Journal of Chromatography B, 877 (2009) 1259–1271

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

Method validation of protein biomarkers in support of drug development or clinical diagnosis/prognosis☆

Jean W. Lee*, Michael Hall

Amgen Inc., Thousand Oaks, CA 91320, United States

ARTICLE INFO

Article history:
Received 1 August 2008
Accepted 13 November 2008
Available online 18 November 2008

Keywords:
Method validation
Protein biomarkers
Drug development
Clinical diagnosis/prognosis
Ligand binding assay
Mass spectrometric method

ABSTRACT

Protein biomarkers are used for various purposes in drug development and clinical diagnosis and prognosis. In this review, a fit-for-purpose method validation approach is discussed that fulfills the needs of exploratory and advanced applications in both the pharmaceutical and diagnostic arenas. Method validation for protein biomarkers is typically applied to ligand binding assays (LBA) although hyphenated mass spectrometric methods can be used as adjunct methodologies to confirm LBA specificity or provide valuable information during early discovery or demonstrative phases of a novel biomarker. Pre-analytic variables of protein biomarkers, such as the purpose of the intended application, analyte(s), biological matrix, availability of reference standard, calibrator matrix, assay platform, and sample collection/handling, must be considered in any method development and validation plan. Method validation for exploratory applications involves basic experiments for assay range finding, accuracy and precision, selectivity, specificity, and minimal stability. For advanced method validation, more rigorous tests with a wider scope are performed. These tests include additional patient population ranges, more runs on accuracy and precision from multiple analysts/reagent lots/instruments, selectivity and specificity tests using patient samples, and stability tests subjected to conceivable conditions over long-term use. Differences in biomarker method validation for drug development vs. clinical diagnosis and issues of using developmental commercial kits are discussed. The co-development of biomarkers for drug development and diagnostics presents collaborative opportunities between the pharmaceutical and diagnostic sectors.

Contents

☆ This paper is part of the special issue "Quantitative Analysis of Biomarkers by LC–MS/MS", J. Cummings, R.D. Unwin and T. Veenstra (Guest Editors).

* Corresponding author. Tel.: +1 805 447 9463; fax: +1 805 499 9027.
E-mail address: jwlee@amgen.com (J.W. Lee).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.11.022

1. Introduction

Biomarkers have routinely been used by physicians in disease diagnosis. Typically, panels of biomarkers (e.g. for cardiac, lipids, diabetes and tumor markers) in serum or urine samples from patients are measured by autoanalyzers using FDA approved commercial kits. Recently, biomarkers have been recognized and accepted as useful tools for drug development [1,2]. The appropriate application of biomarkers to preclinical and clinical drug development reduces time to market, hastens the attrition of undesirable candidate compounds prior to expensive phase III clinical trials, and guides dose selection with early indications of efficacy or toxicity as depicted in Fig. 1 [3–7]. There has been a paradigm shift from "trial and error" to mechanistic based, target-driven drug development using biomarkers to track the biological exposure–drug effect relationship.

A long list of novel putative biomarkers that are not included in routine clinical lab tests has been generated from intense genomic and proteomic research. These biomarkers have been utilized in clinical trials of drug candidates for exploratory, demonstrative or characterization applications as part of the process of developing mechanism-specific biomarkers toward the ultimate goal of surrogacy [8–11]. Development of novel biomarkers has notably increased in the areas of diabetes, cancer, rheumatoid arthritis and cardiovascular disease. For example, depending on the mechanism of action, novel biomarkers could be included in diabetes trials in addition to the commonly employed measurements of blood glucose, glycosylated hemoglobin (Hb_{a1c}) and circulating insulin. Specifically, for peroxisome proliferator-activated receptor gamma (PPARγ) agonists in type-2 diabetes, free fatty acid was linked to insulin resistance as mechanism-specific, and adiponectin was identified as a proximal biomarker correlated with insulin sensitivity [12,13] while Hb_{a1c} was qualified as a surrogate marker [14]. Other mechanisms and the relevant biomarkers (in parentheses) such as lipid and bone metabolism (free fatty acids, leptin, osteocalcin), incretins (glucagon, GLP-1) and inflammation (cytokines, hsCRP, PAI-1, fibrinogen, adhesion molecules) are also involved in the complex disease of diabetes [15–19]. In the area of cancer, measuring circulating PSA, CEA and other tumor marker levels, monitoring mRNA of specific genes, and imaging are used in clinical trials to evaluate drug responses during treatment [20–24].

The use of novel biomarkers has become a prominent component of decision-making processes in drug development. They are used in *in vitro* and preclinical models and early clinical phase for quick hit and early attrition decisions [8,25–27]. The processes of biomarker discovery, characterization, and clinical qualification/validation have been discussed (Fig. 2, right side bar) [11,28]. In exploratory and demonstrative studies, pharmacodynamic (PD) correlations are typically unknown, data are used mainly for internal decision-making, and the output is generally not subject to regulatory review. The extent of method validation can thus be limited to a few basic components to expedite the process and preserve resources without unduly impacting commercialization, depicted as Exploratory Method Validation in Fig. 2 [29]. In contrast, the purposes of the characterization phase are to provide pivotal data to establish linkage to clinical outcome and to monitor patient progress upon treatment. Characterization of a novel biomarker in the translational phase requires data collection to show preclinical sensitivity and specificity and linkage to clinical outcomes in multiple clinical studies in humans. The purposes at this phase are different from those of the Exploratory or Demonstrative phase. The data are often used for critical decisions (such as supporting dose selection and patient stratification, demonstration of drug safety or efficacy, and differentiation of drug candidates), for submissions to be reviewed by regulatory agencies, or for post-marketing patient monitoring [30,31]. In addition, the same method used for characterization would likely be used in the qualification phase toward surrogacy confirmed over multiple drugs of similar mechanism and during surveillance studies. Therefore, biomarker characterization studies would require more intense rigor and cover a wider scope in Advanced Method Validation (Fig. 2), with greater traceability and more detailed documentation than that of the previous phases to meet the study objectives in a defined context of its use.

The knowledge of drug/protein target interactions in disease pathways also contributes to the concept of "personalized medicine" or "target therapy." Clinical biomarker assays are no longer simple results from an autoanalyzer with cutoff values that lead to the diagnosis of a certain disease. More significantly, gene-, protein-, or metabolite-based biomarker profiling of a patient can be used to identify and stage disease (diagnosis), decide treatment, and monitor progress and predict the outcome with certain confidence (prognosis). Mass spectrometric-based proteomic research has generated peptide maps of healthy vs. patients with various disease types at different stages. Proposals have been made to use these patterns and the identified proteins from these peptides for early diagnosis and to monitor progress of treatment [32,33]. Additionally, novel protein biomarkers linked to diseases or drug safety have been unveiled from this proteomic research [34–38].

With wider applications of biomarkers, especially novel biomarkers, in clinical diagnosis and prognosis and at various phases in drug development, questions arise about how the analytical laboratories (e.g. clinical or bioanalytical laboratories) should validate a method to be suitable for the intended applications. Methods for novel biomarkers are generally generated from the innovator laboratories in university or company settings, often using commercial kits developed "for research use only" (RUO) and therefore may not be fully validated (Exploratory Method Validation in Fig. 2). On the other hand, methods that are FDA approved or cleared would be subjected

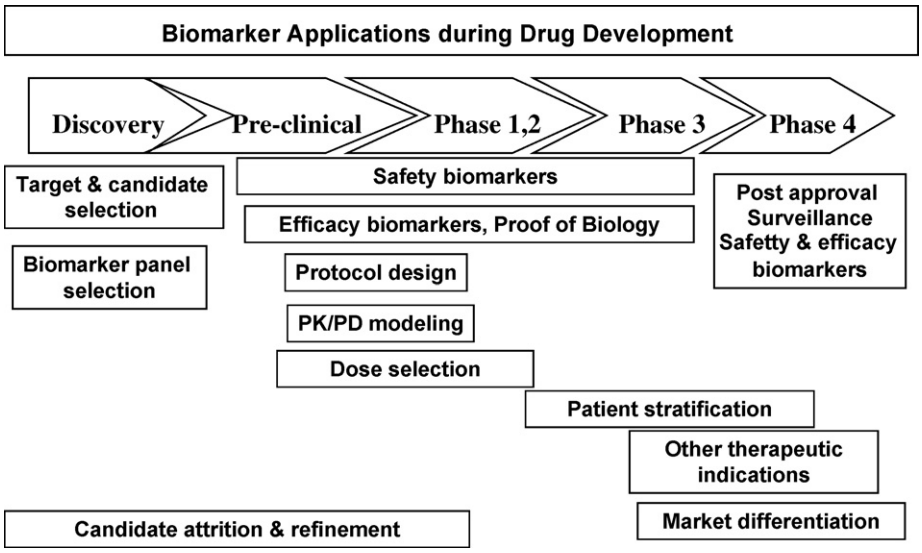


Fig. 1. Biomarker applications during various phases in drug development. The purposes of application are shown in the boxes underneath the various phases. Within a clinical study, biomarkers of exploratory and advanced validation may be included for different purposes such as patient stratification or PD profiling.

to rigorous validation tests similar to the Advance Method Validation. A fit-for-purpose method validation approach at various phases of biomarker application has been described by Lee et al. [29]. Here the discussion will focus on method validation of protein biomarker bioanalysis to meet the diverse purposes during drug development clinical trials and clinical diagnosis/prognosis.

