STUDY OF COMPUTATION, RELATEDNESS AND ACTIVITY PREDICTION OF TOPOLOGICAL INDICES

I. FABIČ-PETRAČ and B. JERMAN-BLAŽIČ
Institut Jožef Stefan, Ljubljana, Yugoslavia

V. BATAGELJ
Department of Mathematics, University of Ljubljana, Yugoslavia

Abstract

A large number of numerical graph invariants (topological indices) have been defined and used in many different fields of chemistry. Some of them are used as chemical structure descriptors in QSAR (quantitative structure–activity relationship) studies. The paper describes the development and implementation of a computer program for the computation of the most often used topological indices: \( n, n_1, n_2, M_1, M_2, X_5, F, x_1, W, p, J, D^{(5)}, GDI, r, T^e_k, T^L_k \). As these indices reflect different aspects of molecular topology, the intercorrelation among them is investigated by applying hierarchical clustering methods. A method based on string comparison techniques is developed for the determination of indices correlated to biological activity for a studied set of compounds. The biological activity prediction on the basis of a subset of topological indices least-correlated amongst themselves is done by applying the nearest neighbourhood approach.

1. Introduction

In the near past, graph-theoretic methods have been largely applied in many fields of chemistry. Recently, we are witnessing the growth of their use as methods for characterization of chemical structures and use in the field of structure–activity correlations. Graph models of chemical structures retain their full topology which largely determines important characteristics of molecules [1]. The topology of the molecule represents the mutual relations of the atoms, i.e. the information of the connectivity between them in the structure [2]. Mathematicians and chemists have undertaken a long search to find suitable parameters derived from the molecular graph which will reflect the molecular topology. During the last two decades, a large number of numerical graph invariants have been defined and applied in different fields of theoretical chemistry, pharmacology, toxicology, and environmental chemistry. Such quantitative measures which reflect the structural features of the molecule are called topological indices. A topological index of a compound under consideration is a graph invariant derived from the chemical graph of that compound. Topological indices express in numerical form the topology of the chemical compound represented. Topological indices were developed for the purpose of obtaining correlations with the physicochemical properties of chemical substances [3] and to express the molecular
similarity or dissimilarity. They were used for classification and for prediction of chemical and biological properties of the compounds. One of the most important developments in recent years has been the increasing use of topological indices in the design of drugs and other biologically active substances. One important question in drug design is to what extent the structure of a compound and its biological activity are correlated. This is also the main task for QSAR (quantitative structure–activity relationship) analysis. The basic assumption is that the molecular properties are determined primarily by structure and that compounds with similar structure lead to similar biological action or other physicochemical properties. Once a relationship between the molecular descriptors and biological activity is known, it is used to predict the activity of yet untested or even unsynthesized compounds [3,4]. In the text that follows, a study of the relatedness of the topological indices is made. Special methods were developed for determination of less correlated indices. In the process of determination, the ordering of the compounds represented by their topological indices and the corresponding biological response value were used. Some property predictions were done by applying the nearest neighbour techniques appropriately accommodated for the studied set of compounds.

2. **Topological indices**

The graph representation of a molecule consists of points representing the molecule's atoms and bonds linking them as straight lines. In graph theory, points are usually referred to as vertices and lines are referred to as edges. In chemical graphs, the hydrogen atoms are often omitted because they normally do not play a major role in determining the structure of a molecule. The topological indices are derived from such graphs, called hydrogen-suppressed graphs or, sometimes, skeletal structures. Graphs with multiple edges (double, triple bond), weighted vertices (heteroatoms), and weighted edges can also be used, although usually they are not considered. The length of any line representing a chemical bond and the angles between the lines are not considered in this approach.

A topological index of a compound under consideration is a graph invariant derived from the chemical graph of the compound. This number characterizes the molecule and it does not depend on how the vertices are enumerated. Indices are designed by transforming a chemical graph into a number and the means by which this was accomplished varies from index to index.

A graph invariant may be a polynomial (characteristic polynomial), a set of numbers (spectrum of a graph), or a numerical value. In particular, the molecules may be described by several quantitative descriptors reflecting different structural aspects. The purpose of defining a topological index is sometimes to represent each chemical structure with a numerical value, keeping it at the same time as discriminatory as possible. This means assigning to every chemical graph a numerical invariant such that two graphs have the same value of that index if and only if they are