A New Method for the Reliable Detection of ^{13}C Multiplets of Fluorine Containing Compounds

Dimitris Argyropoulos, Rostislav Pol, Vladimir Mikhailenko, and Sergey Golotvin

Advanced Chemistry Development, Inc. (ACD/Labs), 8 King Street East, Toronto, ON, M5C 1B5, Canada

Introduction

In modern organic and medicinal chemistry, fluorine is used to enhance the chemical properties of molecules in many desirable ways. When substituted for hydrogen, the increased stability of the C-F bond can delay the metabolism of the molecule and forbid peroxide formation. As well, its heightened lipophilicity has a direct positive effect on the molecule's bioavailability [1]. Consequently, it is estimated that more than 20% of commercial pharmaceutical APIs, 30% of the top 15 prescribed pharmaceuticals, and 30% of agrochemicals contain at least one fluorine atom [2,3]. While advances have been made to selectively introduce fluorine in drug development, more research is required to understand the direct effect of these additions on specific chemical structures. A comprehensive understanding of this can increase the likelihood of developing a successful compound and reduce development time [4].

A challenge in the development of fluorine chemistry is the potential difficulty of interpreting ^{13}C NMR spectra of fluorinated organic compounds. This is because ^{13}C spectra are commonly recorded using only ^{1}H broadband decoupling and the ^{13}C-^{19}F couplings are still present. The C-F coupling constants can be very large (up to 250 Hz or more), which may result in multiplets severely overlapping with other peaks in the spectrum. Additionally, since ^{13}C spectra usually have low signal to noise ratios it is frequent for the lower outer part of a multiplet to lie below the noise level and not be visible. To mitigate this, it is possible to record ^{13}C spectra broadband decoupled from both ^{1}H and ^{19}F but it requires specialized NMR probes and decoupling techniques. Moreover the very broad range of ^{19}F chemical shifts could pose a danger of damage to the probe due to the excessive power that would be required. Consequently, this approach is not considered practical for routine use. As an alternative, more consistent spectral analysis methods have the potential to reduce costs, error and expedite analysis time.

Here we present an analysis method that reliably peak-picks and identifies multiplets in the ^{13}C spectra of organic compounds based on the accurate prediction of C-F multiplets and matching with the experimental spectrum. We show that regardless of whether the final results contain multiple, overlapping multiplets, the expected carbon resonances are reliably identified and assigned for each spectrum.

Method

Traditionally, there has been a discrepancy in the way that peak picking is performed when done automatically by a computer and manually by a spectroscopist. In general, automatic peak-picking and multiplet identification of a ^{13}C spectrum involves selecting the peaks that are above a predefined threshold and checking to see if any three or four of them have a height or integral ratio of 1:2:1 or 1:2:2:1. These are set as a triplet or a quartet, respectively. This very simplistic approach suffers from several drawbacks:

- It ignores outer components of multiplets whose intensities are below the preset threshold
- It may not detect triplets and quartets that are not present in the expected intensity (or integral) ratios because of sub-optimal acquisition parameters (digital resolution)
- The method will fail if the multiplets overlap with other spectral peaks. This is a common problem since C-F coupling constants can be very large, e.g., on the order of ~270 Hz for a CF_3 group. These large coupling constants result in multiplets that can span up to 810 Hz, and ultimately overlap with other peaks
- The method cannot identify doublets coming from carbons with a single F atom attached

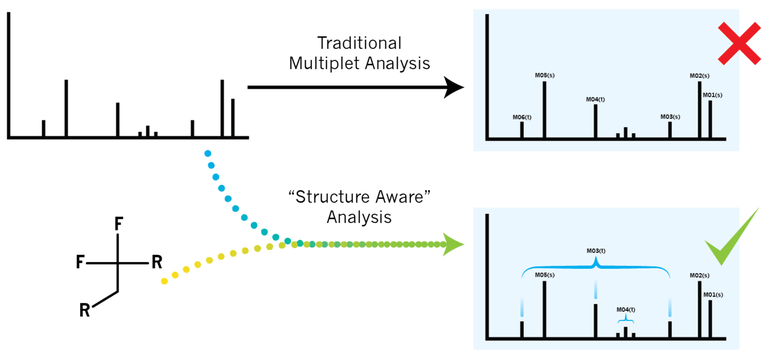


Figure 1 shows how the traditional method for multiplet analysis only takes the experimental spectrum into account when peak picking and assigning multiplets. The multiplet that overlaps with other peaks is not recognized, and the low intensity signals are missed entirely. The "Structure Aware" analysis method properly detects the overlapping multiplets and low intensity multiplet signals using both the experimental spectrum and proposed structure as a reference.

Nevertheless, this method has historically been preferred since it is the most simple to program. Alternatively, manual analysis by a human expert is performed to accommodate for the outlined complications. Prior to peak picking, an experienced NMR spectroscopist would examine the proposed structure and know that the ^{13}C spectrum will have multiplets. This structure aware bias is necessary to accurately examine the spectrum. As a result, a similar "Structure Aware" analysis must be incorporated into the peak picking algorithms of analysis software.

