Advancing purity assessment with Quantum chemistry and AI

Cameron Robertson [1], Gustavo Martos[1], Ralf Josephs[1], Robert Wielgosz[1]
[1]Bureau International des Poids et Mesures (BIPM)

Introduction

Certified reference materials (CRMs) underpin metrological traceability chains, ensuring accurate quantification of organic substances in critical sectors such as pharmaceuticals, food safety, and environmental testing. Over the past decade, qNMR has emerged as an alternative technique for purity assignment to mass balance approaches, offering unique advantages in versatility, semi-nondestructive analysis, and rapidity. Its adoption as a primary method has been further validated through generation of the BIPM Octad CRM project with contributions from global NMIs. This project established eight CRMs to standardize qNMR workflows across organic compounds [1,2].

Recent advancements in quantum chemistry and artificial intelligence (AI) are now poised to bring substantial magnetic strength independent sensitivity and resolution improvements to qNMR methodologies. Quantum mechanics enables precise automated prediction and calculation of qNMR parameters [3], increasing scope of potential precision and uncertainty improvements. Quantum mechanics also opens the door to AI-driven data acquisition and processing manipulation for improving sensitivity and resolution in different ways.

Objectives

The primary goals of this work are:

- Enhance the accuracy of purity assignment.
- Reduce uncertainty of purity assignment.
- Support BIPMs Octad library development.
- Improve qNMR workflow automation and ease of use for diverse user base across NMIs and comparison labs.
- Set new benchmarks for international metrology through computational innovation.

These goals will be achieved through:

- Validating computational workflows through case studies on model systems (e.g. MA/KHP and DMTP/BTFMBA) using cosmic truth https://ctb.nmrsolutions.fi/
- Compare qNMR purity assignments values for QMSA and manual integration methods.
- Refine QMSA and integral methods with support of AI and machine learning (ML) improving data quality.
  - Addressing critical challenges, such as spin system recognition, line shape artifacts, and impurity quantification.

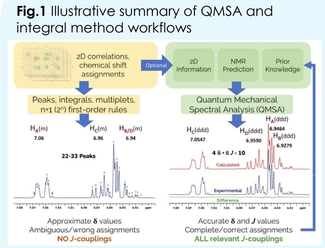


Fig.1 Illustrative summary of QMSA and integral method workflows

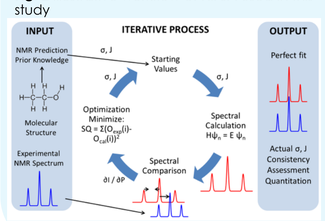


Fig. 2 Illustrative workflow of HiFSA used in this study

Quantum Mechanics based structural analysis (QMSA)

QMSA allows for a different way of interpretating NMR data, from the commonly used *'phenotypic' (visual)* interpretation of NMR to a *'genotypic' (computational)* interpretation, connecting spectra to fundamental NMR theory, reducing spectral complexity to spin parameters of nuclei. Reduction of complexity facilitates large data set compilation, easing transfer for AI and machine learning [3].

In QM-NMR, 2 main methods exist; compound to spectra (C2S) and spectra to compound (S2C), DFT and QMSA respectively. DFT *'predicts'* spectra and QMSA *'calculates'* the compound from spectral decoding of spin parameters (δ, J, N).

We evaluate the use of QMSA for qNMR purity assignment using 1D NMR,^{1}H iterative full spin analysis (HiFSA). Cosmic truth [4], was used to calculate spin parameters from experimental data acquired at BIPM, utilizing DFT as a starting point for matching and validation workflow (C2S-S2C-Assign C2S+S2C- Match C2S+S2C – qNMR C2S – S2C).

Numerical spin parameters

Improved experiment al data for comparison

Artificial Inteliigence (AI) and Machine learning (ML)

Transfer of aggregated computational NMR data in large data sets allows for machine learning training and AI applications. For this project we present two methods that will be explored. GENETICS-AI utilizes NMR theory to create new pulse shapes and sequences to achieve greater spectral performance [4]. The second method presented is a ML tool SHIMNET which corrects shim distortion in spectra however many ML tool like this can be used to correct spectra [5]

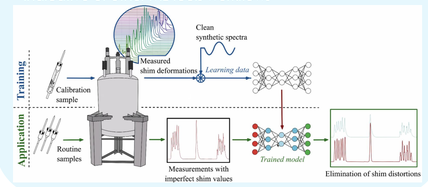
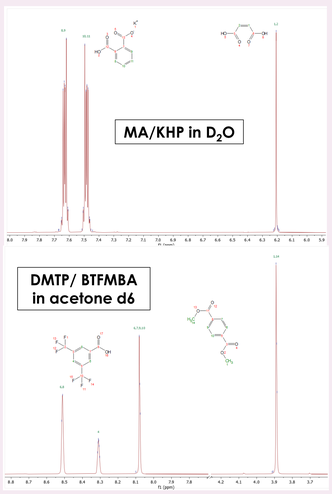


Fig. 4 Workflow of SHIMNET for numerical 'shimming' and indicative of other ML tools workflows

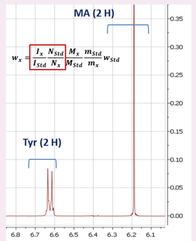
Purity assignment comparisons

Purity assignment of two model systems consisting of MA / KHP in D2O and DMTP / BTFMBA in acetone-d6 via HiFSA was compared to integration methods. The purity results obtained using both approaches agreed within uncertainties under specific conditions. There was ample room for improvement for the QM-method regarding achievable uncertainties as demonstrated by improvements observed by QM guided integrals (GI) and poorer performance when using decoupled data.

