

# ⟨1106⟩ IMMUNOGENICITY ASSAYS—DESIGN AND VALIDATION OF IMMUNOASSAYS TO DETECT ANTI-DRUG ANTIBODIES

## INTRODUCTION AND SCOPE

Anti-drug antibodies (ADA) can be induced when animal or human immune systems recognize a protein drug product as foreign. The administration of biopharmaceuticals can elicit product-specific ADA, and various types of ADA responses can develop in either nonclinical or clinical studies. [NOTE—A list of regulatory documents, white papers, and other relevant references is contained in the *Appendix*.] Although the main focus of this chapter is ADA immunoassay design and validation, the chapter also includes discussion of an overall risk-based immunogenicity assay testing strategy that includes preclinical and clinical studies.

ADA assay results are directly influenced by assay design, assay reagents, how the assay is run, what samples are run in the assay (timing of sample collection, etc.), and how assay data are analyzed. In fact, it essentially is impossible to compare the results of studies that use different ADA assays. Guidance, such as this general information chapter, recommending best practices and considerations for ADA assay development helps ensure that the assays produce results that are meaningful for patient safety and product efficacy.

The primary concern with unintended or unwanted immunogenicity of biological products is whether antibodies produced by patients receiving the product lead to some form of clinical response (e.g., an effect on safety or efficacy). The utility and interpretation of preclinical toxicology studies also can be influenced by the presence of ADA.

The pharmacokinetics (PK) or the pharmacologic activity of the drug can be altered by ADA that either enhance or reduce the clearance of the drug, alter bioavailability, or inhibit or exacerbate the pharmacological action of the drug. If an endogenous counterpart of a drug exists, ADA that inhibit the activity of the product also can bind to and cross-react with an endogenous protein counterpart of the product, potentially leading to a deficiency syndrome. Under some circumstances, ADA can form immune complexes that can induce serum sickness-type clinical responses. Moreover, IgE isotype ADA responses can result in anaphylaxis.

Immunogenicity assessments are playing an increasing role in biopharmaceutical development as part of Product Quality Risk Assessments (PQRAs) and the assessment of the criticality of quality attributes (as described in ICH Q8 and Q9). They also play a role in the demonstration of process and product comparability after manufacturing process changes. Often the manufacturing process for a biological therapeutic will be refined during clinical development, and often changes occur after the sponsor obtains marketing authorization. Such changes, however minor, potentially could affect the bioactivity, efficacy, or safety of a biopharmaceutical, and immunogenicity is a key consideration. Changes in the levels and types of degradation products (oxidized, deamidated, aggregates, or others), isoforms of the protein, and process-related impurities such as host cell protein and DNA could affect immunogenicity and warrant closer evaluation.

## FACTORS THAT AFFECT THE IMMUNOGENIC POTENTIAL OF A THERAPEUTIC PROTEIN

Many factors can influence whether administration of a biological product will induce an immune response in the recipient, including the structure of the protein product itself, product variants, product formulations, the immune status and genetic makeup of the patient, and the dosing route and regimen used in the clinic.

### Protein Structure

The primary amino acid structure of the product and its variants can determine if there are immunogenic epitopes that the patient's immune system recognizes as foreign, leading to an immune response. Amino acid sequences that are not found in human proteins and thus could be recognized as foreign by the human immune system (e.g., those derived from a nonhuman cell line or created by protein engineering, e.g., fusion proteins) not surprisingly can induce an immune response in humans. In addition, chemical modification of amino acids (e.g., oxidation or deamidation) may result in a sequence that can stimulate an immune response, although few data to date have confirmed such occurrences following administration of therapeutic proteins. Truncation of the protein could expose amino acid motifs (neoepitopes) not normally exposed in the native protein, stimulating an immune response.

In general, glycosylation does not appear to play a major role in the induction of immune responses to biological products, although nonhuman (e.g., murine) or nonmammalian (e.g., products derived from plants) glycosylation can induce immune responses. An example is a human monoclonal antibody that contained a terminal galactose- $\alpha$ -1,3-galactose because of posttranslational modification by its murine production cell line. This antibody was antigenic and caused severe hypersensitivity reactions in presensitized individuals bearing cross-reactive IgE. There was no evidence that this glycan induced primary immune reactions in naïve (i.e., not presensitized) individuals. Such examples are rare. In fact, in many cases complex carbohydrates may prevent or reduce the antigenicity of immunogenic proteins by shielding epitopes from binding antibodies.

Aggregated protein has been shown to induce immunogenicity in animal models. Proposed mechanisms of action include either presentation of high molecular weight, repeating subunits (multimers) directly to B-cells thereby inducing a T-cell-independent response or by inducing a T-cell-dependent response via enhanced antigen presentation. Addition of polyethylene glycol (PEG) molecules (PEGylation) to recombinant therapeutics has been attempted in order to attenuate immunogenicity and antigenicity of recombinant proteins by limiting exposure of epitopes. Although no clear evidence establishes that the immunogenicity of proteins is diminished by PEGylation, some clinical data suggest that PEGylation can limit antibody binding (i.e., antigenicity) to the protein backbone. PEG itself can be immunogenic and anti-PEG antibodies should be monitored, especially in subjects with known PEG hypersensitivity. In fact, a background level of anti-PEG antibodies is present in the general

population, leading to the requirement to develop an anti-PEG antibody assay as well as an ADA assay during the clinical development of PEGylated products.

## Process-Related Impurities

Process-related impurities such as endotoxins, host cell DNA, or proteins can act as adjuvants and can provoke an immune response by evoking danger signals via activation of immune cellular receptors such as Toll-like receptors. Some postulate that leachables from primary packaging components also can act as immune stimulators or affect the higher-order structure of the protein product and induce an immunogenic response, although firm data are not yet available.

## Immune Status of the Patient

The ability of the patient's immune system to recognize and respond to a protein product can dictate the level of immune response. Patients who are taking immunosuppressive drugs such as glucocorticoids, cyclosporine, or methotrexate may have a lower likelihood of immune response to a protein product despite its immunogenic potential. Conversely, autoimmune diseases and inflammation may involve the overactivation of an immune system so that a product's level of immunogenicity may be much greater than one would anticipate. The pharmacologic activity of the protein therapeutic itself should also be taken into account. Some protein therapeutics directed against B-cell antigens can deplete peripheral B-cells. Conversely, other protein therapeutics may have immune-modulatory activities, e.g., altering patterns of T-cell trafficking. These activities may affect an individual patient's or a patient population's ability to mount an immune response to a protein product.

A patient's immune status should also be considered when the patient exhibits specific pre-existing ADA, cross-reacting ADA (for example, in the case of murine- or plant-derived products), or antibodies against production cell line-related impurities that might induce a clinical response.

## Genetic Background of the Patient

Molecules that recognize and present protein-derived peptides to the adaptive immune system—the human leukocyte antigen or major histocompatibility antigen molecules—show considerable genetic diversity between individuals and between populations in different parts of the world. This diversity is one reason why different patients may have different immune responses to the same product. Consequently, if clinical studies include only a population of limited genetic diversity, then the immunogenicity profile of a protein therapeutic in that population may not reflect its immunogenicity profile in the larger, more diverse population that would be exposed to product after approval.

## Dose and Route of Administration

The way a therapeutic protein product is used can influence its potential for immunogenicity. Different routes of administration appear to have different effects on immunogenicity, and subcutaneous injection generally is perceived to be a more immunogenic route of exposure than is intravenous administration. The dosing regimen also can influence immunogenicity. A protein therapeutic that typically is administered one time as a single dose (e.g., a thrombolytic protein) is less likely to induce an immune response than is a protein therapeutic that is used in a multidose regimen because the immune system usually requires a prime followed by a boost to ensure a robust response. Patients may have pre-existing sensitization even without any known exposure to a therapeutic protein product, and they may exhibit adverse clinical responses on first exposure. However, products with long half-lives or those that are particularly immunogenic (e.g., those with multiple T-cell epitopes) have been shown to induce ADA after a single injection. A chronic dosing regimen has a greater chance of inducing an immune response because the immune system receives multiple exposures to the product and this can lead to a strong memory T-cell response.

