ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Pharmaceutical and Biomedical Analysis 39 (2005) 364–375

JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS

www.elsevier.com/locate/jpba

Review

Validation of immunoassays used to assess immunogenicity to therapeutic monoclonal antibodies

Dong Geng [1], Gopi Shankar [1], Allen Schantz, Manoj Rajadhyaksha, Hugh Davis, Carrie Wagner *

Department of Clinical Pharmacology & Experimental Medicine, Centocor Inc., 145 King of Prussia Rd., Radnor, PA 19087, USA

Received 23 November 2004; received in revised form 29 March 2005; accepted 1 April 2005
Available online 15 June 2005

Abstract

Immunogenicity has always been an important consideration in the evaluation of pharmaceutical protein biologics. In this article, method validation parameters relevant to enzyme immunoassays are described for assays applied to the analysis of anti-drug antibodies, with special considerations for immunogenicity to therapeutic monoclonal antibodies. Common strategies for experimental investigation of various validation parameters are proposed. In addition, a novel, yet simple, approach is proposed to categorize the validation effort into two mutually interdependent phases, based on the characterization of validation parameters as "system descriptive" or "system controlled". System descriptive parameters are those that must be characterized but need not have pre-specified acceptance criteria for assay validation. In contrast, system-controlled parameters should be understood early in assay development, and optimized and confirmed using a priori acceptance criteria in validation to assure sufficient control over them during routine bioanalysis. This approach not only streamlines the validation process but also eliminates unnecessary redundancies. This validation method can be achieved with proper scientific rigor and remain within the realm of GLP compliance. The authors hope that other research groups would engage in discussions on validation of anti-drug antibody assays in order to establish a consistent approach across the industry and academia.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Immunogenicity; Enzyme immunoassay; Anti-drug antibody; Assay validation; Method validation

Contents

1. Introduction 365
2. Anti-drug antibody (ADA) immunoassays 365
3. Method validation 366
   3.1. Specificity 367
   3.2. Selectivity 367
   3.3. Dilutional linearity 368
   3.4. Accuracy 368
   3.5. Cutoff value 369
   3.6. Sensitivity 369
   3.7. Precision 369
   3.8. Robustness 370
   3.9. Ruggedness 370

* Corresponding author. Tel.: +1 610 651 4524.
E-mail address: Cwagner@cntus.jnj.com (C. Wagner).
[1] D.G. and G.S. contributed equally to this article.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.04.045

3.10. Stability 370
3.11. System suitability controls 371
4. Streamlining the method validation process 371
4.1. Method development and optimization 372
4.2. System-descriptive validation 374
4.3. System-controlled validation 374
4.4. In-study monitoring 374
5. Conclusion 375
References 375

1. Introduction

Monoclonal antibodies have demonstrated utility as biotherapeutic interventions for immune-mediated inflammatory diseases and cancer. The protein sequence of such drugs can be non-human, chimeric, humanized, or fully human. Regardless of their composition, a vital safety concern for regulatory agencies, drug manufacturers, clinicians, and patients, is the potential for anti-drug immune responses (immunogenicity) elicited by drug treatment [1]. Anti-drug antibodies (ADA) may cause adverse events including infusion reactions and hypersensitivity. ADA binding to the drug can potentially cause drug neutralization, altered biodistribution, or enhanced drug clearance rates, which can result in reduced efficacy of the treatment. Thus, immunogenicity testing is required to demonstrate the clinical safety of new biotherapeutic products. The United States Food and Drug Administration (FDA) requires that anti-drug antibody responses be detected, characterized appropriately, and correlated with any pharmacological and/or toxicological observations [2]. Further, with the near-future entry of "follow-on" or "generic" biologic drugs, immunogenicity will become a critical feature in demonstrating product comparability. Immunoassays contribute significantly to our understanding of immunogenicity. Hence, immunoassays should be properly developed and sufficiently validated before testing clinical samples.

At present, there is no perfect assay for determining immunogenicity. However, a number of assay methods are available including ELISA, radio-immunoprecipitation assay, electrochemiluminescence, surface plasmon resonance, and bioassays, each with relative merits and weaknesses that have been discussed in recent publications [3–5]; however, the enzyme immunoassay remains to be the most widely used. Irrespective of the assay format, once a test method is developed and optimized, validation should be performed to assure that the results are meaningful. Despite the fact that the FDA has increasing regulatory expectations on ADA testing, to date there are no specific guidelines for immunogenicity assay validation.

The purpose of this paper is to provide our perspective on the validation of ELISAs used to investigate clinical or non-clinical ADA immune responses to therapeutic monoclonal antibodies with respect to the following questions: What is assay validation and why is it necessary? What are the validation parameters necessary for ADA testing and how can they be approached experimentally? And, can the method validation process be streamlined to improve operational efficiency? Many of these points would be the same for ADA testing for any biologic product, however, some unique considerations apply to immunogenicity testing for monoclonal antibody drugs.

2. Anti-drug antibody (ADA) immunoassays

Despite the emergence of novel technologies such as Biacore® and BioVeris™ (previously called IGEN®), microtiter plate-based ELISAs are still the most widely used format for testing anti-drug antibodies due to their high-throughput efficiency, simplicity, and high sensitivity. Although the general principles would apply to these other methodologies, the specifics of assay validation discussed in this paper focus on ADA ELISAs. Either direct or indirect sandwich-format ELISAs may work well for non-clinical ADA detection (provided positive control antibody is available from the target species) and the aspects of assay validation discussed in this paper apply generally to any type of ADA immunoassay. However, because monoclonal antibody-based drug products are often chimeric, humanized, or fully human, it usually becomes impossible to employ a sandwich ELISA using a secondary anti-human antibody detection reagent to test clinical samples. Hence, ADA ELISAs in our laboratory are most often designed in a bridge-format, which provides high selectivity and pan-species ADA detection capability, making it feasible to implement a single assay format for both non-clinical and clinical studies. Two formats of 'bridge' ELISAs are common (see Fig. 1): in which the drug is either coated directly onto the surface of polystyrene microtiter plate wells, or the biotinylated drug is bound indirectly to the plate surface using streptavidin. In either case, when a test sample is added to the plate, ADA in the sample binds to both the solid phase and solution phase drug. Solution phase drug is labeled so as to permit detection of binding. Common examples of detection systems include enzyme-conjugated drug (Fig. 1, format-1) and a chromogenic substrate, or a combination of biotinylated drug and streptavidin-conjugated enzyme (Fig. 1, format-2)

