L2: Large Molecule Specific Assay Operation

Team members:

Team lead

– Lauren Stevenson (NA)

Other members

– Clare Kinglsey (EU)
– Karolina Oesterlund (EU)
– Marian Kelley (NA)
– Heather Myler (NA)
– Boris Gorovits (NA)
– Yoshiyuki Minamide (APAC, Japan)
– Arumugam Muruganandam (APAC, India)
– Mario Dominguez (LA)

In scope

- Testing of ruggedness and robustness
- Setting up a balanced validation design
- Dilution linearity
- Specificity testing
- Selectivity testing
- Parallelism
- Hook effect

Interdependencies with other teams

– L1 – Assay Acceptance
– A6 – Stability

Out of scope

- Cross validation (A3)
- Approach for spiking QCs for validation (L1)
- Use of drug product, drug substance or reference standard as the entity used in validation/sample analysis (A4)
- Long term stability in matrix (A6)

Current status

- Good consensus achieved on all topics:
  - Balanced validation design
  - Dilution linearity
  - Hook effect (pre-study validation)
  - Robustness and ruggedness
  - Selectivity testing
    - Normal and disease matrix
    - Testing of lipemic and hemolyzed samples
  - Specificity
  - In-study hook effects
  - Parallelism

- However, in an effort to refine recommendations, team is continuing to gather case studies on:
  - In-study hook effects
  - Parallelism
  - Impact of lipemic and hemolytic samples on LBAs

Setting Up Balanced Validation Design

Recommendations:

- Ensure same number of observations in each Precision & Accuracy (P&A) run – number of sets of QCs in each run should be the same
  - Aligned with Team L1, which recommends 3 independent sets each of 5 QC levels (LLOQ, LQC, MQC, HQC, ULOQ) in ≥ 6 P&A runs

- Analyst as variable: Number of analysts in validation should be reflective of sample testing practice, noting that:
  - Commonly one analyst performs bulk of validation with inclusion of second analyst for a subset of P&A runs
  - Possible to justify using only one analyst during validation if study sample analysis will also only use one analyst (small studies)
  - If multiple analysts will test samples, recommend at least 2 analysts in validation

Dilution Linearity and Hook Effect

Recommendations:

Approach

- Spike samples at ≥ anticipated Cmax; if Cmax is unknown, spike at highest feasible concentration
- There should be a minimum of 90% matrix in highest feasible concentration sample
- Make dilution series that includes anticipated sample concentrations both above and spanning the curve range

Acceptance Criteria

- In-range samples should be within 20% theoretical and precision of the cumulative back-calculated concentrations should be ≤ 20%
- If >ULOQ spiked samples read ALQ = no hook effect

Dilution Linearity and Hook Effect

- Other considerations: when a hook effect is present
  - Approaches for sample testing dilution schemes when a hook effect has been demonstrated above certain drug concentrations
  - Example
    - If unblinded: Test high concentration samples at multiple dilutions
    - If blinded: Test all samples at multiple dilutions
    - Data reporting approaches:
      - Mean of all in range values where the higher dilution results in a lower signal
      - Value closest to ED50
  - Recommendation: Several approaches for sample testing and data reporting may be valid; key is to define *a priori* which approach will be employed

Testing Robustness and Ruggedness

Definition

- Robustness/ruggedness terms are often used interchangeably and there has historically been confusion regarding the absolute definition of each term
- It is understood, however, that both parameters are indicators of assay reproducibility under varied conditions
- Therefore, any robustness/ruggedness analysis should address the question of whether the assay will perform well under real life changes in standard laboratory situations

Recommendations

- Robustness/ruggedness testing should generally be incorporated into method development process
  - Typically includes temperature variations, reagent lots, plate lots, analysts, instruments, incubation times
  - Be aware of the needs of your assay and the conditions under which it may be run, including differences in ambient temperatures in different labs; regional differences in serum/plasma sources, etc.
- Demonstrated during validation by virtue of use of multiple instruments/ analysts and typical variations in incubation times
- Cross validation in later stage further demonstrates R&R

Selectivity

Recommendations

Approach

- Test 10 or more individual samples, unspiked and spiked at LLOQ level (required) and higher level (e.g. HQC) (recommended)
- When possible, for disease indications, selectivity should be performed in disease matrix

Acceptance Criteria

- ≥ 80% of unspiked samples should measure <LLOQ
- ≥ 80% (8/10) samples should be within 20% nominal for HQC spikes and 25% for LLOQ spikes
- The *same* 80% of samples should meet criteria

Selectivity

Lipemic Samples

- Recommendation
  - The need to perform selectivity assessments with lipemic samples will be dependent upon the drug, disease indication and assay format
    - Typically, performing selectivity assessments with disease matrix samples will address any effects of lipemia which may be present in that population
    - The team is actively seeking examples where there was an issue caused by lipemia to guide when additional assessments may be recommended

