Defining Your Product Profile and Maintaining Control Over It, Part 4

Product-Related Impurities: Tackling Aggregates

Kurt Brorson and Joseph Phillips

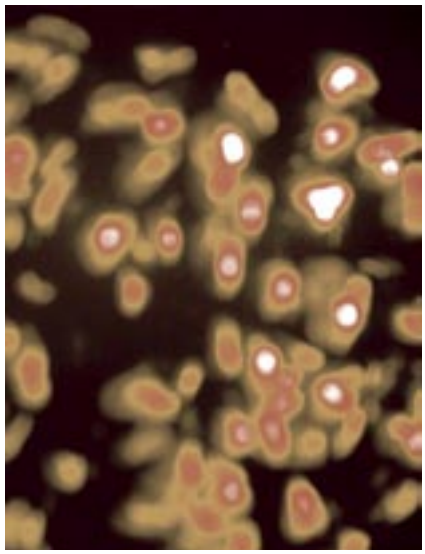
Proteins can self-associate. For some proteins, self-association is natural and required for their function (e.g., tumor necrosis factor, TNF, is a homotrimer). However, proteins isolated from their native environment (such as in a vial or in an in-process fluid rather than in a cell or in plasma) can self-associate into nonnative oligomers, both covalent and noncovalent. This has proven problematic for biopharmaceutical development because this may negatively affect process efficiency and/or subsequent clinical use.

Regulatory Considerations

Regulatory agencies are concerned about aggregate levels in biopharmaceuticals. It is generally understood that aggregate levels can vary between products. Acceptable aggregate levels are determined by various clinical and manufacturing factors. Although "<5%" may be a common acceptance criterion for early IND monoclonal antibodies (MAbs) for which there is limited manufacturing and clinical experience, it should not be regarded as a "magic" number for all cases. Product and process understanding are expected to improve during development and aggregate specifications tightened appropriately. When aggregates form during extended storage, separate stability specifications may be justified. ICH Q6B allows separate stability specifications, but they should be justified by clinical and manufacturing experience.

Product formulation and container–closure choice can be critical for minimization of aggregates. For example, prefilled syringes are a relatively new dosage form for biopharmaceuticals; less experience is available for them than for other dosage forms. However, instances of aggregate formation catalyzed by syringe and/or needle leachables have occurred. It is paramount that stability studies be performed in the proper dosage form and storage conditions. Real-time and accelerated stability studies are useful because in some instances aggregate formation is stability indicating for a particular product.

Products stored in vials also have dosage form considerations. For example, some products stored in vials are subsequently diluted and transferred to an infusion bag before administration. What happens in the IV bag? Do the aggregates break apart? Do they accumulate? Neither? These questions warrant careful evaluation.



FRANK RIX (WWW.ISTOCKPHOTO.COM)

Good Methods for Detecting and Quantifying Aggregates

Multiple assays are available to measure aggregate levels in biopharmaceuticals. Some such assays measure molecule size distributions and thus directly quantify aggregates. Others indirectly measure aggregates by monitoring surrogate properties such as levels of beta-sheets or avidity of molecular interactions. Only two assays, SEC-HPLC and SDS-PAGE, are considered robust routine QC

Product Focus: Recombinant proteins, monoclonal antibodies

Process Focus: Process Development (product characterization), formulation development, CMC/clinical interface

Who Should Read: Manufacturing and process development, project managers, regulatory affairs, and analytical personnel

Keywords: aggregation, self-association, validation, qualification

Level: Basic

Proceedings of the WCBP CMC Strategy Forum, 19–20 July 2004

The sixth Well-Characterized Biotechnology Pharmaceutical (WCBP) Chemistry, Manufacturing, and Controls (CMC) Strategy Forum was held on 19–20 July 2004 at the Lister Hill Auditorium on the NIH Campus in Bethesda, Maryland. The event was sponsored by the California Separation Science Society (CaSSS; www.casss.org) as part of an ongoing series of discussions between industry and regulatory participants exploring current practices in analytical and bioprocess technologies for development and communication of consensus concepts. The topic of this forum was "Defining Your Product Profile and Maintaining Control Over It."

