

Standardization

Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products

Anthony R. Mire-Sluis^{a,*}, Yu Chen Barrett^b, Viswanath Devanarayan^c,
Eugen Koren^d, Hank Liu^e, Mauricio Maia^f, Thomas Parish^g, George Scott^h,
Gopi Shankarⁱ, Elizabeth Shores^j, Steven J. Swanson^d, Gary Taniguchi^{k,†},
Daniel Wierda^l, Linda A. Zuckerman^m

^aOffice of Biotechnology Products, Center for Drug Evaluation and Research, FDA, 5515 Security Lane, HFD-3, Rm 1011, Rockville, MD 20852, USA

^bBristol-Myers Squibb Company, PO Box 5400 Princeton, NJ 08543, USA

^cGlobal Statistical Sciences, Lilly Research Labs, DC 2233, Eli Lilly & Co., Indianapolis, IN 46285, USA

^dClinical Immunology, Amgen Inc., Thousand Oaks, CA 91320, USA

^eWyeth Pharmaceuticals, 401 N. Middletown Road, Pearl River, NY 10965, USA

^fPreclinical and Clinical Development Sciences, PDL Inc., 34801 Campus Drive Fremont, CA 94555, USA

^gProctor and Gamble Pharmaceuticals, Route 320 Woods Corners, Norwich NY 13346, USA

^hLigand Binding Services, MDS Pharma Services, 2350 Cohen Street, St. Laurent, Quebec, Canada H4R 2N6

ⁱDepartment of Clinical Pharmacology, CENTOCOR, Inc., 200 Great Valley Parkway, Malvern, PA 19355, USA

^jDivision of Therapeutic Proteins, CDER, FDA, N29A RM2A01 HFM-538, 8800 Rockville Pike, Bethesda MD 20892, USA

^kClinical Laboratory Operations, Biomarin Pharmaceutical Inc., 371 Bel Marin Keys Blvd., Novato, CA 94949, USA

^lDepartment of Immunotoxicology, Eli Lilly & Company, 2001 W Main Street, Greenfield, IN 46140, USA

^mBioAnalytical Research and Development, ZymoGenetics, Inc., Seattle, WA 98102, USA

Received 14 March 2004; accepted 14 June 2004

Abstract

Most biopharmaceutical therapeutics elicit some level of antibody response against the product. This antibody response can, in some cases, lead to potentially serious side effects and/or loss of efficacy. Therefore, the immunogenicity of therapeutic proteins is a concern for clinicians, manufacturers and regulatory agencies. In order to assess immunogenicity of these molecules, appropriate detection, quantitation and characterization of antibody responses are necessary. Inadequately designed antibody assays have led to the hampering of product development or, during licensure, post-marketing commitments. This document provides scientific recommendations based on the experience of the authors for the development of anti-product antibody immunoassays intended for preclinical or clinical studies. While the main focus of this document is assay design considerations, we provide scientific focus and background to the various assay performance

* Corresponding author.

† Formerly at Genentech Inc., South San Francisco.

E-mail address: mire-sluisa@cdcr.fda.gov (A.R. Mire-Sluis).

parameters necessary for developing a valid assay. Sections on assay performance parameters, including those that appear in regulatory guidances, are contained in this manuscript.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Assays; Antibody; Immunogenicity

1. Introduction

The primary concern with unintended humoral immunogenicity of biological products for manufacturers, regulatory agencies, clinicians and patients is whether antibodies produced by patients receiving the product result in clinical sequelae. However, the assessment of the correlation between immunogenicity and clinical sequelae is entirely dependent on the appropriate detection, measurement and characterization of antibodies against biological therapeutics. The current assays for antibodies in biological fluids are prone to a wide range of technical difficulties in addition to scientific challenges in their interpretation.

Assays can be classified under one of the four types: Definitive Quantitative, Relative Quantitative, Quasi-Quantitative and Qualitative (Lee et al., 2003). Definitive and Relative Quantitative assays are those where the analytical results are expressed in continuous units of the reference standard after calibration from the standard curve. Quasi-Quantitative and Qualitative assays are those for which an appropriate reference standard is not available. With quasi-quantitative assays, the analytical results are expressed in terms of the assay signal in continuous units, whereas with qualitative assays analytical results are discrete (ordinal or nominal). The immunoassays used for screening anti-product antibodies fall under the class of Quasi-Quantitative assays due to the lack of similarity between the standard samples and test samples. This is because the reference standards may not accurately reflect the antibody affinities, proportions and other conditions in the test sample. For the same reasons, the analytes in test samples are not similar amongst themselves. Since the lack of similarity implies non-parallelism, the analytical results determined from dose interpolation (calibration) from the standard curve in the absence of parallelism will represent an inexact approximation (Finney, 1978). Due to this quasi-quantitative nature of the immunogenicity screening assays, we advocate a titer-based

approach in preference to the calibration approach for quantifying the results from test samples, unless parallelism has been demonstrated appropriately in the data analysis.

Numerous challenges to assay design exist and require consideration in the design of immunogenicity assays—as summarized in Table 1. Once a method is optimized, rigorous validation must be carried out to assure that the results they produce are meaningful.

Note: These recommendations are based on the experience of the authors. They reflect scientific concepts that should assist assay developers in forming a rationale for the development of their specific assay. They are not intended to be used in lieu of published FDA guidances and points to consider, or direct discussions with the agency (U.S. Department of Health and Human Services et al., 2001; International Committee for Harmonization. Q2A: Text on Validation of Analytical Procedures; International Committee for Harmonization. Q2B: Validation of Analytical Procedures: Methodology).

It is anticipated that these recommendations are generally applicable to the majority of anti-drug antibody immunoassays. However, it is acknowledged that it may not be possible to implement these recommendations for a number of specialized anti-drug antibody assays and that other approaches may be equally valid (Findlay et al., 2000). In addition, these recommendations are not intended to inhibit the development of novel technology that can improve our ability to detect clinically relevant antibodies. The approaches described are intended to assist the development and validation of immunoassays, but other approaches can be acceptable, as long as they are scientifically justified.