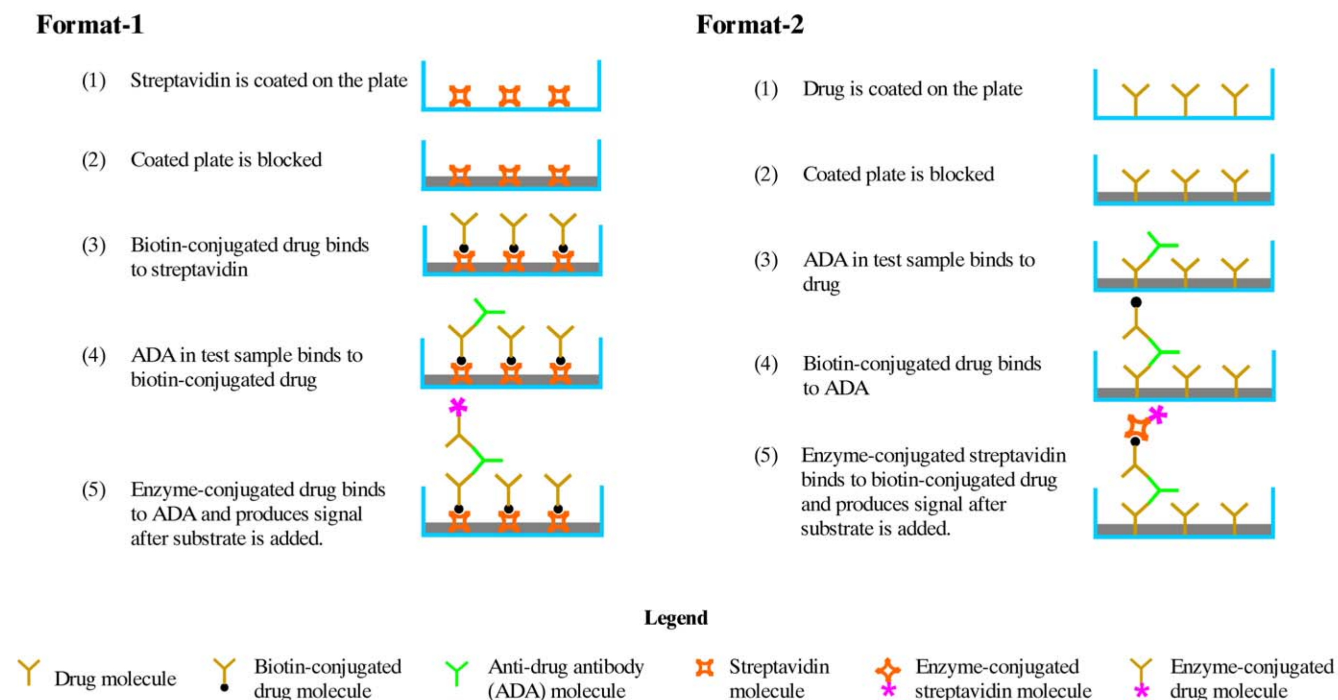

Fig. 1. Two formats of bridge ELISAs.

followed by a chromogenic substrate. The latter format has the benefit of signal enhancement; the former may be used to reduce non-specific or low-affinity antibody binding that may cause substantial interference in certain subject populations (e.g., rheumatoid factors in rheumatoid arthritis patients). The optical density (OD) of the resultant colored product is recorded using a spectrophotometer. Within a unique linear range for any assay, the magnitude of the OD tends to be directly proportional to the ADA level in the test sample. However, this relationship is unique for each sample, and is not accurately quantifiable. This is because each sample represents a unique biologic response individually selected from the patient's repertoire and probably further modified through affinity maturation and epitope spreading. As a result, the assay positive control is never identical to the ADA analyte.

3. Method validation

For any analytical method, its performance and reliability must be demonstrated to ensure a high level of confidence on the test results. Without an understanding of the systematic and random variations in an assay it is not possible to set appropriate criteria that allow accurate differentiation of a truly positive result from a truly negative result. *Validation is the process of demonstrating, through the use of specific laboratory investigations, that the performance characteristics of an analytical method are suitable for its intended analytical use* [6,7]. In the case of ADA methods, suitability includes proof that the assay consistently, reliably, and reproducibly detects drug-specific antibodies in a complex biological matrix such as serum or plasma. Notwithstanding the regulatory obligation to validate assays, it is important to understand whether a newly developed (or "standardized") assay will continue to perform as expected when applied to a large number of heterogeneous samples under actual analytical conditions over time. Without formal investigation, can we assume that the results will be the same when the assay is performed by different scientists, and on different days? How do we know that the results are correct? Would a test sample that is tested after a period of refrigeration or frozen storage produce the same result as a fresh sample? Without a demonstrated understanding of the above issues, how can we rely on the results produced by an assay? And most critically, are we using reliable data to make decisions regarding clinical trials, and providing such information to the FDA? Thus, validation is not just a regulatory burden, but also a critical scientific and business obligation for a drug manufacturer. Therefore, immunogenicity assays for human clinical trials and non-clinical studies have to be carefully selected, developed, validated, and conducted.

The initial step in the development of a validation strategy is to define the analytical objectives of the assay, i.e., what type of test samples will be analyzed, and what is the expected readout of the assay? Then, keeping in mind that *validation refers to the determination and maintenance of assay reliability*, the following questions must be addressed: (1) What interferences might be expected from the sample matrix? (2) What level of accuracy and sensitivity are required? (3) What level of imprecision is acceptable? (4) What are the robustness and ruggedness needs for the routine performance of this assay? Each of the analytical method validation param-

eters specified by regulatory documents such as specificity, selectivity, accuracy, sensitivity, precision, robustness, stability, and ruggedness should be pondered carefully; known and potential variables for each parameter must be identified and their impact investigated. Additional parameters unique to the analytical method should also be considered. The validity of a cutoff value and the minimum required dilution are additional parameters that apply to ADA ELISAs and must also be demonstrated. Considerations for each of these analytical parameters are further discussed below.

It should be noted at this point that specific recommendations on sample sizes (number of test subjects, number of test samples, etc.) have intentionally been omitted in this article because factors such as reagent and sample availability, sufficient statistical power, and the nuances of an assay can affect sample size determination. Therefore, no single statistical recommendation can fulfill the needs of all method validations, although, when resources permit, the general idea that "more is better" is a good general principle to follow. It is strongly recommended that a biostatistician review each method validation plan before execution.