Strategies to advance in this respect (AI/ML/Algorithm refinement) are under discussion and additional progress is expected soon as the software is further refined with training on soon to be open access data sharing from BIPM led comparison studies with CRVs also to develop not only cosmic truth but accessible to other quantum chemistry dependent NMR parameter calculators such as PERCH, GAMMA, ACD, Chenomx, NMRdb, Spinach, ORCA, Mnova, NMRFx, and VeSPA to develop qNMR potential.



Phenotypic Integration



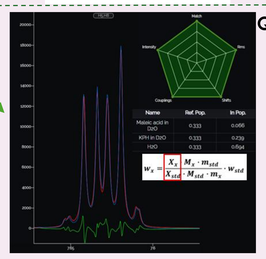
Manual phase and baseline correction

Assignment and nucleotide number determined for peaks used in quantitation.

Integration of 99.999% peak intensity

Impurity correction by orthogonal techniques LC-DAD/MS-MS/UV etc

QMSA Genotypic



QM matching (initial and processed)
Dependent on acquisition variables (Field, tube, line shape), mean and RSD effected.

QM decoupled matching
High overestimation of purity

QM guided integral matching
Independent of variable effects from acquisition.

Impurity correction by simulated spectra of literature identified impurities and deconvolution

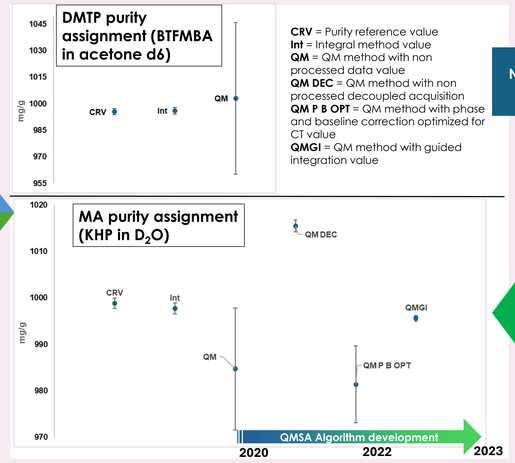


Fig. 5 Purity assignment workflow for manual integration method and QMSA methods for case studies MA/KHP in D2O, DMTP / BTFMBA in acetone d6.

Numerical spin parameters,

>S/N experimental data for QMSA

QMSA database trained ML and AI techniques

GENETICS-AI generation of pulse shapes and sequences alongside parameters to increase signal intensity and quality of data [6]

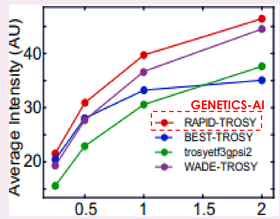


Fig. 6 Comparison of pulse sequence performance related to average intensity of cross peaks generated in 2D TROSY experiments with varying mixing times.

Post processing of experimental data through machine learning to correct data with poor shims to give better matching to calculated spectra (QMSA) via upcoming open access SHIMNET [5]

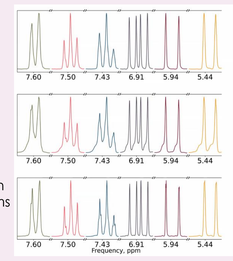


Fig. 7 Spectral comparison between (Top) optimized shims, (Middle) Shims deoptimized by location in probe, (Bottom) Deoptimized shims reoptimized by SHIMNET

Impact on Metrology

The integration of QM and AI/ML into qNMR-based purity assessments will strengthen the missions of international standard bodies. It will be achieved through QMSA being centered on δ and J values rather than integration allows for replicate analysis on any NMR instrumentation, from low (benchtop) to high (BioNMR) fields, to contribute to each other, improving quantitation accuracy and interlaboratory comparability.

Here we showed comparison of QMSA calculated versus experimentally determined values and this can provide numeric measures of consistency and interpretation of confidence, enhancing transparency and reproducibility of analysis. The use of QMSA also shows the ability for semi-automation of the qNMR workflow therefore reducing human bias, a major contributor to uncertainty in comparison studies. Now that computational power is not a barrier to complex molecule QMSA, the limiting factor is the quality of spectral data input into the workflow, this is where anchoring to AI based methods can further refine and develop the workflow through detection and correction of spectral issues therefore further enhancing the potential for accurate and reproducibility metrological activities.

QMSA and AI/ML are powerful tools in improving metrological applications through:

- **Supporting good scientific principles via high quality data.**
- **Facilitating audit reporting and data sharing between modalities due to nature of outputs.**
- **Independent ability to distinguish compounds and structurally related impurities.**

References

[1] Steven Westwood et al ,2019 Metrologia 56 064001D
[2] Steven Westwood et al, 2023 Pure Appl. Chem. 95 1-77
[3] Lucy Botros et al, 2025 Journal of Nat. Prod. Xxx
[4] Manu V. S. et al, 2022 PNAS Nexus 1,133
[5] Jopa S. 2025 *in preparation*
[6] Manu V. S. et al, 2023 Nat Commun 14, 4144