diseases.[11−15] Cardiovascular disease, the leading cause of death worldwide (http://www.who.int/whr/2004/annex/topic/en/annex_2_en.pdf), represents a particularly telling example.

Acute cardiac events and chronic heart failure both have a long history in the patient. There is an unquestionable need for early determination of disorders which might eventually lead to these events and once they occur provide reliable information on the event itself, about the prognosis, and adequate treatment. In the area of infectious disease, increased crowding of domestic animals interacting with the ever denser and more mobile human population create chains of infections within the species but also across species. Early diagnosis and measures to prevent spread of infectious agents are of increasing concern. As the populations of industrialized countries grow to include proportionally older individuals, diseases of aging like cancer and chronic disorders of the central nervous system are of growing importance in our society and need to be better addressed by diagnostics and treatment measures.

It is Essential to Fulfill IVD Requirements and Advantageous to Address IVD Needs

To fulfill the requirements and meet needs of the IVD market (List 1) it is essential to provide valuable diagnostics at acceptable costs. Valuable diagnostics fulfill basic IVD requirements, i.e., they have to accomplish the following:

(1) provide most definite and reliable diagnostic information, that supports decisions on intervention (prevention, treatment or nontreatment)

(2) meet objective specifications to show sufficient *test accuracy*, including both

- precision of the measurement and
- trueness of the measurement, and

(3) exhibit the required level of *diagnostic accuracy*, including both

- diagnostic sensitivity and
- diagnostic specificity

determined by the desired *positive predictive values* and *negative predictive values*, which depend on *disease frequency* (prevalence and incidence).

From a regulatory and IVD industry perspective it appears advantageous to consider the following descriptions and examples, which could facilitate a smoother transition of clinical proteomics findings and developments from basic research to routine diagnostic applications.

Valuable Diagnostic Tests Provide Most Definite and Reliable Information to Support the Decision Making on Interventions

A diagnostic test that does not result in information that supports decisions on interventions, i.e., treatment, nontreatment or prevention, including risk stratification and prognosis, is of limited value, even if the result is interesting from a research viewpoint. The most successful tests are those that provide a clear-cut basis for prevention, treatment alternatives or nontreatment. An excellent example for markers that provide definite and reliable diagnostic information impacting medical practice is cardiac troponins (cTns).[16] cTns have been demonstrated to significantly improve sensitivity and specificity in the detection of acute myocardial infarction (AMI) compared to other diagnostic methods, in particular the combination of electrocardiography (ECG) and the traditional markers.[17]

In this case, a new diagnostic category, known as acute coronary syndrome (ACS), including AMI as the most serious form of the ACS, had to be established to accommodate troponin positive patients who did not meet the WHO criteria for AMI that time.[16] The test was able to detect real disease more sensitively (i.e., smaller infarcts) than previous diagnostic methods. According to recommendations by The Joint European Society of Cardiology and American College of Cardiology (ESC/ACC) Committee a redefinition of myocardial infarction has been proposed and ECG is still required, but no longer sufficient to diagnose AMI.[18] Any detectable cTn above a defined threshold in a patient's circulation is a sign of myocardial damage, which requires medical intervention.[16]

Accuracy of Measurement is an Essential Performance Characteristic which Encompasses Precision and Trueness of Measurement

The accuracy of measurement (test accuracy) is the closeness of the agreement between the result of a measurement and a conventional true value of the measurand. It includes both precision and trueness of measurement.[19,20] Test accuracy is usually not defined by a numerical value in terms of the measurand. It is an essential performance characteristic, because it is usually critical to achieve the required level of diagnostic accuracy (see below, Figure 1). Thus, failure to achieve sufficient test accuracy can under some circumstances endanger patients.

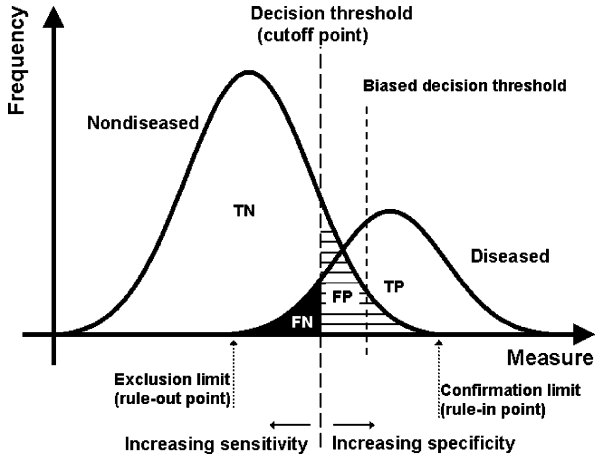

Figure 1. Dependence of the frequency of a nondiseased population and a diseased population versus a measure shows the relationship between sensitivity and specificity, decision thresholds (cutoff points), predictive values, and test accuracy. For simplicity hypothetical Gaussian distributions are displayed. The measure typically denotes results of a surrogate test or a test for contributory diagnostic evidence (e.g., single and multiple analytes or patterns). A pathognomonic test (gold standard), which constitutes definitive diagnostic evidence, is used to assign diseased and nondiseased. Sensitivity is defined as true positive/true positive + false negative; TP/(TP + FN). Specificity is defined as true negative/(true negative + false positive); TN/(TN + FP). Sensitivity and specificity in the populations is determined by the position of the cutoff point. Clearly, there is a reciprocal relationship between sensitivity and specificity. By choosing different cutoff points a single test can be used differently. To confirm the presence of a disease the rule-in threshold (test with high specificity) is used while the rule-out threshold (test with high sensitivity) is used to exclude disease presence. To determine the performance of a test in a population, it is important to know what proportion of positive tests correctly indicate the presence of the disease (positive predictive value, PPV), i.e., TP/(TP + FP), and if a test is negative, what proportion indicates absence of the disease (negative predictive value, NPV), i.e., TN/(TN + FN). Disease frequency, (TP + FN)/(TN + FP), i.e., $AUC_{diseased}/AUC_{nondiseased}$, affects these values. For example, the PPV decreases with decreasing frequency of the disease according to (values have to be multiplied appropriately if PPV is supposed to be displayed in percent):[21] PPV = TP/(TP + FP) = frequency × sensitivity/(frequency × sensitivity) + (1 − frequency) × (1 − specificity)). The impact of test accuracy depends on imprecision that skews and broadens curves (not displayed) and bias that shifts curves (displayed as biased decision threshold). Of note, when the assessment of surrogate tests is performed by using only a population comprising the "sickest of the sick" and the "wellest of the well"[5] one can easily imagine that diseased an nondiseased will be unrealistically separated. The performance of the surrogate test will then usually be overestimated.

Precision of Measurement. Precision of measurement (test precision) is the closeness of agreement between independent test results obtained under stipulated conditions. Precision depends on the distribution of random errors. The measure of precision is usually expressed in terms of imprecision. The

imprecision of measurements is the estimated random component of error deduced from the statistical distribution of repeatedly measured values. In IVD imprecision is mainly expressed by the coefficient of variation (CV: standard deviation divided by the mean).[19,21,22]

Precision of a distinct measurement procedure can be subdivided according to specified conditions.[19]

Repeatability of results of measurements can be defined as closeness of the agreement between the results of successive measurements of the same parameter carried out under the same conditions of measurement. These conditions are called repeatability conditions. Repeatability conditions include the same measurement procedure, the same observer, the same measuring instrument, used under the same conditions, the same location, and repetition over a short period of time. Repeatability is often termed in laboratory medicine "within-run" imprecision, a term now considered obsolete.[19]

Reproducibility relates to change in conditions, i.e., different laboratories, instruments, operators, and measuring systems (calibrants, reagent batches, etc.), a wide range of patient samples exhibiting unusual features and is often termed interlaboratory imprecision or total imprecision. The term total imprecision is likewise considered obsolete.[19]

It is important to note that IVD precision figures are typically given as interlaboratory precisions, a much wider scope of variation than covered by usual proteomics studies, which typically deal with precision within a run in one laboratory. Even so, total interlaboratory test imprecision in IVD is typically in the range of 1.5−10% CV (much less than typical values found in proteomics applications).

