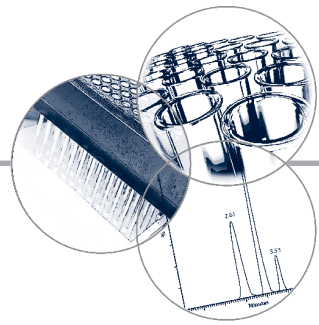

Recommendations and requirements for the design of bioanalytical testing used in comparability studies for biosimilar drug development

With the imminent expiry of patents on a number of biological products on the market, the development of biosimilars (or 'follow-on biologics') creates an increasing opportunity in the biotechnology industry. Although general guidelines on the quality and safety of biological products also apply to biosimilars, there is a need to address specific requirements for developing biosimilar drugs. Since it is critical to show comparability of the biosimilar products to their reference (or innovator) products, developing the appropriate bioanalytical methods to support such preclinical and clinical comparability studies is of great importance. The present work recommends the requirements for the development and validation for both pharmacokinetic and immunogenicity assays to support the biosimilar drug development.

Biosimilars are generic biological drugs, also called 'follow-on biologics' or 'biogenerics'. The first generation of biopharmaceuticals using recombinant technologies started 30 years ago; thus, the impending expiry of patents on some of these products creates a large opportunity. This impending situation has created a rapidly evolving area of product development in the biotechnology industry.

Legislation, recommendations and guidance for the development and commercialization of biosimilars are in place within the EU [1–3], and the USA are set to follow [4,5]. Biosimilars need to demonstrate a sufficient degree of resemblance to their branded original '**innovator**' or 'reference' products. To succeed, development/marketing costs of biosimilars need to be lower than that of the innovators, and the preclinical and clinical testing would also be abbreviated, relative to the innovator. However, no clear guidance or recommendations are available, thus far, to determine the design and requirements of biosimilar **bioanalytical testing**. Good bioanalytical data are required to show comparability between a biosimilar and the innovator, as well as to show comparability between products from different batches. This document provides scientific recommendations for the design of biosimilar bioanalytical testing, including quantitative determination of biosimilar products in matrices for pharmacokinetic (PK)/toxicokinetic (TK) data analysis, as well as the **immunogenicity assays** that include screening, confirmatory, neutralizing and characterization of the immune response.

Quantitative determination of biosimilar products in matrices for PK/TK data analysis

The **validation** of a quantitative assay for PK/TK purposes should follow the US FDA and European Medicines Agency guidances, the white papers and literature references [6–11]. Since the assay under validation here will be used to support comparability studies where both biosimilar and innovator compounds will be used, it is preferable to develop an assay using the same platform technology for both compounds utilizing the same set of assay reagents under the same assay conditions. However, it is not necessary to utilize the same assay platform as the PK assay used in the innovator development program.

A comparability test should be performed using the assay platform of choice to establish that the analytical method is capable of quantifying both the biosimilar product and innovator product within the assay variability limits. This can be achieved by using the biosimilar or innovator product as the calibrator to quantify both biosimilar QCs and innovator QCs in at least the accuracy/precision test. It is also preferred to assess this in the selectivity (matrix interference) test during assay development.

Xiao-Yan Cai[1], Dominique Gouty[†2], Sharon Baughman[3], MS Ramakrishnan[4] & Constance Cullen[1]

[1]Department of Biologics Bioanalytical, Merck, 1011 Morris Avenue, Building U13-1, Union, NJ 07083, USA
[2]Intertek USA dba Alta Analytical Laboratory, 3985 Sorrento Valley Blvd. Suite C, San Diego, CA 92121, USA
[3]Clinical and Preclinical Pharmacokinetics, Intertek Pharmaceutical Services, 6170 Cornerstone Court East, Suite 230, San Diego, CA 92121, USA
[4]Research and Development, Biocon Ltd, 20th KM Hosur Road, Electronics City, Bangalore 560 100, India
[†]Author for correspondence:
Tel.: +1 858 558 2599
Fax: +1 858 558 2600
E-mail: dominique.gouty@intertek.com

FUTURE SCIENCE part of fsg

Key Terms

Biosimilar: Generic biological drugs, also called 'follow-on biologics' or 'biogenerics'.

Innovator: Original biologic therapeutic drug, also called 'reference' product.

Bioanalytical assays: Methods for quantitative or qualitative measurement of a drug/compound of biological significance, in biological fluids.

Validation: Process of ensuring that a bioanalytical method conforms to defined user needs, requirements and specifications under defined operating conditions.

