Biosimilars, lower-priced versions of currently marketed biologic drugs, have vast potential to lower healthcare costs and give more patients access to life-changing treatments.

But developing biosimilar products is challenging, in part due to the difficulty of establishing physicochemical and functional similarity to the originator product. As biological molecules increase in size from simple, non-glycosylated proteins of less than 200 amino acids to glycosylated proteins containing more than 1,000 amino acids (such as monoclonal antibodies), challenges become increasingly complex and expensive.
Biosimilars are highly regulated throughout the world and the exact requirements for demonstrating analytical similarity and therapeutic equivalence differ from one country, or region, to the next. Companies must navigate a regulatory maze in order to produce a globally relevant data dossier in a cost-effective manner. This problem is exacerbated by the fact that similarity standards change as regulators gain experience with biosimilars and identify new issues. Agency decisions, especially in the US and EU, can set precedents that affect the size, scope and cost of biosimilar programs worldwide; companies need to keep up with them and recalibrate their programs accordingly.

To generate evidence of analytical similarity that can potentially smooth a biosimilar’s, not a new drug’s, regulatory path, reduce clinical data requirements (thus paring development costs) and maximize the chances of a successful global product launch, developers should follow the three steps illustrated to the right.

To begin, developers must demonstrate that a biosimilar protein has an amino acid sequence that is identical to the original product, although a very small proportion of protein molecules may display minor sequence differences. These differences are generally due to transcription errors and are so infrequent that they mostly go undetected. Alterations in differences in N- and C-terminal amino acids – may be acceptable, if they can be justified.

Even with a perfectly matched sequence, variants will still exist; typically, these can arise due to modifications of amino acid side chains. It follows that the active substance comprises not one molecular species but a whole population. The makeup of that population always will vary from batch to batch of the same product, let alone for products coming from different manufacturing sources, as is the case with biosimilars.

While most minor modifications will have no impact on safety and/or potency, others can fundamentally alter a product’s quality attributes. As a result, the impact of any differences in variant levels need to be fully characterized, well understood and convincingly demonstrated to be clinically irrelevant.

Complicating matters further, different regulatory agencies set the bar for matching attributes at different heights. The EU’s Committee for Medicinal Products for Human Use (CHMP) does not expect quality attributes to be identical; minor structural and impurity profile differences

ASSESS THE CRITICALITY OF PRODUCT QUALITY ATTRIBUTES

To begin, developers must demonstrate that a biosimilar protein has an amino acid sequence that is identical to the original product, although a very small proportion of protein molecules may display minor sequence differences. These differences are generally due to transcription errors and are so infrequent that they mostly go undetected. Alterations in differences in N- and C-terminal amino acids – may be acceptable, if they can be justified.

Even with a perfectly matched sequence, variants will still exist; typically, these can arise due to modifications of amino acid side chains. It follows that the active substance comprises not one molecular species but a whole population. The makeup of that population always will vary from batch to batch of the same product, let alone for products coming from different manufacturing sources, as is the case with biosimilars.

While most minor modifications will have no impact on safety and/or potency, others can fundamentally alter a product’s quality attributes. As a result, the impact of any differences in variant levels need to be fully characterized, well understood and convincingly demonstrated to be clinically irrelevant.

Complicating matters further, different regulatory agencies set the bar for matching attributes at different heights. The EU’s Committee for Medicinal Products for Human Use (CHMP) does not expect quality attributes to be identical; minor structural and impurity profile differences

ASSESS THE CRITICALITY OF PRODUCT QUALITY ATTRIBUTES

To begin, developers must demonstrate that a biosimilar protein has an amino acid sequence that is identical to the original product, although a very small proportion of protein molecules may display minor sequence differences. These differences are generally due to transcription errors and are so infrequent that they mostly go undetected. Alterations in differences in N- and C-terminal amino acids – may be acceptable, if they can be justified.

Even with a perfectly matched sequence, variants will still exist; typically, these can arise due to modifications of amino acid side chains. It follows that the active substance comprises not one molecular species but a whole population. The makeup of that population always will vary from batch to batch of the same product, let alone for products coming from different manufacturing sources, as is the case with biosimilars.

While most minor modifications will have no impact on safety and/or potency, others can fundamentally alter a product’s quality attributes. As a result, the impact of any differences in variant levels need to be fully characterized, well understood and convincingly demonstrated to be clinically irrelevant.

Complicating matters further, different regulatory agencies set the bar for matching attributes at different heights. The EU’s Committee for Medicinal Products for Human Use (CHMP) does not expect quality attributes to be identical; minor structural and impurity profile differences
may be acceptable, though they must be rigorously justified. Regulations in the US similarly state that minor differences in clinically inactive components may be permissible, but clinically meaningful differences in safety, potency or purity are not. In practice, however, the two agencies interpret the requirements somewhat differently.