Trueness of Measurement. The trueness of a measurement is the agreement between the average value obtained from a large series of test results and a nationally or internationally accepted reference value (true or conventional true value). It denotes the systemic error component, expressed as bias.[19,20]

Reference materials and reference methods are the best guarantee for correct calibration and thus trueness of measurement.[20] A reference method or reference measurement procedure is defined as a thoroughly investigated measurement procedure, clearly and exactly describing the necessary conditions and procedures, for the measurement of one or more property values that has been shown to have trueness of measurement and precision of measurement commensurate with its intended use and that can therefore be used to assess the accuracy of other measurement procedures for the same property(-ies), particularly in permitting the characterization of a reference material.[23]

Of note, reference methods do not necessarily represent internal/procedural controls that may be used together with external controls in laboratories for quality control purposes.[24] Internal/procedural controls are designed into a system to monitor one or more components of errors including errors of the analytical process (e.g., reagent function), operator variance (e.g., sample processing and handling), and environmental factors (e.g., variations in temperature, humidity). External controls are not designed into a test system. They are materials similar to patient samples with expected properties commercially prepared or inhouse prepared samples, proficiency test samples with confirmed results, reference or control organism strains, patient specimens with established values previously tested, and control materials (other than those lots used to calibrate the system, i.e., standards). Reference materials are materials or substances with properties that are established for the use as standards, calibrators, controls, the verification of a measurement method, or the assignation of values, e.g., conventional true values.[20,21] As it is not always possible to determine the true value by a measurement that is associated with a systematic and random error, it is useful to define a single conventional true value that is preferentially internationally accepted, e.g., as is the case with glycated hemoglobin.[25] Instead of dealing with different "true values" depending on the method applied, physicians can rely on a single conventional true value. Thus, the implementation and application of reference materials can help to reduce the risk of uncertain or incorrect data interpretation.

Usually, highly purified chemicals are weighed out for the preparation of a solution with a specified concentration or for the calibration of solutions of unknown strength. A secondary reference material usually contains one or more analytes in a matrix that reproduces or simulates the expected. The values, e.g., conventional true values, have been assigned by a formal process of value transfer from a primary reference material. Certified reference materials (CRM) are usually also matrix based and have one or more of its values certified by a technically valid procedure. CRMs are accompanied by or traceable to a certificate or other document of a certifying body. Standard reference materials are CRMs and a trademark name of the National Institute of Standards and Technology. Tertiary reference materials are typically industry standards, calibrators, and controls that contain one or more analyte in an appropriate matrix. The concentration of the analytes is usually based on secondary reference materials.

Objective Test Accuracy (Both Trueness and Precision of Measurement) Specifications Have to be Met in IVD. For established diagnostic tests, objective test accuracy specifications are available to judge analytical performance characteristics.[26] Specifications may be derived from professional recommendations, e.g., guidelines from international, national expert groups, institutional groups, expert individuals, from regulatory and legislation authorities, from organizers of external quality assessment schemes and proficiency testing programs or simply from published data on the state of the art. For example, for cTn assays a CV of 10% reproducibility at the 99th percentile limit is suggested.[27] The 99th percentile of a reference control group denotes the decision limit. A percentile is a value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. cTns indicate myocardial necrosis when the maximum concentration exceeds the decision limit on at least one occasion during the first 24 h after the index clinical event.

Test accuracy specifications may also be assessed through the impact of the trueness of a measurement expressed as bias and the imprecision of a measurement on diagnostic sensitivity and specificity and the consequences for clinical decision making for specific clinical situations.[28] More general test accuracy specifications refer to the two major IVD scenarios, i.e., (1) monitoring individual patients and (2) diagnosis using reference intervals. Here, test accuracy specifications may best be based on the components of biological variation within subjects (CV_I) and between subjects (CV_G). A common view is that imprecision of a measurement should be below $0.5 \times CV_I$ and bias should be below $0.250 \cdot (CV_I^2 + CV_G^2)^{1/2}$ [26] (Table 1).

Of note, a test has to be sufficiently robust and specific to cope with matrix effects and interferences and still fulfill the required specifications of precision. Interferences or interfering substances are often considered to be components of the

Table 1. General Test Accuracy Specifications Including Both, Imprecision of the Measurement and Trueness of the Measurement (expressed as bias)[26a]

	specifications		
	optimal	desirable	minimum performance
imprecision	$< 0.25 \cdot CV_I$	$< 0.5 \cdot CV_I$	$< 0.75 \cdot CV_I$
bias	$< 0.125 \cdot (CV_I^2 + CV_G^2)^{1/2}$	$< 0.250 \cdot (CV_I^2 + CV_G^2)^{1/2}$	$< 0.375 \cdot (CV_I^2 + CV_G^2)^{1/2}$

[a] Specifications depend on within-subject (CV_I) and between-subject (CV_G) variations.

sample that contribute to the measurement signal, e.g., hemoglobin in specific photometric analyses. Matrix effects do usually not contribute directly to the measurement signal but indirectly through physical or physicochemical properties of the sample, e.g., viscosity, adsorption processes, reaction acceleration, or inhibition.

Test precisions of clinical analyzers may be used as an orientation guide for precision specifications, too. For example, in a multicentric evaluation of analytical performance of 33 protein assays a dedicated immunoassay analyzer exhibited a median total CV of 3.4% with a 2.5th and 97.5th percentile of the imprecision distribution at 1.3% and 13.0%.[29]

It is Advantageous to Address Needs that Contribute to the Overall Performance. Besides the requirements regarding test accuracy, it is advantageous to address needs that contribute to the overall performance characteristics (List 1). Costs associated with a test, including the instrument and consumables, which may be summarized in terms of cost per reportable result, are important factors for the acceptance of a test and have to be balanced in view of potential savings achieved by performing a test. Depending on the field of application, features such as ease of use, degree of automation, throughput, turn-around-time, menu breadth, open channels (for nonproprietary tests), instrument footprint, connectivity to the laboratory information management system, service quality and continuity have to be addressed. In conclusion, tests, assay formats and instruments have to fulfill objective quality specifications and should be at least compatible with the state of the art.

The compatibility of a new test's format with existing diagnostic test processes, i.e., preanalytics, actual analysis, and post-analytics, may also factor into the overall performance characteristics. Preanalytical processes include the order of the test(s) based on the physician's request to support diagnosis of a patient, sample collection, transportation of the sample, accession of the sample into the laboratory, processing of the sample (e.g., centrifugation, sorting, aliquoting, loading of the instrument), reagent setting and calibration of the instrument, and match ordered test(s) and sample. Post-analytical processes include approval of test results, their collection, transmission, appropriate display (e.g., charts) and documentation, waste disposal, sample storage, inventory, and order management, and payment of invoices.

Some of the processes mentioned above are of particular importance for proteomic studies. First, it is crucial to have detailed information on the donor, e.g., on gender; age; dietary, smoking, alcohol consumption and other habits; patient history; medication; pregnancy; post/pre-menopausal; suspicion or confirmed diagnosis of disease(s); clinical laboratory results; etc.

The specimen type used, its collection, transportation, processing, storage, etc., and the respective stability of a marker can dramatically influence the concentration and integrity of a marker.[30] This together with matrix effects from the other constituents of the respective sample have a significant impact on diagnostic performance and obviously on the success of research activities, too.[21,31−33] Biological specimens used for diagnosis include whole blood, serum, plasma, urine, cerebrospinal, synovial, amniotic, pleural, pericardial, peritoneal, cervical, seminal fluids, nipple aspirates, sweat, saliva, exhaled air, various types of solid tissues, feces, etc.[34,35] Serum is the specimen most frequently applied in IVD followed by different plasma types and urine,[36] while the pathology laboratory generally requires tissues and cells, e.g., erythrocytes, platelets, lymphocytes, etc. for cytopathological investigations.