Another regimen that has caused ADA with clinical sequelae is one whereby a product is given for a short period of time, stopped, and then introduced again only after a long lag period. This has the effect of priming the immune system, and the reintroduction can cause immune-related events such as allergic responses. The use of chronic dosing on a regular basis, although it repeatedly provides the drug product, appears to avoid such hyper-responsiveness by inducing some form of tolerance.

## DETERMINATION OF PRECLINICAL AND CLINICAL IMMUNOGENICITY

### Preclinical: Relevance and Scope of Preclinical Immunogenicity

The immunogenicity of many biotherapeutics is greater in preclinical studies and has low predictive value for humans because an immune response to human or humanized proteins tends to be greater in animals than in humans because of the perceived foreignness of the protein in animals and the absence of an endogenous biological counterpart. Even though detection of ADA in animals may not be clinically relevant, researchers must assess ADA for the interpretation of the required toxicity data necessary for regulatory submissions (see ICH S6R1 in the *Appendix*).

Generally ADA assessments, PK data, and pharmacodynamic (PD) data aid in the interpretation of the results and validity of animal toxicology studies. In some instances preclinical immunogenicity data also can be used to compare the relative immunogenicity of products before and after manufacturing changes, although only with the grossest of changes in relative immunogenicity does this appear to be meaningful. Preclinical studies, particularly in higher animal species, are typically not statistically powered to draw conclusions on relative immunogenicity but, when they can, it should be recognized that differences in MHC restriction and natural immune tolerance of animals versus humans may not permit translation of such

information to effects in humans. Furthermore, it should also be acknowledged that strains of preclinical species are limited in genetic diversity compared to the human population.

ADA may alter exposure to an active drug by blocking the therapeutic agent from binding to the target or by accelerating the clearance of the therapeutic agent from circulation, resulting in reduced exposure. ADA may increase the half life of biologic drugs and with this the exposure if the drug-ADA complex is still bioactive. Typically, samples should be collected for possible ADA analysis from all preclinical safety studies in which animals are exposed to the drug for more than 7 days (see references in the *Appendix* for more information). Because the key consideration for including immunogenicity analysis in nonclinical studies is to demonstrate the exposure to the drug for the duration of the study, this also can be achieved by demonstrating drug-mediated modulation of a PD marker. In addition, an absence of an effect on the serum concentration or PK profile of the drug also can indirectly demonstrate the absence of detectable effects of an ADA on drug exposure. However, in some situations the development of an ADA may not significantly affect the clearance of the drug, but instead the ADA may affect the drug's binding to and activity at the target (e.g., anti-idiotypic antibodies). Therefore, lack of an effect on a PD marker and/or on PK also should be considered to ensure activity is not affected by the ADA.

For most nonclinical studies, samples from various phases of the study should be collected, banked, and analyzed with regard to any observed pharmacological or toxicological changes. In fact, most nonclinical toxicology studies do not evaluate the kinetics of ADA development, and samples for the assessment of ADA usually are taken at baseline, end of treatment, and end of recovery periods.

Analyses of samples taken during the dosing phase also can be performed if unusual PK data or toxicological findings are observed at the end of the study. In this case it needs to be taken into consideration that analysis of samples taken during the dosing phase may be complicated by drug interference. Therefore, it is important to take samples at appropriate time-points (e.g., before the next dose) and to have suitable assays with high drug tolerance in place. In some instances when a soluble target may be present in the circulation, understanding of ADA level (titers or relative mass units) can facilitate the interpretation of toxicological findings.

A risk-based approach should be used to determine if further characterization (e.g., demonstrating neutralizing activity) is necessary for preclinical studies. Factors such as the presence of endogenous counterparts, utility of a PD marker, or the PK assay may affect this decision. If data from a neutralizing antibody assay (NAb) are deemed necessary, then this assay could be either a target-binding inhibition immunoassay or cell-based assay, irrespective of the risk level. Immunogenicity evaluation and study data interpretation typically require serial sampling and analysis of serum samples for PK, PD, and ADA. Such repeated and frequent sampling may not be feasible when researchers conduct studies using rodents, particularly mice. In such cases, the study can be designed to allow discrete analyses of toxicity, PK, PD, and ADA endpoints from similarly treated cohorts of mice with sample collection at similar study time points to allow inferential analysis of effects observed across treatment groups.

## Clinical: Relevance and Scope of Immunogenicity Assessments

In clinical studies, ADA detection and characterization is important to understand the safety and exposure and efficacy profile of the therapeutic. Typically, the ADA analysis strategy in clinical studies involves a screening assay for binding antibodies then a confirmatory assay, followed by further characterization in NAb. Immunogenicity data from clinical studies generally are analyzed in the context of their relevance to the PK and PD of the therapeutic and on adverse events. For replacement therapies (e.g., enzyme-replacement therapies, blood clotting factors, and erythropoietin), a comprehensive monitoring program should be designed based on ADA detection combined with multiple safety parameters to monitor the potential for serious adverse events.

Researchers should consider the kinetics of the appearance of ADA during clinical studies because this can affect not only ADA detection but also the potential clinical sequelae. Some products induce antibodies rapidly, but other biotherapeutics can take years before an immune response is detected or can be correlated to any clinical sequelae. Investigators also should consider whether an antibody response is transient or persistent. Therefore, understanding the kinetics of ADA appearance is important. This goal can be achieved by carefully selecting ADA sample collection times and taking care in developing the ADA assay for clinical studies. Samples should be collected during each phase of the study (pretreatment, during treatment, and during any washout phases). The sensitivity and drug tolerance of the ADA assay also must be appropriate for the intended use of the assay. For example, when products have long terminal half-lives (e.g., monoclonal antibodies) scientists should develop ADA assays that are capable of detecting ADA in the presence of product levels that are expected to be present in patient test samples. In addition, during the design of ADA sampling plans, researchers should consider the appropriateness of obtaining samples after a washout period.

The number of patients assessed for ADA in clinical trials and the duration and timing of ADA sample collection are critical factors to understand the incidence and clinical impact of ADA. Other factors may confound ADA analysis, including the nature of the therapeutic itself, the presence of pre-existing cross-reactive antibodies, rheumatoid factors, heterophilic antibodies, soluble targets, or ligands. Samples should be taken to assess the levels of these interfering factors in serum (and other PD markers as applicable) at the same time as ADA samples.

Besides determining the presence of binding antibodies, clinical immunogenicity assessments typically include further characterization of positive samples in titer assays as well as in NAb to determine the potential effect of ADA levels to neutralize the drug's effect or to mediate safety events. In addition, understanding the kinetics of ADA and NAb development by detection of ADA in sequential samples taken throughout the study phases and elucidation of the ADA immunoglobulin class(es) and subclass(es) may aid in better understanding of patient- and treatment-related factors and the mechanisms by which the therapeutic induces ADA development.

## RISK-BASED APPROACH TO ASSESSING IMMUNOGENICITY AND ITS CONSEQUENCES

### Assessing the Potential Risk of Product-Specific Immunogenicity

The concept of risk is defined in ICH Q9 as the combination of the probability of occurrence of *harm* and the *severity* of that harm. In relation to pharmaceuticals, the protection of the patient by managing the risk to safety should be considered of prime importance and thus risk assessments of product-induced immune responses should focus on the potential severity of clinical consequences from ADA responses rather than the probability of occurrence of ADA responses. A few patients with severe or life-threatening ADA-related side effects are of more concern than many ADA-positive patients who have no clinical consequences. Risk mitigation should also be factored into the overall risk-assessment process (e.g., elimination of clinical impact by co-medication).

Although ADA testing strategies can be based on immunogenicity risk assessment, this may not always be feasible during early drug development when reliable assays may not be available.

ADA-induced safety events can range from mild side effects to life-threatening conditions. The potential severity of the consequences of an ADA response should be considered as early as possible. *Table 1* summarizes some but not all of the risk factors that may influence the severity of clinical consequences from an ADA response.