2. Technologies for protein biomarker measurement in translational medicine and new diagnostic tests

2.1. Transition from proteomics technologies

Advances have been made in the development of quantitative proteomics methodologies [39–44]. The performance of several of

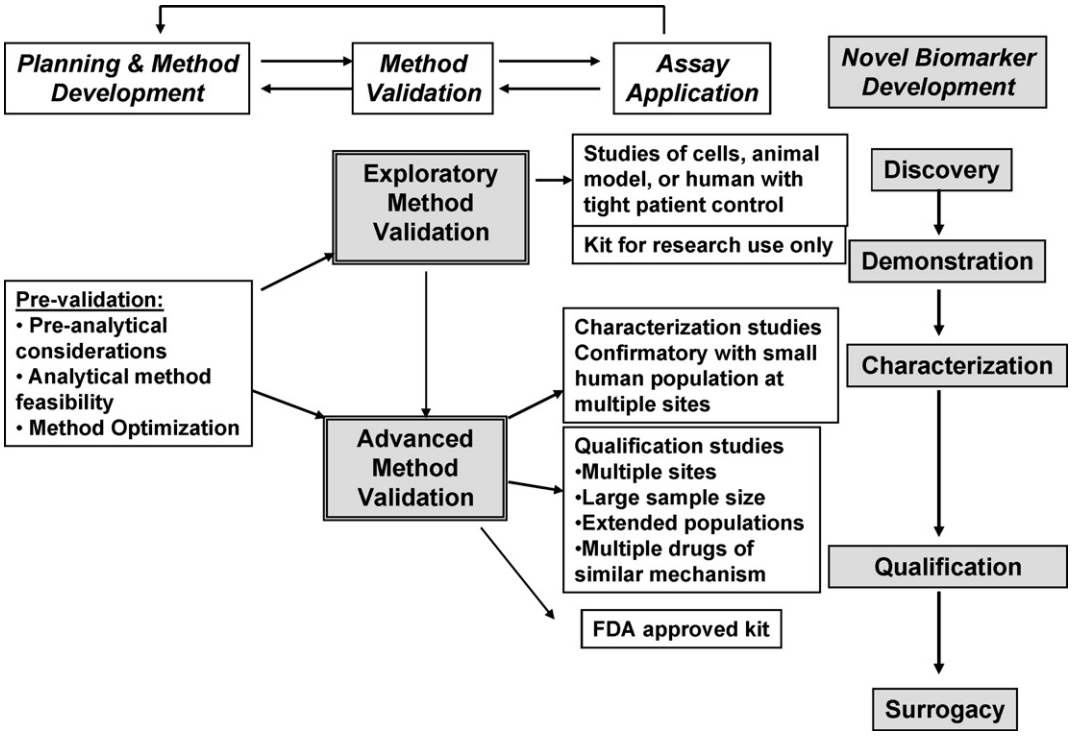


Fig. 2. Iterative processes of biomarkers method validation related to stages of discovery, characterization and qualification. The flow diagram on the right side depicts the process of biomarker development from discovery to surrogacy. Attrition of novel biomarkers is analogous to but not concurrent with drug candidate development, from a few hundred to a smaller potential panel and finally to a final few for a target mechanism. The rigor and scope of method validation are dependent on the intended applications. Lessons learned form exploratory applications contribute to the pre-analytical planning, method refinement and validation for advanced applications.

these methodologies has been recently assessed by a study organized by the Association of Bimolecular Resource Facilities (ABRF) [45]. The analytical methods surveyed by this study fell into two broad categories: (1) gel-based methods and (2) MS-based methods (including both stable isotope-labeling and label-free techniques) [46]. The major conclusions from this study were that the methods are complex and require high levels of expertise for success. Hence, there was large lab-to-lab variability when the same technologies were used. As a result of this complexity and reliance on expertise, these methods are not commonly adopted for preclinical/clinical applications to study dose-response relationships in drug development.

By virtue of the highly specific chemical information that is generated, MS-based techniques do play a major role in biomarker discovery, but validation and routine monitoring of protein biomarkers are most readily achieved by the use of ligand binding assays (LBA) [47,48]. There may be instances, however, where LBA is not a viable option such as the case where no or inadequate affinity reagents are available. In these cases, MS-based approach may be a feasible alternative.

2.2. Mass spectrometry

LC–MS/MS techniques such as selective and multiple reaction monitoring (SRM or MRM) can be used for biomarker quantification. Specific peptides from putative protein biomarkers that have been previously identified from proteomics information or other existing biological knowledge are quantified. Although the response curve (signal vs. concentration) of the mass spectrometer varies from peptide to peptide, absolute quantification is possible by using synthetic isotopically-labeled versions of the peptides [48–51]. Typically, the protein sample of interest is digested with a specific proteolytic enzyme, and the isotope labeled control peptides are added to the mixture. The digest is then separated online by HPLC and analyzed by ESI-MS/MS. In SRM or MRM modes, sensitivity is greatly increased and attomole detection limits have been reported but are not typically obtained [48]. For preclinical studies involving disease models of several animal species, the homolog sequences in different species can also be synthesized and tested. Overall, the MRM–LC–MS/MS approach can be generalized for multiple species and multi-analyte assays. For example, one study tracked the expressional levels of 47 different proteins in human plasma with quantitative CVs between 2 and 22% [52]. In general, however, such MS-based examples for biomarker validation and monitoring are few and the approach is still in its infancy.

2.3. Ligand binding assays

In contrast to MS-based methods, LBA is the “gold standard” by which protein biomarkers are validated and monitored in diagnostic and pharmaceutical laboratories. Therefore, the overall discussion of method validation in this article is specifically relevant to LBA.

LBA methods are low-cost, sensitive and simple with high dynamic ranges and throughput. The most prevalent LBA format, enzyme linked immunosorbent assay (ELISA), requires the availability of antibodies against the candidate biomarker. Selectivity of ELISA has to be optimized, and MS can be used as an adjunct methodology to confirm specificity during ELISA method validation. Furthermore, hybrid immunoaffinity/MS approaches have shown promise as quantitative methods that confer the high resolution and chemical specificity of MS with immunoaffinity-based pre-concentration [53–58].

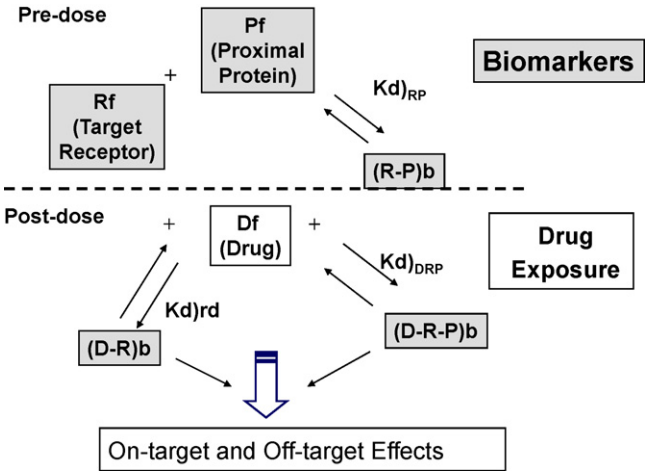


Fig. 3. Binding of target receptor and proximal proteins to a drug with target-mediated mechanism. The shaded boxes are bound and free forms related to biomarker measurements prior to and after drug administration. F: free; b: bound; Kd: binding constant; D: drug; R: Receptor; and P: Proximal binding protein.

2.4. What should be measured?

2.4.1. Multiplex of biomarker panels

In the clinical diagnostics arena, a panel of tests is often requested by a physician to determine the disease status of a patient. The use of multiple biomarkers increases the certainty for diagnosis and predictive power for prognosis. Similarly, during the early phase of drug development (phase 0), a panel of potential biomarkers is tested with different therapeutic candidates at various doses to pick the best candidate based on dose–effect relationship information. Multiplex assays for multiple analyte profiling (MAP) using bead-based (e.g. Luminex) or planar (e.g. MesoScale Discovery) formats can save time and require less sample. Multiplex assay s are best used in the early phase for initial biomarker screening due to disproportionately variable biological ranges of the analytes and non linearity of the assays [59]. After the selection of the few relevant biomarkers, the decision can be made to use either several single-analyte methods or a multiplex of a few compatible analytes to support drug development in later phases.

2.4.2. Single-analyte measurement

Many drug actions are through target mediation of receptors. Highly specific drugs, such as most protein therapeutics, will bind to both tissue membrane-bound and soluble circulatory receptors. The cleaved ectodomain of the membrane receptor may exist in several forms in circulation as soluble receptors with variable binding affinities to the drug. In addition, the target receptor (target biomarker) may bind to proximal protein biomarkers as depicted in the diagram in Fig. 3. Depending on the binding affinities and molecular sizes of the interacting species, LBA quantification may be affected due to competitive binding to the assay reagent(s) or decreased binding due to steric hindrance.

In the scenario that the LBA reagent shares a common biomarker-binding epitope with that of the drug, the unbound (“free”) form of the biomarker is being measured. In the presence of the drug, which is often dosed at a very high molar ratio to the biomarker, the free concentration will usually be too low to be measurable at most time points except at the late recovery phase.

Another scenario is that the LBA reagent recognizes a different epitope or the reagent has a much higher binding avidity than that of the drug to the biomarker. "Total" concentration (drug-bound + free) of the biomarker is therefore measured in the presence of the drug. Since the drug–receptor complex is the driver of the PD effect, the knowledge of which forms are being measured aids with data interpretation and pharmacokinetic (PK)/PD modeling. In addition, determination of the total concentration of the target biomarker may provide information on possible compensatory rise due to induction or membrane shedding [60,61].

Sometimes the information regarding reagent specific binding epitopes or binding constants is not available. Other methods to measure the total concentration can be used. One such method involves alkaline or acidic pretreatment of the sample to dissociate protein binding, with pH neutralization before LBA reagent addition [62,63].

