Automated QC with Compound Structure Verification and Quantitation

Carlos Cobas, Felipe Seoane, Esther Vaz, Stanislav Sýkora†, and Michael Bernstein
Mestrelab Research, S.L Feliciano Barrera 9B – Bajo, 15706 Santiago de Compostela, Spain ; †Extra Byte, Via Raffaello Sanzio 22C, Castano Primo, Italy I-20022.
mike@mestrelab.com

Introduction

There is a growing need to perform NMR analyses faster, and under full automation control. Advances in available hardware have contributed favourably to this objective, and the routine collection of high-resolution NMR data on large compound libraries is now a relatively simple matter. But without reliable software tools to automatically analyse the data the full benefit of this exercise cannot be realised. Furthermore, this analysis may be part of a broader requirement encompassing other analytical steps, access to databases, etc.

We will describe here a software tool that automatically and simultaneously performs these very important analytical tasks:

- Verification of compound structure
- Concentration determination
- Update of spectral database with results

Automatic Structure Verification (ASV)

This functionality has been under development for a few years. We have adopted an approach that can use some or all of the following data:

- ^{1}H NMR spectrum
- 2D ^{1}H-^{13}C HSQC spectrum
- LC-MS and LC-UV data

These individual tests are then combined to derive an overall score that reflects how well the spectroscopic data and putative structure are in accord.

^{1}H spectrum analysis in ASV

By far the most common input data for ASV is a conventional ^{1}H NMR spectrum. This is information rich and the analysis requires the following stages (Fig. 1):

Extraction of the solute spectrum. We use our proprietary and highly efficient Global Spectrum Deconvolution (GSD) algorithm technology to fully deconvolve each peak in the spectrum. Our "Autoclassify" procedure then classifies each peak in the spectrum as to whether it relates to compound, impurities, solvent, artefact, or ^{13}C satellites. Peaks from labile protons are distinguished. Relevant solute peaks only can then be used to construct a synthetic spectrum for further analysis.

Spectrum prediction. We use the Modgraph software suite to predict the NMR spectrum of the molecule. This advanced software is a combination of a number of algorithms and chemical shift databases, allowing us to select the best protocol. The shift databases may be augmented with data for a user's particular compound classes.

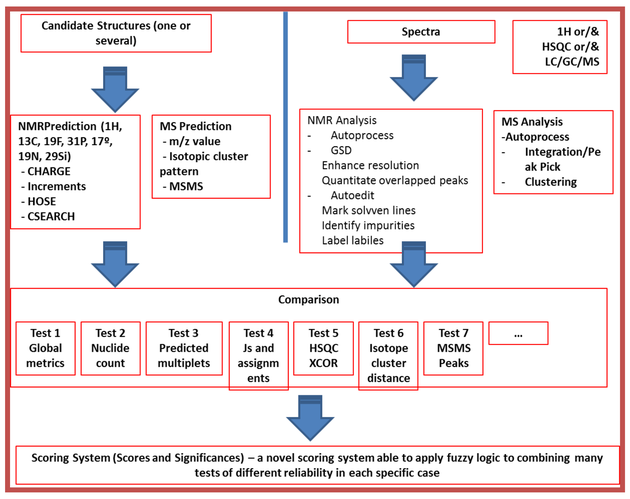

Fig 1. ASV schema.

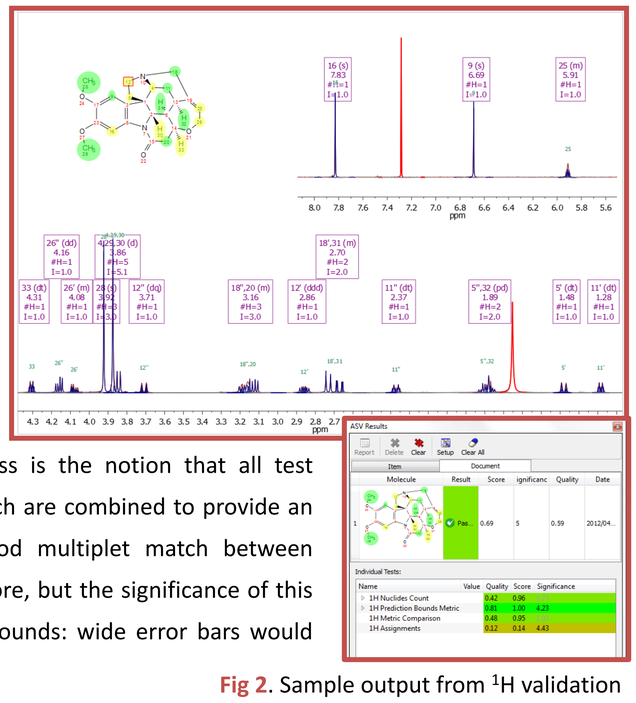

Fig 2. Sample output from ^{1}H validation

Scoring system. Underpinning this complex process is the notion that all test results are assigned a *score* and a *significance*, which are combined to provide an overall *quality*. In a simple example, a very good multiplet match between experimental and predicted may result in a high score, but the significance of this test could be modulated by the prediction error bounds: wide error bars would lower the significance.