The purpose of this two-day forum was to survey which methods are most useful in identifying and measuring process-related and product-related impurities — and identify strategies and specifications to ensure a consistent product profile. The first day concentrated on process-related impurities (Parts 1 and 2 of this article); on day two the focus shifted to product-related impurities (Parts 3 and 4).

Participants on the second day discussed general strategies for maintaining control over product-related impurities. The moderators were Kurt Brorson (Office of Biotechnology Products, CDER, FDA), and Joseph Phillips (senior director, Analytical Sciences, Amgen, Inc., Thousand Oaks, CA; they were joined on the panel by Zahra Sharokh (Transkaryotic Therapies, Inc.), Michelle Fraizer-Jessen (CDER, FDA), and Michael Bond (Centocor, Inc.).

The members of the permanent CMC advisory committee are Siddharth Advant (Diosynth Biotechnology), John Dougherty (Eli Lilly and Company), Rohin Mhatre (Biogen Idec Inc.), Anthony Mire-Sluis (Amgen, Inc.), Wassim Nashabeh (Genentech, Inc.), Nadine Ritter (Biologics Consulting Group, LLC), Mark Schenerman (MedImmune, Inc.), Heather Simmerman (Amgen, Inc.), and Keith Webber (CDER, FDA).

163 people attended the two-day conference, representing industry companies, consultant companies, and the FDA.

assays. Many others are considered suitable for development, product characterization, and comparability studies. Some development assays are considered qualitative or semiquantitative at best. On the other hand, the two routine QC assays measure only subsets of the total aggregate population potentially present in protein solutions. SDS-PAGE quantifies non-SDS dissociable aggregates only, and SEC-HPLC cannot reliably measure very large or weakly associated complexes.

The top four methods for detecting aggregates are size-excursion chromatography (SEC-HPLC), analytical ultracentrifugation (AUC), field-flow fractionation (FFF), and electrophoresis.

Many regard **SEC-HPLC** to be the "workhorse" assay for routine aggregate detection. In an informal survey of the workshop participants, nearly 100% reported using it for lot release. However, there are clear limits on its use as a characterization assay. For example, test articles are generally diluted before column loading. During sample processing, aggregates can break apart even before entering the column. In addition, very large aggregates may never actually enter the column; instead, they are removed by column frits or filters. Band spreading during the chromatography run is another point at which dilution can occur and weakly associated aggregates can break apart. All of these potential sources of insensitivity and variability warrant inclusion of orthogonal complementary assays during product characterization or process comparability studies.

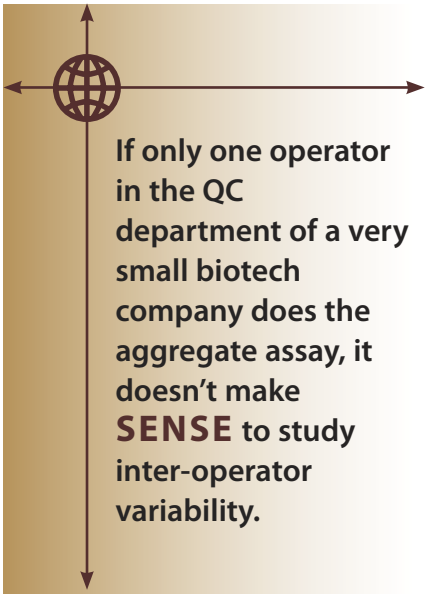
Analytical ultracentrifugation (AUC) has proved to be a useful characterization tool. Although none of the workshop participants said they use AUC for lot release, more than half use it in comparability and characterization studies. A majority of those that use AUC are using it to crossvalidate SEC-HPLC. A chief advantage of AUC is that it can be performed in solution phase; no dilution is required. Thus, it can detect very large aggregates and weakly self-associating complexes. It has also proved to be an excellent formulation tool, allowing companies to develop formulations that minimize even weakly associated aggregates. Conference participants indicated that equilibrium AUC is used less often than sedimentation AUC. Some technical issues surrounding AUC include extensive data manipulation, cell cleaning difficulties, and UV-light induced, false positive aggregate formation.