3.1. Specificity

Specificity is the property of an analytical method to unequivocally detect the target analyte [8], and with minimal or no cross-reactivity to unrelated analytes. It is important to validate specificity during validation, during revalidation triggered by a reagent change, or when a new method is compared with a standard method.

In the case of ADA assays, the target analyte is generally polyclonal, comprised antibodies of various isotype classes, specificities, and affinities; hence, no single positive control can accurately represent the target analyte. It follows then, that the selection of representative analyte(s) plays an important role in the validation process. For clinical assays, human ADA is ultimately the analyte of interest. However, because assay development and validation typically precede clinical trials, ADA positive controls from humans are rarely available during initial assay development and validation. Even when a human positive sample becomes available, one may not have sufficient quantities, or patient consent, to use for this purpose. Furthermore, ADAs tend to be unique to each drug program, leaving the investigator with the challenge of obtaining representative positive samples for each assay development and validation. The situation is further complicated by the fact that ADAs are polyclonal, and so while an ADA assay may function across isotypes, species, etc., assay performance may vary quite noticeably when comparing antibody responses of disparate properties, and when making comparisons across species. Thus, the so-called analyte in an ADA assay is not a defined molecule, but rather a group of different molecules sharing a capacity for drug binding. When establishing specificity, accuracy, and sensitivity, it may be informative to employ several control analytes from different sources or individuals that may reasonably represent the test population. For example, it has been observed that the ADA response generated during a toxicology study in which rabbits are repeatedly treated with a human therapeutic antibody will generate rabbit IgG of high affinity to the human IgG constant region. An improved control analyte for such an analysis might be an affinity-purified antibody from a hyper-immunized rabbit. In contrast, human subjects acutely treated with the same human therapeutic antibody predominantly produce low-affinity IgM or IgG against an epitope of the drug's complementarity-determining region(s) [9]. Clearly, an assay validation based solely upon detection of the high affinity human Fc-specific rabbit analyte might not be representative or adequate to support the validation and subsequent ADA analysis for human clinical trials.

Specificity of analyte binding can be demonstrated by immunodepletion analysis. Positive control(s) or positive sample(s) can be tested in the assay after pre-incubation with the drug (which should inhibit signal in the assay) or with structurally similar molecules such as an unrelated antibody drug molecule or purified human/animal immunoglobulin (which ideally might not influence drug-specific assay signal significantly).

3.2. Selectivity

Specificity in the presence of components expected to be present in the sample matrix (interfering substances), referred to as 'selectivity', can be challenging to validate. In an ADA assay, the specificity remains the same but selectivity can vary between test samples due to the heterogenous and polymorphic nature of samples from higher mammals and humans. Hence, a reasonably sized population of representative samples must be analyzed during validation to assure a proper estimate of assay selectivity. Albumin and gamma globulins, normally the major components of plasma or serum, can be a source of interference in an assay. Other components that may cause interference are specific proteins present in particular disease populations (such as rheumatoid factor) and substances that bind competitively with the analyte (including the drug itself or the drug target). Furthermore, lipemic and hemolyzed sera may also interfere with ADA detection in certain assays. Thus, selectivity is a critical parameter that determines the reliability of an ADA assay, and must be validated. Experimental approaches to common selectivity variables are discussed below; nevertheless, it may be necessary to use an imperfect assay with selectivity problems despite sufficient attempts at method development and optimization. In such a case, these problems should be described in the validation report.

To evaluate selectivity, immunoglobulins or another potential interfering substances may be spiked into positive and negative samples. Positive samples, or mock positive samples (positive control antibody added to a chosen naïve matrix), and negative samples should be prepared with a biologically high but relevant concentration of each potential interfering substance. The samples with and without the

interfering substance are then tested and the mean signal recovered (with added interfering substances) versus the target (no added interfering substance) is calculated. A typical acceptance criterion for recovery is 80–125%, in which case it can be concluded that there is no substantial interference. Otherwise, it should be inferred that the substance significantly interferes with analyte detection.

The effect of the sample matrix (typically serum or plasma) should be evaluated if more than one sample matrix could potentially be tested using the assay. Matching samples (such as serum and plasma) from the same donor should be compared. Mock positive samples may be prepared by spiking the positive control into naïve serum and plasma samples. Naïve donor serum and plasma can also be used as the negative samples. It is likely that positive samples will need to be represented by mock positive samples. The acceptance criterion is that the matrix should not alter the assay outcome, i.e., this validation test passes if positive and negative samples give comparable results when the respective samples in both matrices are compared.

Selectivity investigations should also include a comparison of specificity of the analyte within normal and disease-state sera in view of the possibility that interfering substances may be prevalent in some populations or disease states. It is generally recommended that clinical tests be validated using sera from the target population, whether patients or healthy volunteers. Likewise, non-clinical studies may examine healthy animals or include disease models. If the assay is expected to test specific populations (and if specimens are readily available), then healthy and disease model/patient samples can easily be compared. However, if the specific target indications or disease models have not been conclusively identified at the time of method development and validation, it is recommended that sera from a number of disease states be investigated in addition to normal donor sera. Mock positive samples should be prepared by adding analyte into multiple individual serum/plasma samples from naïve patients from the intended population, or a pooled normal human serum if target disease sera are unavailable. Recovery of the positive signal is evaluated by comparing the mean results of the patient and reference populations spiked with the same amount of analyte. As with the comparison of interfering substances described earlier, a typical acceptance criterion for mean recovery is 80–125% (i.e., mean OD recovered from the patient population compared to the mean OD recovered from the healthy population). If recovery is higher or lower than this range, then the validation report should indicate that there is some degree of interference due to target disease state serum/plasma components.

A unique feature of ADA assay selectivity is that the drug itself can act as an interfering substance. This issue can arise depending on the length of time since the last dose of drug was administered (the "wash-out" period) and its pharmacokinetic profile. The drug may already be bound to ADA, or can compete with the capture of ADA in the assay. The observed effect of this interference is an apparent reduction in the assay signal that could yield a false negative result. Preparing a set concentration of positive controls or samples with varying amounts of experimentally added drug can mimic the presence of pharmaceutically administered drug in a sample. Testing such samples can provide information about the degree of antigen interference that may exist. At present, significant efforts are underway to discover a means of effectively uncoupling or measuring the drug-ADA complex to enable detection, or to detect ADA despite the presence of bound drug. However, this is complicated by an increased elimination in vivo of high molecular weight complexes, thus precluding an accurate quantitation of the induced immune response.