We are proposing a computational method that uses predicted ^{13}C spectra, that is accurate in terms of both chemical shifts and C-F coupling constants. The general scheme, shown in Figure 1, involves predicting the ^{13}C spectrum of the proposed structure and identifying the peaks that will be split from C-F coupling. It subsequently inspects the experimental spectrum for similar chemical shift positions and coupling constants to identify any multiplets. If only some of chemical shift patterns are found then the peak-picking threshold is lowered and another attempt is made to see if the missing peaks are present at a lower intensity. In the case of multiplets that overlap with peaks, the algorithm will initially identify whether the expected multiplet positions and intensities are correct. If additional peaks are present, the algorithm will only pick and group the ones relevant to the multiplet, leaving the others to be assigned accordingly.

Results

To determine whether this new method is effective we tested its ability to accurately assign both 1D and 2D NMR spectra of compounds with a range of fluorinated groups.

Case 1 | ^{13}C Spectrum of Fluorine Compound #1

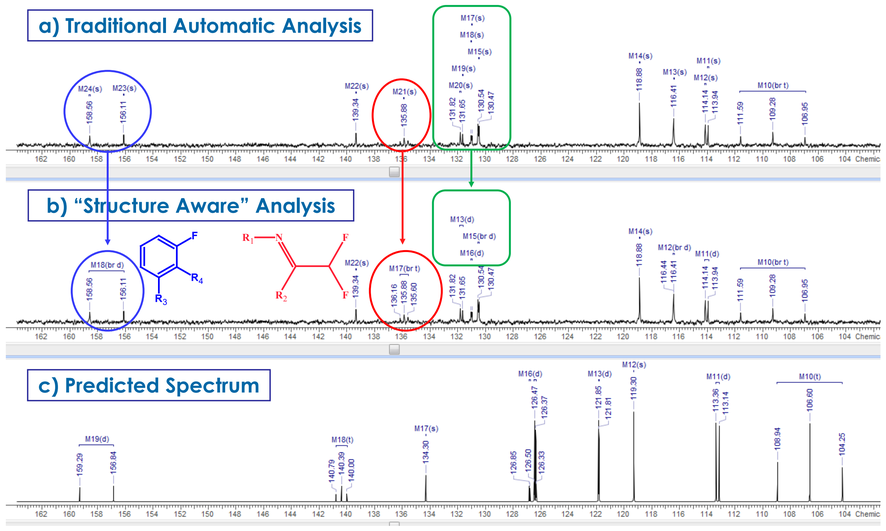


Figure 2 The experimental and predicted spectra of Compound #1. a) Experimental spectrum with multiplet analysis by the traditional automatic method. b) Experimental spectrum after performing "Structure Aware" analysis. c) The predicted spectrum of Compound #1.

Figure 2 (a & b) show the low S/N experimental spectrum for Compound #1, which contains a CHF_2 and an aromatic C-F group. The traditional automated peak picking and multiplet analysis is only able to identify the carbon on the CHF_2 group. It completely neglects the signal of the quaternary carbon attached to it. This peak has a very low S/N and in fact the outer components can just be seen protruding above the noise. Lowering the threshold with this method would not have produced better results since other spikes in the baseline would have been misidentified as peaks. Finally, the doublet of the aromatic C-F is identified as two individual peaks. When the spectrum is analyzed using the "Structure Aware" analysis method, all the multiplets in this spectrum are identified including the aromatic C-F doublet and the low S/N triplet of the carbon next to the CHF_2 group. This is achieved by having accurate prediction. Figure 2 (c) shows the predicted spectrum of Compound #1, which is in excellent agreement with the experimental spectrum.

Case 2 | ^{13}C Spectrum of 3,4,5-Trichlorobenzotrifluoride

Figure 3 (a & b) shows the experimental spectrum of a compound that contains a CF_3 group on an aromatic ring. The challenge with this spectrum is that almost all the signals will be in the same region and be split because of fluorine coupling. This makes identification of the various quartets challenging.