2. Defining the requirements of the assay

This paper focuses solely on the screening assay and on limited aspects of sample quantitation, if

Table 1
Advantages and disadvantages of various anti-drug antibody assay formats

Method	Advantages	Disadvantages
ELISA—Bridging Format (coating with drug and detecting with labeled drug)	<ul style="list-style-type: none"> • Detection of all isotype responses • Not species-specific • High through-put 	<ul style="list-style-type: none"> • Reduced ability to detect low affinity antibodies due to monovalent binding to the coat required and/or loss in multiple washes • Immobilizing drug on solid surface may mask or alter epitopes • Conjugating the molecule may mask epitopes • Conjugating chemistry may degrade or alter molecular structure of drug • Bridging formats are affected more by drug interference than other ELISA formats
ELISA—Direct Format (coating with drug and detecting with labeled anti-Ig)	<ul style="list-style-type: none"> • High through-put • Allows for dual arm binding therefore possibly increasing ability to detect low affinity antibodies 	<ul style="list-style-type: none"> • Immobilizing drug on a solid surface may mask or alter epitopes • Isotype detection determined by conjugate • Species specificity determined by conjugate • Low-affinity antibody may be lost in multiple washes • Detection reagents may be different between control and sample
Indirect Format (coating with a specific MAb or biotin etc. then drug)	<ul style="list-style-type: none"> • Coating plate with specific MAb keeps drug in oriented position • Very consistent coating and epitope exposure 	<ul style="list-style-type: none"> • Requires extensive studies to show that the MAb does not alter epitope accessibility, except for the one bound, that should be shown not to be clinically relevant • Isotype detection determined by conjugate • Species specificity depends upon conjugate • Lower affinity antibody may be lost in multiple washes
Radio Immuno-precipitation Assay (RIP)	<ul style="list-style-type: none"> • Solution phase • Moderate to high through-put • Generally good sensitivity 	<ul style="list-style-type: none"> • Uses radioactivity • Tracer short-lived (^{125}I) (half-life 60 days) • Conjugating chemistry may degrade or alter molecule—radio-label may mask epitope and need reproducible labeling procedure • Decay of radio-label may affect molecule stability • Challenging for antibody therapeutics • Species specificity depends upon precipitating agents • Isotype specificity determined by precipitating agent • May trap labeled drug non specifically in precipitate • Interference in epitope presentation can occur in solution, dependent on the drug and surrounding milieu • Potentially weaker binding interactions relative to solid surface immobilization

(continued on next page)

Table 1 (continued)

Method	Advantages	Disadvantages
Surface Plasmon Resonance	<ul style="list-style-type: none"> • Solution phase (because of dextran mobility) • No detection conjugate required • Detection of antibodies with all forms of association and dissociation rates • Detection of all isotype responses • Not species-specific 	<ul style="list-style-type: none"> • Linking chemistry may affect molecule • Linking molecule to dextran may mask epitopes • Regeneration step may degrade molecule • Reagents/technology vendor-specific • Low through-put • Confirmation of a positive response to a humanized antibody therapeutic is difficult if Ig is the same isotype as product (can use Fab). • Consideration must be given to possible false positives with an antibody product due to the presence of soluble ligand in the sample • Often less sensitive than ELISA or RIP, etc.
Electrochemiluminescence— Bridging Format	<ul style="list-style-type: none"> • Solution phase • Detects all isotype responses • Not species-specific • High through-put • Ability to use high matrix concentrations • Increased surface area for detection • Detection signal consistent during life of TAG conjugate 	<ul style="list-style-type: none"> • Need to prepare two conjugates (biotin and TAG) requiring more material • Conjugating molecule may mask or alter epitopes • Conjugating chemistry may degrade or alter molecule • If target molecule exists in a biological sample being tested and has two of the same antigen epitopes, may cause false positives • Reagents/technology vendor-specific

carried out at the initial screening stage. In order to select an appropriate screening assay format, one must define the intended purpose of the assay. There are various levels of testing that can be carried out on a sample apart from an initial positive/negative assessment. This can include estimates of the levels of antibody, confirmation of positives by adding drug to the sample, isotyping and testing for neutralizing capacity.

Screening assays should be able to *specifically* detect at least IgM and all IgG subclasses. It is recognized that the presence of other classes, such as IgE and IgA may have clinical sequelae, but this is highly dependent on the disease and patient population. Discussions on the utility of detecting these classes of Ig should be initiated with the regulatory authorities prior to the development of the assay.

The assay should also exhibit sensitivity, precision, recovery and reproducibility suitable for the intended purpose of the assay. These requirements will vary depending on the use of the assay (preclinical or clinical), stage of product development and the nature of the product and its target population.

Each of the currently available assay formats has its advantages and disadvantages (refer to Table 1). Should the initial format not perform to expectations

(e.g., poor sensitivity or high backgrounds), more than one assay format may need to be developed and tested, as suggested in Table 1. Suitable assay optimization should be undertaken to ensure that a valid comparison can be carried out. For example, minimum dilution, coat concentration, conjugate concentration and incubation periods should be optimized before comparisons can be made.

Antibodies have different on-rates and off-rates and these, with other binding characteristics, combine to determine how an assay format will affect the ability to detect antibodies. However, it is likely that limiting wash steps or reducing wash fluid speed assists the detection of antibodies with fast off rates that may not be detected due to extensive washing. Orthogonal assay formats and technologies (High-Performance Liquid Chromatography, Surface Plasmon Resonance, Immunoblotting, etc.) are available to further characterize the immune response and can assist in the characterization of the screening assay.

It is difficult to definitively recommend a single assay format, so during assay comparisons, consideration should be given to the nature of the product (e.g., a therapeutic protein or a monoclonal antibody), potential interference in the assay from co-medications, dosing regimen (e.g., potential for product to be

present in the sample) and/or disease-specific issues (e.g., rheumatoid factor). One must also consider whether the selected assay format allows for appropriate epitope exposure.

3. Selecting reagents required to develop the assay

3.1. Positive controls

As soon as enough protein therapeutic is available that is representative of the material to be used in preclinical studies, one should initiate the immunization of one or more animals to develop a positive control. For preclinical studies, the animals under study make the most appropriate candidates for immunization. For clinical studies, cynomolgus monkeys or other non-human primates may be preferable due to the high degree of similarity between their immune repertoire and that of humans. In addition, the occurrence of cross reactivity with anti-human secondary reagents, if confirmed experimentally, allows for the use of the same detection reagent as a human sample. However, selection of a positive control can be format-dependent, since a bridging ELISA format for example, is species-independent. It must be taken into account that not all species may produce antibodies with comparable affinity, avidity, and class/subclass types, thus potentially affecting their performance. Therefore, antibodies from other species need to be compared to those obtained from the species under study (on a risk-based analysis, most importantly for human clinical assays) to ensure comparable binding performance. Furthermore, optimal assay conditions may be different for antibodies from different species; therefore care must be taken to generate an assay that is ideal for the species of interest.