3.3. Dilutional linearity

The linearity of an analytical procedure is its ability (within a given range) to obtain test results that are directly proportional to the concentration of analyte in the sample [8]. Estimations of ADA concentration should be limited to titer estimation due to the inherent dissimilarity between actual samples and the "reference standard". When ADA is estimated as a titer, dilutional linearity is nice to evaluate but not critical. One should ensure that the positive control dilutes linearly within a reasonable range, and that the positive control used to assure system suitability (the day-to-day plate performance or consistency positive control) falls within a linear region of the dilution curve rather than on a plateau or a region of the curve that might include a prozone (hook) effect. If, however, the determination of ADA-positive samples is based on interpolation from a reference standard curve, then it is very important to demonstrate linearity. Linearity is generally expressed as the regression coefficient, the slope, and the *y*-intercept, of the curve within a specified range ("limits").

3.4. Accuracy

Accuracy is the agreement between an experimentally measured value and an accepted reference (standard) or a theoretical "true" value. In other words, it is a measure of the "trueness" of a method [8] and describes systematic error (mean bias) of a test. Accuracy of a new method is estimated by comparing it with the results of another test method of known accuracy and precision (i.e., a "gold standard" method), or through reference material of known or generally accepted composition. ADA assays are unique by nature because there is generally no comparative or standard method available to determine accuracy. ADA immunoassays are quasi-quantitative due to limitations imposed by the selection of generic reference standards intended to represent a variety of potential polyclonal responses. Therefore, the approach here is generally an assessment of spike-recovery, i.e., repeated measurements of the same spiked sample under specified conditions. Therefore, the accuracy of ADA assays is inferred from selectivity.

3.5. Cutoff value

Since ADA immunoassays are quasi-quantitative due to the limitations of the reference standard, sample positivity is usually determined using a threshold limit of non-specific background (an OD value), also referred to as a "cutoff" or "cutpoint". An assay cutoff is determined by analyzing samples from naïve individuals, preferably those afflicted with a target disease, or animals of a disease model. Alternatively, healthy donor samples could be used when it is not feasible to obtain specimens from a specific disease group or disease model. Naïve samples producing unusually high OD results are excluded from data analysis only if the reactivity is shown to be drug-specific. These data are used to determine a cutoff value that is calculated to yield a false positive rate of at least 5%. One might ask why it would be desirable for an ADA assay to have a 5% incidence of false positive results. In fact, the final incidence of false positives is much less because ADA evaluations typically also include a competitive binding step in which *potentially* positive samples must demonstrate specific binding to the drug before a sample is determined to be *truly* positive or negative. This type of multi-step approach allows one to maintain the lowest practical cutoff, thus permitting the detection of lower affinity/concentration antibodies. Ultimately, this facilitates a broader comparison of ADAs relative to safety and efficacy.

In our experience, it is not desirable to have an ELISA assay cutoff less than 0.2 OD units because precision deteriorates as absorbance values fall below that level. If the results fit a normal distribution curve (also called Gaussian distribution or "bell curve"), a common approach to selecting the cutoff is via the mean OD and standard deviation of all naïve sample test results and defining the cutoff as the mean plus 1.645 times the standard deviation. Statistically, this should provide a false positive result of 5% in a naïve population. If the results do not fit a normal distribution, or the cutoff is much less than 0.2 OD units, this type of calculation is invalid and an alternative estimation of the cutoff must be performed, preferably in consultation with a biostatistician.

3.6. Sensitivity

Sensitivity is traditionally defined as the amount of analyte required to produce a significant change in signal versus that obtained in the absence of analyte. There are two measures of sensitivity: detection limit (often referred to as the lower limit of detection, LLOD) and quantitation limit (often referred to as the lower limit of quantitation, LLOQ). Detection limit is useful for 'limit' tests, which measure "positive versus negative" results, i.e., assays that merely substantiate that the amount of analyte is above or below a certain level [10]. Quantitation limit, on the other hand, is useful for quantitative assays. As indicated earlier, ADA assays are quasi-quantitative, but regulatory agencies prefer the validation to report sensitivity in concentration units, not titers. Thus, one is required to use monoclonal antibodies or affinity-purified specific antisera for the determination of the quantitation limit. This is determined via limiting dilution of a reference analyte whereby the quantitation limit is the greatest concentration of the analyte able to reproducibly generate a signal in excess of the assay cutoff. Due to limitations inherent in the use of a representative analyte, it is preferable to use at least two positive control analytes, preferably having different characteristics with regard to affinity, isotype, or clonality for the determination of ADA assay sensitivity. For ADA ELISAs, sensitivity varies proportionally with the cutoff, and so the latter must be validated carefully. Again, it is important to note that the sensitivity of ADA assays will also vary based on the positive controls utilized, and it is impossible to determine the 'true' quantitation limit because no positive control or sample can completely mimic all possible polyclonal ADA responses.

3.7. Precision

Precision is a measure of the degree of reproducibility of the analytical method under normal operating circumstances. In other words, it is the degree of agreement among individual test results when a procedure is applied repeatedly to a homogenous sample [10]. Precision is measured mathematically by the random error, or imprecision, between replicate experiments. It comprises repeatability and intermediate precision [8]. Generally, positive control(s), negative control(s), and a diluent sample (representing infinite dilution of sample during titration) are run on at least three separate days by at least two or three separate analysts. In each run, an analyst prepares at least three plates, with at least three replicates of each sample. The test samples should originate from identical aliquots prepared in advance by one individual. This single comprehensive experiment can yield all of the data required to estimate the various types of assay precision. The assay precision of a quasi-quantitative ADA assay can vary depending on the concentration of positive control one chooses to use. Using a positive control at a high and a low concentration is recommended to demonstrate the precision within the assay range. Negative controls are merely required to generate signals less than the assay cutoff 95% of the time, however, positive controls are evaluated more rigorously.

Repeatability, or intra-assay precision, is the variation between replicate samples on the same plate (well-to-well variation within plate). For each sample, the replicate assay results of each analyst are calculated separately (mean, S.D., %CV) on each individual day of testing, and then repeatability is expressed as the overall average S.D. and %CV. Typically, repeatability for OD results above 0.2 is expected to be $\leq$20%.

Intermediate precision comprises inter-assay precision (variation between separate plates), inter-day precision (day-to-day variation for an analyst), and inter-analyst precision (analyst-to-analyst variation). Intermediate precision is calculated using the mean OD results of the sample replicates.