**Table 1. Factors That May Influence the Severity of ADA-Related Clinical Sequelae**

|  | Lower Risk <sup>1</sup>                               | Medium Risk  | Higher Risk   |
|--|---|--|---|
| Presence of an identical endogenous counterpart (includes families of proteins that share domains) | No endogenous counterpart (e.g., Botulinum toxin)     | Redundant counterpart (e.g., interferons or growth hormones)     | Nonredundant counterpart [e.g., erythropoietin or megakaryocyte growth and development factor (MGDF)] |
| Presence of a structurally related endogenous molecule or domain                                   | No structurally related endogenous molecule or domain | Medium structurally related endogenous molecule or domain exists | Highly structurally related endogenous molecule or domain exists                                      |
| Patient's disease state  | Life-threatening                                      | Moderate-to-severe   | Mild  |
| Potential clinical consequences of immunogenicity  | No clinically meaningful impact on safety or efficacy | Manageable impact to safety and/or efficacy                      | Extensive clinical influence on safety or loss (or dramatic increase) of efficacy                     |

<sup>1</sup> Risk is a consequence of both severity and frequency.

A number of factors may contribute to the incidence of an ADA response, and some but not all of these factors are shown in *Table 2*. During an immunogenicity risk assessment the factors shown in *Table 2* should be considered in conjunction with those in *Table 1*. The clinical consequences of immunogenicity are unpredictable, even with the risk assessments outlined above.

An immunogenicity risk assessment is of real value only when all the factors that influence the likelihood and severity of a potential immune response are carefully considered. Risk assessments should be done in a cross-functional manner, including input from clinicians, safety assessment, PK, bioanalytical scientists, as well as process scientists. Consultations with regulators and clinical safety monitoring boards also may be helpful and should be carried out iteratively during the product development process as clinical data are obtained.

ADA-mediated clinical sequelae and ADA incidence rate are separate entities, but the two factors are interrelated because the number of patients with ADA-mediated serious adverse events may rise with a higher ADA incidence rate.

**Table 2. Factors That May Influence ADA Incidence**

|   | Lower Incidence   | Medium Incidence   | Higher Incidence  |
|---|---|--|---|
| Circulating level of endogenous counterpart | Abundant  | Scarce   | None  |
| Patient's immune status                     | Suppressed  | Normal   | Activated   |
| Exposure: dosing regimen or frequency       | Single dose   | Chronic (maintenance)<br>Multiple dosing   | Episodic dosing   |
| Route of administration                     | Intravenous or oral   | Subcutaneous, intramuscular, mucosal (non-inhalational)                                      | Intradermal or inhalational   |
| Product Characteristics                     | Lowest levels (or absence) of product or process-related impurities (e.g., aggregates or denaturation, fragmentation) | Intermediate levels of product- or process-related impurities, and potential epitope content | Higher levels of product- or process-related impurities (e.g., aggregates; denaturation, fragmentation)     |
|   | Molecular integrity of active substance maintained  |  | Higher level of antigenic epitopes (derived from murine cell lines, contains novel mutated sequences, etc.) |
|   | No or low content of potential T-cell epitopes  |  | Mechanism of action: immune activation  |
|   | Mechanism of action: e.g., immunosuppressant  |  |   |

## Risk-Based Approach to ADA Testing Strategy

The ADA testing strategy should be based on an immunogenicity risk assessment. Appropriately designed, validated, and executed immunogenicity assays and testing schemes provide the data that make risk assessments possible and predict the eventual outcome for patients. The frequency of sampling, neutralizing activity assessments, and qualitative or quasi-quantitative measurements may all depend on perceived risk.

In clinical studies, patient safety is of primary concern, and the extent of ADA characterization necessary depends on the potential risk of ADA-related sequelae. The type of drug should also be taken into account. For example, for a multicomponent fusion protein that contains at least one component with a potentially high risk of adverse events, a domain-mapping method (i.e., reactivity of ADA with individual components) is recommended.

Generally, more extensive ADA testing and characterization should be applied if the risk of clinical adverse effects is high. Samples should be analyzed and characterized based on the timing and incidence of the ADA response as well as the occurrence and severity of clinical side effects. A higher risk of ADA incidence normally does not warrant extensive characterization of ADA, and usually the risk of clinical consequences drives the bioanalytical strategy. Nevertheless, some investigations may be driven by the need to understand the cause of a high ADA incidence (e.g., the reactivity of ADA with aggregated versus nonaggregated drug).

The following step-wise, risk-based testing strategy can be refined depending on the product's level of risk and during the design of the clinical studies to ensure that the maximum amount of data can be gained, including correlations to PK, PD, etc.

### STEP 1

Develop ADA methods that are fit for purpose and are consistent with current industry best practices and regulatory guidance. Incorporate baseline and postdose ADA testing into the clinical study design, together with concurrent testing for PK plus any relevant PD, safety, or efficacy markers that will facilitate interpretation of ADA data. The analysis of results from ADA testing should be built into study analysis plans.

### STEP 2

Test all pre- and post-dosing samples for ADA. Two important tests that should be carried out in all cases are the ADA screening assay and the drug inhibition or immunoglobulin depletion confirmation assay. Report as negative any ADA results below the assay cut-point and with drug levels below the interference levels, as well as those that test negative in the confirmatory assay. Test methods that are capable of sensitive ADA detection despite the presence of trough levels of drug are desirable. In their absence, samples containing drugs above the interference levels should be reported as inconclusive with a statement of possible drug interference. In such cases, ADA analysis could be performed later following a drug washout period to reach a conclusive result (refer to the section *Relative Sensitivity* of this chapter for further information). For the confirmed positive samples, ADA levels should be estimated, preferably by titration, but they can be reported in terms of relative mass units. Certain mass-based technology platforms may necessitate the use of relative mass units.

### STEP 3

Samples deemed positive in *Step 2* should be tested for neutralizing ability and potentially other characteristics, depending on the risk assessment. In high-risk situations, NAb activity should be measured, typically using a cell-based assay. Depending on the drug's mechanism of action, sometimes a ligand-binding NAb assay format can be used if it is adequately proven to specifically detect NAb. Concurrently generated PK/PD/safety or biomarker data should be used to help interpret the clinical relevance of neutralizing antibody activity. In addition, determination of ADA isotypes and affinity may be helpful in the overall assessment of the immune response. Allergic reactions associated with drug administration may necessitate measurement of drug-specific IgE, although detection may depend on the sampling scheme and method design.

## DESIGN OF IMMUNOASSAY-BASED TEST METHODS

Immunoassay methods for ADA detection generally are complex and require a broad understanding of multiple technical challenges. Screening assays that serve as the key first step in the immunogenicity testing scheme are designed to have a certain false positive (rather than false negative) rate in order to maximize the sensitivity for detecting ADA. In addition, using a risk-based approach, it is more appropriate to have 5% false positives rather than false negatives during this initial screening phase. Typically, positive screening samples are confirmed to contain drug-specific ADA in confirmation assays before the determination of the level of ADA (titers) or any additional characterization.

### Screening Assays

In their simplest form, screening ADA assays immobilize the therapeutic protein on a microtiter plate or onto beads to capture ADA (solid phase) or co-incubate a labeled therapeutic protein at a predetermined concentration with the sample containing ADA (solution assay). The bound polyclonal ADA is then detected using a labeled secondary reagent or labeled drug. Limiting the number of wash steps or reducing wash fluid dispensing rates may increase the detection of antibodies with fast off-rates on assay platforms that use wash steps. Generally, screening assays are designed to detect classes of antibodies that may be most relevant to the product's route of administration, e.g., IgA for mucosal routes of administration. Although the most common ADA raised against protein therapeutics are of the IgM and IgG isotypes, other isotypes of ADA including IgE and IgA may require detection based on the clinical response and the route of product administration. In addition, depending on how

rapidly ADA responses develop and the half-life of the therapeutic, it may be feasible to detect the development of ADA initially of the IgM isotype that later affinity matures to an IgG isotype following repeat administration of product.

Because screening assays serve as the first step in the immunogenicity testing program, these assays generally are configured to have moderate throughput and often are automated. The various technology platforms used to develop screening immunogenicity assays have inherent strengths and weaknesses as outlined in *Table 3*. Development of a bioanalytical strategy to use a certain technology platform for assay development should take into consideration the nature of the product (e.g., a therapeutic protein or a monoclonal antibody), potential sources of interference in the assay (e.g., therapeutic concentration anticipated in patient samples, soluble target based on co-medications, and biology), and disease-specific interfering or cross-reacting factors (e.g., rheumatoid factor).