Ligand-binding assays: Method based on the reversible binding of a ligand to a binding protein to quantify the presence of a molecule.

If the bioanalytical comparability between biosimilar and innovator compounds was not demonstrated during assay development, a separate assay validation for each compound, including all elements of validation discussed in this paper, would be needed. The cause for the bioanalytical differences between the innovator and biosimilar drugs could be due to the different methods used to establish the label strength, therefore leading to different starting concentrations in the assay. It should be noted that if separate assays are to be used to support the comparability studies, all samples may need to be tested for both analytes, where bioanalytical laboratories must remain blinded.

An investigation should be completed to resolve the difference, as this observation could imply that the two products are different. The characterization assay of the biosimilar and innovator could show equivalency (potency, receptor-binding assay), however the acceptance criteria for these bioassays (50–150%) are typically wider than the acceptance criteria for the **ligand-binding assays** (80–120%). Therefore, a bioequivalency observed in the potency assay may not ensure comparability for the bioanalytical study.

If the development runs present acceptable bioanalytical comparability for the biosimilar and the innovator, one assay using the biosimilar compound as the assay calibrator can be used during assay validation. Assays can be developed and validated using either the biosimilar or the innovator as the assay calibrator. From our experience, we recommend to use the biosimilar as it is our drug of reference, but any one of them is acceptable. To demonstrate comparability, we do not recommend point-to-point comparison. We recommend that at a minimum, accuracy and precision tests should be conducted using the biosimilar drug as the standard curve to quantify both biosimilar and innovator QCs throughout the entire assay range (from ULOQ to LLOQ). The same assay acceptance criteria should apply for the biosimilar drug, as well as for the innovator. Meeting the accuracy and precision acceptance criteria will demonstrate that both compounds are comparable, since one standard curve is used to quantify both.

The validation should evaluate the assay performance characteristics described in the following sections.

■ Accuracy & precision

It is recommended that the regression model established during method development be confirmed in a minimum of six independent validation runs, and typically in the same runs in which method precision and accuracy are assessed. These accuracy and precision runs should use the biosimilar compound to construct the standard curve (when comparability was demonstrated between biosimilar and innovator compounds during assay development) to quantify the QCs made from either biosimilar or innovator compounds.

For the curve within a run to be acceptable, the %RE (relative error or % bias) of the back-calculated value for at least 75% of the standard points, not including anchor points, should be within 20% of the nominal concentration, except at the ULOQ and LLOQ, where the value should be within 25%.

Accuracy and precision QCs are prepared by spiking both the innovator and the biosimilar at five or more concentrations (anticipated LLOQ, no more than three-times LLOQ, mid-, high- and anticipated ULOQ). It is recommended that at least three independent determinations (replicates) per run be performed for each sample, in a minimum of six runs for either biosimilar and innovator QCs.

For a method to be considered acceptable, it is recommended that the intra- and inter-run precision (% coefficient of variation) and the absolute mean bias (%RE) be no more than 20% (25% at LLOQ). In addition, it is recommended that the method total error (sum the % of the coefficient of variation and absolute %RE) be no more than 30% in order to be consistent with the validation acceptance criteria [7].

The range of quantification should be based on the lowest (LLOQ) and highest (ULOQ) validation samples that meet the target accuracy and precision criteria.

■ Dilutional linearity

The dilutional linearity should be tested. If single dilution testing for future sample analysis is performed, the back-calculated concentration for each diluted sample should be within 20% of the nominal within the linear range (25% at ULOQ and LLOQ). During sample analysis, if a sample is tested at multiple dilutions, the back-calculated concentration for cumulative diluted samples should be within 20% of the nominal value of the original value. The precision of the cumulative back calculated concentration should be no more than 20% (25% at ULOQ and LLOQ).

The presence or absence of hook (or prozone) effect should also be evaluated at the higher QC concentration (>1000×).

■ Matrix interference (selectivity)

Matrix interference should be performed using biosimilar QC spiked samples (spiked at high and low concentrations into at least ten individual matrix samples). The matrix interference testing should also include the blank individual controls that will be tested at the minimum required dilution (MRD).

The recommended target acceptance criteria for selectivity is that acceptable recovery (same as those described for accuracy) is obtained in at least 80% of the matrices evaluated. Disease state matrices can contain components that interfere with the assay; we recommend that selectivity experiments be performed with disease state matrices only if needed.