When AUC is used in comparability and characterization studies, it is important that correlation and consistency be evaluated between assays. This is particularly important if AUC data disagree with SEC data; at least three of the workshop participants have experienced this. In the first case, AUC was necessary during formulation development to track and minimize weak self-association that couldn't be seen by SEC-HPLC. In the second case, AUC detected an increase in aggregates that occurred after a process change that was missed by SEC-HPLC. In the third case, AUC was critical in investigation filtration resistance problems that SEC-HPLC couldn't resolve. In all likelihood, the offending aggregates went undetected by SEC-HPLC because they were 100-mers; they probably never entered the column.

Conventional sedimentation velocity AUC can be useful in formulation studies, in which protein concentration and buffer pH can be varied between AUC runs and the impact on complex formation examined. More advanced gravity sweep sedimentation velocity AUC can elucidate an overview of multiple molecular weight species in a given protein solution.

Field flow fractionation (FFF) is a second solution-phase assay that can detect loosely associated complexes. About 30% of the workshop participants reported using FFF — all for characterization work, with only a small minority using it to crossvalidate SEC-HPLC. The participants felt that this technology has a great deal of promise but first must overcome technical challenges and improve robustness.

Electrophoresis is a common QC lab assay; the majority of participants reported using SDS-PAGE for product lot



release. Like SEC-HPLC, this assay cannot give a comprehensive picture of all particles in a protein solution; it detects covalent non-SDS-dissociable complexes only. Only a small number of participants reported using capillary electrophoresis; they use it only for characterization and crossvalidation work.

Other methods: Dynamic light scattering (DLS) was considered by the participants to be a useful semiquantitative or qualitative assay for measuring particle sizes, but not necessarily for quantitating size distributions. About 25% of the participants use it in characterization studies coupled to SEC-HPLC. None use it for lot release and it is rarely, if ever, used in static mode for product evaluation.

Mass spectroscopy was also considered to be only a qualitative assay. A few participants use it primarily for aggregate characterization. They use both MALDI and electrospray.

Some participants noted that they use uncommon or legacy assays to evaluate aggregates. These assays include UV light scattering (>300 nm), filtration resistance, and hydrodynamic chromatography. Others adapted assays that commonly look at other quality attributes. For example, surface plasmin resonance (SPR), which measures affinity and association constants, can indirectly measure MAb aggregates if their affinity for antigen differs from that of monomers. Similarly, aggregates that have different levels of beta sheets than monomers can be evaluated by FTIR or CD. Other surrogate assays reported by the participants included appearance tests, light obscuration, turbidity, electronic particle size analysis, and thermoanalysis (DSC).

Assay Validation Approach

The majority of participants felt that assays in early phase studies should at least be qualified. Unlike for validation, not every parameter in ICH Q2A and B requires investigation during assay qualification. For example, if only one operator in the QC department of a very small biotech company does the aggregate assay, it doesn't make sense to study inter-operator variability. It was noted that ICH Q2A and B were never intended to be a list of check boxes. A qualified assay, however, should be demonstrated to be suitable for intended use.

The participants regarded preparing aggregates for validation studies to be a complex task. An analogy to the Heisenberg uncertainty principle was invoked to explain the conundrum. During validation, the study perturbs the system. Are artificial aggregates used in a validation study representative of the small amount of native aggregate in your product? Extreme manipulations may be required to make the aggregates or purify them from the product. They can also change over time. Crosslinking noncovalent aggregates in place can prevent reversion to monomers, but the crosslinking chemicals add complexity to the material.

