

Consistent Detection of Multiplets in ^{13}C Spectra Using Structure Aware Algorithms

Dimitris Argyropoulos, Sergey Golotvin, Rostislav Pol and Vladimir Mikhailenko

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



ACD/Labs

Introduction

Automated interpretation of NMR spectra is a highly demanded technique in the current chemical and pharmaceutical industries. Detection of peaks and multiplets, which is an essential step in a successful automated interpretation, becomes quite challenging when heteroatoms such as ^{19}F and ^{31}P are incorporated into the structures. These structural additions give rise to further peak splittings in the 1D ^{13}C and heteronuclear 2D NMR spectra. With up to 30% of the commercial pharmaceutical APIs and agrochemicals containing either a ^{19}F and/or a ^{31}P atom, [2,3], a comprehensive understanding of the structure can increase the likelihood of developing a successful compound and reduce the development time [4].

Here we present an automated structural interpretation method for multiplet detection in such complex peaks. The software starts by predicting the spectrum of the proposed structure and then looks for similar patterns in the experimental spectra. If such patterns match the multiplets the peaks can be defined without a false negative or need for manual inspection.

In this poster, we demonstrate examples of overlapping and low-intensity multiplets arising from ^{19}F and/or ^{31}P couplings in 1D and 2D spectra that are detected by this approach.

Method

Conventionally, automatic peak-picking and multiplet identification of a ^{13}C spectrum, is performed by looking for specific patterns. For example three or four evenly spaced peaks with a 1:2:1 or 1:2:2:1 intensity ratio, are respectively defined as a triplet or a quartet, however, there is still no reliable method for finding doublets.

This conventional approach suffers from several drawbacks:

- It ignores outer components of multiplets whose intensities fall 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)
- They fail if the multiplets overlap with other spectral peaks, a rather common problem since these peaks are usually quite broad and the ^{19}F and ^{31}P nuclei, are usually coupled to more than one carbon.
- They cannot identify doublets coming from carbons with a single ^{19}F or ^{31}P atom attached

Peak interpretation becomes even more challenging in 2D spectra with characteristic diagonal peak patterns. Although in F-containing chemicals the problem can be alleviated by using costly hardware and recording ^{19}F , as well as ^1H decoupled ^{13}C spectra, such an option does not exist for ^{31}P -containing compounds.

The general scheme, shown in Figure 1, involves predicting the ^{13}C spectrum of the proposed structure and identifying the peak splitting due to ^{13}C - ^{19}F (or ^{13}C - ^{31}P) couplings. 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 identified, then the peak-picking threshold is lowered, and another attempt is made for the missing peaks at a lower intensity.

For overlapping multiplets, 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 within the multiplet, leaving the others to be assigned accordingly.