There are several considerations to be taken into account for the selection of positive controls. For example, the timeframe for the need of the positive control can dictate which species is more accessible for the production of a polyclonal preparation (e.g., a rabbit) as opposed to the work required to develop a panel of monoclonal antibodies. The degree of homology between the product and any endogenous counterpart in a selected species may influence the suitability of the positive control. For example, a highly conserved

protein may not induce a sufficient immune response in a non-human primate. However, 'hyper-immune' monkey sera/plasma can often be produced using adjuvants in the immunization protocol.

It is worth considering the immunization of several members of the same species to account for non-responders. A sufficient volume of pre-immunization serum/plasma should first be acquired as a negative control. When a high titer is obtained, as much material as possible should be collected across several bleeds. Some antibodies available from commercial vendors may be used as positive controls, or as an interim reagent until positive serum/plasma is available. An assessment of the quantities of reagent produced from the immunization protocol is useful to ensure enough reagent is available through assay development and the duration of studies. Purification and quantitation of antibody preparations used as controls are generally required for use during the assay optimization and validation process.

3.2. Matrix samples and negative controls

To assess non-specific backgrounds, recoveries, etc., and to create a negative control, samples from untreated individuals should be obtained from the target population, e.g., healthy and disease state. These samples can usually be obtained from commercial sources. Depending on availability and the variability of the data, ideally 15 individual samples should be obtained for animal studies. For clinical studies, typically 50 individual samples should be obtained. In addition, if a number of pools are prepared, then more material will be available for further studies. Based on statistical assessment of data from initial experiments, it may be necessary to increase these numbers.

Obtain as much matrix as possible per individual to allow for repeat testing. If untreated patient samples are difficult to obtain, then a cohort of healthy donor samples should be tested, but compared to a few disease state samples to check for any disease associated phenomenon (such as high background caused by pre-existing antibodies). Consideration of pre-existing antibodies, binding proteins, soluble receptors, etc., can influence the selection of matrix samples and appropriate negative controls.

3.3. Quality controls

The purpose of quality controls (QC) is to monitor the performance and acceptability of the assay during the study. These may differ from the controls used to optimize and validate the assay. In preparing the positive quality controls for preclinical studies, it is preferable to use antibodies from the same species as the study subjects. If the positive control is from a different species, the biological matrix used for diluting the positive controls should be of the same species as that of the study animals. For clinical studies, in the absence of human-derived antibodies, primate material is preferable as it allows for anti-human secondary reagents to be used. As a clinical program matures, the pooling of positive human samples, if available in reasonable amounts to prepare a positive QC is recommended.

A low positive QC provides the best assurance that the assay can detect low positive samples on a routine basis. The low positive QC should have the lowest response that can be reproducibly seen above the cut point response. Some objective criteria must be used in the selection of the concentration of a low positive control. Otherwise, choosing an unreasonably high concentration for a low positive QC may produce an assay signal that is substantially above the cut point, and may therefore never lead to the rejection of an assay run. The simplest acceptance criteria for the low positive QC would be that it typically produces an assay signal above the cut point. The acceptance range can be defined during assay qualification based on inter- and intra-assay variability. A suitable approach would be to select a concentration for the low positive QC that would lead to the rejection of an assay run 1% of the time. This can be done via appropriate statistical analysis and curve-fitting of data from a dilution series of a high positive QC from a few assay runs. An alternative approach is to create a dilution series of a high positive sample in matrix and assess in every assay run whether the dilution that meets the cut point is within a defined range.

A high positive QC should also be included, at levels depending on the dose response relationship of the assay to assure suitability throughout the range of assay responses. It can be useful to include a mid range QC in assays during optimization and validation to ensure consistency in the performance of the assay within that region of assay responses.

3.4. Other reagents

It is often necessary to search for commercial vendors for assay reagents. If there is only a single supplier of a critical reagent then it is advisable to consider some contractual relationship that warns the user about product changes or cessation of production. It is also recommended to test multiple lots of reagent to assure vendor product specifications meet the user's requirements. If the necessary reagents are not available from commercial suppliers, a plan should be developed for in-house production. This may include preparation of antibody fragments (Fab, F(ab')₂, Fc), conjugation of reagents, affinity purification of the positive control, affinity purification of anti-CDR antibodies, etc. At this time, other reagents such as known specific binding proteins, soluble receptors or other possible interfering substances should also be ordered. Consideration should be given to the time requirements to purchase or produce reagents. It is recommended that quality acceptance criteria and standard operating procedures for testing be applied to critical reagents included in the assay, to ensure quality and continuity when replacements are required.

4. Design of the assay

After preliminary assay development considerations have been addressed (e.g., selection of reagents, suppliers, etc.), optimal assay procedures and parameters have to be determined to obtain a valid assay. Examples include "minimum dilution", incubation time, coating antibody concentration, number of wash cycles, etc. The optimization process may be carried out by varying and optimizing one variable at a time and holding other variables fixed. While this might be easy to implement, it is often quite laborious. More importantly, this approach may fail to identify potentially important interactions between certain assay variables. Failure to take into account such interactions into account during the optimization phase could greatly compromise the performance characteristics of the assay (see Fig. 1 for an illustration of some of this inter-dependence). Therefore, we highly recommend the use of appropriate statistical experimental design methods that allow multiple variables to be varied and optimized simultaneously (e.g., fractional factorial

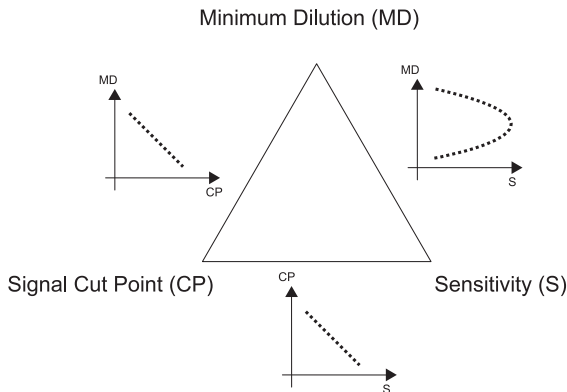


Fig. 1. A diagrammatic representation of the inter-relationship between sensitivity, cut point and minimum dilution as assessed during assay optimization.

experiments, central composite designs, etc.). This in turn allows for the consideration of any interactions between the variables which can be accomplished with commonly available software and/or in collaboration with a statistician. Any software used should be validated for its intended use.