Every specimen type has its pros and cons with respect to the scope of a study and the technologies applied. There is no such thing as a universal or ideal specimen type that could be generally recommended. Often, it is necessary to work with the specimens that are available anyhow. However, if possible, one should avoid choices of specimens and processing procedures that inherently preclude some future analysis. Blood is easily accessible and there are usually no contraindications to a standard blood draw. Serum and plasma display moderate variability if compared to other specimens such as urine. Nevertheless, serum and different plasma types exhibit significant variability. On the other hand, they are considered to mirror best the whole human proteome.[37] Serum allows for various testing including electrolytes, proteins, peptides, metabolites, etc. This is particularly advantageous, if additional testing ("adds on") as a consequence of a first result is required, because it may eliminate the need of additional blood draws and thus saves time. To have maximal freedom of testing and because of limitations regarding the draw of various different specimen types, ideally one specimen type should be applicable for all determination. However, some blood constituents require special collection and storage conditions.[38]

Plasma, in particular heparinized plasma, is preferred when a quick answer is critical, because clotting time is saved. Citrated plasma is mainly applied for coagulation analysis. Anticoagulants, preservatives, and other additives influence the composition of the specimen and may interfere with certain analysis.[21,38] Thus, it is of importance to carefully select the application of these compounds in blood collection tubes or their addition at a later point after the collection process.[30,35]

The collection of a specimen is crucial, too.[32,33] For example, for blood specimens it is important to consider how the blood was drawn. Factors such as timing of collection, position of the patient (lying, seated) and the time period the patient was in this position (impacting hematocrit), stress for the patient associated with the blood collection, type of collection (venipuncture, arterial puncture, skin puncture, etc.), site of collection and its preparation (e.g., cleaning of the site), tourniquet technique (e.g., time of venous occlusion), order of sampling (first, second, third tube), the quantity of blood withdrawn (e.g., dilution of citrated plasma), and obviously the person who

draws the blood impact the composition of the specimen. The same is true for parameters of collection devices or collection sets applied, e.g., needle gauge, single draw or multi-draw needles, and bag or tube types applied. Here, the material (glass or plastic) should be considered as well as other features such as siliconized tubes (potential activation and acceleration of clotting, reduction of adhesion of clots), gel or nongel separators (gel serves as a mechanical barrier, which eliminates the changes that occur when the clot or cells are in contact with serum or plasma), and additives such as clotting coagulants and preservatives.

The main factors that should be considered upon transportation are timing and temperature among others. Processing techniques may also impact the quality of a specimen.[32,33] The coagulation process to gain serum may depend on volumes, clot activator (type, concentration), clotting time, and temperature. G-force, duration and temperature are factors that have to be considered upon centrifugation of serum and plasma. For plasma samples activation of platelets at low temperatures can be an issue. Also, platelet depletion and the technique applied (second centrifugation step or filtration) may be considered.[32] Storage is another important aspect. The influence of freezing and thawing procedures, storage formats, temperature and duration as well as repeated freeze and thawing can significantly impact the quality of specimens.[32,33,39]

In conclusion, to ensure a high degree of consistency and to reduce the risk of error in subsequent analyses, strict adherence to optimized protocols or standard operating procedures is mandatory for each sample throughout the whole process, which has to be subject to thorough quality assurance and control procedures including documentation.[40−45]

Diagnostic Accuracy: The Ability to Discriminate States of Health and Disease

The diagnostic accuracy, i.e., diagnostic sensitivity and diagnostic specificity, of a test is its ability to discriminate between alternative states of health and disease. Frequently, there are more than two states to be differentiated. However, the clinical question can often (but not always) be dichotomized where the objective is to separate patients into two groups based on the presence or absence of a certain disease (Figure 1). In 1971, Lusted postulated that the measure of the performance of the observer measures the worth of a diagnostic test and that receiver (or relative) operator characteristic or receiver operating characteristic (ROC) curves (Figure 2) provide an ideal means of studying observer performance.[46] Since then, ROC curves have been widely used to determine the diagnostic accuracy of a test in IVD and clinical proteomics.

Receiver Operator Characteristic (ROC) Curves: Sensitivity versus 1-Specificity. To outline major challenges associated with laboratory tests and clinical proteomics approaches, basics of diagnostic performance measures are briefly outlined. ROC curves represent graphically the inverse relationship between specificity or 1-specificity and sensitivity (Figure 2). Sensitivity is the ability of a test to detect the condition of interest, for example the number of true positives of an affected group, i.e., detection of disease when it is truly present. Specificity is the number of the true negative results of the unaffected group, i.e., recognition of condition absence when it is truly absent. 1-specificity is the number of false positives. As the reliability of determinations of sensitivity and specificity depend on the size of the chosen population, it is essential to provide the 90% or 95% confidence intervals, because this gives a clearer picture of the strength of the data.[47] Details have been extensively outlined elsewhere.[21,22,47−49]

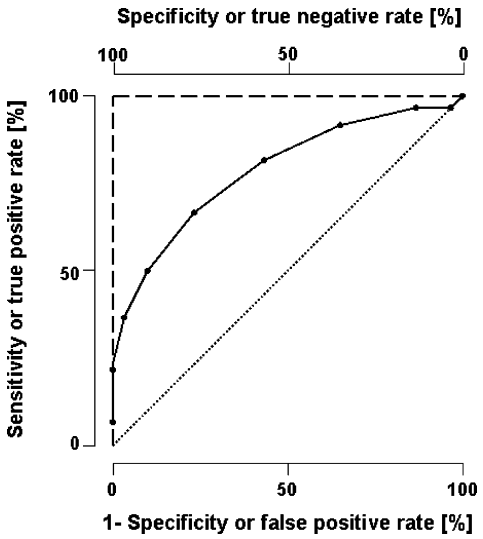

Figure 2. Receiver operating characteristic (ROC) curves are used to determine the diagnostic accuracy of an assay. When population overlap completely, no differentiation between diseased and nondiseased is possible (dotted line), leading to a useless test with an AUROC of 0.5. When there is no overlap, the ROC curve superimposes along the axes (dashed line). This perfect test has an AUROC of 1. Typically, there is a distinct overlap of populations as outlined in Figure 1, which creates ROC curves between the two extremes, e.g., the displayed hypothetical curve (solid line with dots).

AUROC: A Measure of Diagnostic Accuracy. A summary measure of diagnostic accuracy is the area under the ROC curve (AUROC), which incorporates sensitivity and specificity. It represents the overlap between the healthy and diseased population. An AUROC of 1 would represent a perfect diagnostic accuracy (no population overlap) and an AUROC of 0.5 would represent no difference between groups (no diagnostic value). It is rather unlikely that a diagnostic test reaches an AUROC of 1, because it is hardly possible to select a medical decision level or cutoff point that completely discriminates between two groups. Thus, the major challenge of IVD tests lies in the selection of the right cutoff points, i.e., sacrifice sensitivity for specificity or vice versa.

Of note, complex diseases, where it is necessary to differentiate multiple disease states, co-morbidities or mimics, and chameleons usually hamper the finding of efficient cutoff points to achieve high diagnostic accuracy. In addition, these diseases often exhibit complex time-dependent physiological or pathological processes, which may aggravate the finding of single IVD markers useful to differentiate states. For example, it is challenging to find IVD marker(s) to rule-in ischemic stroke within the first 3 h after onset of symptoms, which is the treatment window for thrombolytic therapy. As the acquisition of cerebrospinal fluid is usually contraindicated serum or plasma markers are needed. The crossing of the blood−brain-barrier by potential brain specific markers into the circulation within a short time frame and eventual renal clearance processes challenge the finding of appropriate markers. Finding a rule-in marker for ischemic stroke is also challenging, because of the heterogeneous etiology of stroke. Ischemic strokes have to be discriminated from hemorrhagic strokes (subarachnoidal, intracranial) and transient ischemic attacks and multiple

mimics such as alcohol withdrawal, hypoglycemia, intoxication, seizures (epileptic), amnesia, multiple sclerosis, migraine, tumors, closed head injuries, infections, and inflammatory processes (meningitis, encephalitis, abscesses). This example also clarifies, why laboratory diagnosis is often based on the combination of tests (see below).