**Table 3. Advantages and Disadvantages of Various Assay Types Currently Used to Assess ADA**

| Assay Type   | Advantages  | Disadvantages  |
|--|---|--|
| Direct/Indirect ELISA                                    | High throughput   | Potential for high background  |
|  | Readily available   | May not be specific  |
|  | Easy to automate  | Utility depends on capability to detect different Ig subtypes  |
|  | Inexpensive   | Drug tolerance lower for solid-phase ELISAs  |
|  | Readouts can increase sensitivity (e.g., electrochemiluminescence)                    | Excessive washes can decrease detection of low-affinity ADA  |
|  | Drug tolerance higher for solution-phase ELISAs                                       |  |
| Bridging Format  | Low background  | Difficult to confirm presence of IgM   |
|  | Highly specific   | Need to label product  |
|  | Readily available   | Reduced ability to detect low-affinity antibodies and IgG <sub>4</sub> because of single-arm binding between bound ligand and detector |
|  | Easy to automate  |  |
|  | Inexpensive   |  |
|  | Format can be used across species and detects all isotypes                            |  |
| Surface Plasmon Resonance/Biolayer Interferometry Assays | Can be used across detection platforms (colorimetric, electrochemiluminescence, etc.) |  |
|  | Flexibility to characterize immune response (concentration and relative affinity)     | Expensive technology   |
|  | Higher drug tolerance and can detect low-affinity antibodies                          | Limited suppliers  |
| Radioimmunoprecipitation Assays                          | Inexpensive   | Moderate throughput  |
|  | Highly sensitive  | Radioactive waste  |
|  | Better for high-affinity antibodies   | Need to frequently requalify radiolabeled reagents because of short half-life  |
|  |   | Utility depends on capability to precipitate all relevant antibody present   |
|  |   | Can be less useful for low-affinity antibodies   |

In addition, analysts should consider whether the method does not result in an underestimation of ADA-positive samples because the ADA binding epitopes on the capture reagent are blocked with a tag or by coating on plates or beads. Another point for consideration in the design and development of an ADA assay is the adaptability of the test for both nonclinical and clinical matrices. Although most assay formats can transfer readily from nonclinical to clinical use, there are exceptions. Further, clinical ADA assays should be qualified or validated for use with samples collected from a similar patient population. Here again, although most ADA screening assays may not show unique matrix interferences between one disease matrix and another, there may be exceptions. Immunoassays used for ADA detection generally are quasi-quantitative methods because standardized, species-specific (especially human) polyclonal ADA calibrators generally are not available. The positive controls, typically developed in-house as hyperimmune serum in animals or by phage display serve as surrogates for the drug-induced ADA in treated patients.

As depicted in *Table 3*, some of the more common assay formats currently in use for development of screening assays include plate-based or bead-based enzyme-linked immunosorbent assays (ELISA; see also USP general chapter *Immunological Test Methods—Enzyme Linked Immunosorbent Assays (ELISA)* (1103)) with colorimetric, fluorometric, or luminescent read-outs, plate-based or solution-phase electrochemiluminescent (ECL) or ELISA assays, surface plasmon resonance assays (SPR; see also USP general chapter *Immunological Test Methods—Surface Plasmon Resonance* (1105)), or biolayer interferometry assays, and radioimmunoprecipitation assays (RIPA). In order to differentiate positive from negative responses, assay cut-points are statistically determined using samples collected from the target population. The assay cut-point also helps to determine the assay sensitivity. An incorrectly established cut-point may result in false negatives or too many false positive responses that should be ruled out as drug-specific responses in the confirmation assay. Assay performance typically is optimized during development by evaluation of the following parameters: sensitivity (lowest amount of detectable antibody in a sample demonstrated using surrogate controls); specificity (likely detection of a true positive rather than a nonspecific interaction);

precision (reproducibility of results from multiple analyses); interference (interfering substances in sample matrix, including the administered drug, that affects assay sensitivity); and stability and robustness (likelihood of optimal assay performance over time). After optimizing these parameters, analysts typically validate the method for its intended use. If the initial assay cannot meet the performance goals (e.g., because of poor sensitivity or high backgrounds), then analysts should either improve and validate the first assay format again or develop and validate more than one assay format.

## Confirmatory Assays

Samples that are positive in the screening assay usually are confirmed in a second assay that includes adding a certain fold excess of the therapeutic. This is intended to demonstrate that a positive signal seen in the ADA screening assay is caused by the presence of drug-specific antibodies. Because the cut-point for the screening assay is set to result in the detection of approximately 5% false positives, the confirmatory assay is used to rule out the false positive samples from further analysis. Multiple options are available for performing confirmatory assays. Usually a soluble drug is added to the sample and should compete with the plate-bound immobilized drug for binding to sample ADA. A specific interaction to a soluble drug results in a decrease in the assay signal. As with the screening assay, the cut-point for the confirmation assay is established statistically. Verification of the presence of drug-specific antibody can also be performed using an orthogonal method on a different assay platform that may have different nonspecific binding profiles. Analysts should take care while adopting this approach to ensure an adequately sensitive assay in order to avoid a false negative reaction. Finally, the specificity of ADA to a drug can also be confirmed by depleting all immunoglobulin from a sample (e.g., using a Protein A or G column) followed by reanalysis of the depleted sample. In the latter approach, the depleted sample scores negative in the assay if an antibody caused the original signal. Validation of the confirmatory assays helps ensure that the results from the confirmatory assay are appropriately interpreted.

## Characterization Assays

Following screening and confirmation, the relative level of ADA in a positive sample is assessed by titration. The most common approach is to serially dilute the sample and report the reciprocal of the highest dilution factor at which the sample tests positive, or the titer of the sample. The higher the sample dilution (and therefore the ADA titer value), the higher is the concentration of circulating ADA in that sample. This approach has been used historically to report serological data from vaccine studies. Another less frequently used approach is to express the amount of ADA in the sample in mass units relative to a surrogate standard. Although this approach has the advantage of relating the amount of antibody in the sample to assay sensitivity, analysts should recognize that the calibrator may not be representative of the polyclonal ADA responses under measurement.

Traditional approaches to dilutional linearity testing do not apply to ADA assays. However, when analysts express ADA data in terms of titer values, they also should demonstrate that the positive control(s) displays reasonable (relative) linearity of dilution.

In addition to performing titration, analysts routinely characterize positive ADA samples in neutralization assays to determine the *in vitro* effect of ADA that might reflect the *in vivo* biological or pharmacological activity of the therapeutic product. Additionally, the isotype(s) of the ADAs also can be analyzed. Isotype(s) identification is sometimes performed in a multiplexed manner. Using one fluorescent multiplex platform, analysts mix each sample with multiple secondary reagents that are specific to different immunoglobulin isotypes and are labeled with unique fluorochrome labels.

An SPR-based platform also can be used for this purpose (also see *Immunological Test Methods—Surface Plasmon Resonance* (1105)). The isotype of an ADA can be determined by observing binding of isotype-specific reagents (such as an anti-human IgG<sub>1</sub> Ab) to the ADA that has been captured by immobilized drug. Analysts should take care when identifying and validating the isotype-specific reagents because unexpected cross-reactivity often is observed. Isotyping can help understand the maturation of the immune response. For example, an ADA response that is comprised solely of IgM antibodies is an immature immune response without T-cell involvement and may or may not progress further. In contrast, an immune response comprised of IgG<sub>1</sub> and IgG<sub>4</sub> antibodies represents a more mature response that has already engaged more components of the immune system. ADA titration and characterization assays are validated routinely using many of the same parameters as screening and confirmation assays to ensure consistent assay performance.

## VALIDATION OF IMMUNOASSAYS

Method validation is a process of demonstrating, by the use of specific laboratory investigations, that the performance characteristics of an analytical method are suitable for its intended use (see also USP general chapter *Validation of Compendial Procedures* (1225)). The level of method validation depends on the stage of product development and the risks associated with the product. A partial validation involving assessments of method sensitivity, specificity, and precision requirements with less emphasis on robustness, reproducibility, and stability may be adequate for the earlier stages of clinical development (Phase 1–Phase 2 studies), whereas fully validated methods are required for pivotal and postmarketing studies.

Validation of an assay before use of the method for sample bioanalysis is called *pre-study validation*, and amendments to this process may be made between studies. This process maps out the performance characteristics of the assay and should demonstrate that the method is suitable for its intended purpose when it is subsequently applied to study samples. In contrast, *in-study validation* refers to the monitoring of assay performance during study-phase applications of the assay in order to ensure that the assay remains valid and that the resulting bioanalytical data are reliable.

