

ACD/LABS [ADVANCED CHEMISTRY DEVELOPMENT, INC.]

Building a Foundation for Autonomous Structure Verification with ASV

Sarah Srokosz, ACD/Labs

Across AstraZeneca (AZ), the journey towards automating NMR analysis workflow began two decades ago when they started digitalizing their NMR data. Since then, they have continued to pave the way towards their ultimate automation goals—such as a fully autonomous structure verification system— by improving and expanding upon these digitalization efforts.

| 2004 Digitized NMR Spectra | 2014 Digitalized Live Analytical Data (for some) | 2024 Global Analytical Database (GAD) |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>NMR spectra exist as scanned images or PDF files of data processed and analyzed manually.</p> <ul style="list-style-type: none"> • Hard to find desired spectra • Limited audit trail/file info/metadata • No chemical shifts or <i>J</i>-couplings • Can only do visual comparison (no changes or expansion, poor quality) | <p>Spectrus Platform applications rolled out across global pharmaceutical development.</p> <ul style="list-style-type: none"> • Live data stored in a searchable database • Full parameter details, with metadata (chemical shifts, <i>J</i>-couplings, integrals, etc.) • Interactivity | <p>Oncology R&D and Biopharma R&D discovery teams share searchable database of live analytical data.</p> <ul style="list-style-type: none"> • Automatically captures all raw NMR and LC/MS data • Chemistry across the company now has a FAIR small molecule analytical database |



Amber Balasz

Director of US Analytical, Structural, and Chromatography Team & NMR Specialist
AstraZeneca Oncology R&D
Boston, USA



Richard Lewis

Principal Scientist
AstraZeneca Biopharma R&D
Gothenburg, Sweden

Produced by ACD/Labs based on talks given by Amber Balasz (April 7th, 2024 at the ACD/Labs Software Symposium at the 65th Annual Experimental Nuclear Magnetic Resonance Conference) and Richard Lewis (October 24th, 2024 at ACD/Labs' Structure Elucidation and Verification Virtual Symposium).

Accelerating Structure Verification without Compromising Accuracy

In AstraZeneca's Oncology and Biopharma R&D organizations, there is an effort to accelerate the make phase of the design-make-test-analyze (DMTA) cycle. To help achieve this, they were particularly interested in automating their post-purification workflow.

In this workflow, once analytical data is acquired, chemists must verify the proposed structure before they can register the compound. At the Gothenburg site alone, chemists are verifying the structures of hundreds of compounds per week. Additionally, as more of their chemistry becomes automated, the amount of data chemists must deal with continues to increase.

While there is a substantial need to accelerate structure verification, this cannot come at the expense of accuracy, as having an incorrect structure misleads design teams, which is bad for both the organization and patients.

Keeping Fully Autonomous Structure Verification in Sight

The short-term goal in implementing automated structure verification (ASV) with NMR Workbook Suite™ was to accelerate their structure verification workflow and reduce the burden of structure verification on chemists. In the long term, AZ's goal was to eliminate the need for a human to spend time verifying known structures altogether. This means they are working to build a system that they can rely on to make decisions from analytical data. Relying on this system means that it either needs to produce accurate results 100% of the time, or that it can be wrong occasionally, but it needs to be able to identify these cases, prompting for more data and/or human interpretation.

Optimizing ASV Performance

With both shorter- and longer-term goals in mind, AZ are investigating ways to further optimize the accuracy of their ASV system (i.e., minimize the number of false results).