5. Minimum dilution

The minimum dilution is the dilution that yields a signal close to the signal of non-specific binding (NSB) of assay diluent. This is required because positive samples are subsequently titrated in assay diluent to measure the antibody present, by diluting down to the cut point of the assay. Therefore, one needs to be assured of a true assay signal vs. assay noise. However, matrix interference, recovery, and assay sensitivity should be taken into consideration when selecting the minimum dilution.

Assessing the minimum dilution is also necessary so that interference from matrix components can be minimized. The process must involve the consideration of minimizing background while ensuring target assay sensitivity. The determination of the minimum dilution is not a straightforward process and involves the use of a dilution series from a panel of at least 10 samples from an untreated patient population (or healthy donors if these samples are not readily available). The ideal number of samples required will depend on the variability of the data.

From the dilution series of untreated patient (or donor) samples, the % change in background relative to the assay diluent can be determined for different dilution levels. By taking a few of these dilution levels as candidates for minimum dilution, the % reduction in background, sensitivity, cut point, recovery, etc., of the assay can be determined. The optimal minimum dilution can then be selected by appropriate statistical analysis such that there is an ideal balance between these various assay performance characteristics. Note that this selection must be done simultaneously with the selection of other variables in the assay as noted above in Section 4.

From our experience, an ideal minimum dilution should be aimed to be not greater than a 1/100 dilution (preferably 1/20–1/50) since higher dilutions impact the sensitivity of the assay, although this can depend on the nature of the samples and the assay format.

6. Recovery

Testing for recovery in immunogenicity assays is testing whether components in the matrix being investigated can inhibit anti-drug antibody from binding to drug under assay conditions, thereby affecting assay response. Rather than setting a single recommended recovery level, it is important to understand the recovery profile of the matrix under analysis and provide results that can be scientifically justified. For example, with a very sensitive assay, a lower recovery may be acceptable. It may appear that assessment of recovery is more critical to single-point screening formats than titration formats, where matrix effects would be expected to be diluted out. However, in low level positive samples, only a minimal dilution may be required to reach the assay cut point, thus allowing for interference from residual matrix.

To determine recovery, both low and high concentrations of specific anti-drug antibody preparations (polyclonal or monoclonal) should be added into assay buffer, 5–10 healthy donors and 5–10 treatment population sera/plasma. The responses in assay buffer are then compared to responses observed in matrix. It is recommended to select donors that express both high and low NSB. This will not only test the assay recovery but also check the matrix interference and confirm the selection of the minimum

dilution. Since the antibody preparations are most likely to be non-human-derived, this may not necessarily reflect the true accuracy of detection of human antibodies for clinical studies, however it is an acceptable approach.

7. Assay cut point

The cut point of an assay is the level of response of the assay at or above which a sample is defined to be positive and below which it is defined to be negative. The cut point is therefore determined statistically from the level of non-specific background of the assay and the response of matrix samples, above which a putative positive response can be detected. As was described earlier in Section 4, an ideal balance has to be achieved between minimum dilution, sensitivity and the cut point. The cut point is established during the validation phase.

In determining the assay cut point, use test samples as similar as possible to the actual study samples. For example, samples should be obtained from 'healthy controls' and study subjects with the disease under study if the assay will be used for both of these populations. It may be necessary to define different cut points depending on the population being tested.

During the assay optimization phase, a low number of samples (5–10) can be used to estimate the cut point, minimum dilution, etc., and therefore to decide whether the assay format is suitable for further development. Once the optimization is complete, an adequate number of samples should be analyzed during validation to provide a statistically valid assessment of assay response variability at the selected minimum dilution. The assessment of assay variability should be based on data from at least three independent runs of the assay (preferably on different days), including replicate tests of each sample (e.g., replicate wells in a microtiter plate). The inter-sample variability in each of the three assays should be acceptable and justified for the intended purpose of the assay if this approach (and the assay) is to be valid. Further details on validation of this parameter will be included in a subsequent manuscript.

While determining day-to-day and sample-to-sample variability, remove all sample responses shown to

be due to preexisting specific anti-drug antibodies. In addition, reactive samples that contain rheumatoid factor, soluble receptors, etc., should be eliminated. Analyzing the distribution of the population of responses can be very informative and outliers should be addressed. The assessment and handling of outliers is not an easy task and will require the assistance of a statistician. Immunodepleting outliers or at least the samples that have high responses overall, can identify those samples that contain pre-existing antibodies that cross react with the drug.

For patient populations with high incidences of pre-existing antibodies where it is difficult to obtain a suitable number of true negative samples, a healthy donor population may be used to set the cut point.

An upper negative limit of around 95% for the cut point is recommended. Using a risk-based approach, it is more appropriate to have 5% false positives, than false negatives. There are many ways that this can be attained, such as parametric approaches using the mean plus 1.645 S.D., where 1.645 is the 95th percentile of the normal distribution, or a robust alternative with median and median absolute deviation instead of mean and S.D., respectively, or a non-parametric approach using empirical 95th percentiles. The exact method used will depend on various factors, including the nature of the data, regulatory input, etc.

A screening assay that does not identify any positives whatsoever following a preclinical or clinical trial casts doubt on the ability of the assay to detect low positive samples. A screening assay that picks up 5% positives that are subsequently shown to be due to NSB in a confirmatory (immunodepletion) assay provides assurance that true low positives can be detected.

8. Normalization of assay cut points on an inter-assay basis

It is recognized that assay signal responses will vary between assay runs and that the cut point needs to be modified to take this into account. The assay scientist will need to develop a procedure to normalize this inter-assay variability.