Reliable performance of the assay also depends on all the elements spanning bioanalytical testing and data manipulation, such as assay reagents, analysts, equipment, and computer programs. In essence, the assay is a system comprising several elements other than assay steps and reagents alone. Pre-study validation therefore establishes system suitability (establishment of criteria for control samples that are used to accept or reject runs and imprecision limits for individual samples), and then

in-study validation continues to verify it. Critical changes to methods often require additional validation (partial or full), sometimes leading to the revision of the system suitability criteria.

### Minimum Required Dilution

During assay development the minimum required dilution (MRD) can be defined as the dilution level of the ADA negative sample that results in the highest signal-to-variability ratio (or  $Z'$  factor).

The ability to dilute such samples also should be assessed to ensure that the chosen MRD is adequately distal to any prozone (hook) effects that may have been observed. Although they are rare, some unusual prozone effects may require the test method to include more than one dilution of a test sample to minimize false negative data.

The MRD should be evaluated during the assay development/design phase, i.e., before analysts initiate the validation experiments, so analysts do not need to repeat the evaluation during validation. It can be established using 10 individual drug-naïve ADA negative samples, each tested in 2-fold serial dilutions (e.g., a range of 1:5 to 1:80). The MRD is sometimes defined as the dilution level that results in the highest signal-to-background ratio when typically the background is the dilution matrix. However, this definition ignores the variability in the signal. Therefore it is preferable to define the signal-to-background value and include variability.

One way of doing this is to use a  $Z'$  factor that includes both the intensity of the assay readout and its variability at different dilutions (see the *Appendix* for more information). The  $Z'$  factor for each dilution level is obtained from the formula  $[(\text{mean}(S) - 3 \text{SD}(S)) - (\text{mean}(B) + 3 \text{SD}(B))]/[\text{mean}(S) - \text{mean}(B)]$ , where  $S$  is the assay signal of the diluted sample and  $B$  is the background signal. Thus, the MRD is the dilution that results in a desired value for the  $Z'$  factor and signal-to-background ratio. This metric is widely used for high-throughput screening assays to ensure adequate confidence in the ability to differentiate between truly active versus inactive compounds. An inappropriately large MRD can compromise the sensitivity of an assay.

### Pre-Study Validation

Pre-study validation establishes the following:

1. Assay cut-points
2. System suitability criteria
3. Relative sensitivity
4. Specificity
5. Selectivity/interference
6. Precision
7. Robustness
8. Reproducibility
9. Stability

### Assay Cut-Points

Because of the quasi-quantitative nature of ADA detection methods, the use of a decision threshold or cut-point becomes necessary to discriminate between ADA-positive and -negative samples. Because the screening assay and specificity confirmation assay produce different types of results, separate cut-points are necessary. In some instances, a different cut-point also may be necessary for evaluating titers of confirmed positive samples from the titration assay. Some key points in the evaluation process are summarized briefly below (more information is available in the *Appendix*).

#### SAMPLES FOR CUT-POINT EVALUATION

Samples from an appropriate population should be used for the development of assay cut-points. In some cases it may not be practical or feasible to obtain matrix samples from a population that has a target disease before initiating pre-study validation experiments. Consequently, samples from healthy drug-naïve subjects are used commonly to establish the initial cut-points. This approach is preferred for a conventional Phase 1 study with normal volunteers. When the clinical program progresses beyond Phase 1 and samples from the target disease population become available, it is more appropriate to re-evaluate cut-point data for the target population. If the distribution of assay responses with respect to both the mean and variability are comparable between the target population and the normal volunteers, then the same cut-point can be used. If not, target disease-specific cut-points are more appropriate, and fixed or floating cut-points computed from the data obtained from the baseline samples from a clinical trial can be considered.

#### SCREENING CUT-POINTS

The screening assay cut-point is a signal in the screening assay that identifies a sample that is likely to contain ADA (termed a *screen positive* or *potential positive* sample) versus an ADA-negative sample. A screening assay cut-point is established during pre-study validation based on a systematic and statistically rigorous analysis of assay responses from a panel of individual samples that are considered to be representative of a drug-naïve target patient population.

To determine the screening method cut-points for clinical assays, analysts should use samples from at least 50 drug-naïve individuals for a robust evaluation. If additional indications are targeted, analysts should test at least 20 drug-naïve individuals per indication. If the variability is significantly different from the original indication, then additional drug-naïve individuals may need to be tested. If not, then the original cut-point can be applied to the additional indication. For nonclinical assays a total of at least 25 drug-naïve individuals should suffice. To ensure this cut-point is robust, at least two analysts should perform this



experiment over three days in at least two different plate layouts. A balanced experimental design and plate layouts will help avoid potential confounding between analysts, subject samples, run dates, etc. For clinical ADA assays, if multiple disease-state populations are being tested they should be distributed evenly across the plates to ensure they are properly balanced across plates and runs. Statistical outliers of the sample results should be examined and eliminated, e.g., using outlier box plots defined in terms of quartiles and interquartile range. In addition, confirmed reactive samples (e.g., via immunodepletion) can be excluded as well.

Three types of screening cut-points can be calculated for application during the study phase—fixed, floating, and dynamic—and one of these should be appropriately chosen for study phase bioanalysis.

**Fixed cut-point**—A fixed cut-point is a cut-point that is determined in pre-study validation and subsequently is used for the in-study phase. The fixed cut-point is used for analyses of test samples until there is a need to revalidate or change the cut-point (e.g., because of a critical change in the assay, assay transfer to another laboratory, upgraded instruments, etc.). The cut-point value can be fixed within a given study, for a target population, or across studies and multiple target populations. In order to use this approach, one should statistically demonstrate similar means and variances across runs during pre-study validation. A fixed cut-point can be determined based on the mean + 1.645 SD (standard deviation), which represents the 95<sup>th</sup> percentile of the population under normal distribution (and therefore is expected to identify approximately 5% of the samples as false positives). The standard deviation should include different sources of variation such as the intra-run, inter-run, inter-analyst, and inter-subject variability. If the data are not normally distributed, appropriate transformations (typically log transformations) can be used. If transformation doesn't help, usually it is acceptable to determine the nonparametric 95<sup>th</sup> percentile. However, in preclinical trials it may be considered adequate to use a cut-point at the 99<sup>th</sup> or 99.9<sup>th</sup> percentile because immunogenicity of a protein normally results in high antibody titers. Alternatively, even if the validation data suggest similar means and variances across runs, to account for possible deviations between assay runs during the in-study bioanalysis phase, it would be safer to apply a floating cut-point.

**Floating cut-point**—A floating cut-point is a cut-point calculated by applying an additive or multiplicative normalization factor, determined from the pre-study validation data, to the biological background obtained during the in-study phase (see Appendix G of Shanker et al., 2008, in the *Appendix* for details). The biological background may be represented by the negative control (pool of matrix from subjects that are negative for ADA), the assay diluent, or the predose subject sample (subject-specific cut-point). The method for determining floating cut-point uses the variation estimate from the pre-study validation that includes different sources of variation such as the intra-run, inter-run, inter-analyst, and inter-subject variability. Such a cut-point is recommended when the means of drug-naïve samples are not similar but the variances are similar across runs. When a negative control is used for normalization, analysts also should ensure that the negative control results represent the drug-naïve matrix sample results of the target population. This is accomplished by demonstrating that the signal of the negative control trends directionally with the signal of the individual samples. Alternatively, the use of assay diluent for normalization or pretreatment subject ("baseline") sample results may be more appropriate. However, a pre- versus post-dose ratio might be a better solution.

**Dynamic cut-point**—A dynamic cut-point is a cut-point that changes between the plates in a run, between runs in a study, or between studies, and it does not apply the variation estimates from pre-study validation. The latter characteristic differentiates it from a floating cut-point. This approach is necessary only where means and variances significantly differ between runs. Because this approach entails testing of several individual drug-naïve samples for the evaluation of a run-specific cut-point, it consumes a large portion of the plates from each in-study run and therefore is not practically feasible, especially when analysts use 96-well plates instead of 384-well plates. Differences in variability between assay runs sometimes can be resolved by further optimization of some key steps in the assay protocol or by resolving some analytical issues. In some cases, the differences in variability can be attributed to different analysts or instruments, and use of an analyst-specific or instrument-specific floating cut-point may resolve this issue. If further optimization does not resolve the situation or if the causes are not clear, another practical alternative may be to pool the variability across all runs and use this pooled variance for floating cut-point evaluation during the in-study phase.