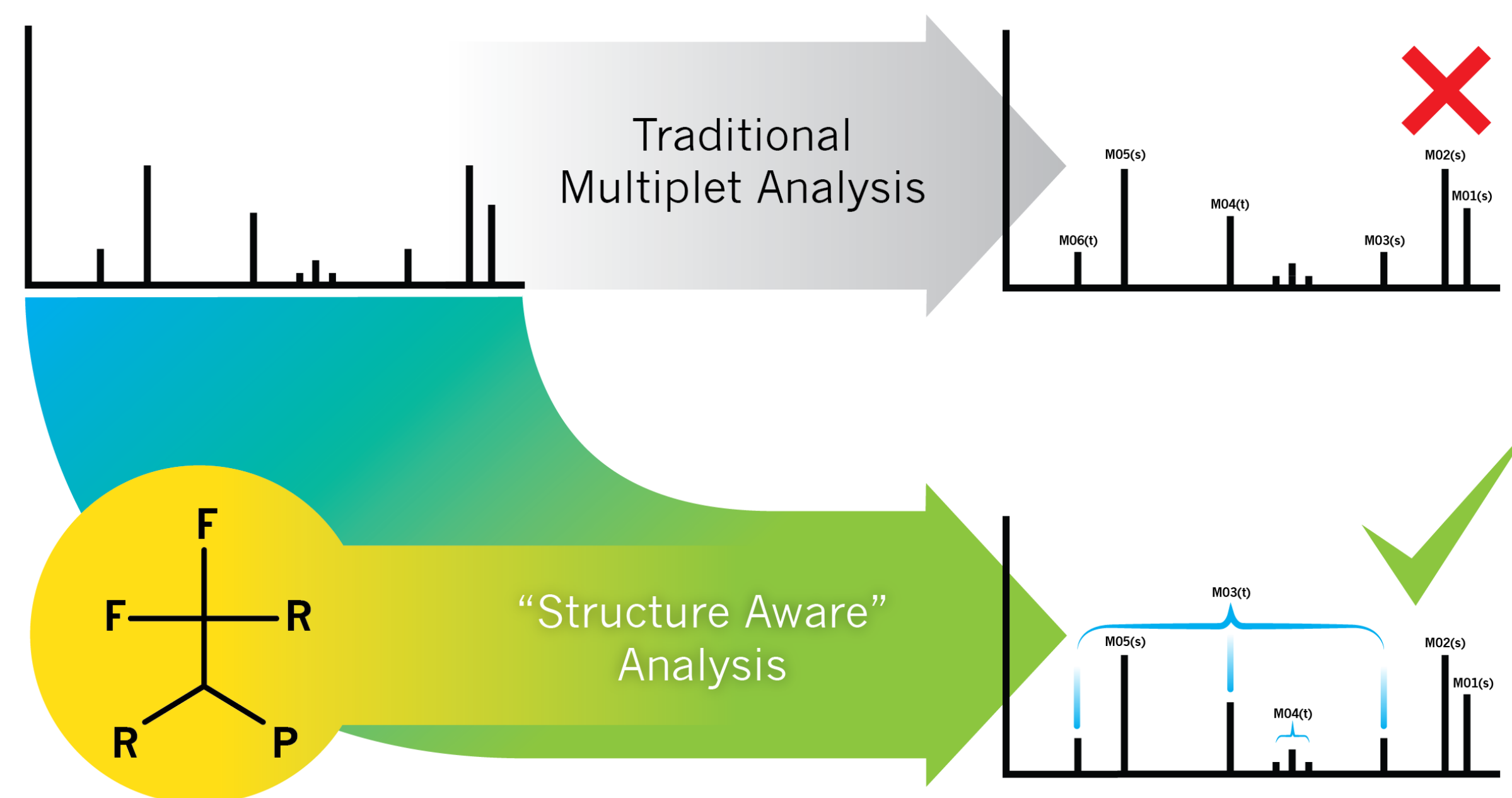


Figure 1 An illustrative overview of "Structure Aware Analysis" method. The "Structure Aware" analysis method properly detects the overlapping multiplets and low intensity multiplet signals using both the experimental spectrum and proposed structure as a reference.

Results

To demonstrate the effectiveness of this method we tested its ability to accurately detect peaks and multiplets in 1D and 2D NMR spectra of compounds with a range of fluorinated and phosphorylated groups.

Case 1 | ^{13}C Spectrum of a Compound with Fluorine and Phosphorus Atoms

In Figure 2 the presence of a P and a F atom two bonds apart causes the carbon peaks that are coupled to both of them, split into multiplets. Since these are single atoms, the multiplets are either doublets or doublets of doublets, with varying coupling constants depending on the distance from the F and P atoms. While the traditional algorithms fail, "Structure Aware" analysis successfully identifies and marks all the multiplets.