Using the response of a negative control plus a predetermined factor to assess the cut point in every assay allows for changes in NSB and assay readouts

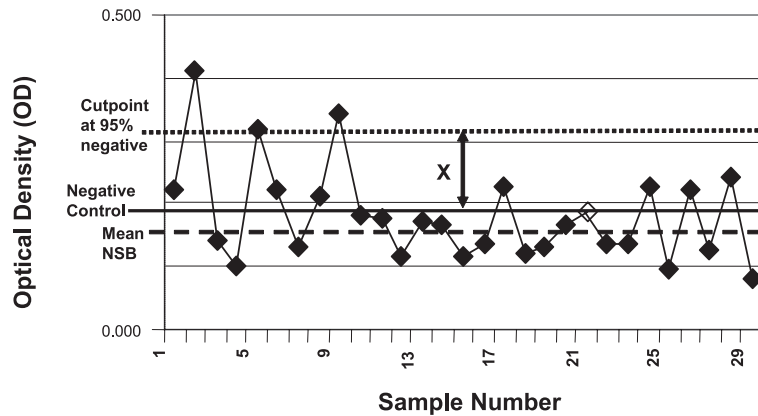


Fig. 2. An illustration of the selection of an assay cut point based on the nonspecific binding (NSB) of negative control matrix samples. Thirty negative controls are assayed (closed diamonds) and a mean NSB optical density (OD) is established (dashed line). An upper 95% confidence limit above the NSB mean is estimated (dotted line) that should result in 5% false positives. In order to account inter-assay variability, a negative control is selected to be incorporated in every assay (open diamond). A normalization factor (X) is derived from the difference in OD of the negative control and the OD of the 95% limit. After each assay, the normalization factor is added to the OD of the negative control to set the 95% cut point of the assay.

over time (Fig. 2). The predetermined factor is the upper 95% limit of the NSB data from the untreated patient (or healthy donor) samples, as described above, minus the mean (or median) response of the negative control pool. However, there should be an upper acceptance limit for the response of the negative control to make sure the assay is performing as expected and that an unusual level of NSB does not compromise sensitivity and validity of results from test samples. This limit can be based on an appropriate (e.g., 99%) upper confidence limit of the data from negative control samples used during validation.

The use of a predetermined factor to determine the cut point for subsequent assay runs is appropriate only if the variability between samples within each assay run is homogenous across assay runs (Fig. 3). This assessment can be made during optimization by applying appropriate statistical tests (for example, using Bartlett's or Levine's test in the analysis of variance model by treating the assay runs as fixed effect). If there is a statistically significant non-homogeneity of intra-assay variability, it is strongly recommended that the assay be redeveloped. It may not always be possible to resolve the problem through redesign. However, it would be possible to continue with the assay through validation by including enough negative controls in each assay to

determine a run specific factor that will define the cut point.

It is also worth considering the inclusion of low positive controls near the cut point of the assay during the assessment of the variability of the negative control. It is our experience that the shift in responses to positive controls are similar to negative controls, but the inclusion of a low positive control in each run with an associated acceptance criteria assures similarity of performance of true low positive samples (as described previously). Normalizing data against a negative control, although recommended here, is only one possible approach to account for assay variability. For example, for assays with minimal day to day variability, a cut point could be defined during assay validation and used throughout further studies. However, such an approach will require rigorous statistical validation to assure acceptability.

During the optimization process, it is useful to select samples with enough material to serve as the negative control for an extended time (most often a pool of matrix—it has been found that large pools tend to act in a similar manner as the study population). It is preferable to include more than one of these, as a back-up negative control, should the primary material run out unexpectedly. If a new negative pool does not match the previous one the

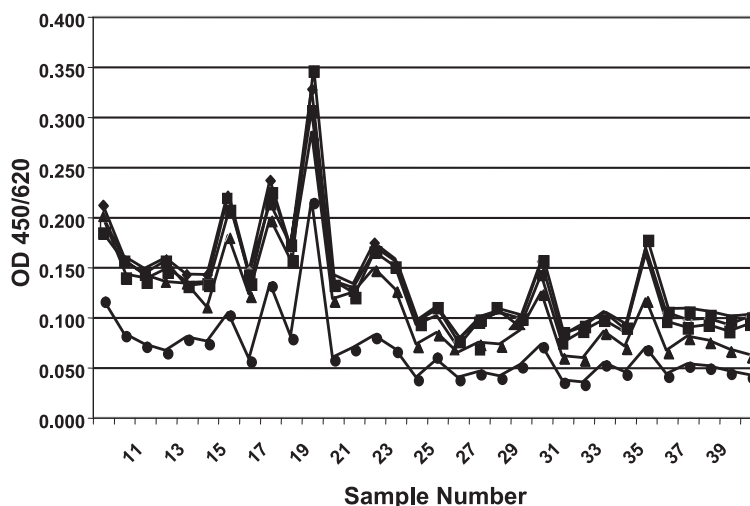


Fig. 3. Data from four independent assays of negative control samples. Forty baseline samples were assayed in four independent assays. The data illustrate that there is variability between the ODs of samples between assays, but that there is minimal variation of relative signals between samples. If an assay exhibits this type of variability the use of a negative control to set the 95% cut point in each assay as illustrated in Fig. 2 is valid.

cut point will need to be re-established relative to the new control.

9. Sensitivity

The purpose of any antibody assay is to detect biologically relevant levels of anti-drug antibodies in biological fluid samples. In order to assure that the assay is suitable for its intended purpose it is critical to assess sensitivity.

A rationale is required using sound scientific principles to provide regulatory authorities with an estimate of assay sensitivity expressed in mass units. It is recognized that it is highly unlikely that a human positive control would be available at the time of assay development (Fig. 4), so alternative controls may be required. It is also understood that polyclonal preparations can contain antibodies of different affinities and in different proportions, so an absolute assessment of sensitivity to all types of antibodies is difficult. Therefore, the assessment of assay sensitivity relies almost exclusively on the characteristics of the positive control.

To determine the sensitivity of the assay, a purified preparation of antibodies specific to the product is required. These can either be an affinity purified

polyclonal preparation, or a monoclonal antibody(ies). It is likely that only the sensitivity of IgG classes of anti-product will be estimated, since it is highly improbable that other classes of product specific Ig will be available for analysis.

The choice of the source and nature of the positive control is dependent on the format of the assay and the secondary reagents used. For example, it is simpler and more acceptable to use a non-human primate preparation in assays that use an anti-human secondary reagent if that reagent cross-reacts with the non-human primate antibodies.

During the development of the assay, consideration should be given to the reactivity of anti-drug antibodies of differing affinities since the sensitivity of the assay will vary between 'high' and 'low' affinity antibodies. Therefore, factors that may influence detection of low-affinity antibodies such as washing, buffer constituents, length of incubations, etc., should be considered during assay optimization. It is recognized that demonstrating which factors directly affect the detection of low-affinity antibodies is a challenging task.