In the case of small molecule biomarkers including peptides, an extraction method using organic solvents or solid phase separation can dissociate protein interactions prior to LC–MS/MS analysis. Such a method provides total quantification. A biomarker may interact with cognate proteins and is similar to a high protein binding drug such as warfarin where quantification of the total concentration of the drug is dependent upon dissociative pretreatment. The understanding of protein binding is required for drug development; however, issues of protein binding for biomarkers have rarely been discussed due to the lack of thorough understanding and adequate analytical tools.

2.5. Yardstick of measurement – standard reference affecting assay type

Unlike drug assays, where samples are quantified against calibrators prepared from highly purified and well-defined reference standards, protein biomarkers are often heterogeneous, existing in multiple forms in the body. Biomarker standards can be produced in pure form by synthetic or recombinant technology, which may not be the same as the endogenous species. Standards can be made by purification from biological matrix as a mixture, which may not be well defined. Depending on the situation, various assay types exist. A *Definitive Quantitative Assay* uses a well-characterized reference standard that is fully representative of the endogenous biomarker. Absolute quantitative values for unknown samples are calculated from a regression function. Only a small fraction of biomarkers such as small molecule bioanalytes and peptides (e.g. steroids, insulin) belong to this type. The majority of biomarker assays are *Relative Quantitative Assays* where the reference standard is not well characterized, not available in a purified form, or not fully representative of the endogenous form (e.g. cytokine immunoassays). The standards are used as calibrators in continuous numeric units for data regression where the endogenous analyte holds a similar dilutional response relationship (parallelism). With the lack of a reference standard or failed parallelism, a *Quasi-Quantitative Assay* may exist if the analytical response is continuous (numeric), with the analytical results expressed in terms of a characteristic of the test sample. Examples are anti-drug antibody assays (where the readout is a titer or percent bound), enzymatic assays (where activity might be expressed per unit volume) and flow cytometric assays [64].

Qualitative Assays of biomarkers are those that generate discrete, discontinuous data in either ordinal (e.g. low, medium and high, scores 1–5) or nominal (yes/no, positive/negative) formats. Qualitative methods are of limited use for PK/PD studies, but can provide valuable diagnostic/prognostic information and patient stratification (staging) such as differentiating marked effects from on/off gene expression, and activation or inhibition in relatively homogenous cell populations in immunohistochemical assay scores.

3. Method validation or qualification to support drug development and diagnostic use

The processes of biomarker method validation share some similarity to those in support of PK studies of well-defined protein therapeutics, where consensus has been reached through white paper publications and AAPS and FDA-sponsored meetings [65–68]. The processes for developing tests for FDA approval are defined [69,70], and the best practices in method validation of diagnostic tests in clinical laboratories are described in multiple publications from National Committee for Clinical Laboratory Standards (CLSI) [71–80]. In addition to the approved kits, new *in vitro* diagnostic tests (IVD) of protein biomarkers are also used in the research environment in both the diagnostic and pharmaceutical sectors. This includes for research use only or "for investigational use" (IUO) kits. For an RUO kit, almost all of the assay performance parameters have to be established by the user since the method often has not been validated or tested with rigor by the kit manufacturer. Thus, the user must be cognizant of the inherent risks and perform due diligence to adequately validate the method for the intended application.

Biomarker method validation and application in drug development are relatively recent, and the operational logistics in conducting the experiments are still evolving. Biomarker validation procedures can be quite different from one company to another or within a company itself at different sites/functions. Therefore, confusion is possible with laboratories trying to decide the best practices to use [27]. In addition, the heterogeneity issues of protein biomarkers require special considerations for method validation that are different from those of drug analyses [29]. The differences in the intended uses of biomarkers in diagnostics vs. those for drug development also require unique logistics and guidelines. Table 1 lists the different applications of biomarkers for drug pharmacokinetic analysis and clinical diagnosis. Clearly, the practices of biomarker method validation and application for drug development do not entirely follow those of diagnostics or PK assays but have their own features to fit specific phases of the application [29]. A different set of assay performance parameters may need to be established by the bioanalytical laboratory by conducting more validation experiments than those of diagnostic or prognostic uses. The variable purposes at various phases of drug development or IVD commercialization require a "fit-for-purpose" approach in method validation and application.

Biomarker method validation strategies can be categorized based on support of exploratory or advanced studies as depicted in Fig. 2. In drug development, exploratory studies include those in *in vitro* cell systems, animal models or in human with stringent controls. Advanced applications include those for the characterization and qualification of a novel biomarker, and/or as a primary PD or safety marker used to support regulatory filing of a drug candidate. A biomarker substudy is usually considered exploratory and is generally used for internal decision-making. The biomarker results may prove useful for multiple drug development programs in multiple therapeutic areas. In some cases, the outcome is negative or uninterpretable, and the hypothesized biomarker becomes no longer viable. Prospectively, biomarker substudies are not intended for inclusion in regulatory submissions; however, the results may validate the biomarker as a useful endpoint for future registration studies. The rigor of method validation increases from exploratory to advanced use.

For example, advanced application in drug development often requires extensive selectivity tests against matrices from various patient populations and specificity tests against the drug compound(s). Furthermore, if a commercial diagnostic kit is used directly for PD assessment of a drug that decreases the biomarker

Table 1
Application of method validation of biomarkers from drug development vs. clinical diagnosis and drug analysis.

	Biomarker for drug development study	Biomarker for diagnosis	Therapeutic drug analysis
Intended application	Safety. Efficacy (Proof of biology, PD). Exposure–effect relationship	Diagnosis (distinguish diseased from healthy). Prognosis (predict clinical outcome)	Pharmacokinetics. PK parameters
Method types[a]	All four types, predominantly relative quantification	All four types	Definitive quantification
Reference standard calibrators	Many are not well characterized or pure. RUO kit standards vary within and between vendors	Vendor consistent and well established. Under GMP	Well defined. Under GMP
Analytes	Endogenous biomarkers, less well defined		Exogenous drug, well defined
Method and reagent source	Developed in-house or from RUO kits	Well established, FDA approved kits from vendor	Developed in-house
Assay selectivity and specificity	May not be specific, prone to matrix effect		Specific for drug compound
Calibrator matrix	Substituted matrix (buffer or depleted biological matrix)		In-study matrix
Validation sample and QC preparation	QCs made by spiked ref standard into buffer or matrix. Sample Controls by pooling clinical samples	Controls from vendor. May not use the exact biological matrix. Common QC pools among labs	Four to five VS levels and three QC levels made in-study matrix
Accuracy	Mostly relative accuracy. QC in every run for acceptance	QC assessment may not be performed in every run for acceptance	Absolute accuracy. QC in every run for acceptance
Assay acceptance criteria	Confidence interval or a variant of 4–6–X[b] rule for each run	Westgard rule. CAP test for lab accreditation	Confidence interval or a variant of 4–6–X[b] rule for each run

[a] Method types are: (1) Definitive quantitative assay – with well-characterized calibrators fully representative of the endogenous biomarker. (2) Relative quantitative assays – calibrators not well characterized, not available in a purified form, or not fully representative of the endogenous form. (3) Quasi-quantitative assay – with continuous analytical responses not defined by calibrators and analytical results expressed in terms of a characteristic of the test sample. (4) Qualitative assays – with ordinal or nominal discrete data.

[b] Out of six QCs, at least four must be within X% of the nominal or target value for the analytical run to be acceptable. The six QCs consist of two each at low, mid and high concentration and at least one should be acceptable.

level, the method may not be sensitive enough to describe the PD profile and would need to be modified for increased sensitivity.

3.1. Pre-analytical considerations

3.1.1. Choice of biological fluid or tissue

Biological fluids from humans and animals span a wide range of matrices, including whole blood, plasma, serum, urine, cerebrospinal fluid and synovial fluid. It is desirable to analyze a biological fluid with biomarker levels that mirror those of the target tissues. Often the knowledge of the direct correlation between biological fluid and tissues is not available, and an assumption is made that the circulatory levels of biomarkers are proportionally related to those of the target tissues. Such an assumption needs to be confirmed later in preclinical and clinical systems. The most common biological fluids are urine and serum (or plasma) as a result of accessibility and ease of collection. Of these, although collection of urine is the most non-invasive type of collection, urine would not be a good choice for monitoring the intact protein biomarker except for patients with dysfunctional glomerulate filtration. However, peptides that are shed from biomarker proteolysis are successfully monitored in urine samples and linked to diseases [54,81–83]. Additionally, biomarker analysis in urine tends to suffer more from matrix interference than serum/plasma and requires normalization with creatinine. Thus, although somewhat more difficult to collect, plasma/serum is generally the most widely collected biological fluid for protein biomarker analysis. It is important to note that with the assumption of equilibrium of the circulation with the target tissue, a large dilution of protein biomarker concentration exists in plasma/serum relative to the target tissue concentration.

3.1.2. Target ranges

Basal levels of putative biomarkers are often disparate between healthy/disease populations and within an individual depending on his/her temporal health status. Therefore, method validation requires finding the target modulation ranges. The initial concentration range of a biomarker in healthy and disease populations is usually available from the literature or a vendor brochure if a commercial kit is available. Data may already exist that define the inter- and intra-donor variability in these populations [84]. If there are statistical differences in the ranges of healthy and disease populations, the method assay range must cover all the expected levels, as well as any additional changes to the levels that result from the presumed modulation of the drug. If the differences are due to temporal intra-donor factors instead of between populations, sample collections for determining suitable baseline levels of each patient are important to monitor changes with time. Given the higher variability of some assays, it may be useful to collect extra sample volumes of pre-dose time points, stored in multiple aliquots, to be assayed with various post dosed samples in the same run to avoid inter-assay variability.

3.1.3. Sample collection

Pre-analytic variables have hindered data utility in proteomic biomarker discovery and validation [85,86]. Inappropriate collection time and other adverse conditions often lead to confounding or uninterpretable biomarker data. To monitor drug effect, the lag time after dosing should be considered for determining optimal sampling times.

