RSC INTEREST GROUP
NMR DISCUSSION

Structure 2026: Automated Interpretation of Spectroscopy Data

Friday 27 March 2026 • The Discovery Centre, AstraZeneca, Cambridge

9.30	**Registration and Coffee**
10.30	**Welcome and Introduction**
10.35	**Prof. Jonathan Goodman , University of Cambridge** Molecules with confidence: DP4 and structural assignment
11.05	**Dr Marvin Alberts, IBM Research** Spectra to Molecule: Language Models for Automated Structure Elucidation
11.35	**Dr Markus Blatter, Novartis** Automated Structure Verification (ASV) in NMR - A full stack implementation at Novartis
12.05	**Lunch, Posters and Vendor Exhibition**
13.30	**Dr Jonathan Martens, Radboud University** Infrared ion spectroscopy for structure elucidation - IR spectra from a mass spectrometer
14.00	**Prof. Kate Kemsley, University of East Anglia** Automated Structural Assignment and Verification using AI-Predicted NMR Chemical Shifts
14.30	**Tea, Posters and Vendor Exhibition**
15.00	**Dr Richard Lewis, AstraZeneca** Automatically verifying compounds in Pharmaceutical Discovery: Which approach; Which data?
15.30	**Prof. Jacqui Cole, University of Cambridge** AI-driven Materials Characterisation
16.00	**Poster Prize and Close**

ACD/Labs®

BRUKER

SciY

magritek

Jason

Structure 2026 – Speaker Abstracts

Molecules with confidence: DP4 and structural assignment

Prof. Jonathan Goodman, University of Cambridge

Calculating the NMR spectrum expected for a molecule is helpful in structural determination but only a part of the process. Starting from a calculated spectrum, what structural information can be determined with high confidence? The widely used DP4 method assigns the probability of assigning an experimental spectrum to each of a list of calculated spectra. This works very well, provided that the assumption that the correct structure is on the list is true. We have now developed DP5 which does not require this assumption: given one experimental spectrum and one trial structure, what is the probability they correspond? We are now developing a step beyond this: starting from an experimental spectrum and the wrong structure, what is the right structure? These studies are now being extended to other analytical measurements including IR spectroscopy.

Spectra to Molecule: Language Models for Automated Structure Elucidation

Dr Marvin Alberts, IBM Research & NCCR Catalysis, Switzerland

Structure elucidation is integral in the day-to-day operation of any organic chemistry laboratory allowing the structure and composition of unknown substances to be determined. Most commonly this is achieved via different spectroscopic techniques. First among them Nuclear Magnetic Resonance (NMR), Infrared (IR) spectroscopy and Mass Spectrometry (MS). While the acquisition of the spectra has been largely automated, the analysis of them is not straightforward making the analysis of spectra, particularly in large quantities, a time-consuming and tedious undertaking.

We have developed AI-driven approaches to tackle these challenges. We demonstrated for the first time that molecular structures can be reliably predicted from IR spectra alone, a task previously impossible for human chemists due to the limited interpretable information present in IR spectra. This approach was extended to the deconvolution of IR spectra of mixtures, where we developed a method to predict the components present in a mixture based solely on the IR spectrum and a molecular formula. However, chemists routinely have access to multiple analytical instruments. To leverage this, we developed a multimodal framework integrating NMR and IR spectra, which achieved performance matching expert human chemists in structure elucidation tasks. Together, these advances point toward a future where AI tools can aid and assist chemists in structure elucidation, freeing them to focus on interpretation and discovery rather than routine analysis.

Automated Structure Verification (ASV) in NMR - A full stack implementation at Novartis

Dr Markus Blatter, Novartis

In medicinal chemistry, fast and reliable NMR data acquisition and interpretation are essential for structure verification and decision-making. Automated Structure Verification (ASV) using NMR data is thus emerging as a key asset in accelerated workflows. We present a full-stack ASV implementation that seamlessly integrates open-access NMR workflows, optimized for speed, robustness, and user acceptance. Performance is enhanced through a discriminative metric system, enabling differentiation of closely related chemical structures. The workflow begins with experiment selection and continues through acquisition and processing parameter optimization to ensure consistent data quality. A scalable data pipeline — featuring parallel export protocols and server-based ASV nodes — provides rapid access to analytical results. To support adoption among experienced users, a tailored reporting system delivers clear and accessible ASV outputs. The modular architecture enables seamless integration of future technological advances, ensuring long-term sustainability and flexibility. Overall, we demonstrate an unattended, orchestrated workflow where all 1D and 2D NMR data generated on open-access instruments are automatically processed and analyzed by ASV (ACD/Labs). This high-throughput verification of chemical structures reduces manual workload and supports standardized, high-quality reporting.

