# When Spectra Speak: Letting AI Read the Fine Print of Biologics

31 October 2025 | Views

Biologic characterisation has historically relied heavily on spectroscopy, which transforms chemical complexityinto comprehensible fingerprints. Through the careful use of AI, supported by strict data procedures and scientific supervision, these fingerprints may be read more quickly and thoroughly.
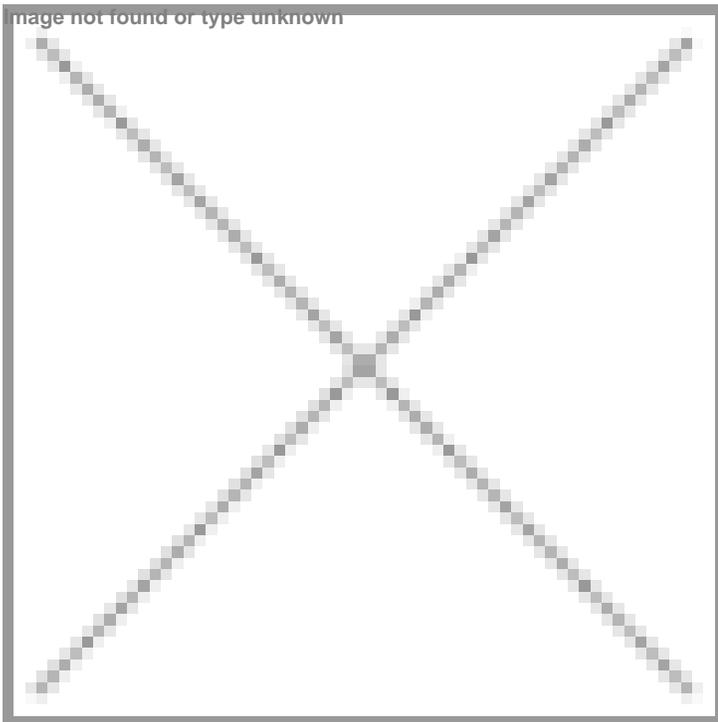
Biologics—complex therapeutic proteins such as monoclonal antibodies, peptide hormones, insulin and its analogs, and engineered enzymes—represent some of the greatest advances in modern medicine. Their ability to precisely target disease pathways has transformed treatments for cancer, autoimmune disorders, and metabolic conditions. Yet the very complexity that makes biologics so effective also makes them difficult to fully characterise and control. Unlike small-molecule drugs, whose chemical structures are relatively simple, reproducible, and easy to verify, biologics are produced inside living cells and therefore carry subtle variations. Small differences in protein folding, glycosylation patterns, or early signs of aggregation can meaningfully alter a drug's safety profile, half-life, or efficacy. Detecting these shifts is critical to ensuring that every batch released to patients meets rigorous quality standards.

Spectroscopy—shining light or other electromagnetic fields onto a sample and observing its response—provides a powerful, multidimensional portrait of biologic structure. Techniques such as Nuclear Magnetic Resonance (NMR), Fourier-Transform Infrared (FTIR) spectroscopy, Circular Dichroism (CD), fluorescence, and liquid chromatography-mass spectrometry (LC-MS) can capture critical details about a protein's shape, chemistry, and structural integrity. The result is a "spectral fingerprint" that reflects stability and subtle compositional differences across batches. However, these fingerprints generate vast amounts of data, often more than traditional methods can comfortably interpret. Scientists have historically focused on a few familiar peaks or markers, leaving broader patterns unexplored. Over the past decade, artificial intelligence (AI) has emerged as a partner in addressing this challenge—helping scientists recognise subtle signals within complex data, provided the underlying

records are complete, accurate, and trustworthy.

The Challenge of Biologics Quality and the Power of Spectroscopy

Biologics differ from small molecules in scale, complexity, and reliance on living systems. A monoclonal antibody, for example, is composed of two heavy and two light chains, each hundreds of amino acids long, folded into intricate three-dimensional structures. Post-translational modifications such as glycosylation add another layer of variability. Every manufacturing choice—cell line, growth medium, purification steps, and storage conditions—can influence these features. Small perturbations matter: a transient pH change during chromatography might shift glycan profiles, or a minor temperature excursion in shipping could trigger aggregation. These micro-differences can affect binding, circulation time, or immune response.