## SPECIFICITY CONFIRMATION METHOD CUT-POINT

Because of the conservative approach of incorporating a 5% false-positive rate into the computation of the screening cut-point, the elimination of false-positive samples via confirmation of drug-specific binding is an important component of ADA bioanalysis. It is also important to understand the level, if any, of the drug itself within the sample.

The amount of change in assay signal that differentiates drug-specific binding from nonspecific binding is referred to as the *specificity cut-point*. The specificity cut-point should be determined by an objective approach in the context of assay variability near the low positive range of the assay. To determine the specificity cut-point, drug-naïve negative samples from at least 25 individuals should be evaluated (however, more are commonly used when available), with and without drug preincubation. Ideally these samples should be the same as those tested during the screening cut-point evaluation, and the unspiked and spiked counterparts of the individual subject samples should be tested together in the same plates.

The mean percent change from the unspiked sample (inhibition) and SD are calculated. The mean inhibition plus 3.09 SD (if a 0.1% false positive rate is desired) represents the specificity cut-point. As in determination of the screening cut-point, specifically reactive samples after preincubation with drug (i.e., those that contain pre-existing antibodies) and statistical outliers should be eliminated in order to make the specificity cut-point more conservative. The analytical process outlined above for the screening cut-point applies for the evaluation of the specificity cut-point as well.

Alternative approaches such as the use of *mock low positive* samples in which the individual drug-naïve samples are spiked with a low concentration of a positive control sometimes is considered for this cut-point evaluation. However this method is subjective and is not recommended because it depends on the concentration of the positive control and the unique affinity/avidity of the positive control that may or may not represent true positive patient samples. Additional sources of information regarding the relative statistical merits of these approaches and methods for verifying the assumptions are listed in the *Appendix*.

## TITRATION METHOD CUT-POINT

The titration method cut-point is a test result value below which further serial dilution of an ADA-positive sample produces negative assay results. Typically, the screening assay cut-point is used as the titration cut-point, but the validation of a separate titration method cut-point can become necessary when the signal from the assay diluent or matrix causes higher results than the screening assay cut-point (because of a blocking effect of serum) or if samples at a dilution higher than the MRD do not generate consistently negative results, i.e., when the screening cut-point falls on the lower plateau of the positive-control dilution curve. In such instances, the same data generated during a screening cut-point experiment can be used to define the titration cut-point using a 0.1% false positive rate threshold criterion (i.e., Mean + 3.09 SD). During bioanalysis, confirmed positive patient samples that fall between the screening cut-point and titration cut-point can be assigned a titer value equal to that of the MRD.

## CROSS-REACTIVITY METHOD CUT-POINT

If applicable, ADA-positive samples that are confirmed to be specific to the drug can be further characterized for cross-reactivity to other related antigens. Like the specificity confirmation assay, a cross-reactivity test method may require a preincubation step with and without the related antigen. Cross-reactivity to the antigen is confirmed when the percent inhibition of signal in the presence of the antigen is greater than or equal to the cross-reactivity method cut-point. The methods for determining the cross-reactivity method cut-point are similar to those for the specificity method cut-point, although it also may be acceptable to apply the drug specificity confirmation cut-point.

## Defining System Suitability

### ASSAY CONTROLS

ADA-positive controls can comprise polyclonal or monoclonal anti-idiotypic antibodies. They should be affinity purified and quantitated to enable assay validation. Each run (or plate) should include at least a low level of positive control (*low positive control*) and a negative control, but the inclusion of a higher level control (*high positive control*) also can be useful in monitoring method performance. Tracking all of these controls over time can help ensure that the method is performing suitably. A low positive control helps ensure that the assay remains as sensitive during study phase bioanalysis as during the pre-study validation.

On the one hand, the low positive control should produce a response that can be seen reproducibly above the cut-point, but sometimes it may result in a signal that is below the cut-point (thereby failing or invalidating the assay). On the other hand, choosing an unreasonably high concentration for a low positive control may produce an assay signal that is substantially above the cut-point, which is inappropriate. To provide objectivity to the selection of a low positive control concentration, it is useful to think in terms of assay rejection rates, i.e., the percentage of assays (plates) that fail because the low positive control produces a result below the cut-point. As an example and in order to understand if the low positive control is sufficiently low, a 1% rejection rate may be a reasonable target for a low positive control. This is calculated as  $\text{mean} + t_{0.01,df} \times \text{SD}$ , where mean and SD are determined using the data from the sensitivity experiment or related assay development data, and  $t_{0.01,df}$  is the critical value determined from the *t*-distribution corresponding to a 1% false positive rate and *df* is the degrees of freedom that depends on the number of samples and runs used in the calculation. This theoretically implies that about 99% of the data from the low positive controls will be at or above the cut-point.

An optional high positive control can be useful for methods that are prone to hook effects, tracking assay performance, reagent qualifications, and troubleshooting. The concentration of the high positive control should be chosen from the upper end of the linear range of the dilution curve, usually just below the upper plateau of the curve.

### SYSTEM SUITABILITY CRITERIA

System suitability criteria using assay controls help ensure that an analytical procedure remains valid for use. Acceptance ranges (system suitability criteria) for quality controls should be established by statistical evaluation of the experimental data acquired during assay validation.

When the floating cut-point approach is deemed necessary and is used for the screening cut-point evaluation, the system suitability criteria or limits can be defined for the ratio of the low positive control to the negative control and for the ratio of the high positive control to the low positive control or a negative control instead of defining limits separately for each positive control. It is also useful to apply acceptance criteria for intra-assay precision (variability of signals of replicates in an assay) for the in-study phase. Although data from assays that fail acceptance criteria during the in-study phase should be rejected, setting criteria for passing or failing assays in pre-study validation experiments should be avoided because these potentially can lead to the exclusion of some validation data, resulting in an inaccurate estimate of analytical error. All assays during pre-study validation should be included, and the only exceptions should be those rejected for an assignable cause (e.g., technical error).

## Relative Sensitivity

No ADA-positive control can be expected to represent the spectrum of humoral immune responses observed in individuals treated with study compounds. The sensitivity of ADA assays is highly dependent on the nature of the positive control reagent(s) so that high-affinity positive controls often produce better sensitivity values than lower affinity positive controls in the same assay. Analysts should consider this when they choose controls, as well as when they estimate assay sensitivity. Moreover, because the drug itself can interfere with ADA detection, the sensitivity of ADA detection becomes progressively worse in the presence of increasing concentrations of drug within the sample. Despite these caveats, the determination of assay sensitivity is valuable when analysts choose an optimal ADA detection method or platform, a low positive control for validation, or assess

the suitability of an assay. The assessment of assay sensitivity in the presence of an interfering drug (drug tolerance) is critical for understanding the suitability of the method for detecting ADA in dosed patients. ADA assay sensitivity should be defined not as a single value, but as a set of at least two values: (1) the concentration of positive-control ADA detected within undiluted matrix in the absence of any drug and (2) the concentration of positive-control ADA detected within undiluted matrix in the presence of drug levels expected at the time points when samples for ADA analysis are taken. Assays should, in general, demonstrate a sensitivity of at least 500 ng/mL for methods applied to clinical studies (or 1000 ng/mL for nonclinical studies) to show suitability for intended purpose—that is, for the detection of clinically meaningful ADA, although assay sensitivity should be justified on a case-by-case basis. It is not useful to express sensitivity in terms of antiserum titers, and thus sensitivity should be assessed using monoclonal antibody or affinity-purified polyclonal preparations. Analysts can evaluate sensitivity by means of two assay runs performed by two independent operators (when feasible) for a total of at least three runs.

To assess sensitivity in the absence of a drug, analysts should prepare mock samples with known concentrations of ADA that are serially diluted (usually 2- to 3-fold serial dilutions) in matrix pooled from drug-naïve individuals and evaluated according to the screening method until the assay results of the dilutions in matrix are below the screening assay cut-point. The lowest concentration of ADA that is consistently found (for example, using a 95% upper confidence limit based on the number of runs or operators) above the screening assay cut-point is determined to be the sensitivity of the assay. Alternatively, it can be the lowest concentration of ADA that is found to be above the screening assay cut-point in all runs by all operators or in 19 of 20 runs (see the *Appendix* for more information).