The assessment of sensitivity is only required for the reporting of the defined performance characteristics of the assay and is not required for routine expression of results of a screening assay—a positive/negative or a titer suffices in that instance. Quantitation of antibody

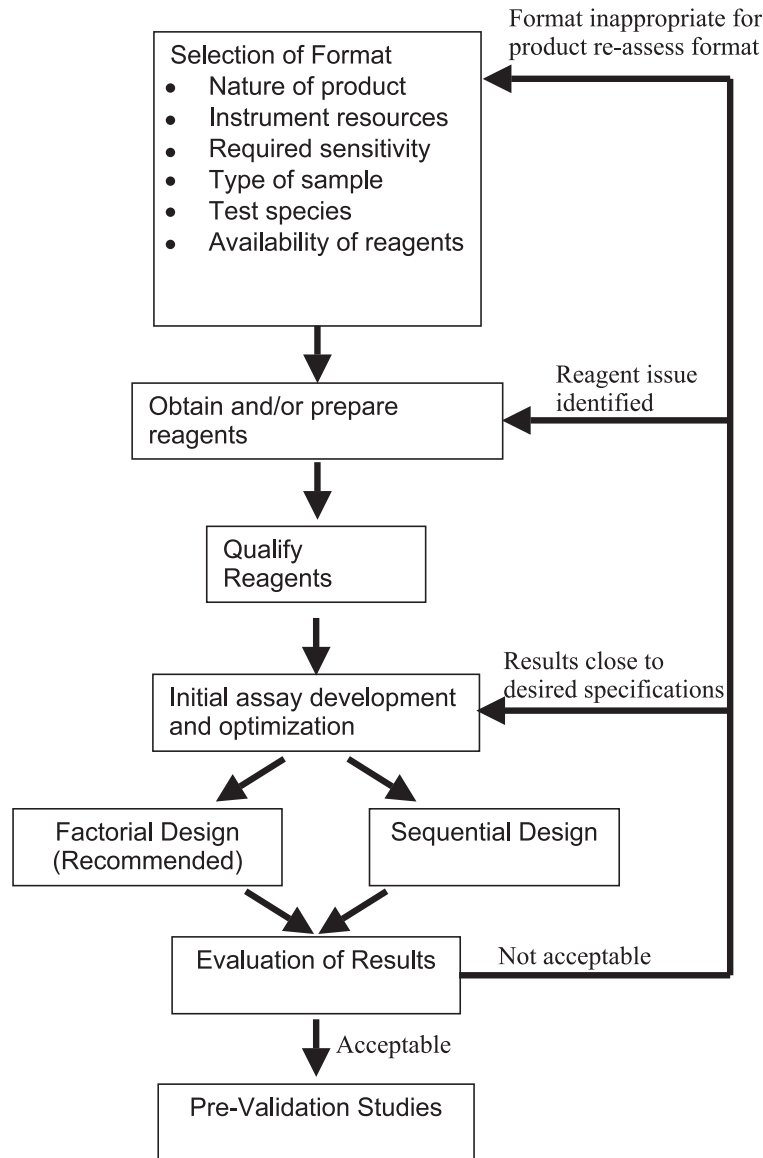


Fig. 4. An assay development flowchart.

levels can be attempted at either the screening or confirmation assay stage, depending on assay design.

Once a suitable preparation of purified antibodies is obtained and the concentration carefully determined, the preparation should be diluted into the concentration of sample matrix expected to be used in the screening assay and diluted to produce a standard curve. For example, if the sample dilution for the assay is 1:50, then the antibody sample should

be assayed in a dilution series containing 2% matrix. One of the matrix samples used in the cut point determination can be used for the dilution, or the negative control, or a low background pool.

Note: For certain assay formats used for clinical studies, the use of animal antibodies (excluding non-human primate) as positive controls will require a secondary anti-Ig detecting reagent that is not the same as the anti-human Ig reagent being used to

test clinical samples. This will require a carefully considered qualification to show that the two detecting reagents can detect Ig at similar concentrations. A comparison of titration curves of a polyclonal human Ig preparation vs. animal Ig coating on plates using their respective conjugates at various dilutions should provide some confirmation of the abilities of the conjugates to detect similar quantities of Ig. In this experiment, the conjugates should provide similar signals at approximately the same dilution being used in the assay. Nevertheless, the results of such experiments do not provide an ideal comparison between animal-derived material and clinical samples, so use of this type of reagent is not recommended.

Assay sensitivity is determined by the concentration at which the antibody preparation produces an assay readout equal to the cut point determined for that particular assay. This should ideally be estimated using an appropriate regression model to interpolate the dilution at which the cut point was reached. The dilution can then be used to determine the mass of antibody detectable at the cut point of the assay. This should be expressed as mass of antibody detectable/ml of matrix.

There is little data available to assess how low levels of antibody can affect patient's long term, or whether levels of antibodies go up or down on a longitudinal basis. This is particularly of concern with products that could induce cross reactive antibodies against endogenous proteins present at low levels. Therefore, it is wise to target a lower threshold to possibly capture events before they lead to serious clinical consequences. Such an approach has been carried out in clinical studies of high risk products, where real time analysis of samples was part of the strategy to minimize the development of serious clinical adverse events. To best insure detection of all clinically relevant antibodies, sponsors should strive for sensitivities near 250–500 ng per ml of serum for antibody assays supporting clinical studies. For preclinical studies, sensitivities of 500–1000 ng/ml are reasonable. This applies to detection of antibodies tested in matrix alone for the purpose of assay validation. Sensitivity of the assay for samples where drug or an endogenous counterpart is present in the sample is addressed in Section 11. Regardless of the

sample or study type, assay sensitivity should be justified by appropriate risk analysis dependent on the nature of the product, its intended use and the sequelae associated with an immune response against the product. Therefore, these numbers are recommendations based on our experience and regulatory interactions.

10. Precision

For immunogenicity assays, evaluation of precision depends on whether the screening assay is a yes/no or a titer format. For the yes/no assay format, estimation of precision addresses whether a sample that is detected as positive will remain positive and a negative remain negative following repeated assay. This will involve assaying of a low positive and negative controls and the assessment of the variability of the assay response. Appropriate statistical methods should then be considered to estimate the false positive rate from a sufficient number of replicates of negative control samples. Similarly false negative rates can be estimated using low positive control samples (it is important to use an appropriate antibody concentration for the low positive control as described in Section 3.3).