Ensuring consistent quality requires more than single-attribute assays. Traditional methods like size-exclusion chromatography, peptide mapping, and bioassays remain indispensable, but each examines only one attribute at a time. A folding disruption may leave aggregation assays unchanged yet show up as distributed shifts across spectra. Spectroscopy provides a complementary view, offering broad, orthogonal insights into molecular integrity:

- NMR probes local chemical environments and can reveal whether regions remain folded or flexible.
- FTIR detects bond vibrations, with amide bands reporting on backbone conformation and secondary-structure content.
- CD summarises overall folding and highlights structural changes induced by stress.
- Fluorescence tracks how aromatic residues like tryptophan become more exposed or buried, signaling unfolding or early aggregation.
- LC–MS often coupled with ion-exchange or two-dimensional separations-can map glycoforms, identify charge variants, and detect truncations with high sensitivity.

Each technique produces hundreds to thousands of data points per sample. Together, they generate high-dimensional datasets that capture multiple critical quality attributes (CQAs). But without systematic approaches, these rich fingerprints can overwhelm manual analysis, leaving blind spots. This is where AI shows promise—assisting scientists in navigating the data's complexity, provided the information feeding these systems is of high quality and well-curated.

AI as a Responsible Partner in Spectral Interpretation

Artificial intelligence can detect patterns that are too subtle or complex for humans to recognise consistently. Instead of focusing on a handful of peaks, AI can review the entire spectrum and identify distributed differences or drifts. This capability enables scientists to classify batches more efficiently, spot anomalies earlier, and monitor gradual changes. Yet AI is only as good as the data it is trained on—if the inputs are incomplete or inconsistent, the insights may be misleading.

Broadly, AI approaches applied to spectral data fall into three categories:

- Exploratory methods simplify large datasets into recognisable patterns, showing clusters of similar samples and exposing outliers.
- Predictive methods use past examples to forecast outcomes. A model might learn which spectral fingerprints correspond to acceptable versus unacceptable batches, and then classify new samples accordingly.
- Anomaly detection methods learn what "normal" looks like and flag anything that deviates significantly.

AI's value grows when it can explain its conclusions. Modern tools highlight which spectral regions contributed most to a decision, helping scientists trace a flagged result back to a specific structural feature or bond type. This transparency is vital in regulated environments, where credibility and reproducibility matter as much as the result itself.

For AI to be trusted, it must be rigorously trained and validated. Models should be built using representative data that capture the full range of expected variability—different instruments, operators, and buffer conditions. Performance must be measured and documented, with models revalidated when conditions shift. Most importantly, AI should be viewed as a decision-support tool rather than a decision-maker. Experts should review every AI alert and confirm with orthogonal assays if necessary. In areas like biosimilar development, AI-based spectral analysis can strengthen evidence of similarity, but it cannot replace the broader evidence package regulators require.

From Rapid Insights to Lasting Confidence

The promise of AI-enhanced spectral analysis lies not only in speed but also in improving decision quality. Properly integrated, AI can accelerate routine analysis, strengthen regulatory submissions, and enhance confidence in product consistency. However, this is only possible if the data are high-quality, complete, and unbiased.

AI-enabled, PAT-driven spectral analysis can substantially shorten review and release timelines in real-time testing settings. In stability programmes, frequent spectral monitoring can uncover early shifts long before conventional assays show changes. In comparability exercises, AI can rapidly analyse thousands of spectra, allowing scientists to focus on a small subset of edge cases. Yet speed should never overshadow transparency. Scientists, regulators, and partners must understand which spectral features drove a decision to act confidently.

Looking ahead, the integration of AI and spectroscopy could extend even further. Digital twins for bioreactors and downstream steps are moving from pilot to practice under PAT/Pharma 4.0 initiatives, though adoption is in early stages. Future systems may even adjust process parameters dynamically to maintain quality. Collaborative approaches, where learning is shared across sites without centralising sensitive data, could help models improve globally. These innovations rest on the same foundation: disciplined data collection, careful validation, and transparent communication of results.

Final Reflections

Spectroscopy has long been central to biologic characterisation, turning molecular complexity into interpretable fingerprints. With the thoughtful application of AI—anchored by rigorous data practices and scientific oversight—these fingerprints can be read with greater depth and agility. The role of AI is not to replace expert judgment but to extend it: surfacing hidden patterns, accelerating routine tasks, and letting scientists focus on the most critical questions. In a field where even tiny molecular differences can influence patient outcomes, combining advanced analytics with strong data stewardship offers a pathway to both rapid insights and lasting confidence.

Authors-

- Dr Anirban Mudi, Lead Platform Product Manager – Next Gen & AI, IDBS, Bengaluru
- Prof. Ashutosh Kumar, Department of Biosciences and Bioengineering, IIT Bombay