Sample integrity must be maintained from collection to the point of sample analysis. Intervening events may include preprocessing, transportation, and short/long-term storage. After

initial analysis, continued sample integrity is required if there is further storage and re-analysis. The measurement of a biomarker after collection, transport, and storage should yield a result as close as possible to that of the uncompromised biomarker *in vivo*. For drug development clinical trials, most of the processes will be followed via the involvements of a central lab that integrates collections from multiple sites, bar-codes, and transports samples to the analytical lab. For routine diagnostic testing being performed in the physician's or hospital laboratory, transporting and long-term storage of samples may be avoided with on-site analysis. However, even under the latter conditions, errors from the pre-analytic phase resulting from variable specimen collection can be higher than those arising from ensuing sample analysis itself [87].

It is important to standardize the techniques for all sample collection and handling early in method development and later in method validation and to remain consistent throughout the duration of the use of the assay [77,80]. The type of venipuncture needle, duration of the blood draw, type of collection tube container, and the type and concentration of anticoagulant may affect the biomarker stability. For example, collection of blood through a small bore needle or high speed centrifugation to pellet cells may cause a shearing effect that can activate endothelial cells leading to the generation of artifacts. The *g*-force and rpm conversion must be well defined for each laboratory's centrifuge to avoid mistakes. For other biological fluids of relatively low protein content (e.g. urine, cerebral spinal fluid), collection tubes, transfer pipettes, and storage containers must be evaluated to ensure that no significant adsorption of the biomarker to the contact surfaces occurs. Although serum is preferred over plasma for most clinical tests, some biomarkers can only be quantified accurately in plasma. Examples are those that are susceptible to proteolysis or are involved in the coagulation pathway or platelet activation. Inhibitors of relevant activation reactions or proteolysis may need to be included in the venipuncture syringe or added to the sample promptly after the blood draw.

3.1.4. Method validation plan

A final analytical procedure must be written, and the validation (or qualification) experiments carried out according to the procedure. Usually a validation plan should be written prior to conducting the experiments. It includes the intended purpose and the type of validation (such as method qualification for exploratory use or full method validation for advanced application). Then the scope of the experiment is delineated. The basic experiments for exploratory use are: assay range finding, accuracy and precision, selectivity, specificity, stability (at least one freeze/thaw cycle and bench top). Compared to exploratory use, the advanced method validation should be more rigorous in order to provide more detailed performance characteristics, with additional tests on patient population ranges, more runs on accuracy and precision from multiple analysts/reagent lots/instruments, selectivity and specificity tests using patient samples, and stability tests subjected to conceivable conditions over long-term use.

Each experiment should include standard calibrators and multiple replicates of quality controls (QC), pooled sample controls (SC) of authentic samples, and various test samples.

3.1.5. Standard calibrators

3.1.5.1. Reference standard material. The recombinant form of a protein biomarker is often provided as the reference material. It does not fully represent the endogenous analyte, which usually exists in various forms in the biological matrix. A well-defined reference standard serves as a scalar for the relative measurement of the endogenous species. The reference material can usually be produced in sufficient amount and characterized with respect to well-defined molecular weight and purity, thereby enabling concentration and molar equivalence calculations. PK data are often reported in mass/volume units; however, unit expression in molar equivalence is more desirable to quantify the biological effect of the biomarker. It would be useful to be consistent in using the same unit for a specific biomarker among studies/laboratories and include the molar and mass/volume conversion factor in the method and for publications. Sometimes the reference standard is a partially purified material from biological sources. In this case, the characterization, chemical components and consistencies of production batches may be less adequate than those of purified recombinant standards.

In addition to the production of sufficient quantities, consistency of productions of the reference standard must be high. The potency of the reference material may vary from one lot to another. In general, the practice in the pharmaceutical industry is to make adjustment in the substock solution so that the standard calibrators behave consistently within a method that is applied to various studies. On the other hand, a kit manufacturer usually has a primary gold standard that the new lot of reference material is calibrated against. Thus, the calibrator concentrations may differ slightly from one lot to another. To support various clinical programs, it is important to procure a consistent source of reference standard material to prepare the standard calibrators and control samples before conducting the experiment. In such cases, the adoption of commercial kits may only use the assay plates and ligand reagents and not the included reference standards [88].

When there is a batch change, it is important for the vendor to collaborate with end users to ensure that the assay performance in the laboratories is not adversely affected. Often times, multiple batches of materials (reference standards and critical reagents) are preferably tested during method validation [23].

Proper documentation of each batch of reference material is necessary. Minimally, documentation of the characterization and stability of a standard, such as a certificate of analysis (CoA) and/or a certificate of stability (CoS) is typically available from suppliers. With respect to drug development, documentation of the stability of the respective lots over the entire time span of the development program should be obtained.

3.1.5.2. Blank matrix for standard preparation. Ideally, standards should be prepared in the same matrix as the intended sample [89,90]. However, it is difficult to find "blank" control matrix for preparation of standard calibrators since most biomarkers are endogenous. An option is to perform an initial screen on a multitude of matrix lots against standard calibrators prepared in a protein-buffer solution to identify a few "blank" matrix lots. These are then pooled for standard preparation. When that is not possible, an alternative "blank" matrix can be prepared by depleting the endogenous analyte using methods such as charcoal stripping, high temperature incubation, acid or alkaline hydrolysis, or affinity chromatography. Alternatively, a protein-containing buffer (e.g. BSA in phosphate buffer) or the matrix from another species with a non-cross-reactive homolog of the biomarker is used. For example, equine serum is often used by diagnostic kit manufacturers. The use of these types of "blank" matrices requires studies of matrix effects and parallelism during method development and validation to understand the impact that differences from patient samples have on assay results [29].

3.1.5.3. Linearity and regression model. The assay linearity is defined by a mathematical function that is fit to the standard calibrators [90]. The assay range is defined by the lower limit of quantification (LLOQ) and upper limit of quantification (ULOQ) that meet *a priori* criteria set before the method validation. No extrap-

olation beyond the limits of the assay range is allowed for sample analysis [89]. The response concentration relationship of a binding assay is curvilinear. Therefore, unlike a linear function with constant variance over the entire range, regression models for LBA are typically described by a 4- or 5-parameter logistic function with non-constant variance. More standard points (6–8) plus additional anchor points are required to adequately define the 4- or 5-parameter function, and weighting factors are used to achieve the best curve fit with the least variance [65].

It is worthy to note that "linearity" has a different meaning in method validation for drug development support vs. diagnosis. In the diagnostic sector, most clinicians prefer to generate a straight line for the calibrators. Linearity is a measure of the degree to which a calibration curve approximates a straight line. Therefore, linearization is often applied for a high order standard curve using log transformation, or only the linear part of a spline fit is used. Since the application for diagnostic kits is to compare the patient sample result against a certain cut off point, curve fitting over the entire standard range may not be as important as that for a drug development application. Therefore, a diagnostic kit may not be directly usable in drug development as it often does not provide sufficient number of standards, rarely anchor points, and no weighting recommendation for curve fitting.

If a commercial kit is to be adopted for a drug development application, more standard points may need to be added, and the appropriate curve fitting is determined from data obtained during method validation. Proper curve fitting is crucial for optimal assay sensitivity and reliable calculation [29,91]. The best curve fitting model offers the best fit for all the standards in the precision profile. In some cases, a less than optimal curve fit over the entire working range with a more focused fit (i.e. minimized %CV) in the low concentration region provides greater assay sensitivity. For example, this strategy would be applied to the measurement of very low abundance biomarkers and for the assay of "free" target biomarkers in the presence of high concentrations of binding therapeutic agents.

3.1.6. Critical reagents

For ELISA methods, antibody pairs are typically used as capture and detection reagents. In general, the more selective antibody is chosen as the capturing agent, especially if it is more readily available than the other member of the pair. Signal generation and detection can be achieved by using a tertiary antibody that is conjugated to a reporter enzyme, such as horseradish peroxidase. Alternatively, a biotinylated detector antibody can be used together with a biotin-binding protein (e.g. anti-biotin antibody or an avidin-type protein) conjugated to a reporter enzyme. The sensitivity of an assay can be increased by varying the number of reporter enzyme molecules on the detection reagents, using multivalent strategies to increase the effective signal from each analyte captured, or utilizing enzyme substrates that exhibit enhanced signal intensity (e.g. chemiluminescence vs. colorimetric).

Some assays use biologically-relevant target receptors or their fragments as binding partners, most often in concert with a specific secondary antibody. This arrangement may improve selectivity for specific ligands (e.g. a cytokine activated from a latent precursor, or a particular subtype of ligand with distinct binding characteristics from its homologs). Since use of a receptor moiety may mimic actual *in vivo* target mediation, the binding selectivity of such reagents can offer added biological relevance to quantification.

Commercial reagents for many protein biomarkers are available. There are established assay kits (FDA approved or cleared for diagnostic use) and also kits for RUO or IUO [69]. The RUOs are often adopted for drug development and must be validated for that intended application. If commercial reagents are not available, antibodies against a specific biomarker can be prepared in-house if the facilities for doing so exist, or antibody production can be outsourced to any of the numerous custom antibody manufacturers. Antibody preparations require astute characterization of parameters such as purity, affinity constant, and cross-reactivity in order to achieve the most robust assay results.

Similar to reference standard material, the critical reagents are required to be of sufficient quantities and consistency. During method validation, it is favorable to assess the quality of multiple preparations of the capture and the detector reagents. If the materials are from commercial sources, negotiation with the vendor to assure a consistent and sufficient supply of the same batch material is desirable if possible. Over the time span of a clinical program, it is vital to keep documentation of the reagents (source, identity, potency or concentration, lot number, and stability). If reagent storage stability is not available, it should be tested by the user. If possible, it is preferable that the same lot of capturing ligand be used throughout a study. If a reagent change is required, qualification experiments need to be performed on the new reagents.

