T16. Prediction of NMR Chemical Shifts. A Chemometrical Approach

K. Blinov

ACD Inc., Moscow, Russia

Keywords: NMR spectra, NMR Spectrum prediction, PLS, Neural

Nuclear magnetic resonance (NMR) spectroscopy plays a key role in determination of unknown chemical structures (Structure Elucidation). Generally, a signal in a carbon NMR spectrum corresponds to the carbon atom in a chemical structure. Main characteristic of a signal in NMR spectrum is its chemical shift, which provides a lot of information about the corresponding carbon atom and its neighborhood. The chemical shift of a carbon atom depends on its neighborhood and so can be predicted based on this information. The best way to choose correct structure from a number of structures corresponding to spectral data is to compare the predicted experimental values with the experimental ones. The littler difference means the corresponding structure. Today, this procedure is widely used in Computer-Aided Structure Elucidation (CASE).

All modern methods to predict a chemical shift can be considered of two types depending on the way of calculating a shift they based on. One of the techniques means that each neighbor atom has its own "increment" value (which depends both on the atom type and the distance to a target atom). The chemical shift of a target atom determined by this technique is the sum of the values of all neighbor atoms. The other way to perform this is to find some atoms with the similar neighbors from database and average their values to assign the obtained shift to a target atom. The first method is fast, but inaccurate, while the second one is usually accurate, but slow. As the result, the usage of both methods is limited in CASE systems.

In our work we tried to develop a fast and still accurate method of chemical shift prediction applicable for CASE. As we had a large amount of experimental data available (near 1.5 million of measured carbon chemical shifts), it seems reasonable to use the fast "increment" method and apply chemometrics methods to extract necessary parameters for this method from experimental data. PLS regression has been used for process experimental data. This method usually produces excellent results when the correct model used. The main goal of our work was to develop an appropriate model of structure representation for prediction of chemical shifts. The models were developed using knowledge about "the nature" of chemical shifts. Several models have been checked and the best one has been kept. Thus, chemometric allows development of a

novel method of chemical shift prediction, which outperforms known methods and can be successfully used in CASE.