■ Sample stability

Stability experiments should mimic, as best as possible, the conditions under which study samples will be collected, stored and processed. The effect of freeze-and-thaw cycles should also be assessed. The stability testing should follow the white papers and regulatory guidance [1,6,7]. If the comparability for accuracy and precision was demonstrated for the biosimilar and innovator compounds, then the biosimilar QCs (at high, medium and low concentrations) will be evaluated.

Immunogenicity assays

The immunogenicity of therapeutic proteins needs to be assessed for safety and efficacy concerns, since small process changes during the production of biologics can lead to changes in immunogenicity rate. It should be noted that immunogenicity rate is difficult to measure, particularly at low incidence. The size will need to be determined with Health Authority input and will depend on the rate of immunogenicity and the risk to patients of developing an immune response (e.g., from autoimmune reactions to self proteins). Generally speaking, a larger size would be required if the rate of immunogenicity incidence is low. It is advisable to work with a qualified statistician and/or regulatory agency to evaluate study plans. It is critical to assess the immunogenicity of the biosimilar compared with the innovator drug. White Papers and regulatory agencies provided guidelines for method development and validation to determine the immunogenicity of therapeutic biologics [3,4,12–16].

Since the assay will be used to support comparability studies, where both biosimilar and innovator compounds will be used, it is preferred to develop an assay using the same platform technology, the same set of assay reagents under the same assay conditions to evaluate antidrug antibodies (ADAs) for both biosimilar and innovator compounds. It is not necessary to utilize the same assay platforms as the immunogenicity assays used in the innovator development program.

As soon as the biosimilar drug is available, immunization of animals to develop a positive control should be initiated. In addition, another positive control against the innovator compound should also be developed. Evaluation of comparability (or lack of) between the two ADA positive controls (ADA against innovator and ADA against biosimilar) should be conducted once they become available. Differences in the starting titers of the positive control antisera against either the biosimilar or innovator compound are possible due to the individual immune response of each animal. This should be taken into account when evaluating the assay performance.

■ ADA screening assay

Multiple assay platforms are available for the detection of ADA, such as direct ELISA, bridging assays, electrochemiluminescence (ECL) assays, SPE acid-dissociation assays and radioimmunoprecipitation assays. Whichever assay platform is selected, the assay needs to address the following questions:

- Can the assay reagents detect both biosimilar and innovator ADAs comparably?
- Can the assay tolerate both biosimilar and innovator drug concentrations comparably?

Both ADA positive controls will be needed to address the above questions. If bioanalytical comparability was demonstrated, the assay can be validated using ADA against biosimilar as the positive control. The acceptance criteria for bioanalytical comparability may be assay dependent, and should be assessed based on the performance of the assay. To demonstrate comparability of anti-innovator antibody and anti-biosimilar antibody in an assay, we recommend that one set of reagents be used to detect both ADA positive controls. For example, if a bridging ECL assay format is used on the MSD platform, the biotinylated and ruthenylated biosimilar drug should be used as capture and detection reagents for the detection of both ADA positive controls. The assay should demonstrate the following:

- The dose–response curve of both ADA positive controls should be comparable (preferably overlay with each other or at least parallel to each other or within the range determined);
- The detection sensitivity for both ADA positive controls should be comparable. It should be noted that because these positive controls are polyclonal antibodies generated in different animals, the immunoreactive strength of these antibodies may not be identical, even if the exact quantity of the affinity purified materials is used in the sensitivity assessment experiments. We recommend that if the sensitivity for both ADA positive controls are within two- to three-fold of each other in mass unit, they can be considered comparable;
- The assay should tolerate both biosimilar and innovator drug concentrations comparably. We recommend that the drug tolerance concentration for each compound should be within two- to three-fold of each other, for drug tolerance to be considered comparable.

If comparability is not demonstrated, separate assays should be validated for biosimilar and innovator ADAs. It should be noted that if separate assays are to be used for future preclinical or clinical comparability studies, interpretation of study data could present additional challenges, since samples from different arms of the study will be tested using different assays.

To validate the screening ADA assay, the general guidance [4,12,14] should be followed to evaluate the following assay performance characteristics.

Minimum required dilution

Minimum required dilution should be tested to establish the optimal percentage of matrix used for the assay. It is important to keep in mind that, in general, higher MRD (more dilutions) can help remove matrix effect, but may also negatively impact assay sensitivity. Careful assessment of both aspects is recommended during assay development.

Screening cut point

The screening cut point is defined as the level of response at or above which a sample is defined to be 'reactive' for the presence of ADA, and below which it is considered negative. A cut point is established during assay validation, usually by running individual matrix samples (usually 50) that represent study target populations. These samples should be tested by multiple runs. When establishing the cut point, the removal of statistically determined outlier values should be considered. A 95% confidence (or 5% false-positive rate) is usually used to calculate the screening cut point using either raw assay response units or signal-to-noise approaches.