For titer assays, a medium or high positive control can be titered and the variability of the titer can be assessed. Using the estimates of intra-assay and inter-assay variability, it will be very informative to express the precision of titer values in terms of a Minimum Significant Ratio (MSR). For example, an MSR of 2 for the titer data will indicate that titer values that are within 2-fold of each other are not statistically different, i.e., in this case, we need to see at least a 2-fold difference between any two titer values to say that they are really different.

Intra-assay (repeatability) and inter-assay (intermediate precision) variability of assay responses should be evaluated on a minimum of 3 different days with a minimum of 2–3 replicates of the same sample in each assay. During precision testing of microtiter plate-based assays, position effects should be investigated using control samples. The degree of acceptable variability at different sample concentrations should be scientifically justified, where, for example, it is most likely that low signals will exhibit higher vari-

ability than those in the middle of the range of the assay readout.

11. Specificity and interference

Specificity is the ability of an analytical method to detect only the target analyte, in this case, anti-product antibodies. It must be noted that the specificity of immunogenicity assays can only be assessed during optimization using the positive control but may not be a true reflection of the specificity of the assay for preclinical or clinical samples. For screening assays, probably the most important consideration is whether components in the sample prevent the assay from detecting a true positive.

The most important substance to test for interference is the drug itself. It is expected that samples containing drug will exhibit interference due to competition for product specific antibodies between endogenous drug and that used in the assay system. Therefore, it is necessary to mix antibody samples with varying concentrations of drug to assess how much drug is required to eliminate or reduce detection in the assay. Interference should be assessed by testing the effect of various concentrations of study drug on the high and low positive controls. An appropriate regression model can then be used to fit these data to interpolate the concentration of study drug that will drop the response below the cut point for both high and low positive samples.

Interference tests are required if samples are to be taken from patients (or animals) when drug is still circulating, e.g., for long-acting drugs or formulations. Data from early pharmacokinetic studies can be used to help determine when it is appropriate to take samples that contain levels of drug that should not interfere with the ability of the assay to detect antibodies. This data should be highlighted for regulatory authorities, especially where long wash-out periods would have to occur due to the extended half-life of the drug before host antibodies to the drug could be evaluated.

It has been recognized that the presence of antibodies can prolong the half life of a drug. It is also known that polyclonal preparations with differing characteristics can respond differently to the presence of drug. However, experiments such as those described above can be informative.

Certain proteins similar to the analyte might be detected in the assay; others present in the sample (e.g., hemoglobin, lipids, soluble receptors) might prevent the analyte from being detected. Immune complexes also need consideration. Thus, testing for interfering factors requires a scientific rationale. The acceptable degree of interference should be scientifically justified depending on the nature of the samples and the significance of the immune response.

12. Accuracy

Accuracy is the closeness of test results obtained by the analytical method to the theoretical value. Since antibody screening assays are quasi-quantitative due to lack of a valid reference standard, the approach here is more aligned to that of assessing recovery (as discussed previously).

13. Linearity

Definitions of linearity contained in regulatory guidances do not apply for quasi-quantitative assays due to the lack of availability of a valid reference standard, unless determination of positive samples are based on comparisons to a reference standard curve. For yes/no screening assays, linearity assessment can be the illustration of an increase in signal resulting from an increasing concentration of anti-product antibody.

For any format of the assay, it is always necessary to assess the shape of the dose response curve of a high positive control in the assay to make sure there are no unusual prozone or 'hook' effects or other such phenomena that could compromise the validity of the assay.

14. Stability

14.1. Antibody samples and controls

Storage conditions for assay samples and reagents should be investigated as part of assay optimization, prior to validation. Stability should be evaluated to specify the optimal conditions for sample and reagent

storage before, during and after analysis. Determination of the number of freeze/thaw cycles, storage times and temperatures should be based on the expected use of the reagent and the storage of the samples. Quality Controls containing high and low levels of antibody should be used to evaluate QC and sample stability during optimization.

It is recognized that the stability of QC samples does not necessarily reflect the stability of preclinical or clinical samples, if the QCs are not of the same species. Due to the difficulty of predicting sample stability, it is recommended to reduce the number of freeze/thaw cycles and to aliquot samples at the first thaw if possible. It is also recommended that several aliquots be made after the initial sample blood draw and serum/plasma preparation, to lessen the requirement for thawing large aliquots. One should consider the material of the containers used for samples, e.g., glass vs. polypropylene to assess the impact on sample integrity. If a sample requires processing prior to analysis, the stability of the process intermediate should also be evaluated.

14.2. Reagents

The stability of the coat and conjugate antibodies should be investigated. For example, the stability of coated and stored blocked plates in microtiter-based assays, coated chips in a surface plasmon resonance assay, etc., should be considered.

Acceptance criteria for the stability of reagents are dependent on the influence of that reagent on the performance of the assay and should be justified.

15. Robustness

Robustness is the capacity of a method to remain unaffected by small experimental variations and deliberate changes in method parameters such as timing and temperature of selected steps in the assay procedure. Robustness can be explored as a follow-up after assay optimization to identify critical steps and parameters in the assay procedure (e.g., length of time for color development), ideally using factorial experiments. Parameters to test during robustness testing should be based on knowledge of the assay and associated risks.

16. Future considerations

Parameters such as ruggedness and system suitability discussed in regulatory guidances (U.S. Department of Health and Human Services et al., 2001; International Committee for Harmonization. Q2A: Text on Validation of Analytical Procedures; International Committee for Harmonization. Q2B: Validation of Analytical Procedures: Methodology) are more aligned to the validation process. This will be discussed in a subsequent manuscript along with more details on the other validation parameters introduced here.

Acknowledgements

The authors would like to thank the members of Genentech's BioAnalytical Sciences group who helped during the original concept of recommendations for developing immunogenicity assays or provided data, particularly Mark Hartmangruber, Cecilia Leddy and Bryan Sandlund. We would also like to thank Wendell Smith for helpful discussions. In addition, we would like to thank the FDA immunogenicity task force for their valuable input—Patrick Swann, Alexandra Worobec, Marilyn Lightfoote, Janice Brown, Andrew Chang, Ken Hastings, Susan Kirshner, John Langone and Dale Slavin. This work was sponsored by the Ligand Binding Assay Bioanalytical Focus Group of AAPS.