For inter-assay precision, the mean ODs of replicate samples from each of an analyst's plates are used to calculate an inter-plate mean OD result, the S.D. and %CV. The average inter-plate S.D. and %CV of the various analysts on the various days is calculated and this represents the overall inter-assay precision of the method. For inter-day precision, the inter-plate mean OD results of each analyst over the days are used to calculate the inter-day S.D. and %CV per analyst. Then these values from each analyst are combined to derive the average inter-day precision. Finally, for inter-analyst precision, the inter-plate mean OD results of all analysts are used to calculate the inter-analyst S.D. and %CV each day. Then values from each day can be combined to derive the average inter-day precision. Typically, intermediate precision ≤30% is considered acceptable.

3.8. Robustness

Robustness is an indication of the reliability of an assay, assessed by the capacity of the assay to remain unaffected by small, but deliberate, variations in method parameters [8]. The focus of robustness is to analyze assay consistency under relevant, real life changes in standard laboratory situations. Parameters to test during robustness testing should be based on knowledge of the assay and associated risks. One should assess which conditions are likely to vary in an assay, in a particular laboratory, and design appropriate tests to examine the parameters that are deemed critical. These may include changes in microtiter plate lots, incubation times, temperature, number of washes, reagent lot and concentrations, or instrumentation. Critical changes would be detected because they would lead to failed control values. To some extent, such changes may occur during the previously described validation steps, and as such are understood prior to robustness experimentation; however, one should make deliberate, controlled, and finite changes in order to describe these parameters more fully.

3.9. Ruggedness

The term ruggedness is absent from ICH method validation guidances [8,11], while the validation monograph in the United States Pharmacopeia [10] describes it as including inter-laboratory precision as well as inter-analyst precision. The ICH documents deal with inter-laboratory variation simply as another element of precision, termed 'reproducibility'. Ruggedness is best defined as a combination of precision and robustness between laboratories. It can be imagined as robustness testing wherein the test parameter is an entire laboratory unit with its own characteristic imprecision. Thus, ruggedness is useful for assessing the "transferability" of an assay, i.e., the validity of testing samples in two or more laboratories. When a single laboratory performs the assay, however, ruggedness becomes a non-issue. Investigations of ruggedness are complex, involving many other parameters for a valid inter-laboratory comparison.

3.10. Stability

Stability studies evaluate assay performance under the intended storage conditions. Ideally, the conditions should mimic the expected sample and reagent handling conditions, storage temperature(s), and varying lengths of storage time. While regulatory guidances on method validation [8–10] do not address this type of stability testing, it is an important aspect of method validation. The stabilities characterized should include the controls, test samples, the coated assay plate (if applicable), and other critical reagents. Some manufacturers provide expiration dates for purchased reagents, and these may be adopted in lieu of in-house stability validation. Comparing freshly prepared controls and samples with those that have been stored under conditions to be used for non-clinical or clinical studies can determine the stability of samples and controls. Common stability parameters include: freeze–thaw cycles, 4 °C storage, −70 °C storage, and the stability of diluted samples and critical reagents. Real clinical or non-clinical specimens might not exist during validation, or may be otherwise unavailable for validation testing. In this case, spiking analyte-negative matrix with the positive control can generate a mock positive sample. Multiple positive samples at different concentrations are preferable, if available. The laboratory should determine the mean recovery of stored versus fresh positive controls and any positive samples. The acceptance criteria for the positive control(s) and sample(s) are that they must generate positive results, with a mean recovery of 80–125% after storage. We have never observed a negative sample becoming positive for ADA during storage, and recommend that a single negative control sample is sufficient for the stability study. The acceptance criterion for this control sample is that it reproducibly generates a negative result.

In a freeze–thaw study, at least three aliquots of each sample are prepared and stored frozen (−70 °C, or other relevant temperature) for at least 24 h, then thawed. The thawing procedure may be assay specific, at 4 °C, room temperature, or 37 °C, but should be clearly defined. When completely thawed, the samples should be refrozen to the original temperature for at least 12 h before repeating the thawing process, and at least one replicate aliquot of each sample should go through at least three freeze–thaw cycles. The freeze–thaw cycles should be staggered so that all samples are tested together against samples freshly prepared in an identical manner to their frozen counterparts.

In a study of sample stability at 4 °C, the samples or controls should be thawed and then stored at 4 °C for a relevant period of time, such as 28 days. In this example, a minimum of two time points (days 0 and 28) are needed, but intermediate time points such as days 7 and 14 are also recommended. Multiple aliquots of each sample and the controls should be prepared and stored at −70 °C, the exact number of aliquots depending on the number of time points one plans to evaluate in the study. The first set of aliquots (positive and negative controls, and at least one sample/mock sample) are thawed

and held at 4 °C from day 1: these will become the 28-day stability samples. Similarly, sets of aliquots are thawed on days 7, 14 and 21 and also held at 4 °C, representing the 21-, 14- and 7-day stability samples. The last set of aliquots is thawed on day-28, and serve as freshly thawed (“day-0”) comparator samples. All samples are then tested together, and recoveries of the spiked positive control in the various stability test samples are calculated relative to the comparator samples.

Antibody samples and controls are generally assumed to be stable as long as they are stored frozen at −70 °C. However, for a validated assay, this must also be demonstrated. Experimentally, this is demonstrated in the same manner as the 4 °C stability study described above, except that aliquots are instead stored at −70 °C and individual aliquots analyzed periodically for extended intervals up to several years. The same process could be applied to samples stored at −20 °C. Note that assay “drift” may be difficult to distinguish from loss of stability of the control samples at −70 °C.

In some analytical situations, it is more convenient to dilute samples one day prior to analysis and store overnight at 4 °C. However, it is important to demonstrate the stability of diluted samples if the laboratory plans to store them for extended periods of time preceding analysis. To test the stability of diluted samples, positive and negative samples or mock samples are prepared at the final dilution used for analysis, and stored under the conditions that will be used to store actual test samples (for example, 4 °C overnight). The stored samples are analyzed side-by-side with freshly diluted samples and the results compared.

For operational efficiency, it may be desirable to store batches of coated and blocked assay plates for later use. Plate stability is assessed using one batch of coated plates to test positive and negative samples (assuming frozen samples to be stable, these would be prepared and frozen ahead of time) at consecutive time points during plate storage (e.g., through 7 days at 4 °C). The use of freshly prepared sample at each time point is recommended only if sample stability is an issue. At a minimum, the first and last (e.g., days 0 and 7) time points must be evaluated, but analyzing more time points in between would allow one to observe trends.

