

Shift Assignment Errors in the Published Literature - the Impact on Computational Shift Prediction

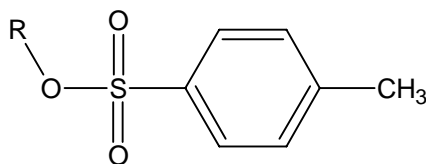
December 23rd, 1998

Dear Barry,

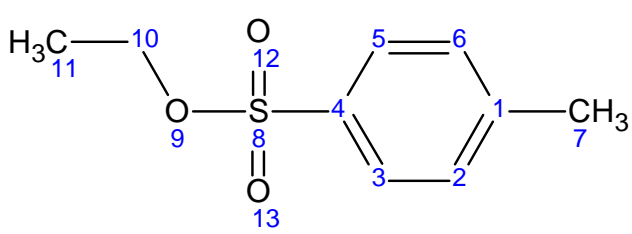
As you are well aware, our NMR shift prediction software has been available for over four years and has been well accepted by the marketplace. To date, we have prediction software and chemical shift assignment databases available for H1, C13, F19 and P31 nuclei. One of the common concerns for anyone considering the implementation of our software is both the quality of our prediction algorithms as well as the quality of the database which was used as the basis of our correlation algorithms and is updated on an ongoing basis with fresh data from the literature. As all NMR spectroscopists will be aware, there are "considerable" mis-assignment errors in the literature resulting from simple typographical errors, poor interpretation of experimental data and errors which have been carried through the literature when based on an initial poor assignment.

Our quality checking procedures include careful examination of the shift assignments and comparison with those entries already existing within our database. Following entry of the data into our database and two stages of typographical error checking we will then look for inconsistencies across the compiled database. In this way we have identified many thousands of poor assignments. Where necessary we will work with our collaborators to obtain fresh experimental data in order to clear up confusions.

One example of a substructure specific error which we have unearthed as a part of our database quality checking is shown below. We have noticed there have been a number of contrary assignments made for the quaternary aromatic carbons for alkoxy substituted tosylates of the form below:



For example, for the ethyl substituted system, the predicted C-13 spectrum is given in the table below

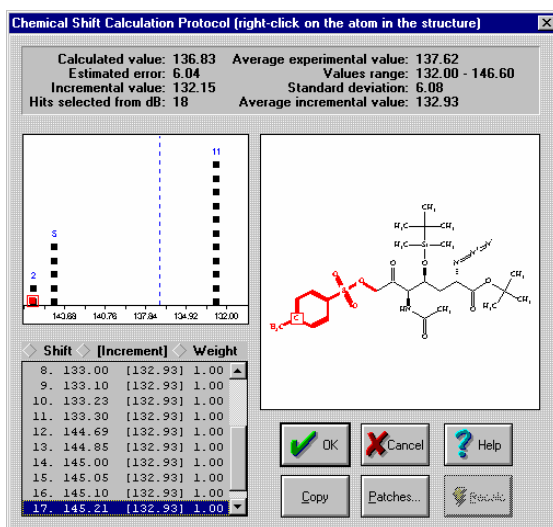


Carbon No.	CHn	Chem. Shifts	Conf. Limits
1	C	136.83	6
2	CH	129.36	0.9
3	CH	128.55	0.9
4	C	143.92	5
5	CH	128.55	0.9
6	CH	129.36	0.9
7	CH ₃	21.48	0.3
10	CH ₂	67.75	0.3
11	CH ₃	14.9	0.2

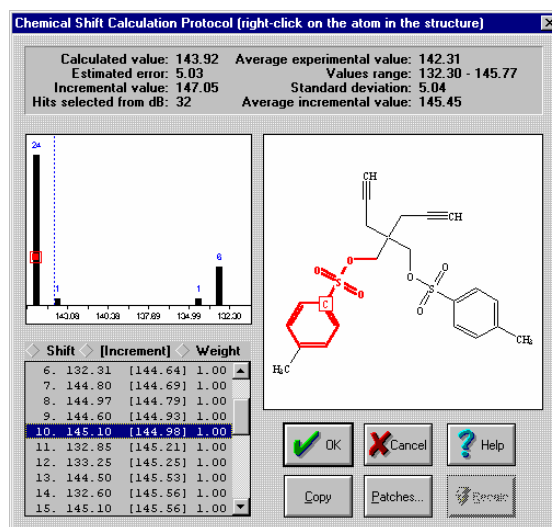
Notice the large error in chemical shift (95% confidence interval) for C-1 and C-4 (6ppm and 5ppm respectively) compared to the small confidence limits for the other predicted chemical shifts. In general C-13 chemical shifts are predicted to better than +/- 3ppm with our C13 NMR prediction program. Within the program it is possible to directly view assigned fragments from within our database of 67,000 assigned structures. These fragments are shown within a *Calculation Protocol* window. The displays for both C-1 and C-4 are shown on the next page. As

can be seen in each of the windows there are a number of representative fragments for both nuclear environments (each of the black squares represents a structure and by passing the mouse over a square displays the related structure and highlights the associated fragment in red). Consideration of the calculation protocol windows shows that for each of the two nuclei, C-1 and C-4, there are TWO distinct columns of related structures displayed, one around 132ppm and one around 145ppm. The conclusion from this data is that ongoing assignments for C-1 and C-4 have been confused in a number of cases. These confusions have not arisen from one particular publication that we can identify. For example the shifts at ca. 132.30 for C-4 are were obtained from the following references:

Tetrahedron 1990, vol 46, page 3061
 Tetrahedron 1994, vol 50, page 11039
 Tetrahedron 1992, vol 48, page 9753
 Tetrahedron 1984, vol 40, page 905
 Khim. Prir. Soedin., 1991, page 707
 Helv. Chim. Acta., 1990, vol 73, page 2090
 Chem. Ber. 1979, vol 112, page 2815



Carbon-1 Calculation Shift Protocol



Carbon-4 Calculation Shift Protocol

In order to clarify and correct the appropriate shift assignment for such structures we obtained NMR spectra for a related structures and using a combination of H-1, C-13, HMQC and HMBC experiments we were able to produce corrected assignments. These corrections have been included in our prediction algorithms thereby impacting the predictions for such related structures in the future. This is not the only example of such a problem that we have identified with literature data and have had to resolve. However, it is an indication of some of the approaches we are taking to building a high quality and appropriate database to be used as reference data for prediction and searching by structure and substructure.

We welcome clarification from any of your readers of any literature assignments that they have noted to be incorrectly carried throughout the literature.

Yours sincerely,
 Antony Williams