Confirmatory (specificity) cut point

The specificity confirmation test is usually a competitive immunodepletion test using a representative number of individual naive serum samples, representing future study target populations (at least ten from each type) spiked with an excess amount of drug (the concentration of drug used in this test is usually based on the results from the drug tolerance test) or with a control (no drug added). The percentage inhibition values from these individual serum samples, treated with or without drug, are used to calculate the confirmatory cut point. A 99 or 99.9% (1% or 0.1% false-positive rate) confidence is usually used to calculate the confirmatory cut point.

Sensitivity (LOD) of the assay

Sensitivity of the ADA assay should be evaluated using affinity or Protein A/G purified ADA at concentrations near the assay cut point. The assay sensitivity is defined as the lowest ADA concentration for which at least 95% of the responses are greater than or equal to the assay cut point. The sensitivity for screening ADA assay should achieve approximately 250–500 ng/ml, as such antibody concentrations have been associated with numerous clinical studies [4].

Drug tolerance

It is expected that samples containing drug will interfere with the detection of ADA due to competition for binding to product-specific antibodies between the drug and that used as a capture reagent in the assay system. It is common practice to assess the effect of various concentrations of drug on the detection of the ADA positive control, during assay validation. However, it is important to keep in mind that sampling time is critical for proper immunogenicity assessment during clinical studies. Drug interference can be minimized if samples are collected at a time when the therapeutic protein has decayed to a level where it no longer interferes with the assay results. Data from PK studies are useful in establishing optimal sample collection times.

Matrix effect

Matrix effects should be assessed by spiking the ADA positive control at high and low concentrations into a minimum of ten individual sera, and comparison to the pooled serum spiked with the same concentrations of ADA.

Precision

It should be noted that there are two types of precision for the ADA assay. One is the precision between replicate raw response units, such as ECL raw counts or optical density values. Another is the precision for the final reported titer value. Usually the precision for the replicate raw count values is much tighter than the one obtained for titer determination. This is expected since the titer is where the dilution of ADA crosses the assay cut point. The analytical assay variability is higher when analyte concentration is close to the detection limit of the assay.

Titer determination may not be necessary for preclinical immunogenicity assays.

Stability

Stability experiments should mimic, as best as possible, the conditions under which study samples will be collected, stored and processed. The stability testing should include bench top storage, as well as freeze–thaw evaluations, following the White Papers and regulatory guidances [4,12,14].

Neutralizing ADA assays

For clinical studies, once a test sample is confirmed to be ADA positive, the sample may be evaluated in the neutralizing-antibody (NAb) assay to see if it is neutralizing the biologic activity of the drug. Although it is a preferred choice by the regulatory agency to have a cell-based NAb assay [4], other assay formats (e.g., immuno-based assays assessing the binding of drug to its target) may also be used, especially when appropriate cell lines are not available during early phase drug development. If a cell-based assay exists for the innovator compound, it might be preferred to evaluate this assay platform during NAb assay development for the biosimilar. Validating cell-based NAb assays is more technically challenging due to higher variability and a longer turnaround time for these types of assays. Efforts have been made in the past few years to address NAb assay optimization and qualification [15,16]. The NAb assays to support the biosimilar program should consider the same questions as in the ADA screening assay, which are:

- Can the assay system detect both biosimilar and innovator ADAs comparably?
- Can the assay tolerate both biosimilar and innovator drugs comparably?

The assay performance characteristics to be evaluated during assay validation are similar (or as close as possible) to the ones mentioned previously for the ADA screening assay.

One aspect worth mentioning is the confirmatory test for the NAb assay. It was recommended by the most recent FDA draft guideline that this aspect of the NAb assay should be evaluated [4]. There are usually two approaches and neither is perfect. One is to demonstrate that the specificity of signal change is due to the NAb present in the sample, instead of some other nonspecific agent which can elicit the effect all by itself. To test this, the drug target (e.g., the cytokine in the case of mAb drug against that cytokine) can be removed from the assay system. If the signal change in a sample was indeed due to the presence of NAb, removing the cytokine from the assay system will reverse the effect caused by NAb (e.g., for a cell-based proliferation NAb assay, in the presence of cytokine, the assay signal increases, the presence of drug against that cytokine decreases the signal, NAb in the sample neutralizes the function of the drug, therefore restores the signal induced by the cytokine leading to increased assay signal. When the cytokine is removed from the assay, the cell can not proliferate in the presence of drug or NAb, therefore the assay signal is decreased). However, if the previous signal change in the test sample (increased assay signal) is due to a nonspecific agent, removing the drug target (cytokine) from the assay system will not have any effect (therefore the assay signal remains to be high because other nonspecific factor in the sample is causing the signal increase). The second approach is to perform immuno-depletion, by removing all immunoglobulins from the test sample by passing it through a Protein A or G column. If the previous signal change was due to NAb, the effect would be reversed. Whichever approach is used, a confirmatory cut point should also be established to ensure 99 or 99.9% confidence in confirming the specificity of the assay.