Infrared ion spectroscopy for structure elucidation - IR spectra from a mass spectrometer

Dr Jonathan Martens, HFML-FELIX & Radboud University

Infrared ion spectroscopy (IRIS) is an emerging mass spectrometry-based technique that combines the sensitivity and selectivity of mass spectrometry with the structural specificity of infrared spectroscopy. In IRIS, mass-selected ions are irradiated using a tunable infrared laser while isolated inside an ion trapping mass spectrometer. The resulting wavelength-dependent photofragmentation yields an infrared spectrum that provides direct information on functional groups and molecular connectivity, making IRIS highly complementary to conventional MS and tandem MS (MS/MS).

A key advantage of IRIS is its ability to deliver structural information directly from complex samples and at low analyte concentrations, where nuclear magnetic resonance spectroscopy is often impractical. In contrast to MS/MS fragmentation data, which remains difficult to predict reliably in silico, infrared spectra of candidate molecular structures can be accurately predicted using routine computational chemistry methods. Comparing these predicted spectra with experimental IR spectra of unknowns, provides tentative structure assignments without the need for chemical standards. More recently, machine-learning models for IR spectral prediction have further accelerated interpretation and opened routes toward more automated workflows.

In this presentation, I will introduce the principles of IRIS with a focus on practical implementation, applicability, and data interpretation. While the technique was originally developed using high-power, widely tunable infrared free-electron lasers available at large-scale facilities such as the FELIX Laboratory, ongoing advances in infrared laser technology are increasingly enabling IRIS on compact, table-top systems suitable for routine laboratory environments. Several case studies from biomarker discovery in inherited metabolic disorders will be presented to illustrate how IRIS resolves structural ambiguities that persist after LC-MS/MS analysis, including the differentiation of isomeric metabolites and the identification of unknown compounds.

Automated Structural Assignment and Verification using AI-Predicted NMR Chemical Shifts

Prof. Kate Kemsley, University of East Anglia

In this talk, we present a fully automated framework for the structural assignment and verification of small molecules. The pipeline begins with full or partial multiplet analysis of high-resolution ^{1}H, ^{13}C, and HSQC experiments to extract chemical shifts, integrals, and substitution classes. Atom-level ^{1}H and ^{13}C predictions are generated using ensembles of graph convolutional neural networks of the message-passing class [1]. Trained on chemical shift annotations from hundreds of thousands of compounds, these models offer high prediction accuracy across a wide range of atomic environments. Assignment proceeds via a metaheuristic genetic algorithm, in which the optimal solution minimises the predictionobservation differences while accounting for prediction confidences, degrees of substitution, and integral information. The resultant assignment scores are expressed probabilistically using density functions that model empirical prediction-error distributions derived from a diverse reference collection of ~1500 NMR datasets. The approach is flexible and can handle spurious, poorly resolved, and missing signals in the experimental data. Further, all steps are automated and scalable to large-scale assignment and cross-verification.