One way to increase accuracy is to add more analytical data. However, they want it to work in high-throughput settings, so this is not realistic. While they use ASV in conjunction with MS data to confirm molecular formula, this doesn't help ASV with the difficult task of distinguishing between structural isomers, which is important when analyzing reaction products. So, AZ recently undertook in-house performance testing to investigate other changes to the system or input data that could help them get closer to their goal. (Table 1)

Table 1. Summarized results of in-house ASV optimization experiments.

| ASV Factor | Results and Conclusions |
|-----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input Data | <p>In isolation, ^{13}C data provided best accuracy, compared to ^1H or HSQC data, possibly because of better resolution, more robust predictions due to broader shift range of ^{13}C vs ^1H spectra, and because ^{13}C shifts are less influenced by 3D conformational effects compared to ^1H.</p> <p>While none of their tests in this area produced sufficient accuracy for an autonomous system, it did lead them to conclude that they should always include HSQC data and add additional weighting biases to ^{13}C shift assignments.</p> |
| Single Structure Verification (SSV) vs. Combined and Concurrent Structure Verification (CCV) | <p>Because they are focused on distinguishing isobaric structural isomers resulting from synthetic reactions, they tried to approach things more like a human would. Instead of looking at a single structure and trying to evaluate how well it corresponds to the data with nothing else to compare it to, they looked at how accurate the ASV system was at distinguishing pairs of compounds using the differences in the match factor (MF) score from ASV.</p> <p>They found that when using HSQC or ^{13}C data in this way, the system approached the level of accuracy required for an autonomous system.</p> |
| Peak Picking Mode | <p>MFs improved for automatic peak picking as more data was included, but manual peak picking outperformed automatic peak picking, regardless of what data was included in the dataset.</p> <p>However, as it is not practical to manually pick peaks in a high-throughput environment, this underscored the importance of optimizing the advanced parameter settings to allow the system to better accommodate a wider variety of projects/circumstances without significantly increasing risk.</p> |
| Prediction Algorithm | <p>Using the neural network or Hierarchically Ordered Spherical Environment (HOSE) code algorithm provided similar performance.</p> |

An ASV System for Current and Future Goals

Equipped with these insights, AZ began an ASV system pilot across their global discovery organization in December 2023. (Figure 1) This system automatically:

1. Pulls in a 1D ^1H spectrum and a carbon-edited HSQC from the GAD
2. Processes and analyzes the spectral data
3. Creates a customized review-ready report that ranks a Chemformer-generated set of chemically-probable predicted reaction products

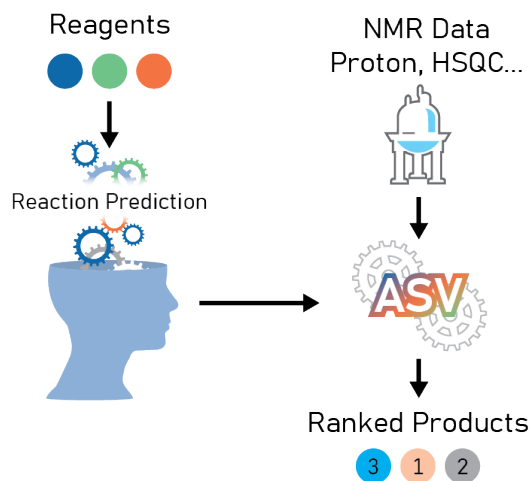


Figure 1. Post-purification structure verification workflow at AstraZeneca.

Chemists review the report and can interactively adjust anything they want in the live data. They can even add or subtract structures from the verification set. This allows them to start analysis at an “edit & review” mindset instead of starting at the beginning with raw data, which accelerates this step of the workflow without increasing the risk of misleading downstream work with incorrect structures.

The Future of ASV is Bright

AstraZeneca is excited about the future of ASV. They are focused on the next steps towards an autonomous system for structure verification, like implementing the ability for their ASV system to make “smart” suggestions to and decisions for the chemist. They believe the key to advancing towards their long-term goals is using all the information available (e.g., synthetic, analytical data) and believe the future state of ASV likely mixes approaches. So, in the meantime, enabled by their previous digitalization work, they continue to expand their in-house testing and optimization with this in mind and are currently receiving promising results from their investigations of incorporating other kinds of analytical data, such as IR.

“Working together, we can implement some automation now while continuing to improve in the future.” – Amber Balazs