2D ^{1}H-^{13}C HSQC spectrum

Using data from a 2D heterocorrelation experiment provides some challenges in identifying and peak picking real cross-peaks, and artefacts. The data allow a number of ^{1}H spectral features to be checked, such as labile proton signal identification, and the derived ^{13}C chemical shifts can be scored against values from a predicted spectrum.

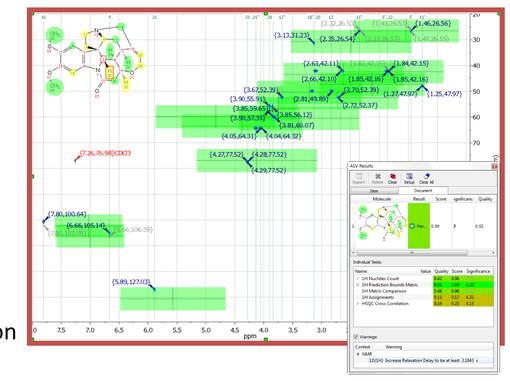

Fig 3. Sample output from ^{1}H-^{13}C HSQC validation

LC-MS data

These data may be used alone for ASV or with NMR data. The first stage in this analysis is to detect whether or not a mass ion can be found that matches the compound formula (together with permitted adducts). The isotopic cluster is evaluated, and a match score derived. Next the extracted ion chromatogram is generated and used with the total ion chromatogram (TIC) and the LC-UV spectrum to determine whether or not this ion corresponds to a significant component in the solution. These data are then analysed to determine whether or not the LC-MS support the putative structure, and appropriate scores and significances are apportioned.

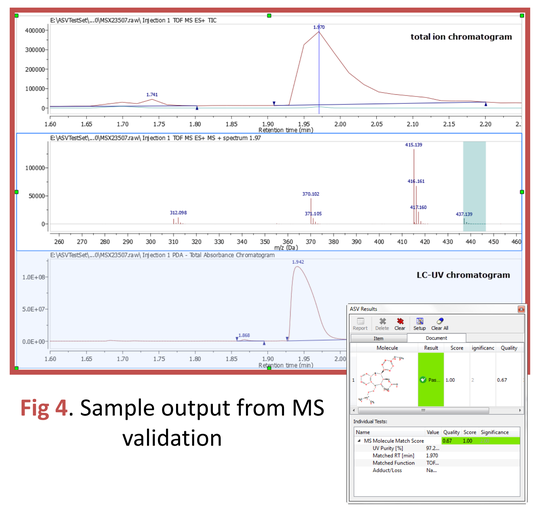

Fig 4. Sample output from MS validation

Ways to run ASV

We have built a qNMR "engine" that can be used in all common methods of compound concentration determination, such as absolute integration, and reference peaks. Each multiplet in the spectrum is automatically ranked as a candidate for qNMR analysis, and an overall concentration and accuracy are automatically determined.

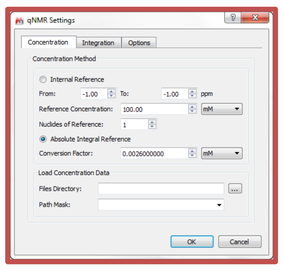

Fig 5. Setup of qNMR parameters

Quantitation

1. Using the User Interface (UI). This is the simplest approach, and a reporting table provides test results, and points to areas which might have contributed to a low score for a test.
2. Batch operation. Here the user directs the software to the available spectral data, and putative structures. Information for qNMR is supplied if this analysis is to be performed.
3. Reporting option. These include spectrum saving, generation of a summary results plot, and the possibility to update a Mestrelab database with the results.
4. Preliminary work is going into the use of ASV in a more automated way, perhaps relying the analysis being included in a pipeline process. This might use technology such as Pipeline Pilot, or Knime.

Conclusions

We have devised a powerful and flexible software functionality that provides a total NMR analytical solution. This takes the form of robust and sophisticated algorithms to determine whether or not a structure is in concord with the available spectroscopic data. Quantitation can be applied, and the final data may be added to a spectroscopic database for further interrogation.

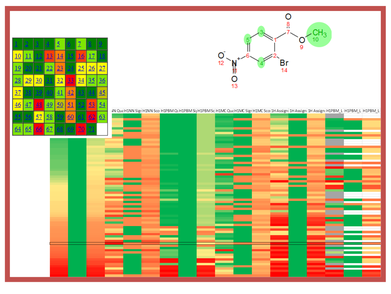

Fig 6. Some ASV reporting operation and output

This document's DOI: 10.3247/SL4Nmr12.003