The percent signal (OD) recovered from positive samples is calculated each day as a percentage of the result derived from the freshly coated (day 0) assay plate. If multiple positive samples are tested, the mean percent recovery is also calculated. A typical acceptance criterion for the positive sample is 80–125% recovery of the signal from the freshly coated plate.

Lastly, the stability of other critical assay reagents should be considered. For reagents stored according to the manufacturer’s recommendations, stability testing is not necessary. Key reagents synthesized by the testing lab, such as antibody conjugates, require stability testing. This can be performed in an analogous manner to that described above for sample stability. If reagent stability is unknown or found to be a problem, it is recommended that critical reagents be frozen at −70 °C in small, preferably single-use, aliquots that can be discarded a short time after thawing.

3.11. System suitability controls

Because ADA assays often lack a standard curve, positive and negative control ranges are validated and monitored during the study to assure the continued reliability of the assay under routine performance. During validation, the assay controls should be tested multiple times by multiple operators over a period of time. The overall results should be used to calculate a mean ± 2S.D. or mean ± 3S.D. range for each of the positive and negative controls (QC samples), comprising the acceptance range for routine assay runs.

4. Streamlining the method validation process

In order to proceed in a logical fashion, ADA assay performance criteria must be defined early in the course of assay development [4]. In the final stages of assay development, the method becomes locked in because it appears to meet those performance criteria. At that point, it is typical to conduct informal pre-validation studies, prior to commencement of formal validation, in order to more fully verify assay performance and obtain unprejudiced (or “more accurate”) pre-specified acceptance criteria for validation parameters. Generally, subsequent formal validation studies, performed for the purpose of meeting good laboratory practice (GLP) regulations, generate data that are just “more of the same”. This “validation after validation” redundancy seems to waste a significant amount of resources. If a laboratory consistently operates according to GLPs, it would seem reasonable to use pre-validation data, or even late-stage development results to help satisfy certain validation criteria, saving both time and effort. Thus, a focused and more efficient approach is warranted, and our proposal is described below.

Fig. 2 depicts a generic, high-level outlook of the process from method development to bioanalytical testing of clinical and non-clinical samples. To facilitate improved efficiency in the method validation process, it is proposed that validation parameters be categorized into ‘system-descriptive’ versus ‘system-controlled’. The word ‘system’ is applied here to emphasize that a valid assay comprises much more than the methodological steps alone; in fact, the laboratory equipment, electronics, assay performance controls, the samples to be analyzed, and other variables, constitute an integrated system (the assay).

System-descriptive parameters are assay features that “are what they are”, and once established, are not expected to significantly vary for a well-optimized method. Hence, they must be characterized but need not have pre-specified acceptance criteria for assay validation. In contrast, changes in *system-controlled parameters* can affect system descriptive parameters, potentially producing unreliable assay results. Even though an SOP is followed, assay results (thereby

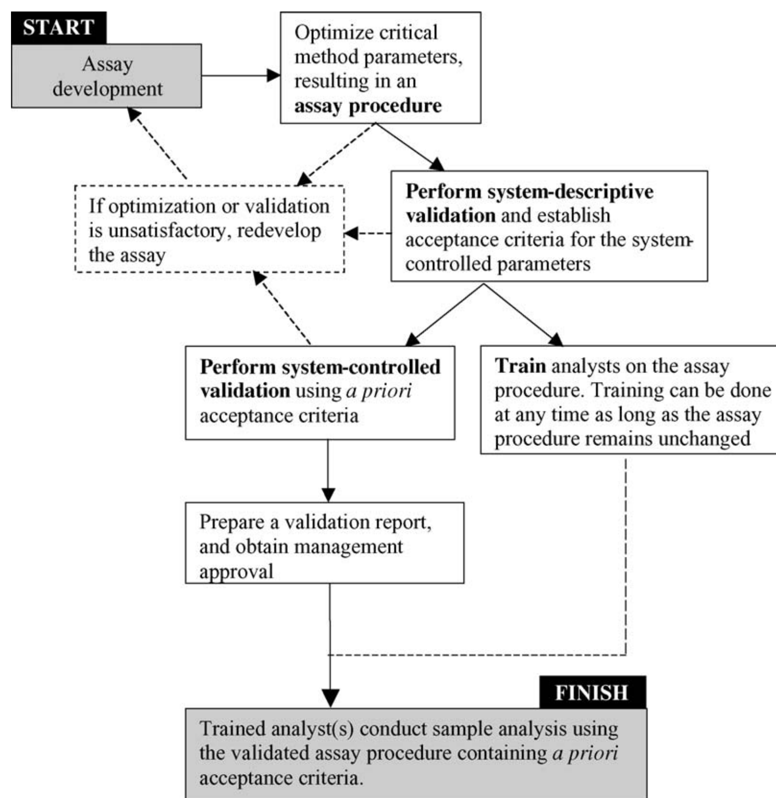

Fig. 2. The typical process from method development to bioanalytical testing of clinical and non-clinical samples.

assay reliability) can depend on the particular laboratory, the analyst(s), equipment, reagent lots, incubation times and temperatures, etc. System-controlled parameters should be understood early, described in the SOP, and confirmed using a priori acceptance criteria in validation to assure sufficient control over them during routine bioanalysis.

For ADA assays, we propose that specificity/selectivity, minimum required dilution, dilutional linearity, and accuracy, cutoff value, sensitivity (which depends on cutoff), and stability are system-descriptive validation parameters. On the other hand, precision, robustness and ruggedness should be considered system-controlled. Based on this, the lifespan of an ADA immunoassay could be categorized into four phases as shown in Fig. 3: method development and optimization, system-descriptive validation, system-controlled validation, and in-study monitoring. Table 1 indicates the parameters and the stages in which they might be validated.

4.1. Method development and optimization

Requisite pre-validation method development must be performed to culminate in an assay what is expected, with a high degree of confidence, to be “validate-able”. In other words, an optimal method should be chosen after comparing data between multiple methods with respect to specificity, minimum required dilution, non-specific background, signal-

Table 1
Validation parameters and the stages of validation

	GLP method development and optimization	System-descriptive validation	System-controlled validation
Specificity	X		
Selectivity		X	
Dilutional linearity	X		
Accuracy	X	X[a]	
Cutoff value		X	
Sensitivity		X	
Precision			X
Robustness			X
Ruggedness			X
Stability	X[b]	X	
System suitability controls		X	

[a] Confirmation only.
[b] If applicable, using the optimized assay.