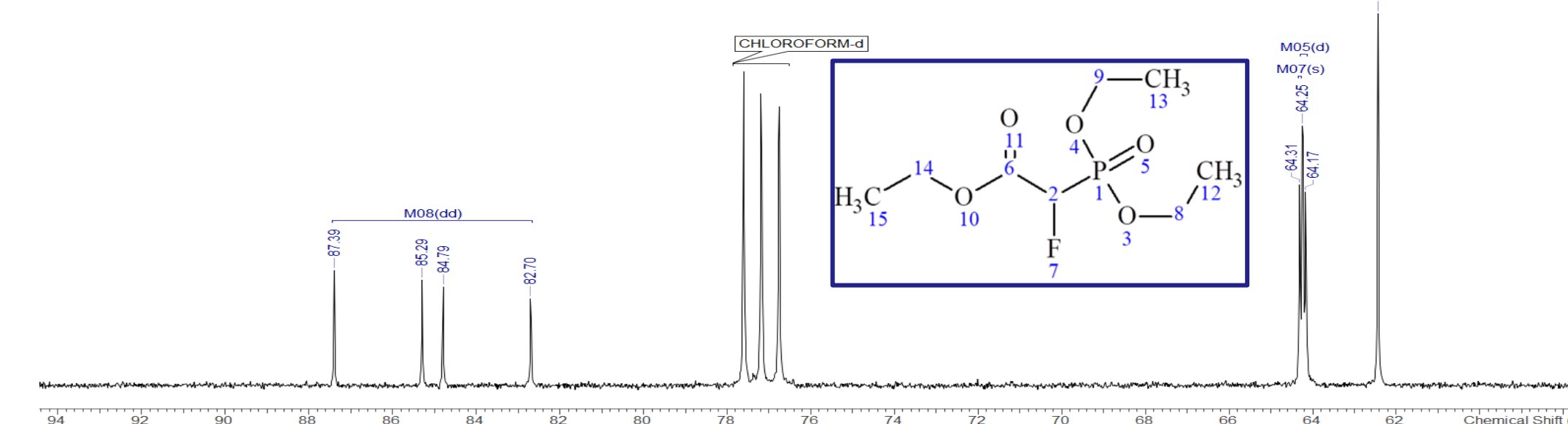
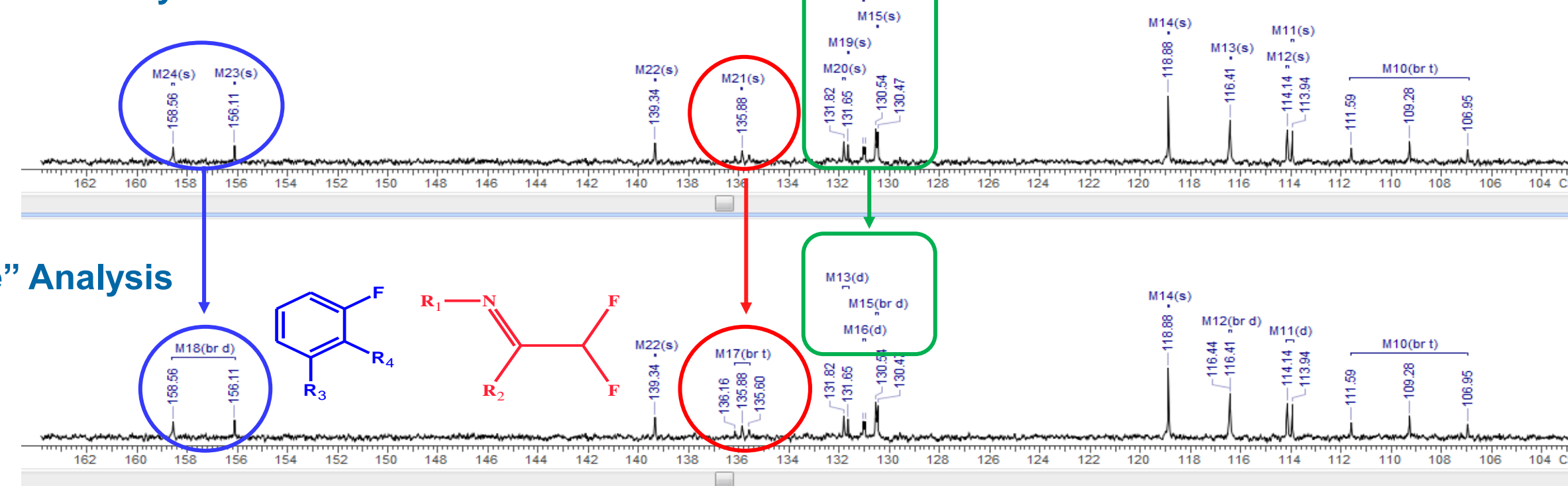


