APPLICATION NOTE

ACD/Labs

ADVANCED CHEMISTRY DEVELOPMENT, INC.

Removing User Bias from Structure Verification by NMR

Dimitris Argyropoulos
Sergey Golotvin
Rotislav Pol
Vladimir Mikhailenko
Yalda Lighati Mobarhan

Introduction

Chemical structure verification by NMR is one of the most fundamental, and yet challenging practices in synthetic chemistry. With the overwhelming expansion in the volume of data and size of compound libraries, automated structure verification using NMR data is becoming an invaluable tool for timely and unambiguous characterization in the synthetic workflow. ACD/NMR Workbook Suite is the only commercially available software that can perform structural verification at different levels for added confidence and flexibility in the analysis.

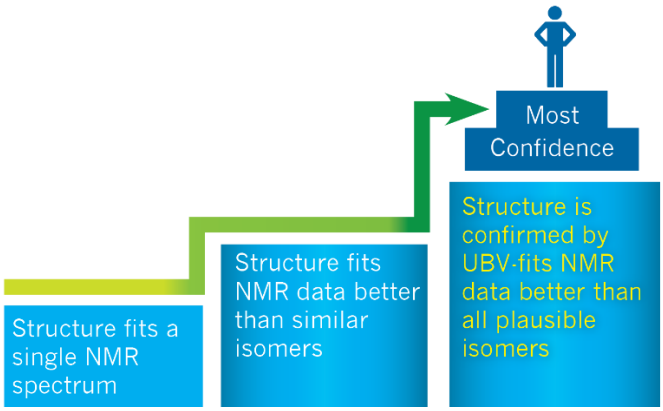


Figure 1. ACD/NMR Workbook Suite performs structural verification at different levels

In the simplest workflow, referred to as Single Structure Verification (SSV), a proposed structure selected based on the knowledge and expected chemistry of the sample is submitted along with the experimental NMR data. A single 1D or 2D NMR spectrum,[1] or a combination of 1D and 2D spectra is required for this method. The software evaluates the match between the proposed structure and the datasets in the NMR project and reports a match factor (MF). While it can be confidently confirmed when a proposed structure fails this "NMR filter", without alternative structures presented as potential "better fits", there is an increased chance for false positives.

In order to confidently reduce the false positive rate, in the next protocol referred to as Combined and Concurrent Verification (CCV), the advanced algorithms concurrently verify the fit between a specified number of generated isomers with the proposed structure.[2] The MF rankings will indicate if any of these alternative structures are more consistent with the NMR data. However, since this protocol matches the experimental data against a selected number of isomers defined by the user or an algorithm, the initial bias from the chemist may still affect the results

Get the Most Confidence in Structure Verification with Unbiased Verification (UBV)

For an absolute level of confidence, Unbiased Structure Verification (UBV) can be activated with the click of a button. This workflow is based on the most advanced structure generators available in the market, coupled to the best ranking system used in a Computer Assisted Structure Elucidation (CASE) system, to reduce the user bias in selecting a best structure. This is specifically beneficial to synthetic chemists, as it can reduce the time and effort spent on characterizing molecules.

[1] Golotvin, Sergey S., Vodopianov, Eugene, Pol, Rostislav, Lefebvre, Brent A., Williams, Antony J., Rutkowske, Randy D., Spitzer, Timothy D. (2007) *Mag. Res. Chem.* 45(10), 803-813.
[2] Golotvin, Sergey S., Pol, Rostislav, Sasaki, Ryan R., Nikitina, Asya, Keyes, Philip. (2012) *Mag. Res. Chem*. 50(6), 429-435.

Since version 2019, NMR Workbook Suite offers UBV for datasets with the following minimum requirements:

1. NMR datasets containing at least 1D ^{1}H, 2D-HSQC, and HMBC spectra, and a proposed structure which will only be used to automatically derive the molecular formula[3]
2. Fully assigned ^{13}C chemical shifts. ^{13}C data can be provided either by a 1D spectrum or by HMBC
3. MW ≤ 800 Da
4. The molecule should contain less than 20 carbon atoms or less than 22 carbon and nitrogen atoms in total

Prior to running UBV, the NMR data need to be processed, and the resonances assigned to the proposed structure. Based on these data a CASE (Computer Assisted Structure Elucidation) run is initiated automatically in the background and the UBV workflow can be selected for an ultimate level of confidence.