Predictive Values Display How a Test Will Perform in a Population and Denote the Probability of the Patient to Have the Diagnosed Condition

To assess the diagnostic accuracy of a test when applied to a population, the effect of disease frequency has to be considered. Critical parameters of the diagnostic performance of a test are predictive values (PVs). The PV of a positive test result (PPV) is the percentage of patients with positive test results that are diseased. The PV of a negative test result (NPV) is the percentage of patients with negative test results that are nondiseased. PVs are not measures for the intrinsic diagnostic accuracy of a test, they are a function of both the sensitivity and specificity of a test and the frequency of a disease (see text of Figure 1). Whether prevalence or incidence is more adequate to determine predictive values is determined by the particular application. The incidence of a disease is the number of cases that arise during a distinct period of time in a given population. The prevalence is the frequency of the disease at a certain point in time in a given population.

As the decrease of disease frequency goes along with a decrease of the PPV, even tests with high sensitivity and specificity fail as screening tests when applied to the general population with a low disease frequency. For example, the comparatively low prevalence of distinct cancers is one of the reasons, why it is so difficult to implement cost-effective cancer screening tests as routine screening tools. Even the prostate specific antigen (PSA) test, the only FDA-approved serum IVD cancer screening test to-date is questioned to be useful as a prostate cancer screening tool, because of the limited predictability and the absence of prognostic value. Elevated PSA values do not tell, if the cancer is aggressive or not and if any surgery would save years of life at the cost of significant morbidity. In view of these aspects and because of the heterogeneity of cancer, it is an extremely difficult charter, to find routine screening markers for distinct cancer types. But as cancer is one of the leading causes of death (http://www.who.int/whr/2004/annex/topic/en/annex_2_en.pdf) and as early, presymptomatic diagnosis dramatically improves the odds of successful treatment, there is undoubtedly an unmet diagnostic need.

Clearly, the results of a diagnostic test cannot be assessed properly, without knowing the probability of the condition before the test is performed.[5] However, the PVs are post-test probabilities of the condition (revised or posterior probabilities). When a patient's history, signs and symptoms, or results of independent diagnostic tests performed previously are available, the pretest probability (the probability of the condition before the test is applied) has to be taken into account. The determination of the pretest probability, e.g., by Bayes' theorem, is beyond the scope of this article. It has been extensively reviewed elsewhere.[5,21,49,50]

Different Approaches to Perform and Combine Multiple Tests

According to the aspects mentioned above, two or more tests are often performed to diagnose patients. There are two basic approaches of multi marker testing in the diagnostic laboratory: (1) the series approach, i.e., test A is applied first and all those specimens with a positive result are retested with test B or vice versa, etc., or (2) the parallel approach, i.e., tests A and B are measured at once. The advantage of serial testing is the typically positive impact on pretest probability and consequently on PVs and its cost-effectiveness, in particular for the screening of large populations for a rare condition, because most patients receive only one test. The potential disadvantage is the delay in treatment while waiting for the results of subsequent tests.

There are different ways in which test results can be combined, e.g., by the OR rule or the AND rule.[49] For parallel testing, the rules are as follows:

OR rule: The diagnosis is positive, if either test A or test B is positive. Both tests must be negative for the diagnosis to be negative.

AND rule: The diagnosis is positive only, if both test A and test B are positive. Either test A or test B can be negative for the diagnosis to be negative.

When the serial testing process is performed the common decision rules are as follows:

OR rule: If the first test is positive, the diagnosis will be positive; otherwise, the second test is performed. If the second test is positive, the diagnosis will be positive; otherwise, the diagnosis will be negative.

AND rule: If the first test is positive, apply the second test. If the second test is also positive, the diagnosis will be positive; otherwise the diagnosis will be negative.

The sensitivities and specificities for the OR rules and the AND rules are the same for the parallel and the serial approach.[49,51] For the OR rule, the sensitivity is higher than either test individually ($Se_A + Se_B - Se_A \times Se_B > Se_A \cup Se_B$), but the specificity is lower than either test individually ($Sp_A \times Sp_B < Sp_A \cup Sp_B$). With the AND rule it is vice versa. The specificity is higher than either test individually ($Sp_A + Sp_B - Sp_A \times Sp_B > Sp_A \cup Sp_B$), but the sensitivity is lower than either test individually ($Se_A \times Se_B < Se_A \cup Se_B$).

The AND approach appears attractive to exclude the presence of a disease, because sensitivity is increased, though as a tradeoff, specificity is lower. When negative, these AND testing profiles may have a high NPV. Therefore, they are useful for clinicians to rule-out a variety of diagnoses. In the diagnostic workup, the OR approach may be preferentially performed. This approach is very predictive of the presence of a particular disease, confirmation or rule-in, because of the high specificity. Of course, when more than two tests are applied a flexible combination of series and parallel testing and the OR or AND rule may also be performed. Nevertheless, even the use of multiple tests of independent markers does not resolve the tradeoff between sensitivity and specificity[21] and issues associated with positive and negative predictive values.

Here, it should be noted that a set of observations can also be considered as a single multivariate observation (patterns or profiles) and interpreted as such by a multivariate comparison. A multivariate reference or decision region can prevent the addition of false-positive results, which is usually the case for the performance of multiple, univariate comparisons. The generation of algorithms to efficiently combine diagnostic information, e.g., by neuronal networks, has been shown to increase diagnostic accuracy.[52]

The contribution of the diagnostic accuracy of single or multiple laboratory tests has to be seen in concert with

Table 2. Purposes for the Application of IVD Tests or Respective Marker Types

marker type	examples
Acute markers are used when an acute disease event occurs and should help in the process of differential diagnosis to provide the information necessary for a specific treatment.	cTns that are used to rule-in AMI and B-type natriuretic peptide (BNP) and N-terminal pro BNP (NT-proBNP) to rule-out heart failure in patients with acute dyspnoe.[16,78]
Screening markers identify diseased preferentially in an asymptomatic stage within a population to start treatment as soon as possible, which usually goes along with a high treatment success. Generally, screening markers are applied in population subgroups with increased risk and disease frequency to achieve adequate PVs.	The detection of protein in urine, in particular albumin, is a screening tool for renal diseases.[79]
Primary risk assessment markers are used to assess the risk of a healthy individual to suffer from a disease in the future.	Total cholesterol, triglycerides, LDL cholesterol, HDL cholesterol, high sensitivity C-reactive protein (hsCRP), among others are markers used for primary cardiovascular risk assessment.[53,55]
Secondary risk assessment, prognostic or progression markers are used to determine how the disease may potentially develop, the risk of a diseased patient to suffer recurrent or other disease events, etc.	Cardiac troponins are used in the secondary risk assessment for AMI patients as an indicator for the risk of adverse outcomes, i.e., morbidity and mortality.[16] As primary risk assessment markers, secondary risk assessment markers help to determine aggressiveness of a treatment and balance the benefits and side effects.
Disease staging or classification markers help to classify different disease states.	Determination of certain proteins in urine, kidney and urinary tract diseases can be classified into prerenal, renal, and postrenal.[79]
Treatment response stratification markers are used to predict the likeliness of a response to a pharmacon before its application to prevent adverse effects and to initiate the most effective therapeutic treatment.	Hemostasis markers may be determined before a distinct anticoagulation therapy is started.[22] Viral resistance markers and antibiotics resistance testing in bacteria are further examples.
Treatment or therapeutic monitoring markers are used to determine and monitor the effectiveness of a treatment.	Blood lipids are applied to follow the impact of exercise, nutrition, and eventually therapeutic interventions.
Therapeutic drug monitoring is based on the determination of pharmaca administered to treat a patient. In this sense, it is not a marker type, but listed for the sake of completeness.	Immunosuppressive drugs, like sirolimus, everolimus, tacrolimus, and cyclosporin A, have to be monitored closely to prevent graft rejection and to minimize adverse therapy effects.[75]
Compliance markers provide information on treatment compliance.	Glycosylated hemoglobin A (HbA1c) can be used to monitor insulin therapy compliance[80] and carbohydrate deficient transferrin (CDT)[81] may be applied to control compliance with alcohol withdrawal therapy.