To assess sensitivity in the presence of a drug, two alternative experimental approaches could be considered: (1) titrate the drug into undiluted matrix containing set concentrations of a positive-control ADA (e.g., 250, 500, or 750 ng/mL). Report the highest concentration of the drug at which ADA remains detectable. (2) Alternatively, because immunogenicity samples often are taken at drug trough time points, prior knowledge of the anticipated trough drug concentration range could be used to determine the assay sensitivity in the presence of the expected concentrations of the drug.

## Specificity

Specificity refers to the ability of a method to detect ADA that specifically binds the drug molecule, its domains, or components. The assay is developed and optimized based on the ability of the positive-control ADA to specifically bind the drug. During validation, results of the specificity confirmation assay support assay specificity.

## Selectivity and Interference

The selectivity of an ADA assay is its ability to identify a positive control in biological matrix samples that may contain potential interfering substances and is an important concern for ADA detection assays. Such matrix effects typically arise from nonspecific binding interactions between a matrix-based factor and the ADA or from specific binding of unknown factors. During validation, analysts assess the selectivity of the ADA assay by looking at the recovery of analyte (represented by a positive control sample) from matrix samples that contain the potential interferent(s). One caveat here is that the selectivity of an ADA assay, as assessed using the positive control, may not reflect the selectivity of the assay when it is used with actual nonclinical or clinical samples.

Interference is the property of a factor (most commonly the drug itself and its target, if soluble) to affect assay results positively or negatively. It should be evaluated using a low positive ADA test sample that is spiked into a sample matrix from drug-naïve patients. Each potential interfering factor should be tested at a physiologically or pharmacologically relevant range of concentrations. The highest concentration of the interfering factor that does not alter the classification of the test sample (e.g., an ADA sample that remains positive relative to the screening assay cut-point) is defined as the tolerance of the assay to that interfering factor. For therapeutics that have a long terminal half-life, the main interferent in an ADA assay is the drug itself. As discussed previously, the drug tolerance of an assay should be interpreted as the sensitivity of the method in the presence of interfering drug.

Other endogenous interferents include oligomeric drug targets, or the target's soluble receptor may interfere with ADA detection. In addition, certain sample pretreatments performed to reduce drug interference can release drug target from drug-target complexes, leading to subsequent interference problems. Hence, analysts should carefully evaluate pretreatment steps such as acid dissociation during assay development to mitigate the risk of inaccurate data.

## Precision

Precision is a measure of the variability in a series of measurements for the same material run in a method. The acceptance criteria for the precision of ADA assays should be within the range commonly expected for immunoassays. These criteria also should be appropriate for the assay platform and should be fit for purpose. During assay validation, precision should be determined in experiments that are run at the level of intended use during the study phase (e.g., number of plates, samples per plate, etc.).

The acceptance criteria for precision should be within the range commonly expected for immunoassays. These criteria also should be appropriate for the platform used, guided by assay development data and experience with the technology platform and assay method. Additional information is found in the *Appendix*.

### SCREENING AND CONFIRMATORY METHOD PRECISION

For ADA screening assays with numeric readouts (as opposed to categorical yes/no readouts), assay precision can be determined using data from at least six independent assay runs of the assay positive controls (low positive and high positive controls). Typically, estimates of intra-assay precision (interreplicate variability, also called intra-assay repeatability) and interassay precision (also called interassay repeatability, or intermediate, total, or overall precision) are reported as percent coefficient of variation (%CV).

Intra-assay precision (repeatability) is the variability of assay results when the same material is tested multiple times within the same run. Interassay precision (also termed intermediate or total precision) is the variability of assay results when the same sample is tested in separate runs, over separate days, and by multiple operators (or only one operator if the study phase bioanalysis is intended to be performed by only one operator). These are expressed as %CV of ADA signals. Data from the replicates of negative and positive controls from each of all the runs tested during the pre-study validation phase are pooled and analyzed within the framework of random-effects ANOVA, resulting in estimates of intra-assay %CV and interassay %CV. Analysts should consider positional effects by varying sample position on microtiter plates because these effects (e.g., edge effects) can influence the assay precision. One should use the same number of test and control sample replicates during validation as are used in the assay during routine use.

Similarly, intra- and inter-assay precision estimates for the confirmatory assay can be derived using the percent inhibition data of the spiked versus unspiked low positive control samples from multiple assay runs (at least six) in the pre-study validation.

## TITRATION ASSAY PRECISION

In order to determine the precision of a titer, two or more analysts should assay serial two-fold dilutions of five or more mock high positive control samples in at least six runs. Mock high positive control samples can be obtained by spiking individual negative sera from the target population with a high positive sample. The titer then is determined by interpolation of each of the dilution curves, and the overall mean and SD are calculated. Then intra- and inter-assay precision (%CV) can be determined.

A recommended but more rigorous approach is to use these data to define a minimum significant ratio (MSR):  $MSR = 10^{t \cdot \sqrt{SD^2}}$ , where SD is the overall standard deviation (intra-run plus the interrune variation) of the titers in common (base 10) log scale, and  $t$  is the threshold from Student's  $t$ -distribution with  $n - 1$  degrees of freedom ( $n$  = number of runs). The calculated MSR reflects the smallest fold-change in the titer values that can be considered as statistically significant ( $P < 0.05$ )—i.e., if MSR = 5, then titers that are different by more than five-fold can be considered significant. In addition to serving as an indicator of the level of variability in the titers of the positive control, this MSR evaluation also can be an approximate criterion for comparing samples with confirmed pre-existing antibodies in baseline versus posttreatment samples in order to assess treatment-induced immunogenicity. The MSR applies only if the titer is interpolated and does not apply to endpoint titers.

## Robustness

Robustness is an indication of the reliability of an assay. It is assessed by the capacity of the assay to remain unaffected by small but deliberate variations in method performance that would be expected under relevant, real-life changes in standard laboratory situations. The choice of robustness variables to test during validation should be based on the knowledge of the assay and its associated risks. Some common variables are microtiter plate lots, incubation times, temperature, and reagent lot and concentrations. Study samples or positive control samples can be used to test assay robustness. The use of acceptance criteria for system suitability controls during robustness validation (computed from the assay development and optimization data or validated system suitability control acceptance criteria) is recommended. Continuous monitoring of an assay during validation and beyond with strict records of key assay parameters (e.g., incubation times, pipetted volumes of critical reagents, operators, etc.) may help identify some of the robustness factor interactions if sufficient data are accumulated.

## Reproducibility

Assay reproducibility according to USP general chapter *Validation of Compendial Procedures* (1225) and ICH Q2(R1) *Validation of Analytical Procedures: Text and Methodology* is the reliability of a method when performed in two or more laboratories. In the context of method transfers and interlaboratory method validity demonstrations, assay reproducibility is the same as a cross-validation.

Reproducibility is applicable only if an assay will be run by two or more independent laboratories during in-study phase bioanalysis. Reproducibility is an assessment of the transferability of an assay, i.e., the validity of testing samples in two or more laboratories and the comparability of data produced by them. Reproducibility assessments do not consider routine changes in an assay such as interequipment or interanalyst imprecision. Such contributors to variability (often referred to as intermediate precision factors) are part of the reproducibility variability.

Study phase acceptance criteria for system suitability controls are established in the originator laboratory (see below) during the original assay validation process. The performance of these controls can be compared across multiple laboratories. When only a single laboratory performs the ADA assay, however, reproducibility need not be validated until the method will be transferred to another laboratory.

## Stability

It is useful to understand the optimal storage conditions for assay samples, controls, materials, and reagents, and they should be investigated as part of assay optimization before validation. Later, during assay validation, stability studies should evaluate assay performance following intended storage conditions. Ideally, stability testing conditions should mimic the expected sample, material, and reagent handling conditions, storage temperature(s), and varying lengths of storage time.

## MATERIAL AND REAGENT STABILITY

ADA assays are stability indicating with respect to the applicable materials and reagents, and thus separate tests for reagent stability usually are not required for assay validation. During study phase bioanalysis, assay materials and reagents are presumed to be stable if the system suitability controls meet validated acceptance criteria. However, analysts should validate the stability of plates that have been prepared in advance (e.g., coated with capture antibody and blocked) and stored.