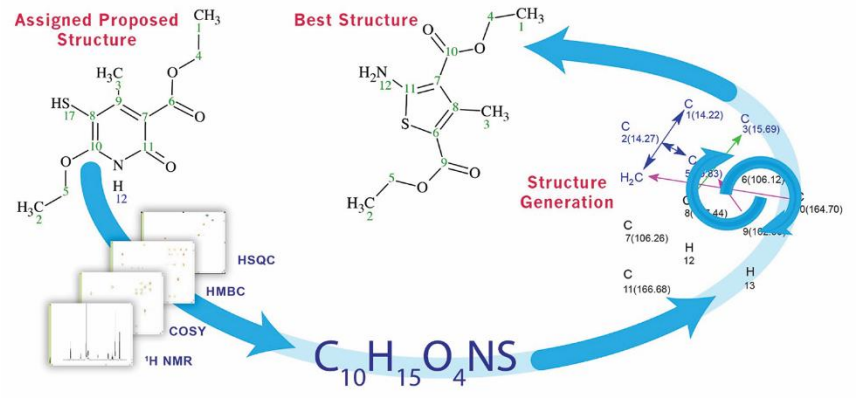


Figure 2. Structure generation scheme in Unbiased Verification workflow

With UBV the advanced structure generator system outputs all possible structures compatible with NMR data without any input by the user, and ensures the best matching structure is assigned. However, this process can be time consuming and to alleviate the computational cost, the user may define one or more molecular fragments that are known to be present prior to structural generation. The resulting structures are ranked by a computationally calculated Match Factor and the mean deviation of predicted versus experimental chemical shift values.

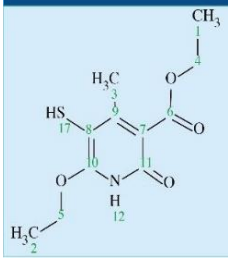
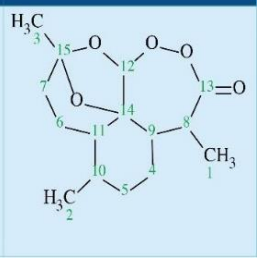
[3] Elyashberg, Mikhail E, Williams, Antony J., Blinov, Kiril. A., "Contemporary computer-assisted approaches to molecularstructure elucidation", RSC, Cambridge, 2012.

Experimental and Results

The UBV method can outperform the previous verification protocols (SSV and CCV), by automatically identifying the correct "best match" with no bias from the user to affect the verification results. As a proof of concept, several datasets were tested and despite deliberately starting with an incorrect proposed structure in each example, the structure most consistent with the NMR data was identified. Two such examples are presented here.

Two commercially available compounds, with the proposed structures as seen in Table 1 were selected, and their corresponding NMR datasets were acquired.

Table 1. "Incorrect" proposed structures #1 and #2 used for evaluating UBV

Proposed Structure #1	Proposed Structure #2
	

The single structure verification process resulted in acceptable MFs of 0.83 and 0.75 for both proposed structures. Subsequently, CCV also ranked these proposed structures as the best matches for their representative datasets. Finally, UBV was performed to evaluate these proposed structures to more confidently verify the correct structure. For each example, UBV generated hundreds of alternative structures, among which the "best structure" with a significantly better fit with the data (than that of the proposed structures) was selected.

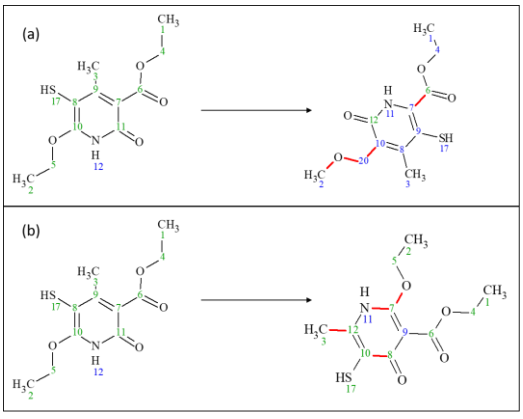
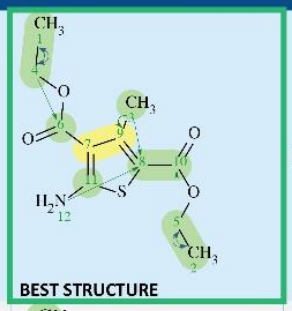
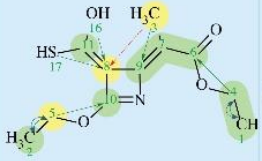
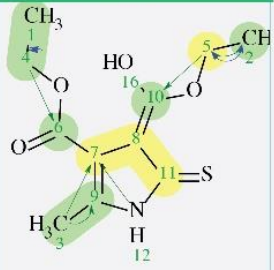
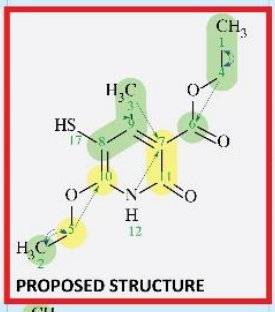
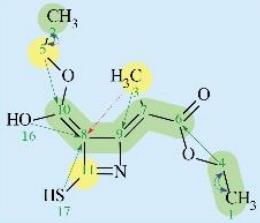
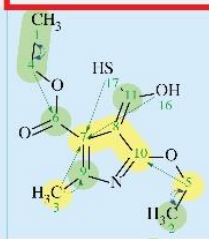
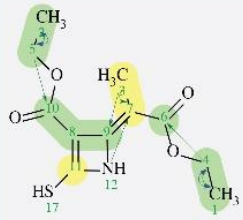
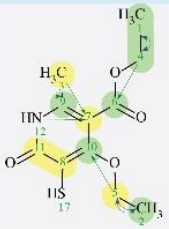