Selectivity

Hemolyzed Samples

- Recommendation
  - The need to perform selectivity assessments with hemolyzed samples will be dependent upon the characteristics of drug, its target, disease indication and assay format
  - The team is actively seeking examples where there was an issue caused by hemolysis to guide when these assessments may be recommended
    - Examples gathered to date indicate that issues due to hemolysis are rare and have not been observed with mAb therapeutics. However, insulin and related therapeutics are more likely to be sensitive to hemolysis

Selectivity – Other Issues

- When endogenous analyte is present
  - Choosing standard curve matrix - when possible, screen for low endogenous level pool
  - Employing a subtraction method to evaluate spike recovery may be possible if endogenous levels are measureable (≥ LLOQ)
  - May need to sacrifice LLOQ level when endogenous levels are detectable, but not quantifiable (< targeted LLOQ)
  - If endogenous level is quite high, then may need to question if your assay range is appropriate

Selectivity – Other Issues

- Operational Question
  - Should selectivity samples be tested fresh (on the same day as they are spiked) or after they have been frozen and thawed once (to mimic real samples)?
  - Recommendation
    - Samples may be tested fresh (after allowing for adequate mixing/ molecular interactions) or after one freeze/thaw (assuming that F/T stability has been demonstrated)
    - Note: F/T is not considered to more accurately reflect real samples as spike and recovery is not how real samples are generated; allowing for one F/T cycle before assessment of selectivity samples simply allows for practical flexibility

In-study Hook Effect

Examples:

- Aggregation-prone drug
- ADA+ samples
- C. Kalensky, M. Stuart, R. Olech and M.T. Rock, Analysis of Samples from Non Clinical and Clinical Trials of Biologics Where High Dose Prozone Effects are Present, presented at AAPS NBC (2008)
  - All "solved" with increased dilution
  - No examples of mAb drugs

Recommendation:

- The scientist should be mindful of any special characteristics or handling procedures related to their drug and how this might lead to potential hook effect in study samples even when one is not observed during validation
- In study, the scientist should be reviewing the data closely in order to "catch" any hook effect (as well as loss of parallelism) that may manifest
- *A case by case risk assessment is needed to determine whether in-study hook effect evaluation should be done; in general, atypical results suggesting an in-study hook effect should be appropriately investigated and documented*

In-study Hook Effect

- If a therapeutic is considered at risk of in-study hook effect
  - Possible approach to test for in-study hook effect
    - Perform a preliminary run of serial dilutions of several study samples at each dose level prior to analysis of study to define dilutions that should be applied in study sample analysis
    - Will need to define *a priori* how these data will be handled

Parallelism

- Routine parallelism assessments currently not being broadly implemented industry-wide (some/few are doing routinely)
- Currently, more questions than answers around when to perform the assessment, how to perform it, and how to report the data

- Recommendation
  - The need to perform parallelism assessments will depend upon the characteristics of the drug, its binding partners and specific assay reagents
  - The team continues to seek examples where non parallelism has been observed to guide when assessment may be recommended
    - Examples gathered to date indicate that non-mAb therapeutics, especially peptides, may be more likely to have issues of non parallelism
    - No examples yet identified for mAb therapeutics

Specificity Testing

Recommendations

Approach

•What to test

- Potentially cross-reacting molecule(s); Note: more relevant for non mAb drugs than mAb drugs as selectivity assessments already assess recovery in presence of mg/mL levels of antibodies
- Concomitant medications: As appropriate – test those that are specified in the protocol to be co-administered or stable regimen; No need to test OTC drugs
- Other: Consider potential impact of ADAs, circulating soluble target, etc. on assay performance

•How to test

- Spike maximum concentration anticipated in study into (1) blank matrix, (2) LLOQ QC; (3) HQC

Acceptance Criteria

•Unspiked samples should measure <LLOQ

•Spiked samples should measure within 20% nominal for HQC spikes and 25% for LLOQ spikes

Outstanding Questions and Loose Ends

- Defining what constitutes lipemia – how standardize?
- Defining what constitutes hemolysis – how standardize?
  - 04May12 – team agreed that since we do not recommend routine analysis for lipemia and hemolysis, we will not provide specific definitions. In our manuscript we will speak to the lack of a unified definition of such samples and may give examples of approaches people have used when testing of such samples was deemed necessary/appropriate.
- In-study hook effect
  - How perform the assessment? And when?
  - How handle multiple results for a single sample?
  - 04May12 – Team agreed that we would not provide a definitive approach but that we would offer examples from the literature and our own experience as to how in-study hook effect might be investigated and the data handled, emphasizing that different approaches will be acceptable. We will highlight that what will be performed for the investigation and how the data will be interpreted and reported should be documented before any investigation experiments are performed.
- For assessments such as specificity, should analysis be done with a single sample or done in triplicate and require 2/3 to pass (to avoid repeat analyses due to technical errors with spiking)?
- 04May12 – Team agreed that we would not specify the exact approach but would offer options.

Next Steps

- Drafting more comprehensive recommendation language on aligned topics
- Ongoing: continue to collect examples to guide when it would be recommended to evaluate parallelism and lipemic and/or hemolytic samples
- Goal – final recommendations on all topics in June 2012