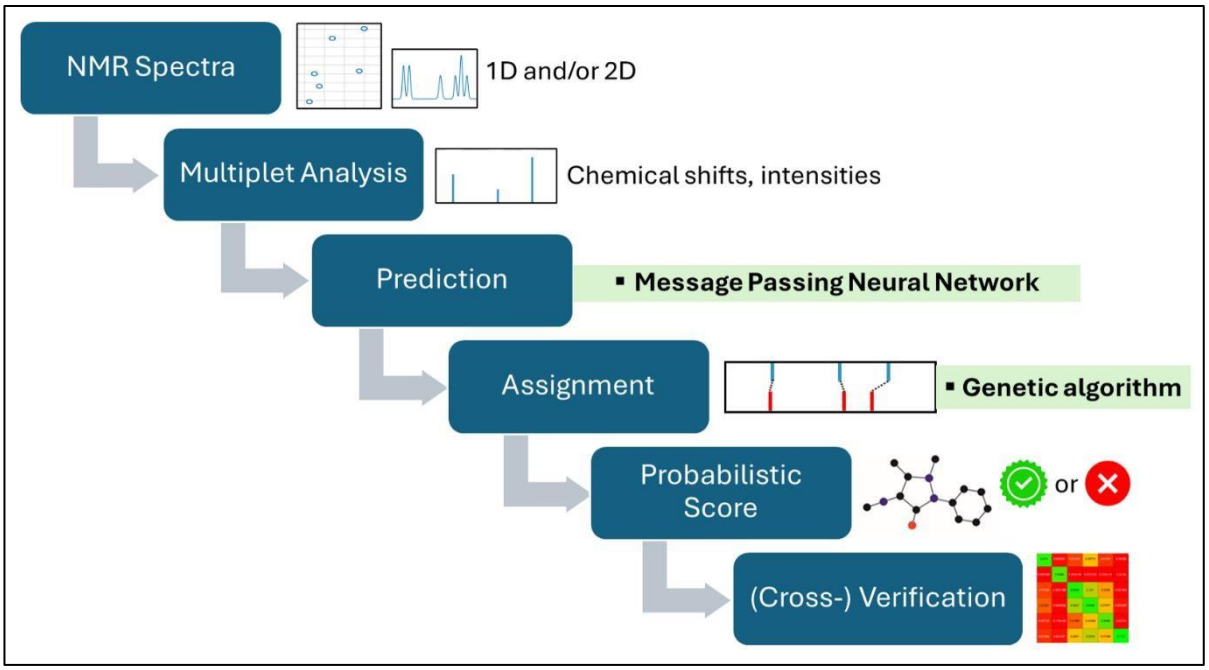

Figure 1. Schematic of the assignment and verification pipeline

[1] **D Williamson, S Ponte, I Iglesias, N Tonge, C Cobas, & EK Kemsley.** (2024). *Chemical shift prediction in ^{13}C NMR spectroscopy using ensembles of message passing neural networks (MPNNs)*, Journal of Magnetic Resonance, 368, 107795. doi: 10.1016/j.jmr.2024.107795

Automatically verifying compounds in Pharmaceutical Discovery: Which approach; Which data?

Dr Richard Lewis, Biopharmaceuticals R&D, AstraZeneca, Gothenburg

Software to interpret analytical data has been in development for decades. Despite some successes they have not produced the step change in productivity needed to keep up with synthetic output that is increasingly coming from automated methods. However, current advances in AI and other methods make this field the most exciting and innovative that it has ever been.

The talk will start with an overview of methods for structure verification and determination and the use cases for each. We make the case that one of the early methods called automated structure verification (ASV) still has relevance amongst newer AI and machine-learning tools.

Previously, ASV has been exclusively applied to NMR data. Here we describe our work expanding ASV to infra-red data. Infra-red is a sensitive technique providing a structural fingerprint that requires less material than proton NMR. We describe an automated algorithm that scores experimental infrared spectra against ab initio calculated spectra. We investigate how well the infrared data can distinguish between isomeric compounds and compared this result to the performance of proton NMR. We find that infrared is as powerful as proton NMR at distinguishing the isomeric compounds and furthermore that the combined techniques outperform either proton NMR or infrared alone suggesting a complementarity in the information content of both. As part of this work, we developed a data visualization technique (the structure classification characteristic, SCC) which neatly summarizes in a visual and numeric manner how well particular analytical data can distinguish between candidate compounds. This can be applied to any analytical data that can score compounds and facilitate identification of the most valuable, time-efficient analytical data to collect.

The talk concludes by showing how machine learning models give valuable input to the ASV process by suggesting reaction products or suggesting alternative products consistent with the analytical data. We describe the currently implementation at AstraZeneca and how we are expanding this to IR data.

AI-driven Materials Characterisation

Prof. Jacqueline Cole, University of Cambridge

This talk showcases the development and application of machine-learning models that automate the (a) spectral classification of raw data from molecular spectroscopy; (b) determination of size and shape of nanostructures from small-angle scattering data.

The former demonstrates the power and efficiency of convolutional neural networks (CNN) to automate the spectral classification of raw data acquired from Fourier transform infra-red (FTIR), 1H and 13C nuclear magnetic resonance (NMR) spectroscopy. The raw spectrum of an unknown organic chemical is the sole input to the CNN model.

The latter presents the development of a new transformer model, SASformer, that can classify the size and shape of a nanostructure using a raw small-angle scattering (SAS) profile as the sole input. This overcomes a key scientific bottleneck in the materials characterisation of nanostructures because conventional SAS data analysis requires the human to present guesses of the shape and size at the input stage to data analysis. The SASformer can automatically use the raw SAS profile of an unknown nanostructure to inform the human of the likely shape and size prior to conventional SAS data analysis. This data-driven scattering-model classification and parameter regression enables new capabilities in structure determination of nanomaterials as well as enhancing efficiency in SAS data analysis.

Overall, this portfolio of AI-driven classification of spectroscopy and scattering data offers several contributions towards the full automation of materials characterisation.

[1] G. Jung, S. G. Jung, J. M. Cole, “Automatic materials characterization from infrared spectra using convolutional neural networks”, Chem. Sci., 2023, 14, 3600-3609.

[2] S. Liu, J. M. Cole, “Automated Determination of the Molecular Substructure from Nuclear Magnetic Resonance Spectra Using Neural Networks”, J. Chem. Inf. Model. 2025, 65, 16, 8435-8447

[3] B. Yildirim, J. Doutch, J. M. Cole, “Multi-task scattering-model classification and parameter regression of nanostructures from small-angle scattering data”, Digital Discovery, 2024, 3, 694-704