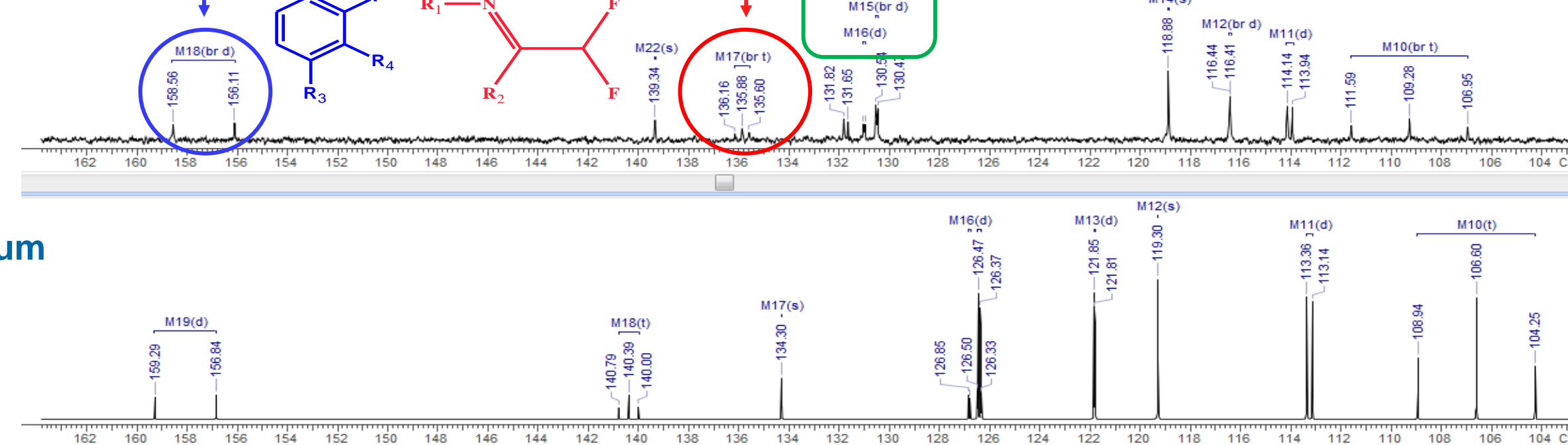
Figure 2 Expansion of the 1D ^{13}C spectrum of Triethyl 2-fluoro-2-phosphonoacetate in CDCl_3 , showing the extensive couplings observed.

Case 2 | ^{13}C Spectrum of a Fluorinated Compound

a) Traditional Automatic Analysis



b) "Structure Aware" Analysis



c) Predicted Spectrum

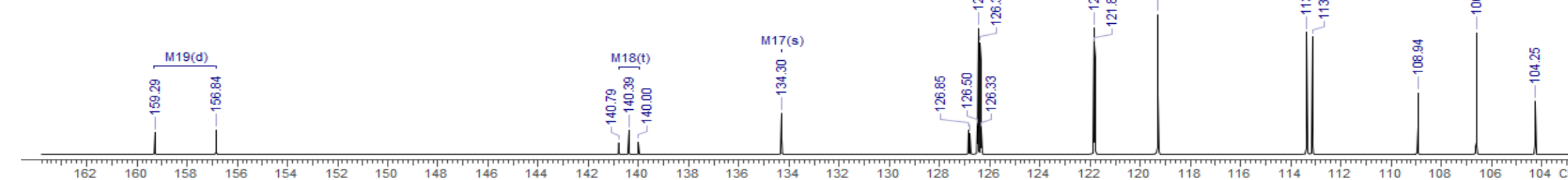


Figure 3 The experimental and predicted spectra of the fluorine compound. a) Experimental spectrum with multiplet analysis by the traditional automatic method. b) Experimental spectrum after performing "Structure Aware" analysis. c) The predicted spectrum of the compound.

Figure 3 (a & b) show the experimental spectrum of a compound with a CHF_2 and an aromatic C-F group. The traditional analysis is only able to identify the CHF_2 carbon and neglects the adjacent quaternary carbon. This peak has a very low S/N with its outer components just above the noise level. Lowering the threshold would not be helpful either, since the spikes in the baseline would be misidentified as peaks. Finally, the doublet of the aromatic C-F is identified as two individual peaks. Using the "Structure Aware" analysis, all the multiplets including the aromatic C-F doublet and the low S/N triplet of the carbon next to the CHF_2 group are identified. This is achieved by accurate predictions shown in Figure 3 (c), which is in excellent agreement with the experimental spectrum.

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

Figure 4 (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.