complementary diagnostic information to support precise treatment decision making, too. As outlined above, patient information from the physicians' examination are crucial. For example, parameters such as age, smoking, high blood pressure, family history of myocardial infarction, diabetes mellitus, etc., in concert with laboratory markers such as LDL cholesterol, HDL cholesterol, and triglycerides, among others, are major factors included in algorithms for the risk assessment of acute coronary events.[53]

Requested Diagnostic Information Determines the Required Diagnostic Accuracy

As outlined above, the disease type, its complexity, and its frequency within a given population, determine diagnostic accuracy requirements and how to perform and combine multiple tests to achieve appropriate PVs. Thus, diagnostic accuracy requirements are also determined by the requested diagnostic information or marker type, because this determines disease frequency and degree of differentiation between disease states. Different diagnostic applications of disease-specific proteins have been outlined comprehensively by Zolg and Langen.[9] Main IVD marker types are presented in Table 2.

In most cases of laboratory diagnosis, the same analytes can also be used for different diseases and as different marker types (Table 2) to obtain different diagnostic information. For example, CRP is the classic acute phase protein.[54] High concentrations of CRP indicate acute inflammatory processes.[54] whereas persistent low concentrations of CRP are used as a primary cardiovascular risk assessment marker.[55] cTns are applied as acute markers for the event (AMI) itself, to determine the size of the damage, and they also provide prognostic information.[16]

Tests for Definitive Diagnostic Evidence, Surrogate Tests, and Tests Providing Contributory Diagnostic Evidence

Tests are also classified by their role in the diagnostic process[56] as pathognomonic tests, surrogate tests, and tests that provide contributory diagnostic evidence. Obviously, this classification may not be entirely clear-cut in practice.[47] Pathognomonic tests (tests for definitive diagnostic evidence) are defined (or assumed) to uniquely define a disease and are used in the classification of diseased and nondiseased. Therefore, these tests are used as "gold standards" to assess the diagnostic accuracy of alternative tests (Figure 1). By definition a pathognomonic test displays the highest attainable diagnostic accuracy. Clearly, this assumption can only be validated by comparison to a less well-established test, which may subsequently be shown to better uniquely define the disease.[47] Usually, the pathognomonic test, e.g., a biopsy, is expensive, elaborate, time-consuming, unpleasant for the patient, or associated with a morbidity or even mortality risk. A surrogate test is used as a substitute for a pathognomonic test. A test for

contributory diagnostic evidence contributes to, but does not establish itself, a diagnostic decision. Typically, surrogate tests or tests for contributory diagnostic evidence are inexpensive, less unpleasant for the patient, and display little association with morbidity or mortality. On the other hand, they are associated with diagnostic inaccuracy.

Discovery, Confirmation and Exclusion Tests

Pathognomonic tests, surrogate tests or tests providing contributory diagnostic evidence may be classified as outlined in Table 2. Their use for the discovery, confirmation or exclusion of a disease allows for another level of classification.[57] A discovery test is used in disease screening (Table 2), a confirmation test to confirm (rule-in, test with high specificity) a suspected disease, and an exclusion test to exclude (rule-out, test with high sensitivity) the presence of a suspected disease. By applying different decision thresholds (cutoffs or cut-points) more than one of these purposes can be served by a single test and the respective marker (Figure 1). When a disease is serious and should not be missed, a disease is treatable, false positive results do not have serious psychological or economic consequences for the patient, e.g., neonatal screening for phenylketonuria or hypothyroidism, an exclusion test is used. When a disease is not treatable or curable, the knowledge that the disease is absent is reassuring to the patient, and false positive results have serious psychological or economic consequences for the patient, e.g., multiple sclerosis and advanced carcinoma, the confirmation test is used.[47]

Potential Implications on Applied Clinical Proteomics

If a smooth transition of clinical proteomics findings and developments to routine diagnosis is desired, then it is important to fulfill IVD requirements and address needs (see above and List 1). The requirements and needs should be considered differently depending on the objective.

Considerations for the Search for New IVD Markers

The choice of the disease, the desired diagnostic information or marker type (Table 2), and the performance of any existing diagnostic procedures determine the diagnostic accuracy that has to be achieved by tests based on a single or multiple new marker(s). It is also advisable to consider that the initial search for new IVD markers should allow for transition to routine IVD instruments (typically as immunoassays) and preferentially lead to serum or plasma tests of the marker candidates, if the marker is supposed to be tested in the clinical laboratory. This does not necessarily mean that serum or plasma is always the primary choice of specimen for the initial search for new markers. Concentration issues and other considerations may propose to start with solid tissues or other fluids.[58,59] If a new marker requires a new technology platform, then its use will be substantially delayed while the clinical laboratory community evaluates the platform and investigates the economics of the test. Typically an investment in a new platform is not justified if it provides only a single new test.

The preanalytical processes, i.e., acquisition of the sample and its handling, should be transferable to the routine environment. Thus, acceptance of a marker whose routine detection required a novel method of serum collection would be inhibited in comparison to an equivalent marker measurable in standard serum or plasma. The single immunoassay test per tube is ideal for optimization of individual assay performance to reach specifications, regarding specimen types, concentration range, precision of the measurement, trueness of the measurement (calibration and traceability to reference materials), correlation to other methods (in particular reference methods), cross-reactivity, matrix effects and interferences, e.g., hemolysis (which generates a red color interfering with some colorimetric assays), icterus (in which high levels of the yellowish pigment bilirubin interfere), lipemia (in which turbidity from lipoprotein particles interferes with some optical measurement techniques), interfering antibodies in the sample (which can, for example, bind directly to one of the antibodies in a sandwich immunoassay and give a false signal) and other substances. Standardization, quality control and assurance including detailed and appropriate documentation, if possible according to standard operating procedures[40−45] is of particular importance for preanalytical processes to provide samples of high quality, but obviously important for the subsequent diagnostic processes, too.

For an initial search for marker candidates, a phase I study, it may be legitimate to focus on the most prominent disease states and keep the sample number to a minimum. The number of samples needed for a phase I study depends on the conjectured diagnostic accuracy and the ratio of diseased samples to controls.[48]

Proteomics appears to be a useful tool for such initial studies, because it provides information on a huge number of proteins and protein variants at once. To facilitate transition to routine diagnosis with its requirements for diagnostic accuracy and test accuracy, it is advisable to switch to an assay format or instrument for subsequent study phases, which is routinely used in the diagnostic laboratory, e.g., switch from MS to ELISA.

This also applies to multi-marker approaches, if a reasonable number of markers to be combined is not exceeded. The tests can be performed serially or in parallel and the reported results can be combined by the AND or OR approach as well as multivariate analyses with high flexibility including other test types, e.g., electrolytes, functional hemostasis test, etc. and of course additional diagnostic information, e.g. age, gender, body weight, etc.

The validation, including subsequent study phases, will establish the value of emerging markers, alone or in combination, and show if they are robust enough to move to IVD. Validation may be defined as documented evidence that the specific, unequivocally described method, process, or technology, operating within the established parameters, performs effectively and reproducibly to produce a result that meets predetermined specifications and quality attributes within given variations with a high degree of statistical assurance.

In phase II studies, accuracies of two or more diagnostic tests are often compared. Patients difficult to diagnose with early or atypical disease, other conditions that might interfere with the test, and controls with conditions that mimic the disease of interest challenge the candidates.