The FDA’s first-ever biosimilar approval in March of 2015 (of Sandoz’s Zarxio, a copy of Amgen Inc.’s blockbuster, Neupogen) provides a case study for the real-world consequences of regulatory nuance. In its review of Zarxio, the FDA rolled out a four-tier critical quality attribute (CQA) ranking system that ranged from “very low” to “very high”.

Apart from amino acid sequence (which, as already mentioned, must be an absolute match), the FDA required a demonstration of equivalence for the remaining CQAs ranked as very high to be within ±1.5 standard deviations (SD) of the results obtained for the reference product. This represents a much more stringent approach than has been used by the CHMP, where rigid classifications and statistical constraints have not been applied.

Furthermore, in both the EU and US, release specifications generally allow a difference of 3 SDs between batches.

Taking protein content as an example, regulatory agencies and pharmacopeias generally allow +/- 5 to 10% variation. For Zarxio, on the other hand, the FDA’s requirement meant that 90% of batches had to fall within 2.26% of the values observed for the reference product batches tested.

What did the FDA’s stringent standards mean for Zarxio? In order to win the agency’s approval, Sandoz had to compare 15 commercial scale batches of their product against the same number of both EU- and US-sourced reference products. Zarxio met the ±1.5 SD requirement, but in some cases only marginally.

What will the FDA’s approach mean for even more complex proteins? Determining which attributes rank as highly critical is far more complex for a large glycosylated protein—such as a monoclonal antibody that exerts both target and Fc receptor binding activity—than it is for a simple mono-functional protein like Zarxio. First, there are far more attributes to compare and thus a greater statistical probability that a difference will occur by chance. Further, the impact of any differences is more difficult to assess. For example, in some instances, Fc functionality is of limited importance; in others, it is critical, and in still others its relevance is the subject of ongoing controversy. Resolving this last situation could require a detailed review of the scientific literature and even studies to dispel areas of residual uncertainty.

Figure 1 (on following page) introduces five broad product attribute categories and how individual quality attributes within these categories might be ranked.
In order to maximize the ability to detect differences between a licensed reference product and the potential biosimilar, analytical biochemists must probe potential differences with multiple techniques, exploiting different molecular properties. For example, in order to explore and measure differences in biological activity, a variety of assays may be needed to capture information about ligand and receptor binding, downstream signaling (e.g. proteins activated by phosphorylation) and resulting effects such as cytotoxicity, apoptosis, mitotic effects, enzymatic reaction rates and responses in \textit{in vivo} systems.
Once detected, any difference must be justified. If the difference is not likely to be justifiable, companies must avoid the temptation to flog a dead horse by trying to justify the unjustifiable and focus their resources on re-engineering the product to address the problem.

Differences in the formulation (e.g., buffer and pH differences due to patent restrictions) or design (e.g., the delivery device or container closure system) may be allowed, provided there is no clinically meaningful difference between the products, and the applicant can demonstrate adequate performance data. See Figure 2 for a sampling of product attribute differences between biosimilars and various reference medicinal products (RMPs) that have been deemed justifiable by EU regulators. [At this time there are insufficient precedents in the US to compile a comparable chart.]

In its assessment of Zarxio, the FDA ranked protein content and potency among the most highly critical CQAs. But the impact of protein content depends on the steepness of the dose response curve. For filgrastim (the active ingredient of Zarxio) this relationship in general is rather flat, with reports that a doubling of dose may drive only a 20% increase in pharmacodynamics effect. For some other proteins, such as insulin, the situation would be very different.

Potency is an inherent property of any biological substance and even a relatively minor difference in this attribute would raise questions as to the cause and whether it could produce secondary effects.

Changes in folding or aggregation, which may enhance immunogenicity or reduce potency, may only be discoverable with orthogonal state-of-the-art technology, including carefully designed ultra-centrifugation, field flow fractionation techniques, microscopy and hydrogen deuterium exchange. Failure to detect a mismatch in these structural attributes early in the development process could mean that the first clear warning signals won’t appear until pharmacokinetics/pharmacodynamics (PK/PD) trials or, in an even more costly scenario, late-stage clinical testing.