References

1. Hagmann, William K., "The Many Roles of Fluorine in Medicinal Chemistry" *J Med Chem*, **2008**, 51, 4359-69.
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", *Pharma Chem*, **2005**, 26-7.
4. Gilli, Eric P., Eastman, Kyle J., Hill, Matthew D., Donnelly, David J., Meanwell, Nicholas A. "Applications of Fluorine in Medicinal Chemistry", *J Med Chem*, **2015**, 58, 8315-59

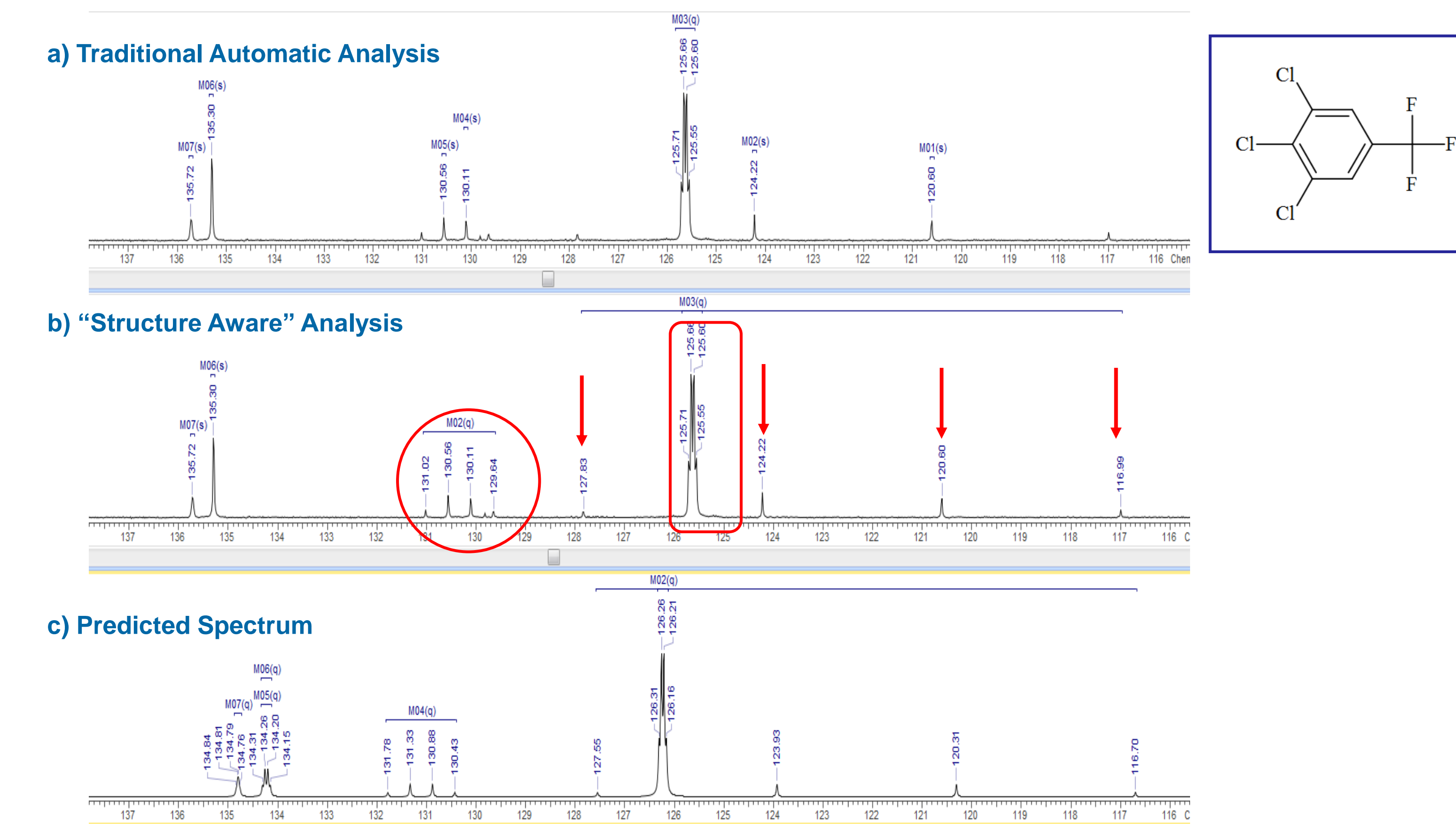


Figure 4 The experimental and predicted spectra of 3,4,5-Trichlorobenzotrifluoride. a) Experimental spectrum with multiplet analysis by the traditional automatic method. b) Experimental spectrum after performing "Structure Aware" analysis. c) The predicted spectrum.

The traditional method only identifies the quartet from the carbon adjacent to the CF_3 group, while "Structure Aware" analysis reliably identifies all three quartets. Interestingly the peaks at ca. 134-135 ppm do not appear as multiplets but rather as broad singlets, partly due to insufficient digital resolution. However, this did not impede the analysis.