The hurdle in phase III studies is usually to assess and compare diagnostic accuracy of different tests. Here, the value of a test is determined in a defined but generalizeable population, i.e., in a clinical trial with a large cohort of patients with the selection of appropriate controls.[48] The costs of such studies are a major issue when one considers that the margins in the diagnostic industry are much smaller than the margins in the pharmaceutical industry. Also, the pressure of the funding bodies appears to be greater for diagnosics.

Transition of Technologies Applied in Proteomics Research, Namely Protein Arrays and Mass Spectrometry, to Routine IVD

One of the major challenges of technology transition from proteomics research to IVD is the test accuracy, both precision and trueness (see above). To prove test accuracy, it is necessary that the technologies applied enable the application or establishment of guidelines and standard operating procedures[32,60] in concert with quality control, quality assurance, and standardization measures, e.g., by applying reference material.[61] The establishment of or the correlation to reference materials and reference methods are the best guarantee for correct calibration and thus trueness of measurement[20] to meet specifications, e.g., those of Table 1. This is particularly important for the establishment of common reference intervals and cutoff points to allow for harmonized clinical decision making and disease management. Patient monitoring over time is one example that does clearly illustrates the requirement of high test accuracy, in particular trueness of the measurement that has to be maintained over time. The application and implementation of reference materials in proteomics as currently applied in IVD may be one of the first steps to account for this challenge. A review on reference materials and the benefits of their application in clinical proteomics is beyond the scope of this article, though.

The parallel measurement of many different proteins is the inherent characteristic of technologies applied in proteomics research. The yet unresolved questions are due to the extremely different concentrations of the proteins to be measured in a very complex matrix like serum. A vivid example in inflammatory processes is the simultaneous determination of the abundant haptoglobin (Hp) (adult serum or plasma reference intervals calculated in relation to IFCC/BCR/CAP Certified Reference Material 470 are 0.03−0.20 g/L for Hp 1−1and Hp 2−2, and 0.04−0.20 g/L for Hp 2−1) and the traces of interleukin 6 (serum or plasma reference interval < 10 ng/L).

An interesting feature of the technologies applied in proteomics research, is the application of small sample volumes down to a few microliters. However, this is currently not critical in routine IVD since most current-generation clinical analyzers require less sample and reagent volumes than was previously the case.

The decision to apply single laboratory tests on standard instruments optimized for each protein or microarray or MS technologies will probably depend on the number and type of different proteins that need to be measured in parallel to provide a diagnostic result, the diagnostic accuracy and the test accuracy.

Challenges for Protein Array Technologies. High density protein arrays may be suited for high throughput screening and discovering initial correlations, but low density arrays seem to be more likely to reflect some aspects of IVD requirements and needs.[62,63] Performance, manufacturing, intellectual property issues, regulations and application challenges should be addressed for a smooth introduction of this technology into the IVD market.

Performance specifications (see above) like test precision, compatibility of multiple assay components, compatibility of concentration ranges of the different analytes, cross-reactivities, matrix effects, and interferences appear to be especially challenging when different proteins are to be measured in parallel in a single vessel. To achieve test trueness, significant efforts will be necessary to generate adequate and more complex controls and standards, and ultimately reference materials.

The manufacturing process will be more complex, but one may expect an increase of the overall efficiency with respect to the manufacturing time and costs through consolidation of multiple tests in a single package. However, lot to lot variability and stability of spotted arrays could be an issue. Quality control will be more difficult and troubleshooting of failures becomes more complicated. Final quality control failures for single analyte performance will compromise the entire product. The batch processes used to manufacture beads, e.g., for "bead arrays", at least partially circumvent some of these issues.

Third-party patents frequently have to be accessed before commercial use of technologies, tests, reagents (e.g., antibodies or antigens) applications, etc. Availability of licenses and costs may restrict the number of components that can be employed for a given multi-marker array and thus hamper its application in IVD. The inhibiting effect of "royalty-stacking", well-known in the pharmaceutical industry may become especially significant for multi-marker arrays.

From a regulatory perspective more complex evaluations and validations, which include the assay format and the instruments themselves, will require large cohort studies and significant registration efforts. Recalls of the whole system due to inappropriate performance of a single component affect customers' access to multiple tests. Different reimbursement rates and usage of the single analytes, the combination of markers, or a distinct microarray, also in view of regional distinctions, could also hamper the application of microarrays. The flexibility of the customer to perform serial testing or only testing a subset of tests is impaired with regard to financial trade offs, too.

Mass Spectrometry in IVD: Challenges Associated with the Detection of Proteins and Protein Profiles and a Successful History of the Determination of Low Molecular Weight Compounds. Some of the aspects outlined for arrays (see above) also apply to MS technologies, in particular to those that use arrays or chips such as surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) MS. Recently, Aivado et al.[64] demonstrated that automation, analysis of replicates, and standardization of the drying process significantly decreased imprecision of SELDI-TOF MS protein analyses. After optimization CVs ranged from 25.7% to 32.6% depending on the signal-to-noise ratio threshold used. Even though lower imprecisions for SELDI-TOF MS protein analyses, e.g., CVs less than 10% have been reported,[65] in view of precision requirements (see above and Table 1) further improvements are required for the application of SELDI-TOF MS in protein IVD.

However, as the SELDI-TOF MS approach is thought to generate profile signatures that correlate to a disease, it will be interesting to see to what extent profile signatures and algorithms improve diagnostic accuracy.[2,3] This improvement will need to be substantial to compensate for the lack of accuracy in the measurement of individual analytes, as well as many additional issues associated with SELDI-TOF MS analysis and studies that have been extensively discussed elsewhere.[66−71]

The concept of profiling has a long history in laboratory medicine (e.g. serum protein electrophoresis), and thus it is not a shift in paradigms[3] but it could rather be expected that novel profiling approaches, e.g., via MS technologies, would also contain diagnostic information. Especially, as it may appear unlikely that a single marker will be sufficient to adequately diagnose complex diseases in its early phase (see above).

On the other hand, the number of distinct discriminators commonly found in profiles[65,67,69,71] reveals that a few signals/markers are usually sufficient for diagnosis. This may eliminate the need for complete pattern or profile information for routine IVD applications—the needles within the haystack are of interest, not the haystack itself that actually obstructs the view on the needles. Consequently, a limited number of single immunoassays whose individual results are combined by an algorithm would be sufficient or even advantageous for accurate diagnosis (see above). Ultimately, even single markers have been and may be found in future that are useful for the diagnoses of complex diseases.

Besides the MS approach of profile signatures without identifying measured components, MS is widely used to detect distinct, identified analytes. Identifying analytes allows for the development and implementation of reference materials for standardization, calibration, recovery determination, quality control, and quality assurance.[66] For the detection of distinct low molecular weight molecules MS has been used as a diagnostic instrument in IVD for many decades,[68] e.g., for newborn screening to detect inborn errors of metabolism,[72,73] toxicology and forensic applications,[74] immunosuppressive drug testing,[75] drugs of abuse,[76] and doping of athletes.[77] In these applications, appropriate internal standards are used to achieve appropriate test accuracy. MS analyses are even applied as reference methods. For example, the determination of a distinct glycated peptide is used as an IFCC (International Federation of Clinical Chemistry) reference method for the determination of hemoglobin A1c with inter-laboratory CVs of 1.4% to 2.3%.[25] This form of MS technology, quite distinct from the use of unidentified feature patterns, is a potential alternative for use in the validation and small-scale application of novel diagnostic assays.