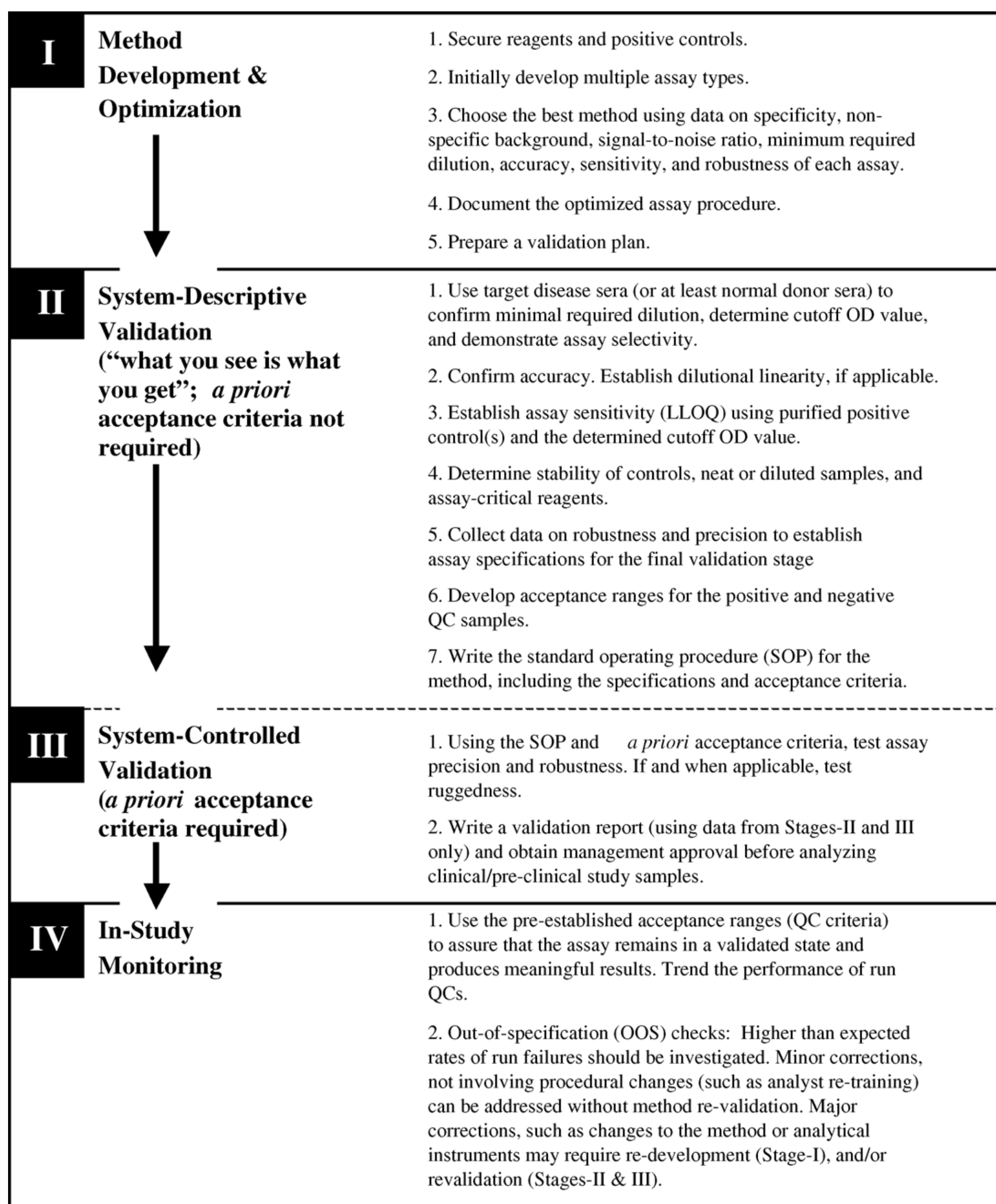

Fig. 3. The lifespan of an ADA immunoassay.

to-noise ratio, accuracy and sensitivity, and consideration of other factors including technical difficulty and throughput. Fig. 3 lists only a few important steps in method development; for guidance on assay development strategies refer to Mire-Sluis et al. [4]. Method development efforts must result in an assay protocol that can be executed during the initial validation phase.

It is a common industry practice to execute method development and optimization to determine acceptance criteria for the various validation parameters. Method development work should provide preliminary estimates of specificity, accuracy, minimum required dilution, and sensitivity of the method. In our opinion, system-descriptive validation data can be collected from the GLP-compliant parts of the method development work. For example, specificity is inherent to the analyte(s) chosen to build an assay; therefore, it need not be re-demonstrated during validation. Table 2 shows an example of specificity data for an ADA assay of a new biological product in our laboratory during method development and subsequently in a formal validation. During method development studies, one operator performed the experiment twice over two days. In the validation stage, two operators performed the experiment twice each over two days. Identical results were found, as expected, between method development and validation stages. Validating something that was already established wasted resources and two additional analyst-days. Similarly,

Table 2
A comparison of method development vs. validation sample data of assay specificity

	%Inhibition (specificity)					
	mAb#1		mAb#2		mAb#3	
	Method development	Validation	Method development	Validation	Method development	Validation
Protein-XA	99.2	99.7	99.0	99.5	96.1	96.7
Protein-XB	98.9	97.5	57.9	63.1	56.6	53.6
Peptide-XC	100.8	100.5	22.8	25.0	5.0	3.3
protein-XD	99.1	97.5	98.8	98.1	98.3	95.9
Control protein-Y	0.2	−2.4	−3.4	2.7	−12.5	−4.9
Control protein-Z	4.0	9.8	2.7	4.2	−0.2	2.3

there may be other parameters that need not be validated after GLP method development.