References

- Finney, D.J., 1978. *Statistical Method in Biological Assay*, 3rd edn. Charles Griffin, London.
- Findlay, J.W.A., et al., 2000. Validation of immunoassays for bioanalysis: A pharmaceutical industry perspective. *Journal of Pharmaceutical and Biomedical Analysis* 21, 1249–1273.
- International Committee for Harmonization. Q2A: Text on Validation of Analytical Procedures. Available at <http://www.fda.gov/cder/guidance/index.htm>
- International Committee for Harmonization. Q2B: Validation of Analytical Procedures: Methodology. Available at <http://www.fda.gov/cder/guidance/index.htm>
- Lee, J.W., Smith, W.C., Nordblom, G.D., Bowsher, R.R., 2003. Validation of assays for the bioanalysis of novel biomarkers. In: Bloom, J.C., Dean, R.A. (Eds.), *Biomarkers in Clinical Drug Development*. Marcel Dekker, pp. 119–149.

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 at <http://www.fda.gov/cder/guidance/index.htm>

Glossary

Accuracy: Accuracy is a measure of the closeness of agreement of a test result obtained by an analytical method to its theoretical true value, or the accepted reference/standard value.

Confirmation assay: A secondary assay, following a screening assay, which confirms the positive result (and may include assessment of specificity and/or titer).

Cut point: The cut point of the assay is the level of response of the assay at or above which a sample is defined to be positive and below which it is defined to be negative.

False positive: Positive test result from a truly negative sample.

False negative: Negative test result from a truly positive sample.

Factorial experiments: A multi-variate approach where two or more variables are varied simultaneously to achieve an optimal solution after taking into account of the interaction between these variables. This approach is typically more efficient than one-factor-at-a-time for optimizing assays and/or when analyzing multiple variables of an assay that are inter-dependent.

Fractional factorial experiments: A type of multi-factorial experiment in which only a subset of factor level combinations is tested. These experiments are very efficient for screening a large number of factors prior to optimizing the most important factors.

Hook effect: Also known as prozone effect, it describes an artifact seen when the analyte is in very high concentration, beyond the dynamic range of the method. This excess amount cannot be measured accurately and yields spuriously low results. To detect hook effects, samples are tested undiluted (or at the optimal dilution) and after higher dilutions.

Immunodepletion assay: A form of confirmation assay wherein the target (anti-product antibody)

is specifically depleted by inhibition of reactivity during a pre-incubation step with an excess of product. It demonstrates the presence of antibodies that bind to the product.

Interference: Presence of entities in samples that prevent the target analyte from being detected or accurately quantified. It is a significant bias in the measured analyte concentration due to the effect of another component or properties of the specimen.

Linearity: Linearity is the ability of an assay within a given concentration range to produce test results that are directly proportional to the amount of analyte in the test sample. It defines the part of the dose-response curve (range) of the assay in which there are no unusual prozone or 'hook' effects or other such phenomena that could influence the validity of the assay.

Matrix: The type of sample (e.g., serum, plasma, or other biological fluid) in which the target (anti-product antibody) is present.

Minimum dilution: The minimum dilution is the dilution that yields a signal close to the signal of non-specific binding of assay diluent.

Nonparametric methods: A statistical approach that makes no assumptions about a distribution. Nonparametric methods are a class of methods that are used for a broad non-parameterized set of underlying distributions. Mean, median and percentiles of a sample are nonparametric estimates of the corresponding population quantities.

Optimization: The process of developing an assay, prior to validation, wherein the variables affecting the assay are elucidated (e.g., antibody concentration, incubation time, wash cycles, etc.).

Parametric methods: Parametric methods are a class of methods that assume a particular parameterized set of underlying distributions for the data (for example, normal distribution). These methods may use mean or robust estimates such as median within the framework of the assumed underlying parametric distribution such as the normal distribution.

Precision: Precision is a quantitative measure (usually expressed as standard deviation and coefficient of variation) of the random variation between a series of measurements from multiple sampling of the same homogenous sample under the prescribed

conditions. Precision may be considered at 3 levels: repeatability, intermediate precision (within laboratory variations, such as inter-analyst, inter-assay, inter-day), and reproducibility. For immunogenicity assays, evaluation of precision depends on whether the screening assay is a yes/no or a titer format. For the yes/no assay format, estimation of precision addresses whether a sample that is detected as positive will remain positive and a negative remain negative following repeated assay.

Quality controls: A set of negative and positive controls included in each assay run to monitor, and maintain consistency of assay performance (system suitability).

Recovery: The quantified closeness of an observed result to its theoretical value, expressed as a percent of the nominal (theoretical) concentration. Recovery is often used as a measure of accuracy.

Regression model: A statistical model relating a dependent variable to one or more independent variables.

Repeatability: Repeatability is the precision of repeated measurements within the same analytical run under the same operating conditions over a short interval of time. It is also termed intra-assay precision.

Reproducibility: Reproducibility expresses the precision between laboratories. It is useful for assessing the “transferability” of an assay and/or the validity of testing samples in two or more laboratories. It is also referred to as ruggedness.

Robust estimates: Robust estimates are estimates that down-weight the influence of extreme values in summarizing the data via the use of median, trimmed mean, median absolute deviation, etc.

Robustness of the assay: Robustness is a measure of

the capacity of the assay to remain unaffected by small, but deliberate changes in method parameters and provides an indication of its reliability during normal run conditions.

Ruggedness: same as reproducibility.

Screening assay: The primary assay that differentiates study samples into potentially negative or positive for the presence of anti-product antibodies.

Sensitivity: The lowest concentration of anti-product antibody that the assay method can reliably differentiate from background.

Specificity: The ability of the assay to determine the analyte unequivocally in the presence of components that may be expected to be present in the sample.

System suitability: System suitability is comprised of quantitative or qualitative criteria that assure that a validated assay performs as expected in day-to-day assay runs. System suitability tests are used to verify that the reliability of the system is adequate (in other words, remains in a validated state) for the analysis to be performed. This quality control system alerts the analyst when the assay is out of specification and assures that the results obtained in a given run are meaningful.

Titer: It is the reciprocal of the highest dilution of the sample that tests positive in the method. It is a common practice to express titer as the common logarithm of the highest dilution of the sample that tests positive in the method.

Validation: It is the confirmation via extensive laboratory investigations that the performance characteristics of an assay are suitable and reliable for its intended analytical use. It describes in mathematical and quantifiable terms the performance characteristics of an assay.