Figure 3. Automated isomer generation scheme in Combined and Concurrent Verification workflow. The red lines indicate the point of change. (a) Isomers are generated by switching the substituents and moving the heteroatoms along the chain. (b) Isomers are generated by moving substituents around the ring and moving the heteroatom within the ring.

Example 1: Proposed Structure #1

As depicted in Table 2, UBV resulted in 5 additional structures with higher match factors than those previously assigned by CCV.

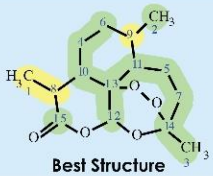
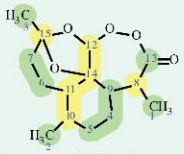
Table 2. UBV results for Proposed Structure #1. The "Best Structure" is a better fit to the data

Rank	Structure	MF	$(d_N(^{13}C+^1H)$ (ppm)	Rank	Structure	MF	$(d_N(^{13}C+^1H)$ (ppm)
1	 BEST STRUCTURE	0.91	2.836	5		0.85	3.46
2		0.83	3.166	6	 PROPOSED STRUCTURE	0.83	3.46
3		0.87	3.27	7		0.89	3.60
4		0.85	3.40	8		0.85	3.62

Example 2: Proposed Structure #2

1D ^{13}C, ^{1}H, and 2D HSQC, HMBC, and COSY spectra were used for structure verification. In version 2019 and later of ACD/NMR Workbook Suite a ^{13}C NMR spectrum is not required. CCV was run with 10 automatically generated isomers. The MF of the proposed structure was 0.75 compared to 0 for all the generated isomers. Subsequently, UBV was run using this proposed structure, requiring a total time of 5 min 40 sec for the generation of an additional 227 unique structures. Five generated structures matched the spectra better than the proposed structure, and the "best structure" selected by UBV was in this case the correct artemisinine structure.

Table 3. UBV performed on Proposed Structure #2. The "Best Structure" with a MF of 0.83 and lower deviations (dN), is a significantly better match to the data.

Proposed Structure			
Method	**Match Factor (MF)**		
SSV	0.75		
CCV	0.64		
Structure	**Rank**	**Match Factor (MF)**	**$d_N(^{13}C+^{1}H)$, ppm**
 Best Structure	1	0.83	3.17
 Proposed Structure	6	0.75	5.69

Expedite Computation Time with a Reliably Defined Fragment

To expedite the computation time, in Proposed Structure #2 the molecular fragment in Figure 4, was defined before running UBV. This considerably reduced both the number of structures from 277 to 57 and the calculation time 340 s to 15 s.

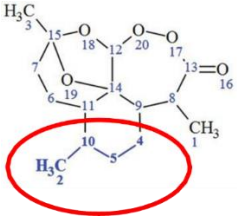


Figure 4. The defined fragment of Proposed Structure #2.

Discussion & Conclusions

UBV represents the most efficient workflow for rapidly recognizing correct structures, by thoroughly investigating all possible structural alternatives compatible with an NMR dataset and a given molecular formula (derived from the proposed structure). By fulfilling the minimum requirements, the method can be applied to evaluate any proposed structure that passes standard NMR verification, and yet potentially represents a false positive. This is especially beneficial to synthetic chemists as an ultimate level of confidence in structure verification is established without false positive errors to ensure the synthesized product is characterized correctly. Furthermore, the improved productivity, high reliability, and reduced chemist supervision involved enables a greater return on investment for analysts and NMR labs.