4.2. System-descriptive validation

Once an optimized assay is available, the system-descriptive validation phase can begin. In practice, it is an extension of the method optimization phase, as long as the established assay procedure remains unchanged, and as mentioned earlier, validation data could be obtained from prior development work. As listed in Fig. 3, several critical validation parameters can be described in the system-descriptive validation stage. While some target disease sera may have been tested in the method development stage, it is in the system-descriptive validation stage that a large number of samples should be analyzed using the optimized assay, resulting in the best approximation of a valid cutoff value. If selectivity was not demonstrated during GLP-compliant parts of the method development, it must be characterized in this stage. For most ADA assays, accuracy can be inferred from spike-recovery selectivity results; separate experimentation to demonstrate accuracy is redundant. Sensitivity should be determined using the validated cutoff value. Using all the data obtained, acceptance criteria for the system-controlled final validation parameters (precision and robustness), and the system suitability controls (the positive and negative QC samples run on each plate during analytical testing) should be initiated at this stage. Sometimes stability is tested using the newly developed assay (rather than an alternate, established, method), in which case it is recommended that critical system-descriptive validation parameters be understood prior to commencing stability studies. This approach will avoid repeating stability studies should the validation be terminated due to unsatisfactory results. System-descriptive validation work should culminate in the development of an SOP that includes assay acceptance criteria, raw data flow, storage of data, data manipulations, analysis and/or interpretations, etc. Once properly understood, these validation parameters need not be re-tested in a formal validation. For example, in our laboratory, polyclonal (pAb) and monoclonal (mAb) positive control antibodies were tested multiple times during pre-validation and subsequently in a formal validation. During pre-validation studies, one operator tested the positive controls across 38 runs over 30 days. In the validation stage, two operators tested the positive controls across 37 runs over 29 days. As shown in Fig. 4, both pAb and mAb controls gave identical results between pre-validation and validation studies, clearly showing that time, effort, and money were wasted. Thus, when thorough pre-validation work is performed, repeating the same work in a formal validation can be redundant, inefficient, and expensive.

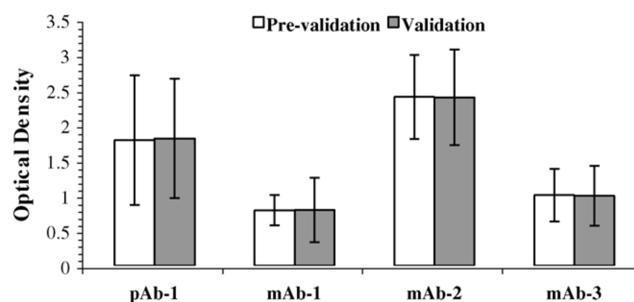

Fig. 4. A comparison of pre-validation vs. validation data of assay positive control ranges is shown as an example. The columns represent mean OD and the error bars represent 2S.D.

4.3. System-controlled validation

Given that a significant amount of validation was achieved in the prior two phases, the final validation phase prior to the bioanalysis of clinical samples could be relatively short. As shown in Fig. 3, it comprises the use of the SOP and Pass/Fail acceptance criteria to test precision and robustness (and ruggedness, if applicable), culminating in a validation report approved by company management. It is critical to assure that clinical and non-clinical sample analyses do not commence without an approved pre-study validation and documented training of the analysts.

4.4. In-study monitoring

A critical component of the assay lifespan that is often unrecognized is the need for in-study monitoring and periodic re-validation. The acceptability of analytical data corresponds directly to the criteria used to validate the method. Thus, once the analytical method has been validated for

routine use, its accuracy and precision should be monitored to assure that the method continues to perform satisfactorily [7]. Monitoring the run-to-run performance of QC samples assures that the assay is performing as well as it did during validation. There are no true reference standards or calibrators that can be utilized to test accuracy in each run of an ADA assay. However, the use of QC samples (one or more positive controls, a matrix negative control, and a diluent negative control) approximates monitoring of accuracy. Establishing acceptance criteria for repeatability, and monitoring inter-day and inter-analyst results of QC samples can assure that the precision calculated from the validation data is maintained. Hence, the validation of a method actually does not end until the method is ultimately retired from analytical use.

5. Conclusion

Immune response assays are designed to detect the presence of ADAs. In order to accept the data generated by these assays, these assays must be rigorously validated. For the novice, this article presented the discrete steps required of a rigorous ADA assay validation protocol. For the expert in this field, we have proposed practical and effective solutions to increase operational efficiency. The latter requires a scientist to understand the nuances of their system to ensure an efficient, yet thorough, assay validation. While we classified validation parameters as "system descriptive" or "system controlled", such classification may vary for different methods. For instance, for an assay where accuracy varied considerably, it may be necessary that spike/recovery controls be included in every plate. In this example, accuracy will need acceptance criteria, and therefore, may be a system-controlled parameter for that assay. Hence, the general outline provided in this publication is intended to promote streamlining of the method validation process and to eliminate unnecessary redundancy and wastage of resources. The authors believe that this can be accomplished within the realm of GLP compliance, and welcome further discussion.

References

[1] H. Schellekens, Clin. Ther. 24 (2002) 1720–1740.

[2] Proceedings of the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Tripartite Guideline, S6: Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals, July1997 (available online at http://www.fda.gov/cder/guidance/index.htm).

[3] M. Wadhwa, C. Bird, P. Dilger, R. Gaines-Das, R. Thorpe, J. Immunol. Methods 278 (2003) 1–17.

[4] A.R. Mire-Sluis, Y.C. Barrett, V. Devanarayan, E. Koren, H. Liu, M. Maia, T. Parish, G. Scott, G. Shankar, E. Shores, S.J. Swanson, G. Taniguchi, D. Wierda, L.A. Zuckerman, J. Immunol. Methods 289 (2004) 1–16.

[5] C. Pendley, A. Schantz, C. Wagner, Curr. Opin. Mol. Ther. 5 (2003) 172–179.

[6] V.P. Shah, K.K. Midha, S.V. Dighe, I.J. McGiveray, J.P. Skelley, A. Yacobi, T. Laylogg, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Putman, S. Spector, J. Pharm. Sci. 81 (1992) 309–312.

[7] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine, Guidance for Industry: Bioanalytical Method Validation, May 2001 (available online at http://www.fda.gov/cder/guidance/index.htm).

[8] Proceedings of the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Q2A, Text on Validation of Analytical Procedures (available online at http://www.fda.gov/cder/guidance/index.htm).

[9] A.R. Schantz, J. Ford, N. Brown, C. Wagner, AAPS Pharm. Sci. 3 (2001), S1 (available online at http://www. aapspharmaceutica.com/search/abstract_view.asp?id=917&ct= 01Abstracts).

[10] The United States Pharmacopeia and National Formulary 2002 (USP 25/NF 20), Chapter <1225> Validation of Compendial Methods, United States Pharmacopeial Convention Inc., Rockville, MD, USA.

[11] Proceedings of the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Q2B, Text on Validation of Analytical Procedures—Methodology (available online at http://www.fda.gov/cder/guidance/index.htm).