3.2. Method validation experiments

3.2.1. Accuracy and precision

3.2.1.1. Validation samples and quality controls. Validation samples (VS) are used to provide assay performance characteristics such as accuracy and precision while quality controls are used for run acceptance (see Section 4). VS containing a biomarker at concentrations in the expected assay working range are used to estimate intra- and inter-run accuracy and precision. Usually they are prepared by spiking the reference material into blank matrix at five levels: LLOQ, about three times the LLOQ (low QC), mid QC, high QC and ULOQ.

Ideally, VS and QC are prepared in the intended matrix. Thus, VS/QC can be prepared by spiking reference standard into a matrix pool containing a low concentration of the biomarker. Alternatively, they can be prepared in the standard protein-buffer to conduct the accuracy and precision experiments. The conduct of accuracy and precision experiments for biomarkers is similar to that for PK with 6 runs to determine inter-assay variability and 2–5 replicates of VS in each run for determination of intra-assay variability [65,67]. The number of VS and assay runs may be less for exploratory method validation, as justified by the intended use of the method and following the validation plan.

Accuracy (expressed as %bias) is the agreement between the measured result and its theoretical true value. Precision is a quantitative measure (usually expressed as SD and %CV) of the random variation between a series of measurements from the same homogenous sample. Total error (TE) is an approximation of the sum of all systematic bias and precision. TE reflects the closeness of the test results obtained by the analytical method to the true value of the analyte for assay acceptance.

The practice of method validation in developing a diagnostic kit for FDA approval follows the guidelines of CSLI [74–76]. The common practice among laboratories is to use the FDA approved kit with minimal validation at the lab, such as only qualifying the analyst and instruments with a few runs. QC for FDA approved kits are provided with defined ranges (reference intervals) for acceptance that is determined by the mean and SD from results of multiple laboratories. Extensively historical QC data such as those in the form of Levey Jenning plots are used to track assay performance within and among participating laboratories for various kits and platforms. Such information is valuable for the user to determine if the assay is under control.

For drug development practices under GLP, the spiked QC prepared by the bioanalytical laboratory is used for assay acceptance.

The criteria are set based on TE of the method determined from the accuracy and precision experiments [65].

Most novel protein biomarker use RUO kits that may or may not include QC. Even if they do, the QCs often do not represent the authentic samples or the values and acceptable ranges may not be well defined. It is the user's responsibility to set up proper control samples to monitor assay performance.

3.2.1.2. Sample controls. If the VS/QC are prepared in an alternative matrix, or if the reference standard material does not fully represent the biomarker forms in the authentic samples, the accuracy and precision performance characterized by the VS data need to be complemented with data from sample controls. From the results of an initial screen of a multitude of matrix lots for blank matrix (Section 3.1.5.2), lots that exhibit high and low biomarker concentrations can be found for pooling. SC pools at high and low levels are aliquoted, and their levels determined in method validation experiments and pilot studies from approximately 30 runs. An acceptance criterion of mean ± 2SD can be used. In addition, since these samples reflect the biomarker forms in the authentic samples, it is preferable that tests of stability and reagent lot variability are performed with SC, instead of the spiked VS.

The common thread to compare precision and relative accuracy among multiple studies by different analytical laboratories is the SC data. Method validation is a continuous process that includes the SC in-study performance. In addition, SC can be used as part of the conformance samples to compare method/reagent changes. For example, Fig. 4 shows SC in an exploratory biomarker application using an RUO kit. The box plot shows lot-to-lot variability with the last two lots showing significantly different results than those from the other lots for both SC pools. The side bar table in Fig. 4 shows higher CV at the low pool, indicating that the assay variability increases at the low concentration region, which would impact data assessment of low concentration samples.

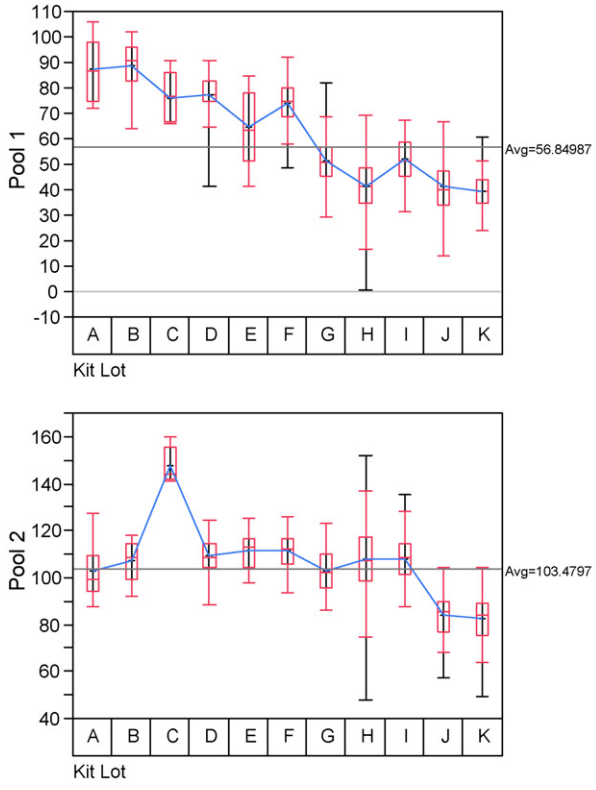
3.2.2. Sensitivity

To provide data for PK/PD studies, assay sensitivity is defined by the LLOQ, which is the lowest concentration that is measurable with acceptable levels of bias and precision and total error. Low concentration clinical samples may fall below the LLOQ. In this case, the method is not sensitive enough and should be optimized further. Sensitivity enhancement may not be possible, however, and extrapolation to values below the LLOQ and above the limit of detection (LOD) may be necessary. If the LOD, instead of LLOQ, is used as the lower limit for data inclusion, the investigator should justify this choice, be aware of the risk of higher variability in the LOD to LLOQ range, and interpret the data with caution.

LOD is often used as the analytical "sensitivity" of the assay in a diagnostic kit. A common practice to determine the LOD is to use the extrapolated concentration from a response signal of +3SD (or −3SD for a competitive immunoassay) of the mean background signal from 30 or more blank samples. An alternative statistical approach to determine the LOD has been recommended by CLSI [78]. It evaluates the Limit of Blank using sufficient blank samples and the LOD from low concentration samples with normally distributed analyte concentrations. The LOD determined by this approach would avoid analytical false negative (type II error) and false positive (type I error) values. Due to variable assay conditions, each bioanalytical laboratory should determine the LOD using the CLSI approach, instead of using the LOD stated in a kit brochure.

3.2.3. Selectivity and specificity

Specificity reflects the ability of a method to distinguish between the analyte of interest and other structurally similar com-



Kit lot no.	N	Pool 1 Mean	Pool 1 %CV	Pool 2 Mean	Pool 2 %CV
All	**450**	**56.9**	**32.6**	**103.5**	**14.5**
A	10	87.3	15.1	102.7	11.3
B	16	88.6	11.6	107.4	7.7
C	5	76.3	13.4	147.8	5.2
D	63	77.6	9.9	109.5	6.9
E	13	64.7	22.3	111.7	7.7
F	53	73.7	12.6	111.4	6.7
G	74	51.0	18.9	103.2	8.6
H	67	41.2	26.9	107.6	14.3
I	64	51.8	16.6	107.9	10.0
J	35	41.3	25.6	84.4	12.0
K	50	39.5	18.4	82.6	12.8

Fig. 4. Sample controls in an exploratory biomarker application using a "for research use only" kit. The graphs of 2 SC pools show box plots of mean, standard deviation and range for each of 11 kit lots (A–K). Pool 1 had higher bias from the overall mean than Pool 2, with a bias shift from previous kit lots starting from lot G. The side table shows that the low pool (Pool 1) has higher %CV (32.6%) than the higher pool.

ponents of the sample. Results from a non-specific assay can appear as false positives and/or an overestimation of the analyte concentration. Selectivity is the ability of the method to discriminate the analyte unequivocally from components that may be expected to be present in the sample and that alter assay results. Specificity and selectivity for LBA are method (platform)-dependent. The CLSI Working Group defines "interference" from a known source and "matrix effect" from an unidentified source [71,79]. There is no specific guideline or consensus in the pharmaceutical sector on how specificity and selectivity experiments should be conducted in method validation. The following selectivity and specificity tests for LBA are from our experience and interactions from the Ligand Binding Assay Bioanalytical Focus Group of the Association of American Pharmaceutical Scientists [29,92].

3.2.3.1. Selectivity tests. Matrix components may inhibit or enhance binding of protein biomarkers to assay reagents. Signal suppression from plasma binding proteins often occurs, resulting in a negative bias. Since health status and collection conditions may lead to variability of binding protein types and levels, selectivity tests should be conducted on several individual lots from the target population against the calibrators that are prepared in a protein-buffer or a single matrix pool [27,93,94].

Initially, matrix effect is detectable from a spike recovery test by adding the reference material to individual matrix lots from the target populations (e.g. 10–40 lots). The amount that is added must be high enough to differentiate the spiked material from the endogenous component. For example, the spike concentration cannot be substantially lower than the baseline concentration, and the spiked volume should not exceed 5% of that of the individual matrix lot. Spike recovery is calculated after subtraction of the baseline unspiked value and compared to the nominal spike concentration or the mean of the test lots (in case of a spiking bias). In general, at least 80% of the test lots are required to recover within the acceptance criteria determined from the accuracy and precision runs.