Participants reported that in some cases aggregates could be isolated from process samples. This is preferable, but not always possible. In most cases, samples are generated through forced degradation. To be meaningful, those artificial aggregates should be demonstrated to represent aggregates in product. They can be examined through multiple assays. SEC-HPLC can compare covalent patterns, and SDS-PAGE can examine similarity of other general characteristics. It is also informative to include an assay that can quantify large and noncovalent aggregates as well. The method of generation can affect aggregate physical characteristics. In one case study at the forum, product freeze–thaw samples were more representative of native aggregates than were pH or heat-induced aggregates.

Setting Specifications for Aggregates

Setting specifications for aggregates is also a complex task. Multiple factors requiring consideration include manufacturing experience, clinical experience, and the nature of the aggregates. One important question to consider: How bad are they really?

The nature of the aggregates may complicate specification setting as well. A few companies reported having separate assays and specifications for both reversible and nonreversible forms. Having separate specifications for stability studies and release assays is permitted by ICH Q6B when aggregates slowly change over time and when justified by clinical experience. The participants agreed that there is no "magic number" (e.g., <5%) that can apply in all cases.

Are Aggregates Always Bad?

It is probably impossible to produce concentrated protein solutions that are totally free of self-association. The real questions facing the participants were

- At what level do aggregates become problematic?
- Do different types of aggregates pose different levels of concern?
- Are there clinical circumstances in which they are more problematic?

It was unclear to the participants whether very weak self-association matters in all cases. For example, if complexes fall apart after dilution in an intravenous bag, a patient may never be exposed to them. Noncovalent complexes with low association constants or fast rates of dissociation may fall into this category. On the other hand, if they are injected subcutaneously in a high-protein concentration formulation, a patient may be exposed to those weakly

associating complexes. Other important clinical considerations are frequency of dosing and patient population (e.g., whether immunosuppressed).

Most participants felt that the total aggregate load in a dose is probably a more important product-quality factor than the percentage of the protein formulation that is aggregated. This was surmised because the immune system is usually most sensitive to absolute amounts of antigen. It was agreed that the mechanism of biopharmaceutical immunogenicity is poorly understood. Perhaps experience from the vaccine industry, where the goal is to generate immune responses to foreign proteins, may shed light on this area.

Aggregates have been associated with other problems beyond immunogenicity. Examples include injection-site reactions, induction of TNF, and complement activation. Aggregation during bioprocessing can also dramatically lower process yields and/or increase process times. For example, if a product intermediate aggregates before a filtration step, it can clog the filter. Aggregates can also precipitate in a column resin. This not only decreases step yield, but it can complicate cleaning of the resin and housing unit.

Interestingly, in some cases aggregates can result in higher product activities. Nonnative aggregates of fibroblast growth factor (FGF) can be more potent than monomers. Aggregation of Von Wilbrands factor is also important for product activity. For another example, aggregation may be desirable for a MAb anti-idiotype vaccine.

Aggregation also can increase product activity. In one case example, two batches of anti-APO-1 antibody were made: an early crude prep and a second highly purified prep. In the apoptosis potency assay, only the first batch had activity. The subsequent investigation revealed that the first batch was 54% aggregated, whereas the second batch was 1% aggregated. The aggregation was traced to the low pH of the protein G elution. The second batch had fewer aggregates because later steps in the more extensive purification process removed them. It is now known that cross-linking of receptors often is important for signal transduction, such as the apoptotic signal mediated by APO-1. In another example involving an antitumor antibody, aggregates caused high levels of variability in binding activity as measured by ELISA.