Conclusions

As candidate diagnostic markers begin the process of adaptation to IVD use, much time and effort can be saved by recognizing at the outset the important criteria by which IVD tests are judged. An IVD test is not simply a better or cheaper measurement than can be obtained with the technologies used for discovery: it is also a measurement that has a fairly definite meaning in the context of the patient's overall clinical picture and especially treatment. In many ways, new markers should be considered separately from new measurement technologies: a new marker could be brought into use much more rapidly as a test on an existing, validated platform (e.g., immunoassay) than on a novel technology platform whose acceptance is naturally slowed by economic (new clinical lab investment), regulatory, and other factors. This fact argues strongly that candidate IVD markers should be identified and fully characterized as part of the discovery process, and should then transition to a technology base better suited to adoption in the clinical laboratory. The opportunity to combine immunoassay technologies and MS, which would be applied as a sophisticated detector that can unambiguously identify ligands appears to be attractive. On the other hand, the costs associated with the application of MS as a detection unit in comparison to other detection technologies are currently considered prohibitive and most MS instrumentation falls far short of the robustness, automation, and user-friendliness required for routine laboratory operation. Nevertheless, while immunoassays dominate IVD protein measurements presently, it appears that the challenges of multiplexing such tests (e.g., on arrays) are sufficiently daunting that quantitative mass spectrometry may have value as an additional format for multiplexing protein measurements in the future given aggressive technology development.

References

(1) Forsman, R. W. *Clin. Leadersh. Manag. Rev.* **2002**, *16* (6), 370−373.
(2) Petricoin, E. F.; Liotta, L. A. *Clin. Chem.* **2003**, *49* (4), 533−534.
(3) Petricoin, E. F.; Liotta, L. A. *Clin. Chem.* **2003**, *49* (8), 1276−1278.
(4) Vitzthum, F.; Schwarz, H.; Behrens, F. *Mol. Cell Proteomics* **2004**, *3* (10 Sup), 52.
(5) Sox, J., H. C.; Blatt, M. A.; Higgins, M. C.; Marton, K. I. *Medical Decision Making*. ed.; Butterworth-Heinemann: Boston, 1989.
(6) McNeil, B. J.; Adelstein, S. J. *J. Nucl. Med.* **1976**, *17* (6), 439−448.
(7) Forsman, R. *Clin. Leadersh. Manag. Rev.* **2000**, *14* (6), 292−295.
(8) Morimoto, T.; Hayashino, Y.; Shimbo, T.; Izumi, T.; Fukui, T. *Int. J. Cardiol.* **2004**, *96* (2), 177−181.
(9) Zolg, J. W.; Langen, H. *Mol. Cell Proteomics* **2004**, *3* (4), 345−354.
(10) Jungblut, P. R.; Zimny-Arndt, U.; Zeindl-Eberhart, E.; Stulik, J.; Koupilova, K.; Pleissner, K. P.; Otto, A.; Muller, E. C.; Sokolowska-Kohler, W.; Grabher, G.; Stoffler, G. *Electrophoresis* **1999**, *20* (10), 2100−2110.
(11) McGregor, E.; Dunn, M. J. *Hum. Mol. Genet.* **2003**, *12* (2), 135−144.
(12) Van Eyk, J. E. *Curr. Opin. Mol. Ther.* **2001**, *3* (6), 546−553.
(13) Granger, C. B.; Van Eyk, J. E.; Mockrin, S. C.; Anderson, N. L. *Circulation* **2004**, *109* (14), 1697−1703.
(14) Stanley, B. A.; Gundry, R. L.; Cotter, R. J.; Van Eyk, J. E. *Dis. Markers* **2004**, *20* (3), 167−178.
(15) Allard, L.; Lescuyer, P.; Burgess, J.; Leung, K. Y.; Ward, M.; Walter, N.; Burkhard, P. R.; Corthals, G.; Hochstrasser, D. F.; Sanchez, J. C. *Proteomics* **2004**, *4* (8), 2242−2251.
(16) Bodor, G. S. *LabMed. Int.* **2004**, March/April, 13−14.
(17) Adams, J. E.; Sicard, G. A.; Allen, B. T.; Bridwell, K. H.; Lenke, L. G.; Davila-Roman, V. G.; Bodor, G. S.; Ladenson, J. H.; Jaffe, A. S. *N. Engl. J. Med.* **1994**, *330* (10), 670−674.
(18) Alpert, J. S.; Thygesen, K.; Antman, E.; Bassand, J. P. *J. Am. Coll. Cardiol.* **2000**, *36* (3), 959−969.
(19) Clinical and Laboratory Standards Institute (CLSI) formerly National Committee for Clinical Laboratory Standards (NCCLS) homepage: harmonized terminology database www.nccls.org.
(20) Büttner, J. *Eur. J. Clin. Chem. Clin. Biochem.* **1995**, *33*, 975−1022.
(21) Burtis, C. A.; Ashwood, E. R. *Tietz Fundamentals of Clinical Chemistry*, ed.; W. B. Saunders Company an imprint of Elsevier Science: Philadelphia, London, New York, St. Louis, Sydney, Toronto, 2001.
(22) Thomas, L. *Clinical Laboratory Diagnostics*, ed.; TH−Books Verlagsgesellschaft: Frankfurt, 1998.
(23) International Standards Organization (ISO), *Terms and Definitions Used in Connection with Reference Materials (ISO Guide 30 (E/F))*, 2 ed.; ISO: Geneva, 1992.
(24) Bandy, M. *IVD Technol.* **2005**, March, 24−9.
(25) Jeppsson, J. O.; Kobold, U.; Barr, J.; Finke, A.; Hoelzel, W.; Hoshino, T.; Miedema, K.; Mosca, A.; Mauri, P.; Paroni, R.; Thienpont, L.; Umemoto, M.; Weykamp, C. *Clin. Chem. Lab. Med.* **2002**, *40* (1), 78−89.
(26) Fraser, C. G.; Petersen, P. H. *Clin. Chem.* **1999**, *45* (3), 321−323.
(27) Apple, F. S.; Wu, A. H.; Jaffe, A. S. *Am. Heart J.* **2002**, *144* (6), 981−986.
(28) Klee, G. G. *Clin. Chem.* **1993**, *39* (7), 1514−1518.
(29) Lammers, M.; Gentzer, W.; Boni, P.; Fitzner, R.; Guarmeur-Jardel, C.; Ledue, T. B. *Multicentric performance evaluation of the BN ProSpec plasma protein analyzer system.*; AACC Congress, Chicago July 29 − August 2, 2001.
(30) Holland, N. T.; Smith, M. T.; Eskenazi, B.; Bastaki, M. *Mutat. Res.* **2003**, *543* (3), 217−234.
(31) Haab, B. B.; Geierstanger, B. H.; Michailidis, G.; Vitzthum, F.; Forrester, S.; Okon, R.; Saviranta, P.; Brinker, A.; Sorette, M.; Perlee, L.; Suresh, S.; Drwal, G.; Adkins, J. N.; Beer, I.; Omenn, G. *Proteomics* **2005**, in press.
(32) Rai, A. J.; Gelfand, C. A.; Haywood, B. C.; Warunek, D. D.; Skobe, C.; Schuchard, M. D.; Mehigh, R. J.; Cockrill, S. L.; Scott, G. B. I.; Tammen, H.; Schulz-Knappe, P.; Speicher, D. W.; Vitzthum, F.; Habb, B. B.; Siest, G.; Chan, D. W. *Proteomics* **2005**, in press.
(33) Biobank, T. U. *Sample Handling and Storage Subgroup Protocol and Recommendations*, ed.; UK Biobank: Manchester, 2004.