“POTENCY IS AN INHERENT PROPERTY OF ANY BIOLOGICAL SUBSTANCE AND EVEN A RELATIVELY MINOR DIFFERENCE IN THIS ATTRIBUTE WOULD RAISE QUESTIONS AS TO THE CAUSE AND WHETHER IT COULD PRODUCE SECONDARY EFFECTS”

Importantly, for the US and EU, reference medicinal product (RMP) must be sourced from the region in which marketing approval is sought. Three-way chemistry, manufacturing and control (CMC) testing and PK/PD bridging studies are generally required to justify use of RMP from an alternate region in a Phase 3 program. It is crucial to compare the biosimilar molecule to multiple batches of RMP in order to understand the extent of variability. Complicating this exercise, reference product samples with different batch numbers may still have come from the same fermentation run.

Figure 2 displays differences between the biosimilar and RMP (grouped into seven attribute categories) that have been accepted by EU regulators for epoetin, filgrastim follitropin alfa, infliximab and somatropin biosimilars (indicated by color coding).
SLIGHT DIFFERENCES IN QUALITY ATTRIBUTES MAY BE ACCEPTABLE

Figure 2. Examples of differences between the biosimilar and reference product that have been accepted by EU regulators for epoetin, filgrastim, follitropin alfa, infliximab and somatropin biosimilars (indicated by color coding).
CLARIFY THE STRUCTURE-FUNCTION RELATIONSHIP

The most rigorous methodology for developing biosimilars is a risk-based, stepwise totality of evidence approach that applies careful, cross-disciplinary analysis to every aspect of product characterization. In recent years, the trend in regulatory thinking has moved towards setting more pragmatic clinical and non-clinical (animal) requirements while at the same time strengthening requirements for CMC data. However, there is no evidence, to this point, of the extent to which FDA will put this thinking into practice.

For mAbs and other large, complex molecules, a comprehensive, scientifically robust CMC program is required for indication extrapolation (e.g., to allow a biosimilar clinical program targeting just one indication to support licensure in all indications for which the reference product is approved).

If there are no discernible clinically meaningful differences at either the structural or biological level, and if this finding is supported by solid PK/PD equivalence data, there should be a high level of confidence that a biosimilar will behave similarly to the reference product. Given that, extensive clinical testing in these other indications should not be necessary. Indeed, CHMP has taken this further and for less complex biosimilars, such as filgrastim, has accepted demonstration of therapeutic equivalence in all indications based entirely on PD data generated in healthy subjects and a small non-comparative patient study.

For mAbs and other complex biosimilars, the totality of evidence concept, and any resultant streamlining of clinical programs, strongly relies on convincing regulators that the analytical methods the sponsor has used possess sufficient sensitivity and selectivity to provide convincing and reliable evidence of high similarity. And when a difference is detected, the sponsor will need to demonstrate to the satisfaction of regulators that it has no clinical relevance.

In the June 2013 European Public Assessment Report on Remsima® (a biosimilar version of infliximab, an anti-tumor necrosis factor alpha monoclonal antibody), a number of differences between the biosimilar and the original product were noted, but all were considered to be of no clinical consequence: a slightly higher proportion of part-assembled antibodies were shown not to impact binding affinity or potency; a difference in C-terminal lysine levels was shown to be transient (because the C-terminal lysine is rapidly cleaved following administration). Finally, a lower level of glycans lacking the sugar fucose was observed, which translated into a lower binding affinity to one of the Fc receptors (FcγRIIIa); these discrepancies were addressed through a series of complex and well-conceived in vitro studies, conducted under representative pathophysiological conditions, which provided evidence that the relatively minor differences were not clinically relevant.
For biosimilar developers, the lure of tapping into a huge potential market—the size of which a variety of sources estimate will range from $5 billion to $25 billion by 2020—is irresistible. For healthcare systems squeezed by spiraling costs, the spread of lower-priced versions of expensive biologic therapies could cut prescription drug spending dramatically. And for patients worldwide, a thriving biosimilars industry promises to make life-changing drugs more affordable and accessible. But for these possibilities to be realized, regulators must adopt a pragmatic, balanced and well-considered approach in setting biosimilarity requirements.

If evidentiary standards expand beyond the scope of identifying clinically meaningful differences, developers may struggle to meet requirements, thereby increasing costs and reducing competition. In effect, such a development would make it more difficult to bring these needed products to patients.

Of course, regulators are guardians of patient safety first and foremost. Finding equilibrium between these competing priorities — making biosimilars more easily and quickly accessible while ensuring their efficacy and safety — is critical. In the meantime, biosimilar developers must navigate a fast-changing landscape while still conducting efficient, scientifically sound and globally relevant development programs. Following these steps will help them do that.
WHEREVER YOUR JOURNEY TAKES YOU, WE’RE CLOSE BY.

CORPORATE HEADQUARTERS
195 West Street
Waltham, MA 02451
USA
+1 781 487 9900

Offices across Europe, Asia and the Americas

www.PAREXEL.com

© 2015 PAREXEL International Corporation. All rights reserved.