Recent developments of LSD, an automatic structure elucidation software

Bertrand PLAINCHONT, Jean-Marc NUZILLARD
and Vicente de Paulo EMERENCIANO

URCA, CNRS UMR 6229, Moulin de la Housse, BP 1039, 51687 REIMS Cedex 2
USP, Instituto de Química, CP 26077 CEP 05513-970, São Paulo, Brasil

The aim of the LSD program\(^1\) is to find all possible molecular structures of an organic compound that are compatible with its spectroscopic data. Structure building relies on connectivity data found in 2D NMR spectra, without any reference to a chemical shift database\(^2\). The measurement protocol that is required by LSD includes the recording of 1D \(^1\)H and \(^{13}\)C as well as 2D COSY, HSQC and HMBC spectra. The status of each atom must be defined. It includes the atom symbol, the hybridization state and the number of attached hydrogen atoms. This part of the data set is most often easily deduced by the user from elementary chemical shift knowledge. The status of the heteroatoms is deduced by the user from the molecular formula. Carbon-carbon bonds are inferred from COSY and HSQC data while HMBC and HSQC data provide connectivity relationships through one or two bonds for non-hydrogen atom. The constraints imposed by atom status and 2D NMR data may be enforced by other atom neighborhood relationships.

A recent development of LSD is related to the handling of sub-structural information. Each solution of a problem may be validated according to the result of a single substructure search or to any combination of such results. This makes possible to keep in solution set only the structures that contain a four-membered ring and to eliminate cyclopropanes. Molecular substructure sets may be hierarchically organized to be used by LSD, with possible applications in the field of natural product chemistry. Tools for the conversion of molecular structures to LSD substructures have been written in order to facilitate the implementation of substructure libraries, like the one of the SISTEMAT knowledge base\(^3\). Another recent development of LSD is the possibility of defining the interval of coupling path lengths (measured in bonds) that is associated to an HMBC correlation. The coupling path length of a strong intensity HMBC correlation can thus be forced to never be strictly greater than three. This would save time and reduce the number of solutions if very long range HMBC correlations are allowed. The same improvement will be considered for COSY correlations.

\(^1\) http://helios.univ-reims.fr/Labos/UPRESA6013/LSD


\(^3\) J.-M. Nuzillard, V. P. Emerenciano. Automatic structure elucidation through data base search and 2D NMR spectra analysis. Natural Product Communications 2006, 1, 57-64.