It should also be noted that the sensitivity of cell-based NAb assays may be worse than that of ADA screening assays. Therefore, having a negative result for a sample may not mean that the ADA in the sample was not neutralizing.

■ Characterization assays: isotyping

It is important to evaluate if the confirmed positive samples from the screening ADA assay is indeed due to immunoglobulin and, if so, what type of antibody. For example, it would be more alarming if the antibody is of the IgE class, as it could have potentially serious safety outcomes.

The isotyping assays are characterization assays and therefore may not require validation.

Conclusion

To succeed, development/marketing costs of biosimilars need to be lower than the branded original and the preclinical and clinical development paths should be faster. The challenges in this fast turn-around reside in the fast development and validation of immunoassays while keeping the science and assay-performance characteristics optimal. Since it is critical to show comparability of the biosimilar products to the innovator products, developing the appropriate bioanalytical methods to support such preclinical and clinical comparability studies is of great importance. If comparability of the biosimilar products to the innovator products is proven, we recommend to only validate one assay; whereas, if comparability or interchangeability is not proven, the validation of two assays (one for the biosimilar and one for the innovator) will be needed.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as:
▪▪ of considerable interest

1 European Medicines Agency. *Guideline on Similar Biological Medicinal Products.* EMA Guideline reference CHMP/437/04 (2005).

▪▪ **Regulatory guideline for biosimilars.**

2 European Medicines Agency. *Guideline on Development, Production, Characterization and Specifications for Monoclonal Antibodies and Related Products.* EMA Guideline reference EMEA/CHMP/BWP/157653/2007 (2008).

3 European Medicines Agency. *Guideline on Immunogenicity Assessment of Biotechnology-Derived Therapeutic Proteins.* EMA Guideline reference EMEA/CHMP/BMWP/14327/2006 (2007).

4 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Guidance for Industry: Assay Development for Immunogenicity Testing of Therapeutic Proteins.* (2009).

5 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Guidance for Industry: S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals.* (1997).

6 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Guidance to Industry: Bioanalytical Method Validation* (2001).

7 DeSilva B, Smith W, Weiner R *et al.* Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm. Res.* 20, 1885–1900 (2003).

8 Viswanathan CT, Bansal S, Booth B *et al.* Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand-binding assays. *Pharm. Res.* 24(10), 1962–1973 (2007).

9 Shah VP, Midha K, Dighe S *et al.* Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. *Pharm. Res.* 9, 588–592 (1992).

10 Findlay JW, Smith WC, Lee JW *et al.* Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J. Pharm. BioMed. Anal.* 21,1249–1273 (2000).

11 Dunn JR. A statistical-based quality control/quality assurance system for immunoassay analysis. *Am. Clin. Lab.* 10, 7–8 (1993).

12 Mire-Sluis AR, Barrett YC, Devanarayan V *et al.* Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Methods* 289, 1–16 (2004).

13 Koren E, Smith HW, Shores E *et al.* Recommendations on risk-based strategies for detection and characterization of antibodies against biotechnology products. *J. Immunol. Methods* 333, 1–9 (2008).

14 Shankar G, Devanarayan V, Amaravadi L *et al.* Recommendation for the validation of immunoassays used for detection of host antibodies against biotechnology products. *J. Pharm. Biomed. Anal.* 48(5), 1267–1281 (2008).

15 Gupta S, Indelicato SR, Jethwa V *et al.* Recommendations for the design, optimization, and qualification of cell-based assays used for the detection of neutralizing antibody responses elicited to biological therapeutics. *J. Immunol. Methods* 321, 1–18 (2007).

16 Wei X, Swanson SJ, Gupta S. Development and validation of a cell-based bioassay for the detection of neutralizing antibodies against recombinant human erythropoietin in clinical studies. *J. Immunol. Methods* 293, 115–126 (2004).