(34) Vitzthum, F. *Entwicklung und Untersuchung Automatisierungsgerechter Physikalisch-Mechanischer Desintegrationsverfahren für eine Nukleinsäure-Gestützte, Humanmedizinische Infektionsdiagnostik [Development and Evaluation of Automated Physical and Mechanical Disintegration Technologies for Nucleic Acid Based Diagnosis of Infectious Diseases]*, ed.; Fraunhofer IRB Verlag: Stuttgart, 2000.
(35) Landi, M. T.; Caporaso, N. *IARC Sci. Publ.* **1997**, *142*, 223−236.
(36) Glover Leonard & Redshaw Management Consultancy: *Laboratory workload segmentation audit of medium to high specimen throughput laboratory locations in Europe*; Glover Leonard & Redshaw: Brussels, Bronxville, 1997; p 71.
(37) Anderson, N. L.; Anderson, N. G. *Mol. Cell Proteomics* **2002**, *1* (11), 845−867.
(38) Winsten, S.; Gordesky, S. E., Transportation of specimens. In *Selected methods of clinical chemistry*, ed.; Faulkner, W. R.; AACC: Washington, DC, 1982; Vol. 9, p 11−15.
(39) Aziz, N.; Nishanian, P.; Mitsuyasu, R.; Detels, R.; Fahey, J. L. *Clin. Diagn. Lab. Immunol.* **1999**, *6* (1), 89−95.
(40) National Committee for Clinical Laboratory Standards (NCCLS), *Blood collection on filter paper for neonatal screening programs: approved standard LA4-A3*. 3 ed.; NCCLS: Wayne, PA, 1997.
(41) National Committee for Clinical Laboratory Standards (NCCLS), *Procedures for collection of skin puncture blood specimens: approved standard H4-A4*. 4 ed.; NCCLS: Wayne, PA, 2000.
(42) National Committee for Clinical Laboratory Standards (NCCLS), *Evacuated tubes and additives for blood specimen collection: approved standard H1-A4*. 4 ed.; NCCLS: Wayne, Pa, 1996.
(43) National Committee for Clinical Laboratory Standards (NCCLS), *Procedures for collection of diagnostic bllod specimens by venipuncture: approved standard H3-A4*. 4 ed.; NCCLS: Wayne, Pa, 1991.
(44) National Committee for Clinical Laboratory Standards (NCCLS), *Procedures for the handling and transport of domestic diagnostic specimens and etiologic agents: approved standard H5-A3*. 3 ed.; NCCLS: Wayne, Pa, 1994.
(45) National Committee for Clinical Laboratory Standards (NCCLS), *Routine urinalysis and collection, transportation, and preservation of urine specimens: approved guideline GP16-A*. ed.; NCCLS: Wayne, Pa, 1995.
(46) Lusted, L. B. *Science* **1971**, *171* (977), 1217−9.
(47) Henderson, R. *Ann. Clin. Biochem.* **1993**, *30*, 521−39.
(48) Obuchowski, N. A.; Lieber, M. L.; Wians, F. H., Jr. *Clin. Chem.* **2004**, *50* (7), 1118−25.
(49) Zhou, X.-H.; Obuchowski, N. A.; McClish, D. K. *Statistical Methods in Diagnostic Medicine*, ed.; John Wiley & Sons: New York, 2002.
(50) Stamm, D.; Büttner, J. Beurteilung Klinisch-Chemischer Analysenergebnisse. In *Lehrbuch der Klinischen Chemie und Pathobiochemie*, 3 ed.; Greiling, H., Gressner, A. M., Eds.; Schattauer: Stuttgart, New York, 1995; p 73−114.
(51) Griner, P. F.; Mayewski, R. J.; Mushlin, A. I.; Greenland, P. *Ann. Intern. Med.* **1981**, *94* (4 Pt 2), 557−592.
(52) Voss, R.; Cullen, P.; Schulte, H.; Assmann, G. *Int. J. Epidemiol.* **2002**, *31* (6), 1253−1264.
(53) Assmann, G.; Cullen, P.; Schulte, H. *Circulation* **2002**, *105* (3), 310−315.
(54) Tillet, W. S.; Francis, T. *J. Exp. Med.* **1930**, *52*, 561−571.
(55) Ridker, P. M.; Buring, J. E.; Shih, J.; Matias, M.; Hennekens, C. H. *Circulation* **1998**, *98* (8), 731−733.
(56) Feinstein, A. R. *Clinical Epidemiology: The Architecture of Clinical Research*, ed.; WB Sauenders Company: Philadelphia, 1985; p 1−812.
(57) Feinstein, A. R. *Clinical Biostatistics*, ed.; The CV Mosby Combany: St. Louis, 1977.
(58) Lescuyer, P.; Allard, L.; Zimmermann-Ivol, C. G.; Burgess, J. A.; Hughes-Frutiger, S.; Burkhard, P. R.; Sanchez, J. C.; Hochstrasser, D. F. *Proteomics* **2004**, *4* (8), 2234−2241.
(59) Zimmermann-Ivol, C. G.; Burkhard, P. R.; Le Floch-Rohr, J.; Allard, L.; Hochstrasser, D. F.; Sanchez, J. C. *Mol. Cell Proteomics* **2004**, *3* (1), 66−72.
(60) White, N.; Zhang, Z.; Chan, D. W. *Clin. Chem. Lab. Med.* **2005**, *43* (2), 125−126.
(61) Petricoin, E.; Wulfkuhle, J. D.; Espina, V.; Liotta, L. A. *J. Proteome Res.* **2004**, *3*, 209−217.
(62) Stoll, D.; Templin, M. F.; Schrenk, M.; Traub, P. C.; Vohringer, C. F.; Joos, T. O. *Front Biosci.* **2002**, *7*, c13−32.
(63) Hornauer, H.; Klause, U.; Müller, H.-J.; Vieth, F.; Risse, B. *Laborwelt* **2004**, *4*, 38−39.
(64) Aivado, M.; Spentzos, D.; Alterovitz, G.; Otu, H. H.; Grall, F.; Giagounidis, A. A. N.; Wells, M.; Cho, J.-Y.; Germing, U.; Czibere, A.; Prall, W. C.; Porter, C.; Ramoni, M. F.; Libermann, T. A. *Clin. Chem. Lab. Med.* **2005**, *43* (2), 133−140.
(65) Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; al., e. *The Lancet* **2002**, *359*, 572−577.
(66) Hortin, G. L. *Clin. Chem.* **2005**, *51*, 3−5.
(67) Diamandis, E. P. *Clin. Chem.* **2003**, *49* (8), 1272−1275.
(68) Diamandis, E. P. *Mol. Cell Proteomics* **2004**, *3* (4), 367−378.
(69) Baggerly, K. A.; Morris, J. S.; Coombes, K. R. *Bioinformatics* **2004**, *20* (5), 777−785.
(70) Coombes, K. R. *Clin. Chem.* **2005**, *51*, 1−2.
(71) Sorace, J. M.; Zhan, M. *BMC Bioinformatics* **2004**, *4* (24), 1−13.
(72) Venditti, L. N.; Venditti, C. P.; Berry, G. T.; Kaplan, P. B.; Kaye, E. M.; Glick, H.; Stanley, C. A. *Pediatrics* **2003**, *112* (5), 1005−1015.
(73) Chace, D. H.; Kalas, T. A.; Naylor, E. W. *Clin. Chem.* **2003**, *49* (11), 1797−1817.
(74) Dooley, K. C. *Clin. Biochem* **2003**, *36* (6), 471−481.
(75) Streit, F.; Armstrong, V. W.; Oellerich, M. *Clin. Chem.* **2002**, *48* (6 Pt 1), 955−958.
(76) Niedbala, S.; Kardos, K.; Salamone, S.; Fritch, D.; Bronsgeest, M.; Cone, E. J. *J. Anal. Toxicol.* **2004**, *28* (7), 546−552.
(77) Frison, G.; Tedeschi, L.; Favretto, D.; Reheman, A.; Ferrara, S. D. *Rapid Commun. Mass Spectrom.* **2005**, *19* (7), 919−927.
(78) Weber, T.; Auer, J.; Eber, B. *Curr. Pharm. Des.* **2005**, *11* (4), 511−525.
(79) Hofmann, W.; Regenbogen, C.; Edel, H.; Guder, W. G. *Kidney Int. Suppl.* **1994**, *47*, 111−114.
(80) Bowker, S. L.; Mitchell, C. G.; Majumdar, S. R.; Toth, E. L.; Johnson, J. A. *CMAJ.* **2004**, *171* (1), 39−43.
(81) Wick, H. D. *Schweiz Rundsch Med. Prax* **1999**, *88* (42), 1711−1714.

PR050080B