The next test of matrix effect is parallelism to evaluate if the biomarkers in the biological matrix lots have similar immunoreactivity as the reference material in the standards. Authentic samples of high biomarker concentrations from at least three individuals are diluted with the standard matrix (standard zero) with at least three different dilution factors. Fig. 5 shows examples of the parallelism test for two different biomarkers, each tested with serum samples from four individual lots. There was no effect of dilution on the calculated concentrations in biomarker A (upper panel) and parallelism was demonstrated. For biomarker B (lower panel), the observed concentrations increased with higher dilution, indicating the existence of matrix interference which was alleviated with dilution at approximately 2.5-fold (0.4 1/dilution factor).

When parallelism cannot be performed due to the lack of authentic samples with sufficiently high concentration of the analyte, a dilutional linearity test can be performed using at least three individual lots spiked with a high concentration of reference standard near the ULOQ. Parallelism and dilutional linearity experiments are in general not included in method validation for diagnostic kits. Therefore, these experiments should be included in method validation adopting a commercial kit in support of drug development.

The presence of matrix effect does not necessarily halt method validation. Further investigation on representative unacceptable lots should be performed (such as testing other dilution factors that may alleviate the effect). When the expected biomarker levels are very low, high sensitivity is required, which precludes dilution or other sample pretreatment to minimize matrix effect. In this case, a compromise is to validate a quasi-quantitative method.

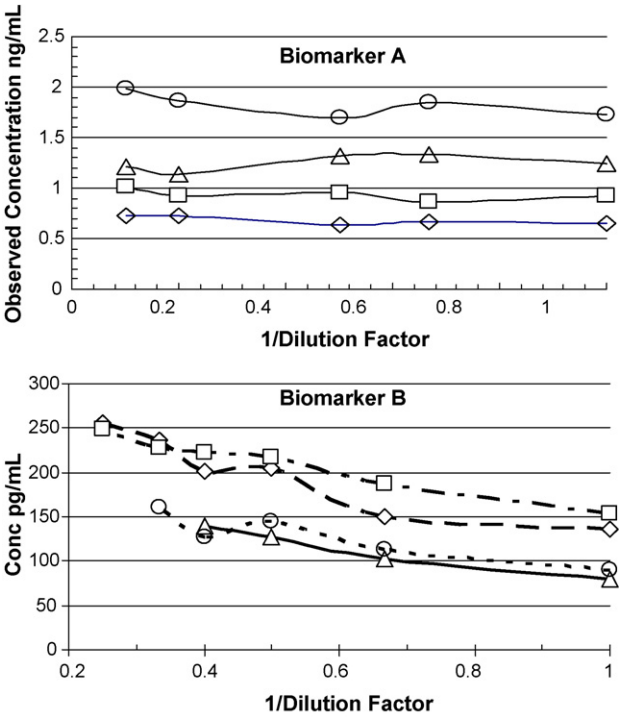


Fig. 5. Parallelism experiment using authentic samples. Clinical samples of relatively high concentrations were subjected to various dilutions with the buffer solution that was used for standard preparation. Four different individual serum lots were tested for biomarker A (Upper panel) and biomarker B (lower panel). Each symbol represents each lot. The observed concentration on the *Y*-axis was the regressed concentration times the dilution factor.

3.2.3.2. Specificity tests. As opposed to drug assays, the goal of specificity tests for biomarkers is not always to demonstrate absolute assay specificity. Instead, it is to provide information about what is being measured in order to fulfill the purpose of the intended application.

Depending on the binding reagents in the method, LBA may measure free, partially bound, bound or total analyte (Fig. 3). If an LBA reagent shares a common biomarker-binding epitope with a drug, the unbound, "free" form is measured. Interference is therefore expected in the presence of the drug compound [92]. Specificity for the biomarker in the presence of the drug should be tested using SC, without and with the addition of the drug compound at various concentrations spanning the expected therapeutic range. In addition to dosed drug, biomarker homologs or endogenous molecules of the same family may contribute to specificity problems. If reference material of the potential interfering agent is available, specificity can be tested using various amounts of the test material spiked into various levels of VS or SC [92].

The LBA method may be designed to measure "total" concentration (drug-bound + unbound) of the biomarker by using a reagent recognizing a different epitope or by adding a dissociation pretreatment step prior to analysis [62,63]. In that case, the lack of interference from various drug concentrations must be confirmed.

For a well-characterized biomarker measured by a definitive quantitative assay, any *in vivo* truncated forms need to be tested for specificity. This applies to various regulatory peptides [95]. MS methods can be used to confirm specificity for an LBA method as well as detect differences in isoforms of a biomarker in diseased populations [50,96,97].

3.2.4. Stability

Reagent stability should be evaluated during the pre-study method validation and extended with in-study validation data. Sample stability is evaluated with respect to the type of biological fluid, collection/processing procedures, storage temperature and time, freeze/thaw cycles and bench top conditions. For advanced applications, extensive stability testing is most preferably performed with SC that experience the conceivable conditions to which the study samples may be subjected. The in-study SC chart can be used for trend analysis of long-term storage stability and to assess lot-to-lot variability of key assay reagents [88].

4. Assay acceptance criteria

For PK and PD applications, the 4–6–*X* rules are common practices used for the acceptance criteria [67,68,98]. In a typical assay batch of 96 samples, two QCs in each of three levels (low, mid, and high concentrations) are included. Out of six QCs, at least four must be within *X*% of the nominal or target value for the analytical run to be acceptable. For each level, at least one should be acceptable. For methods to support PK studies, the value of *X* is usually 15% for LC–MS methods, and 20% for LBA [67,89,98,99]. Instead of a fixed "*X*", it has been recommended to determine "*X*" based on the performance data from method validation [65].

No guidance or consensus has been given for acceptance of biomarker assays. The fit-for-purpose approach of biomarker method validation requires consideration beyond the method performance. Other factors to be considered are the intended use of the data and the biological modulation of the biomarker in the study populations such as normal vs. disease. The process of setting acceptance criteria for a protein biomarker measurement follows an evolving path. During the exploratory phase, the acceptance criteria are usually set according to the QC total error from the initial method validation. After application in pilot studies, the biological data from subject samples can then be used to refine the initial acceptance criteria. For example, an assay with 50% total error may still be acceptable if a 2-fold treatment effect is observed.

The same acceptance criteria being used for a given method based on its performance may be convenient for an analytical laboratory. However, the intended purpose of the application and the possible outcomes in the specific application should be considered. For example, the effect of one population/indication may be different from another (e.g. change from a 2-fold treatment effect into only 30%), which may require a more stringent method and/or enroll more subjects to increase the predictive power in the application for the other population/indication.

For diagnostic applications, instead of pre-study method validation, the in-study method performance data are used to set assay acceptance criteria. Cumulative data such as the QC charts are used to determine if the assay is under control. Statistical trends such as the Westgard rules are commonly used to detect performance beyond 2SD of the performance mean [100,101].

5. Co-development of biomarkers for drug development and diagnostics

Co-development of biomarker assay kits for both sectors has gained support by the FDA [102]. There have been only a few examples of co-development. For example, HercepTest® was co-developed with Herceptin for patient entry criteria (DAKO HercepTest™). In addition, an immunoassay kit for the extracellular domain (p97–115 KD) of the HER2/neu receptor in human serum was developed for clinical patient monitoring [103].

It is important for the diagnostic sector to recognize the market of biomarkers and invest in kit assay development and commercialization to support the use in both post-approval data collection and diagnosis/prognosis. Commercial genotyping assays and gene arrays have been developed, bringing valuable information in pharmacogenomics [102]. Commercial development of protein biomarker panels is more challenging, where many of these will remain as research kits. For a single drug class, the diagnostic kit for a pivotal single biomarker may be an easier target. Guidelines have been issued on clinical evaluations of immunoassays to support IVD products or evaluation of clinical performance and new therapeutic agents [76]. The application of an assay system to monitor post-approval patients would require 510(K) clearance from Office of in Vitro Diagnostics Device Evaluation and Safety of the FDA for clinical use. There are several considerations for protein biomarker kit commercialization:

- *Standardization of reference material*: Substantial differences among manufacturers or even among lots from one manufacturer are observed for protein biomarkers due to the variable sources and process of purification, e.g. recombinant, synthetic or native [23]. It is difficult to define which form or combination of forms of the biomarker should be used as the standard reference material. There have been continuing efforts to establish 'gold' standards for tumor markers that have been deemed as predictive to provide diagnostic and prognostic assessment for patients [104–107]. For novel biomarkers, similar collaborative efforts from the diagnostic and pharmaceutical industries would be needed to render gold standards as reference materials.
- *Standardization of SC*: For biomarker characterization and evaluation, the bioanalytical laboratories often use SC prepared in-house to assess the assay performance. However, if the assay is going to be used in a diagnostic laboratory, standardized SC with the defined target values must be available from a repository or commercial sources for clinical laboratory certification and tracking as well as providing the ability to pool the statistics within and among laboratories [101].
- *Standardization of clinical and assay protocol*: The clinical protocol needs to include patient conditions, specimen collection and processing. Assay protocol must include the reagent sources and maintenance for performance continuity. Conformance test protocols should also be in place with SC and clinical incurred samples for laboratory comparison or reagent/instrument qualifications.
- *Information depository and sharing*: The following information should be accessible to facilitate utilization by participating laboratories: (1) the expected biomarker ranges in normal individuals and targeted populations, (2) the expected temporal modulation after therapeutic treatment, (3) stability information of the analyte in biological matrices, and (4) the conditions for sample collection and storage to preserve analyte integrity.

