Tautomerism Databases

Marc C. Nicklaus

Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, NCI-Frederick, Frederick MD 21702, United States

A compound exhibits tautomerism if it can be represented by two or more structures that are related by a formal intramolecular movement of a hydrogen atom from one heavy atom position to another. The existence of multiple tautomeric forms of the same molecule is still an important, and in many ways unsolved, issue in drug discovery and chemoinformatics. Equally surprisingly, there exist very few publicly available data sets of systematically collected information about molecules capable of undergoing tautomerism, i.e. of experimental or computational determination of tautomer ratios, tautomeric interconversion rates, dependence of these properties on conditions etc., that could be used to, e.g., build better predictive models for the tautomerism of small molecules. We present a database of such data extracted from literature (73 publications including a number of reviews). The current version of this database, which is in the last stages of curation before it will be made available to the public, focuses on experimental data. It has 1,873 entries, each consisting of n-tuples of tautomers studied in a particular set of experimental conditions (pH, solvent, temperature, experimental method), with n ranging from 1 to 8. It comprises records about 3,898 tautomer structures.