## SAMPLE HANDLING AND STABILITY

ADA samples typically are collected in a serum or plasma matrix. For samples stored at or below  $-20^{\circ}\text{C}$ , the stability of ADA are universally accepted, so this sample storage condition may not require validation. It is generally accepted that an ADA sample in serum or plasma will be stable after three freeze–thaw cycles and up to 2 years when stored at  $-70^{\circ}\text{C}$ .

## Documentation of Pre-Study Validation

Typically three types of assay-specific documents are created during pre-study validation: an assay validation plan or protocol, an assay method description, and an assay validation report.

An assay validation plan or protocol is recommended before analysts initiate pre-study validation experiments. This document should state the intended purpose of the method, a detailed description of the immunoassay and reagents or materials, a summary of the performance characteristics that will be validated, and a priori acceptance criteria for precision, robustness, stability, and, when appropriate, reproducibility. Some experimental detail and data-handling procedures should be presented in the validation plan because these details provide a clear guidance to the validation analysts and ensure better control over the resulting data.

A method description typically is established after pre-study validation but before the study. This provides a detailed description of the reagents, controls, and equipment needed to run the assay, together with a step-by-step operating procedure and information about processes for data reduction and interpretation. The point at which such a description becomes a Standard Operating Procedure (SOP) is specific to each manufacturer's quality system.

When validation is completed, manufacturers generally conduct a technical peer review of validation data, followed by a validation data audit. An assay validation report is created after the validation work is completed. This documents all of the study validation data, together with information about the methods and batches of reagents that were used. An audited report is approved by management and then is archived.

## In-Study Validation

In-study validation (monitoring the maintenance of system suitability) and revalidation are critical components of any bioanalytical method. Hence, the validation of a method actually does not end until the method is retired from analytical use.

For in-study performance of quantitative bioanalytical methods, acceptance criteria for precision and accuracy generally are required. Because accuracy is *not* applicable for ADA methods, monitoring the performance of quality control samples reassures analysts that the assay system is suitable for its intended use, i.e., that the assay remains valid and is performing as well as it did during pre-study validation.

The use of a low positive control ensures the assay remains sensitive. Generally during study sample analysis the intra-assay (interreplicate) precision of the results of positive controls, as well as test samples (with assay signal at or higher than the screening cut-point), are controlled using system suitability acceptance criteria to ensure that meaningful data are consistently obtained. Results below the cut-point, however, may not be required to meet CV limit criteria.

## LIFE CYCLE MANAGEMENT

Management of the performance of immunogenicity assays from initial clinical development through subsequent product life cycle requires a comprehensive understanding of the strengths, weaknesses, and capabilities of the method format, as well as of the critical assay reagents and assay performance characteristics. In addition, a well-defined plan for critical reagent production, characterization and qualification, qualification of suppliers of critical reagents, and characterization and qualification parameters for reagents produced in-house (aggregate level and labeling efficiency) help manage the risk of maintaining the assay and transferring the method to other laboratories.

When there are changes in critical method components, equipment, or the population that is studied with a particular ADA assay, an assay revalidation may be required. The revalidation may cover some or all validation characteristics (i.e., it may be a partial or whole assay revalidation). Use of lots or batches of assay critical reagents that are different from those used in pre-study validation do not require assay revalidation, but they must be supported by appropriate experimental qualification data for the new reagent to ensure maintenance of system suitability.

Another critical aspect of life cycle management is the development of a strategy to bridge clinical data between an existing and an improved assay format. Such changes typically occur in a product's life cycle because of postmarketing commitments or other needs. To facilitate comparison and cross-validation of the existing method to the revised versions, analysts should retain sufficient aliquots of the original lots of critical assay reagents. In addition, archiving of analyte-spiked samples as well as blinded patient samples is useful to bridge between reagent lots and methods in order to minimize drift in assay performance. Analysts should develop a written plan outlining the sort of changes to the existing assay or critical reagents that will warrant an assay qualification versus a cross-validation or full validation. A quality management document should include details such as the number of assays that must be performed, the number of analysts that will be used, required training for analysts, acceptance criteria to demonstrate equivalence between existing and revised methods, data analysis, and reporting method. This information demonstrates the robustness and consistency of the assay following changes. Quality controls that ensure assay equivalence include %CV, tolerance limits,  $\text{EC}_{50}$  values of slope, titer level, and signal-to-noise ratio. One approach commonly used to demonstrate equivalence of two immunogenicity methods is the demonstration of  $\geq 90\%$  concordance in archived sample results between the existing and revised methods.

Analysts should use archived samples with a range of positive values as well as an appropriate number of negatives to verify that a new assay segregates samples into positive and negative categories in the same manner as an existing one.

Another important consideration for life cycle management of critical assay reagents is the monitoring of long-term reagent stability under different storage conditions. A detailed stability testing plan includes storage temperatures (4°C, –20°C, and –70°C), aliquot volume, freeze–thaw cycles, and acceptable performance characteristics for assay qualification, and results should be documented. In this context, it may be prudent to archive patient samples to demonstrate the long-term stability of the polyclonal ADA response in actual patient samples.

## APPENDICES

### Appendix 1: Nonclinical Immunogenicity Testing

Ponce R, Abad L, Amaravadi L, et al. Immunogenicity of biologically derived therapeutics: assessment and interpretation of nonclinical safety studies. *Reg Toxicol Pharmacol*. 2009; 54:164–182.  
 ICH. S6(R1) Preclinical safety evaluation of biotechnology-derived pharmaceuticals. Geneva, Switzerland: ICH; 2011.

### Appendix 2: Quality Attributes and Immunogenicity Risk Assessments

CMC Biotech Working Group. A-Mab: a case study in bioprocess development. 2009. [http://www.casss.org/associations/9165/files/A-Mab\\_Case\\_Study\\_Version\\_2-1.pdf](http://www.casss.org/associations/9165/files/A-Mab_Case_Study_Version_2-1.pdf). Accessed 01 December 2011.  
 Shankar G, Pendley C, Stein KE. A risk-based bioanalytical approach for the assessment of antibody immune responses against biological drugs. *Nat Biotechnol*. 2007; 25:555–561.  
 ICH. Q8(R2) Pharmaceutical development. 2009. [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Quality/Q8\\_R1/Step4/Q8\\_R2\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf). Accessed 14 December 2011.  
 ICH. Q9 Quality risk management. 2005. [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Quality/Q9/Step4/Q9\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf). Accessed 14 December 2011.

### Appendix 3: Design and Validation of Immunogenicity Assays

Mire-Sluis AR, Barrett YC, Koren E, et al. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J Immunol Methods*. 2004; 289:1–16.  
 Koren E, Quarumby V, Taniguchi G, et al. Recommendations on risk-based strategies for detection and characterization of antibodies against biotechnology products. *J Immunol Methods*. 2008; 333:1–9.  
 Shankar G, Devanarayan V, Amaravadi L, et al., Recommendations for the validation of immunoassays used for detection of host antibodies against biotechnology products. *J Pharm Biomed Anal*. 2008; 48:1267–1281.  
 Büttel IC, Chamberlain P, Chowers Y, et al. Taking immunogenicity assessment of therapeutic proteins to the next level. *Biologicals*. 2011; 39(2):100–109.

### Appendix 4: Statistical Methods

Zhang JH, Chung TDY, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomolecular Screening*. 1999; 4:67–73.  
 Devanarayan V, Tovey MG. Cut-points and performance characteristics for anti-drug antibody assays. In: *Detection and Quantification of Antibodies to Biopharmaceuticals: Practical and Applied Considerations*. Tovey MG, ed. Hoboken, NJ: John Wiley & Sons; 2011.

### Appendix 5: Regulatory Guidances for Clinical Immunogenicity Studies

FDA. Draft guidance for industry: assay development for immunogenicity testing of therapeutic proteins. 2009. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM192750.pdf>. Accessed 01 December 2011.  
 EMEA. Guideline on immunogenicity assessment of monoclonal antibodies intended for in vivo clinical use. 2012. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2012/06/WC500128688.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500128688.pdf). Accessed 30 July 2012.  
 EMEA. Guideline on immunogenicity assessment of biotechnology-derived therapeutic proteins. 2015. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2015/10/WC500194507.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/10/WC500194507.pdf). Accessed 15 October 2015.