6. Perspectives

The application purpose of a biomarker and its method validation cannot be isolated without connection to various aspects in drug development. The decision process is inter-related with knowledge gained from disease pathways and patient data from exploratory and advanced studies. There is a need for the development of software to handle multiple biomarkers monitored during drug development. Optimal software would integrate multiple biomarkers' PD profiles with PK data to allow models describing physiological compartments of exposure and effects on the disease and host biology. Preferably, knowledge of protein biomarkers is

integrated with those of genomics, glycomics and metabolomics. Biokinetics of the target, proximal, and distal biomarkers should be tracked in various patient populations. Thus, clinical validation of a biomarker goes through learning phases from continuously updated knowledge. The foundation of such knowledge rests upon reliable quantitative methods appropriately validated at each phase of application.

LBA has been the major method in protein biomarker application. Other method types must be included for biomarker investigation to supply a complete picture of biology and chemistry for the biomarker's interaction with the drug candidate and proximal proteins. Western blot, 2-D gel electrophoresis, MudPIT LC–MS/MS and imaging have been used for biomarker research. Quantitative applications are being developed [46,48,52]. Technology integration of laser-microdissected cryostat sectioning, ProteinChip, gene microarray, immunohistochemistry, multiplex binding assays, and hyphenated MS methods (e.g. FACS-MS, MALDI-MS and affinity-MS) will continue to impact biomarker discovery and the application to translational medicine. Since these technologies have evolved in a research environment, translation for the application to preclinical and clinical samples requires the cooperation of the scientists from both discovery and clinical realms. Bioassays coupled with LC–MS to quantify reaction products can provide direct relevant chemical and biological information [108]. Affinity techniques can be coupled to MALDI and LC–MS/MS. Specific receptor proteins or antibodies can be covalently attached to a solid phase as a capture device for selective enrichment of low concentration biomarkers, and the eluant from the solid phase can subsequently be interrogated by MS for chemical information and quantification [53,54]. Advances in molecular imaging will continue to contribute to diagnostic and drug development in tumor specific biomarkers [109,110]. Flow cytometers can be used to select and enrich blood cell populations. The isolated cells can be manipulated and subsequently interrogated by LBA or MS methods [111,112].

Multiple technological tools including cell-based, immunoaffinity and biophysical methods will contribute to the integral knowledge of protein biomarker actions in target cells and biological fluids and the concomitant impact of drug intervention. Therefore, in addition to the need for integrating software, there is also a need for integrating methodologies. The development and validation of integrating software and technologies would enhance the overall knowledge of protein biomarkers and add greater support to understanding diseases and improving patients' lives.

Acknowledgements

The authors are grateful to the Biomarker Subcommittee in the Ligand Binding Assay Bioanalytical Focus Group of the American Association of Pharmaceutical Scientists for ideas and collaborations in the "fit-for-purpose" method validation concept. We also thank Binodh DeSilva for critical review of this manuscript; and Yuling Wu, Laura Brunner, Daniel Burns, Lennie Uy, and Jin Wang for their contributions.

References

[1] FDA, Innovation or Stagnation? Challenge and opportunity on the critical path to new medical products, 2004.
[2] E. Zerhouni, Science 302 (2003) 63.
[3] W.A. Colburn, J. Clin. Pharmacol. 37 (1997) 355.
[4] J.A. DiMasi, R.W. Hansen, H.G. Grabowski, J. Health. Econ. 22 (2003) 151.
[5] P.R. Jadhav, M.U. Mehta, J.V.S. Gobburu, Am. Pharm. Rev. 7 (2004) 62.
[6] G. Levy, Clin. Pharmacol. Ther. 56 (1994) 356.
[7] C.C. Peck, W.H. Barr, L.Z. Benet, Pharm. Sci. 81 (1992) 600.
[8] T.D. Bjornsson, Eur. Pharm. Rev. 1 (2005) 17.
[9] M.S. Pepe, R. Etzioni, Z. Feng, J. Natl. Cancer Inst. 93 (2001) 1054.
[10] J.A. Wagner, Dis. Markers 18 (2002) 41.
[11] J.A. Wagner, S.A. Williams, C.J. Webster, Clin. Pharm. Ther. 81 (2007) 104.
[12] J. Berger, J.A. Wagner, Diabetes Technol. Ther. 4 (2002) 163.
[13] J.A. Wagner, J. Clin. Endocrinol. Metab. 87 (2002) 5362.
[14] U. Krishnamurti, M.W. Steffes, Clin. Chem. 47 (2001) 1157.
[15] C.S. Chu, K.T. Lee, M.Y. Lee, H.M. Su, W.C. Voon, S.H. Sheu, W.T. Lai, Am. J. Cardiol. 97 (2006) 646.
[16] S.M. Haffner, Am. J. Med. 120 (9, Suppl. 1) (2007) S10.
[17] Y.H. Lee, R.E. Pratley, Curr. Diab. Rep. 5 (2005) 70.
[18] S. Liu, L. Tinker, Y. Song, Arch. Intern. Med. 167 (2007) 1676.
[19] M. van Doorn, M. Kemme, M. Ouwens, E.J. van Hoogdalem, R. Jones, H. Romijn, M. de Kam, R. Schoemaker, K. Burggraaf, A. Cohen, Br. J. Clin. Pharmacol. 62 (2006) 391.
[20] M. Khatami, Cell Biochem. Biophys. 47 (2007) 187.
[21] I. Quraishi, M. Rishi, M. Feldman, M.J. Wargovich, B. Weber, Appl. Immunohistochem. Mol. Morphol. 15 (2007) 45.
[22] M.A. Rubin, J. Urol. 172 (2004) S18.
[23] F.C. Sweep, H.A. Fritsche, M. Gion, G.G. Klee, M. Schmitt, Int. J. Oncol. 23 (2003) 1715.
[24] L.R. Yu, T.D. Veenstra, Expert Rev. Mol. Diagn. 7 (2007) 507.
[25] J.W. Lee, W.C. Smith, G.D. Nordblom, R.R. Bowsher, Validation of Assays for the Bioanalysis of Novel Biomarkers, in: J.C. Bloom, R.A. Dean (Eds.), Biomarkers in Clinical Drug Development, Marcel Dekker, New York, 2003, pp. 119–149.
[26] FDA, Using Disease, Placebo, and Drug Prior Knowledge to Improve Decisions in Drug Development and at FDA, 2006.
[27] J.W. Lee, R.S. Weiner, J.M. Sailstad, R.R. Bowsher, D.W. Knuth, P.J. O'Brien, Pharm. Res. 22 (2005) 499.
[28] F. Goodsaid, F. Frueh, AAPS J. 9 (2007) E105.
[29] J.W. Lee, V. Devanarayan, Y.C. Barrett, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J.A. Rogers, R. Millham, P.J. O'Brien, J. Sailstad, M. Khan, R.S. Weiner, C. Ray, J.A. Wagner, Pharm. Res. 23 (2006) 312.
[30] S.A. Stoch, J.A. Wagner, Int. J. Pharm. Med. 21 (2007) 271.
[31] S. Kummar, R. Kinders, L. Rubinstein, R.E. Parchment, A.J. Murgo, J. Collins, O. Pickeral, J. Low, S.M. Steinberg, M. Gutierrez, Nat. Rev. Cancer 7 (2007) 131.
[32] U. Manne, R.G. Srivastava, S. Srivastava, Drug Discov. Today 10 (2005) 965.
[33] M. Roessler, W. Rollinger, L. Mantovani-Endl, M.L. Hagmann, S. Palme, P. Berndt, A.M. Engel, M. Pfeffer, J. Karl, H. Bodenmuller, J. Ruschoff, T. Henkel, G. Rohr, S. Rossol, W. Rosch, H. Langen, W. Zolg, M. Tacke, Mol. Cell. Proteomics 5 (2006) 2092.
[34] A. Guo, J. Villen, J. Kornhauser, K.A. Lee, M.P. Stokes, K. Rikova, A. Possemato, J. Nardone, G. Innocenti, R. Wetzel, Y. Wang, J. MacNeill, J. Mitchell, S.P. Gygi, J. Rush, R.D. Polakiewicz, M.J. Comb, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 692.
[35] W.K. Han, J.V. Bonventre, Curr. Opin. Crit. Care 10 (2004) 476.
[36] M.W. McIntosh, Y. Liu, C. Drescher, N. Urban, E.P. Diamandis, Clin. Cancer Res. 13 (2007) 4422.
[37] S. Pan, M. Shi, J. Jin, R.L. Albin, A. Lieberman, M. Gearing, B. Lin, C. Pan, X. Yan, D.T. Kashima, J. Zhang, Mol. Cell. Proteomics. (2007) 1818.
[38] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J.M. Ren, J. Yuan, C.E. Bakalarski, J. Villen, J.M. Kornhauser, B. Smith, D. Li, X. Zhou, S.P. Gygi, T.L. Gu, R.D. Polakiewicz, J. Rush, M.J. Comb, Cell 131 (2007) 1190.
[39] C.E. Bakalarski, W. Haas, N.E. Dephoure, S.P. Gygi, Anal. Bioanal. Chem. 389 (2007) 1409.
[40] P.A. Everley, C.E. Bakalarski, J.E. Elias, C.G. Waghorne, S.A. Beausoleil, S.A. Gerber, B.K. Faherty, B.R. Zetter, S.P. Gygi, J. Proteome Res. 5 (2006) 1224.
[41] P.A. Everley, J. Krijgsveld, B.R. Zetter, S.P. Gygi, Mol. Cell. Proteomics 3 (2004) 729.
[42] D.S. Kirkpatrick, S.A. Gerber, S.P. Gygi, Methods 35 (2005) 265.
[43] J.P. Lambert, M. Ethier, J.C. Smith, D. Figeys, Anal. Chem. 15 (2005) 3771.
[44] M.P. Washburn, D. Wolters, J.R.r. Yates, Nat. Biotechnol. 19 (2001) 242.
[45] C.W. Turck, A.M. Falick, J.A. Kowalak, W.S. Lane, K.S. Lilley, B.S. Phinney, S.T. Weintraub, H.E. Witkowska, N.A. Yates, Mol. Cell. Proteomics 6 (2007) 1291.
[46] T. Wehr, LC–GC 25 (2007) 1.
[47] Z.J. Sahab, S.M. Semaan, Q.A. Sang, Biomarker Insights 2 (2007) 21.
[48] T.D. Veenstra, J. Chromatogr. B 847 (2007) 3.
[49] E. Kuhn, J. Wu, J. Karl, H. Liao, W. Zolg, B. Guild, Proteomics 4 (2004) 1175.
[50] T. Oe, B.L. Ackermann, K. Inoue, M.J. Berna, C.O. Garner, V. Gelfanova, R.A. Dean, E.R. Siemers, D.M. Holtzman, M.R. Farlow, I.A. Blair, Rapid Commun. Mass Spectrom. 20 (2006) 3723.
[51] D.R. Barnidge, M.K. Goodmanson, G.G. Klee, D.C. Muddiman, J. Proteome Res. 3 (2004) 644.
[52] L. Anderson, C.L. Hunter, Mol. Cell. Proteomics 5 (2006) 573.
[53] M. Hall, J.W. Lee, C. Spahr, H. Lu, R. Ortiz, 56th Annual ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, USA, 2008, p. ThP 482.
[54] E.E. Niederkofler, K.A. Tubbs, K. Gruber, D. Nedelkov, U.A. Kiernan, P. Williams, R.W. Nelson, Anal. Chem. 73 (2001) 3294.
[55] U.A. Kiernan, R. Addobbati, D. Nedelkov, R.W. Nelson, J. Proteome Res. 5 (2006) 1682.
[56] U.A. Kiernan, D. Nedelkov, R.W. Nelson, J. Proteome Res. 5 (2006) 2928.
[57] J. Song, M. Patel, C.N. Rosenzweig, Y. Chan-Li, L.J. Sokoll, E.T. Fung, N.H. Choi-Miura, M. Goggins, D.W. Chan, Z. Zhang, Clin. Chem. 52 (2006) 1045.