Case 4 | 2D NMR Analysis of a Fluorinated Compound

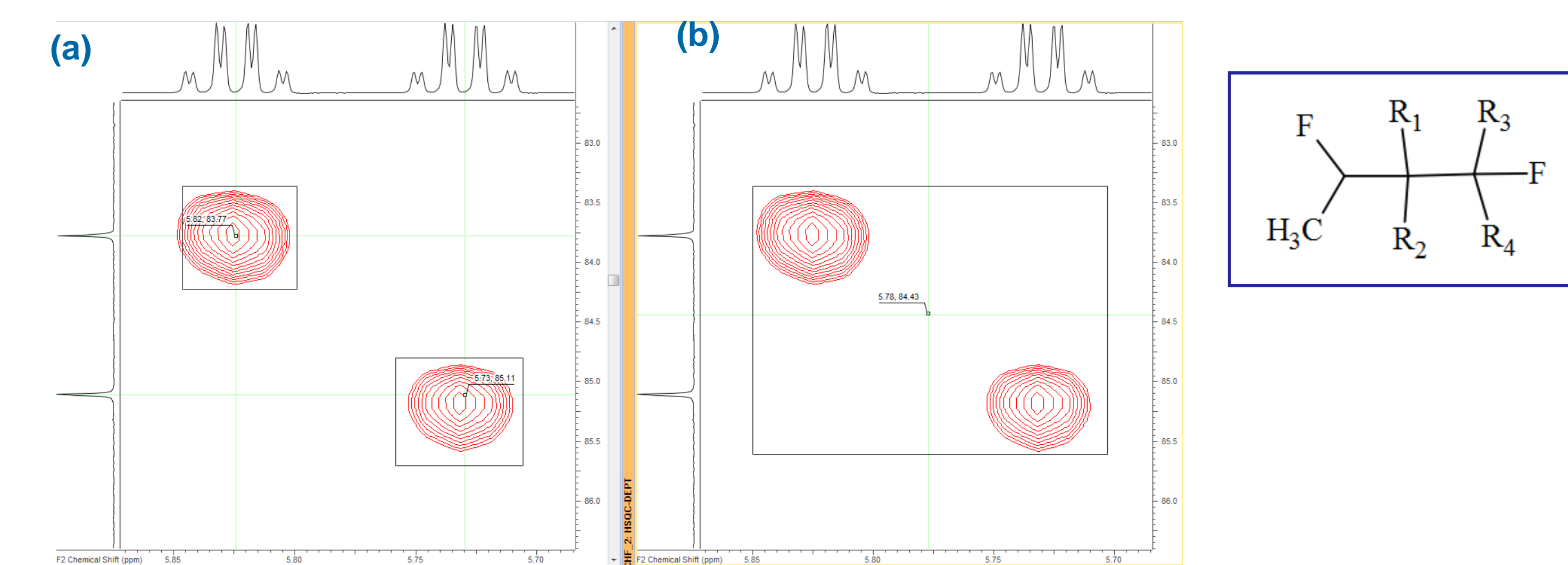


Figure 5 a) HSQC of fluorine compound after the traditional automatic analysis. b) HSQC of fluorine compound after "Structure Aware" analysis.

Next, we applied the method to a 2D HSQC NMR spectrum. The challenge here, is that only parts of the multiplets are visible. For example, a doublet signal in both ^1H and ^{13}C , will only show two peaks in the HSQC instead of 4.

The two missing peaks correspond to forbidden transitions with a change in quantum number of 0 or 2. Therefore, only the allowed transitions are visible. As shown in Figure 5, the traditional peak-picking identifies the HSQC pattern as two individual peaks. This is partly because the starting doublet of quartets of doublets in the ^1H spectrum and the doublet in the ^{13}C spectrum were not identified. However, "Structure Aware" analysis correctly identifies the multiplets in both the ^1H and ^{13}C spectrum, as well as the HSQC peaks as corresponding to a single correlation.

Conclusions

It is evident from the examples that there is a need for more reliable detection of ^{13}C multiplets of fluorine and phosphorus containing compounds and that the proposed method is significantly more accurate than the traditionally used methods. This automated method solves problems that were impeding analysis of large scale of fluorinated and phosphorylated compounds.