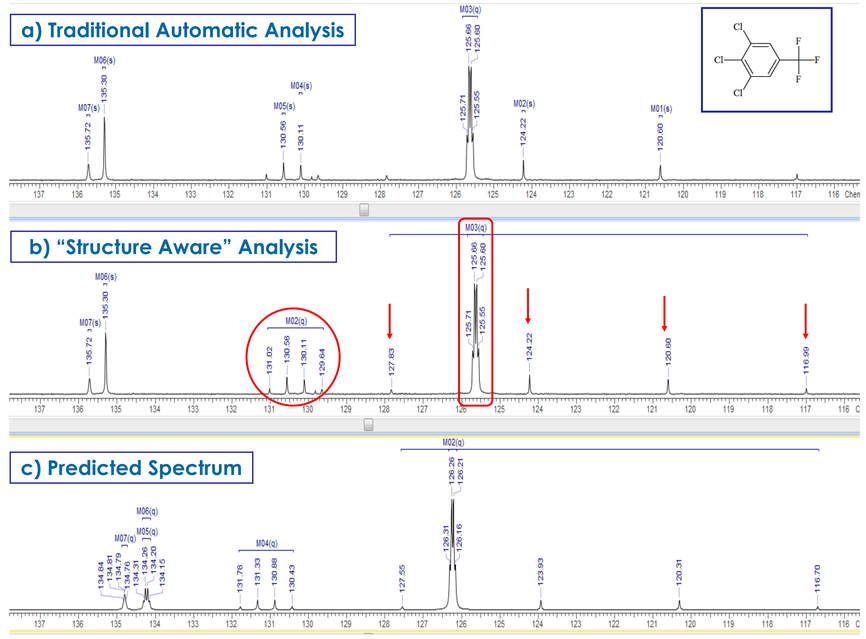


Figure 3 The experimental and predicted spectra of 3,4,5-Trichlorobenzotrifluoride. a) Experimental spectrum with multiplet analysis by the traditional automatic method. b) Experimental spectrum after performing "Structure Aware" analysis. c) The predicted spectrum of Compound #2.

The traditional analysis method only identifies the quartet belonging to the carbon directly connected to the CF_3 group. However the "Structure Aware" method reliably identifies all three quartets and marks them as such. It is interesting also to see that the peaks at ca. 134-135 ppm do not appear as multiplets but rather as broad singlets, partly because of insufficient digital resolution. That being said, this did not impede the analysis.

Case 3 | 2D NMR Analysis of Fluorine Compound #3

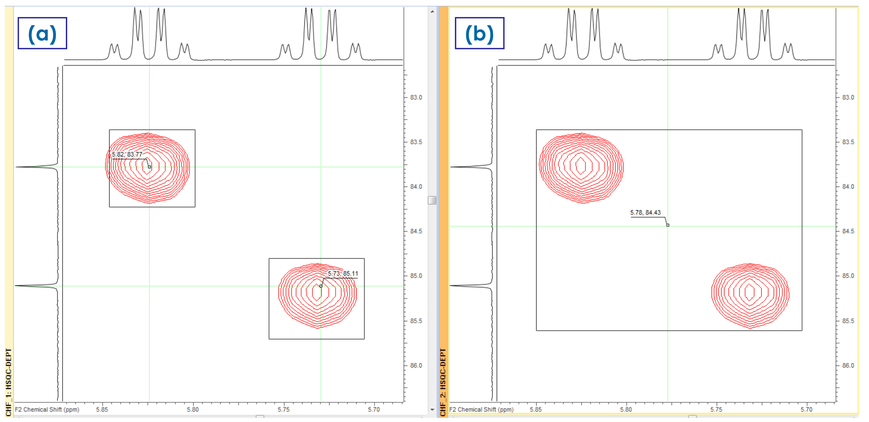
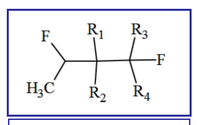


Figure 3 a) HSQC of Compound #3 after the traditional automatic analysis. b) HSQC of Compound #3 after "Structure Aware" analysis.

Cases 1 and 2 show that this "Structure Aware" analysis method identifies multiplets with a significantly higher level of accuracy than the traditional method. To determine how broadly applicable this method is, we tested whether it can be used with 2D heteronuclear NMR spectra, e.g., HSQC spectra. The challenge with analyzing these spectra is that only parts of the multiplets are visible. For example, a signal which is a doublet in both ^{1}H and ^{13}C will only show two peaks in the HSQC, not 4.



Compound #3

This is because the 2 missing peaks correspond to transitions with a change in quantum number of 0 or 2, which are forbidden. Instead only the allowed transitions are visible in a very characteristic diagonal pattern. This is shown in Figure 4. For this spectrum, the traditional way of peak-picking identifies the HSQC pattern as two individual peaks. This is partly because the starting doublet of quartets of doublets in the ^{1}H spectrum and the doublet in the ^{13}C spectrum were not identified. The "Structure Aware" method, though, correctly identifies the multiplets in both the ^{1}H and the ^{13}C spectrum and then properly identifies the HSQC peaks as corresponding to one single correlation.

Discussion and Conclusions

It is evident from the three cases shown that there is both a need for more reliable detection of ^{13}C multiplets of fluorine containing compounds and that the method proposed is a significantly more accurate automated method than what is traditionally used. This solves a problem that was impeding automatic analysis of large scale datasets that contain fluorinated compounds.

References

1. Hagmann, William, K. "The Many Roles of Fluorine in Medicinal Chemistry" Journal of Medicinal Chemistry, **2008**, 51, 4359-4369.
2. Emsley, John, "Nature's building blocks: An A–Z guide to the elements (2nd ed.)", Oxford University Press, **2011**, p. 178.
3. Swinson, Joel, "Fluorine – A vital element in the medicine chest", PharmaChem. Pharmaceutical Chemistry: **2005**, 26–27.
4. Gilli, Eric P., Eastman, Kyle J., Hill, Matthew D., Donnelly, David J., Meanwell, Nicholas A. "Applications of Fluorine in Medicinal Chemistry", Journal of Medicinal Chemistry, **2015**, 58, 8315-8359