[58] K.A. Tubbs, U.A. Kiernan, E.E. Niederkofler, D. Nedelkov, A.L. Bieber, R.W. Nelson, Anal. Chem. 78 (2006) 3271.
[59] C.A. Ray, R.R. Bowsher, W.C. Smith, V. Devanarayan, M.B. Willey, J.T. Brandt, R.A. Dean, J. Pharm. Biomed. Anal. 36 (2005) 1037.
[60] A. Manni, D. Trout, M.F. Verderame, S. Washington, D. Mauger, L. Demers, Breast Cancer Res. Treat. 68 (2001) 139.
[61] C. Heeschen, S. Dimmeler, C.W. Hamm, E. Boersma, A.M. Zeiher, M.L. Simoons, Circulation 107 (2003) 524.
[62] M. Moxness, S. Tatarewicz, D. Weeraratne, N. Murakami, D. Wullner, D. Mytych, V. Jawa, E. Koren, S.J. Swanson, Clin. Chem. 51 (2005) 1983.
[63] H. Salimi-Moosavi, D. Burens, J. Miller, J. Malella, L. Zhang, J. Wang, D. Li, J.W. Lee, B. DeSilva, G. Doellgast, 2008 AAPS National Biotechnology Conference, Toronto, Ontario, Canada, 2008.
[64] A.R. Mire-Sluis, Y.C. Barrett, Y. Devanarayan, E. Koren, H. Liu, M. Maia, T. Parish, G. Scott, G. Shankar, E. Shores, J. Immunol. Method 289 (2004) 1.
[65] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, D. Tacey, H. Hill, A. Celniker, Pharm. Res. 20 (2003) 1885.
[66] J.W.A. Findlay, W.C. Smith, J.W. Lee, et al., J. Pharm. Biomed. Anal. 21 (2000) 1249.
[67] M. Kelley, B. DeSilva, AAPS J. 9 (2007) E156.
[68] K.J. Miller, R.R. Bowsher, A. Celniker, J. Gibbons, S. Gupta, J.W. Lee, S.J. Swanson, W.C. Smith, R.S. Weiner, Pharm. Res. 18 (2001) 1373.
[69] FDA, Code of Federal Regulations, Title 42, vol. 3., 2001.
[70] FDA, Code of Federal Regulations 21CFR809.10, 2007.
[71] CLSI, Document EP7-P, 1986.
[72] CLSI, Document EP6-P, 1986.
[73] CLSI, Document EP9-A, 1995.
[74] CLSI, Document GP10-A, 1995.
[75] CLSI, Document EP5-A, 1999.
[76] CLSI, Document I/LA21-A, 2002.
[77] CLSI, Document H18-A3, 2004.
[78] CLSI, EP17-P, 2004.
[79] CLSI, Document EP14-A2, 2005.
[80] CLSI, Document H3-A6, 2007.
[81] P. Dennen, C.R. Parikh, Clin. Nephrol. 68 (2007) 269.
[82] J. Gao, L.A. Garulacan, S.M. Storm, G.J. Opiteck, Y. Dubaquie, S.A. Hefta, D.M. Dambach, A.R. Dongre, Methods 35 (2005) 291.
[83] E. O'Riordan, T.N. Orlova, V.N. Podust, P.N. Chander, S. Yanagi, M. Nakazato, R. Hu, K. Butt, V. Delaney, M.S. Goligorsky, Am. J. Transplant. 7 (2007) 930.
[84] D. Borderie, C. Roux, B. Toussaint, M. Dougados, O.G. Ekindjian, B. Cherruau, Clin. Biochem. 34 (2001) 571.
[85] R.E. Banks, Clin. Chem. 54 (2008) 16.
[86] R.E. Ferguson, D.F. Hoschstrasser, R.E. Banks, Clin. Appl. 1 (2007) 739.
[87] P. Bonini, M. Plebani, F. Ceriotti, F. Rubboli, Clin. Chem. 48 (2002) 691.
[88] J. Wang, J.W. Lee, D. Burns, J. Miller, L. Brunner, M. Peterson, B. DeSilva, 2008 AAPS National Biotechnology Conference, Toronto, Ontario, Canada, 2008.
[89] FDA, iFederal Reg 66, 2001, p. 28526.
[90] ICH, iQ2A, 1994.
[91] J.W.A. Findlay, R.F. Dillard, AAPS J. 9 (2007) E260.
[92] J.W. Lee, H. Ma, AAPS J. 9 (2007) E164.
[93] J.W. Lee, D. Figeys, J. Vasilescu, Biomarker Assay Translation from Discovery to Clinical Studies in Cancer Drug Development: Quantification of Emerging Protein Biomarkers, Elsevier, 2007.
[94] J.W. Lee, P. O'Brien, Y. Pan, R. Xu, Development and Validation of Ligand Binding Assays for Biomarkers, John Wiley and Sons, New York, 2008.
[95] R.L. Sukovaty, J.W. Lee, J. Fox, K. Toney, D.I. Papac, A. Thomas, T.A. Grover, D.S. Wells, J. Pharm. Biomed. Anal. 42 (2006) 261.
[96] H.Y. Li, G. Doellgast, M. Rose, Personal communication, 2008.
[97] R. Peracaula, G. Tabares, L. Royle, D.J. Harvey, R.A. Dwek, P.M. Rudd, R. de Llorens, Glycobiology 13 (2003) 457.
[98] V.P. Shah, K.K. Midha, J.W.A. Findlay, H.M. Hill, J.D. Hulse, I.J. McGilvary, G. McKay, K.J. Miller, R.N. Patnaik, M.L. Powell, A. Tonnelli, C.T. Viswanathan, A. Yacobi, Pharm. Res. 17 (2000) 1551.
[99] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman, S. Spector, Pharm. Res. 9 (1992) 588.
[100] J.O. Westgard, Ann. Clin. Biochem. 40 (2003) 593.
[101] J.O. Westgard, P.L. Barry, M.R. Hunt, T. Grove, Clin. Chem. 27 (1981) 493.
[102] FDA, Draft Drug-Diagnostic Co-Development Concept Paper, 2005.
[103] W.P. Carney, R. Neumann, A. Lipton, K. Leitzel, S. Ali, C.P. Price, Clin. Chem. 49 (2003) 1579.
[104] P.E. Barker, Ann. N. Y. Acad. Sci. 983 (2003) 142.
[105] J.P. Basuyau, M.P. Blanc-Vincent, J.M. Bidart, A. Daver, L. Deneux, N. Eche, G. Gory-Delabaere, M.F. Pichon, J.M. Riedinger, Br. J. Cancer 89 (Suppl. 1) (2003) S32.
[106] M.E. Hammond, P.E. Barker, S. Taube, S. Gutman, Appl. Immunohistochem. Mol. Morphol. 11 (2003) 103.
[107] B. Rafferty, P. Rigsby, M. Rose, T. Stamey, R. Gaines Das, Clin. Chem. 46 (2000) 1310.
[108] P.A. Everley, C.A. Gartner, W. Haas, A. Saghatelian, J.E. Elias, B.F. Cravatt, B.R. Zetter, S.P. Gygi, Mol. Cell. Proteomics 6 (2007) 1771.
[109] F.J. Leong, A.S. Leong, J. Postgrad. Med. 50 (2004) 62.
[110] R.M. Stephen, R.J. Gillies, Pharm. Res. 24 (2007) 1172.
[111] H. Ghattas, B.M. Darboe, D.L. Wallace, G.E. Griffin, A.M. Prentice, D.C. Macallan, Trans. R. Soc. Trop. Med. Hyg. 99 (2005) 675.
[112] K. Mitsui, H. Doi, N. Nukina, Methods Enzymol. 412 (2006) 63.