Abstract

Progress in Chemometrics Research; Pomerantsev A.L. (Ed.);
Nova Science Publishers, New York, 67-87 (2005)

Varmuza K.:

Global and local chemometric models of spectra-structure relationships.

The recognition of chemical structure information from spectral data challenged chemometricians since the pioneering time of this discipline. A lot of work has gone into the development of spectral classifiers based on multivariate classification methods. Actually, such classifiers are capable to detect with low error rates the presence or absence of some substructures automatically from IR or MS data. Together with the known molecular formula and using appropriate software for isomer generation, a systematic structure elucidation of organic compounds is possible in some cases.

However, because of the complexity of spectra-structure relationships, the high diversity of chemical structures, and the large amount of data this global approach is limited in practical applications. Routinely used for the identification of unknowns are spectra similarity searches in spectroscopic databases. Result is a hitlist which contains the reference compounds exhibiting spectra that are most similar to the spectrum of the unknown.

A method - based on binary substructure descriptors - has been developed for evaluating the structural similarity between the unknown and the hitlist compounds. This method has been applied to improve search methods in IR and MS databases with the aim to yield subsets of compounds that contain useful chemical structure information about the unknown, even in cases the unknown is not in the database.

Spectral and structural data from the hitlist compounds can be used to create local models by PCA or PLS. The resulting scatter plots support the spectroscopist in spectra interpretation and finding useful relationships between spectra and chemical structures.