It was clear from remarks from a few participants with previous experience in the plasma industry that this sector has experienced issues related to aggregates as well. For example, Factor IX made before recombinant DNA technology was 25–50% aggregated, whereas recombinant Factor IX was much less. This turned out to be an important marketing point of the new products. Perhaps one origin of the common "<5% aggregate" specification comes from experience with IVIG. Older preparation methods often resulted in batches with >5% aggregates. The presence of aggregates seemed to correlate with adverse events, whereas newer batches, generally with <5% aggregates, had fewer adverse events. In contrast, the same participants remarked that aggregate levels in human serum albumin didn't correlate with adverse events. It is hoped that more dialogue about aggregates with individuals from this sector will take place, because it would seem to be very valuable for the biotech industry.

Evidence for Immunogenicity

Evidence for enhanced immunogenicity from aggregates seems to be rather sporadic. Only two examples in which aggregates possibly affect product immunogenicity were brought up at the workshop. In the first instance, tetramers in an early formulation of an antibody product seemed to correlate with increased immunogenicity during IND studies. The tetramers were subsequently eliminated by a formulation change before marketing. In the second example, removal of a CEX chromatography polishing step from a MAb process led to an approximately two-fold increase in aggregates — up to 5% aggregates. It also correlated with a higher rate of antiproduct antibodies. However, other product quality changes also occurred, such as higher levels of host cell proteins, which may very well also have been a culprit for the increased immunogenicity.

The bottom line is that a few poorly reported, anecdotal cases don't prove a

Spelling Them Out

Here are some abbreviations used in this article.

AUC: analytical ultracentrifugation

CD: circular dichroism

CDR: complementarity-determining region

CE: capillary electrophoresis

DSC: differential scanning calorimetry

FFF: field-flow fractionation

FGF: fibroblast growth factor

FTIR: Fourier transform infrared spectroscopy

HAHA: human antihuman antibody

HACA: human antichimeric antibody

HIC-HPLC: hydrophic-interaction HPLC

HPLC: high-performance liquid chromatography

IEX-HPLC: ion-exchange HPLC

MAb: monoclonal antibody

MALDI-TOF: matrix assisted laser-desorption ionization–time of flight

PEG: polyethylene glycol

RP-HPLC: reversed-phase HPLC

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC-HPLC: size-exclusion HPLC

trend. It behooves industry to be forthcoming about its experiences in this area to benefit the public health and the overall drug development process.

Does Immunogenicity Always Matter?

The conclusion of the workshop was devoted to potential implications of immunogenicity. This occasionally seems like a mythological question. The first example cited above implies that the answer is "sometimes." In this case, very high titers of antiproduct antibodies correlated with serum-sickness–type reactions. The nature of the antiproduct antibodies is a consideration. For example, neutralizing antibodies are more likely to affect product efficacy than are nonneutralizing antibodies. It has been the experience of the FDA that antirecombinant protein antibodies are largely nonneutralizing, whereas anti-MAb responses are largely neutralizing. This occurs because the neo-epitopes on most MAbs are the CDR surfaces. Thus, HAHA and HACA are mostly antiidiotypes, neutralizing the binding site.

The participants felt that clinical data are paramount when assessing antiproduct immune responses. For example, a low titer immune reaction probably does not represent a major clinical issue. Other factors that the FDA considers include availability of other treatments (e.g., microbial replacement enzymes) and existence of redundant pathways (e.g., EPO). Monitoring nonneutralizing antibody responses is still important because they can effect pharmacokinetics and efficacy. In addition, it was pointed out that antibodies can promote Fc-mediated antigen-presenting cell uptake, leading to even higher immune responses.

*Corresponding author **Kurt Brorson** is a staff scientist at the Office of Biotechnology Products, CDER, FDA, Bethesda, MD; kurt.brorson@fda.hhs.gov. **Joseph Phillips** is senior director, Analytical Sciences, Amgen, Inc., Thousand Oaks, CA.*

Disclaimer

The content of this manuscript reflects discussions that occurred during the CMC Forum workshop in addition to the personal viewpoints and experiences of the authors. This document does not represent officially sanctioned FDA policy or opinions and should not be used in lieu of published FDA guidance documents, points-to-consider documents, or direct discussions with the agency.