Chapter 38
The "ETA" Indices in QSAR/QSPR/QSTR Research

Kunal Roy
Jadavpur University, India

Rudra Narayan Das
Jadavpur University, India

ABSTRACT

Descriptors are one of the most essential components of predictive Quantitative Structure-Activity/Property/Toxicity Relationship (QSAR/QSPR/QSTR) modeling analysis, as they encode chemical information of molecules in the form of quantitative numbers, which are used to develop mathematical correlation models. The quality of a predictive model not only depends on good modeling statistics, but also on the extraction of chemical features. A significant amount of research since the beginning of QSAR analysis paradigm has led to the introduction of a large number of predictor variables or descriptors. The Extended Topochemical Atom (ETA) indices, developed by the authors' group, successfully address the aspects of molecular topology, electronic information, and different types of bonded interactions, and have been extensively employed for the modeling of different types of activity/property and toxicity endpoints. This chapter provides explicit information regarding the basis, algorithm, and applicability of the ETA indices for a predictive modeling paradigm.

INTRODUCTION

Chemicals are the inseparable components of human life, and thus they can affect a wide variety of industrial, agricultural as well as household processes if not monitored properly. Elucidation of the chemistry of compounds is an essential task of the designer so as to enable easy tuning of their properties. The diverse manifestation of chemicals can be systematically explored when the study of their chemistry is suitably attended with the related biology (including toxicology), mathematics and accompanying statistics. Since the past, mathematics has been considered as an effective media of natural science (Wigner, 1960) and has been employed in different branches of science to derive logical relationship employing abstract ideas. Chemistry has a very good coherence with mathematics that leads to the quantification

DOI: 10.4018/978-1-5225-1762-7.ch038

of chemical information by providing suitable platform for the generation of various algorithms. How different chemicals act differently and even the same chemical elicits different responses give the hint to nurture the 'chemistry' of compounds and explore the possible quantitative phenomena involved. The attempt to quantify the chemistry of compounds has lead to the development of several independent parameters termed as predictor variables or descriptors. These descriptors are further utilized to bring under a quantitative relationship with the responses of the chemicals commonly termed as the quantitative structure-activity/property/toxicity relationship (QSAR/QSPR/QSTR) studies. Mathematics plays a great role in deriving such predictor variables as well as in the development of correlations.

Through the journey of QSAR analysis, several predictor variables have evolved by exploring the concepts of chemistry, mathematics and physics. However, the 'graph theory' has contributed much to the development of initial ideas about descriptors. The graph theory, one of the branches in mathematics, can be traced back to Euler's Königsberg bridges problem (Euler, 1736) which led to the idea of solving problems in different fields of science namely the study of electrical circuit by Kirchhoff (Kirchhoff, 1847), chemical isomer enumeration by Cayley (Cayley, 1875) etc. In a true mathematical sense, the term 'graph' also denotes cartesian plot of data, although it was Sylvester (Sylvester, 1878) who coined the term from a contemporary chemical perspective. Application of the graph theory in chemistry leads to the concept of 'chemical graph' which was supposed to have its origin in the late eighteen century holding the hand of the Scottish chemist Cullen (Crosland, 1959). William Cullen introduced the application of 'affinity diagrams', although it remains controversial with similar type of contributions by Black (Crosland, 1959). Later Higgins (Higgins, 1789) employed the use of graphical diagrams to denote the forces between atoms in a molecule. It was Dalton (Dalton, 1808) in the nineteenth century who started the concept of ball-and-stick representation of chemical models. The idea of diagrammatic presentation of chemical compounds was very primitive during that age and through various notable discoveries the chemistry got explored gradually. The famous scientist Kekulé brought the idea of three-dimensional orientation (Fischer, 1974) and later idea of chemical valence came likewise. Considering the scope of this chapter, we would like to limit this discussion at this point.

It was in the twentieth century when the chemical graphs were explored properly, and utilizing the concepts of topological distance measures the scientific fraternity started developing theoretical variables to correlate activity of chemicals with their chemistry. Although the history of QSAR starts back in the past when Cross (Cross, 1863) reported his idea of possible correlation between chemical composition and toxicity of alcohols, the concept of theoretical descriptor did not come in action properly until the year 1947 when the Wiener index (Wiener, 1947) and the Platt number (Platt, 1947) were introduced as the first two graph theoretical descriptors for modeling boiling point of hydrocarbons. Following these milestone achievements in the realm of quantification of chemistry paradigm, a new wave was hit towards the exploration of graph theoretic parameters, and in the 1960s and 1970s the pathway was systematically imputed by the following pioneering contributions namely Simon's minimal topological difference (MTD) concept (Simon, 1974), Balaban J (Balaban, 1982), Gordon and Scantlebury index (Gordon & Scantlebury, 1964), Schultz molecular topological index (MTI) (Schultz, 1989), Hosoya Z (Hosoya, 1971), Randić's branching parameter (Randić, 1975), Kier and Hall's molecular connectivity, Kappa shape, sub-graph count, flexibility and electrotopological state (E-state) atom indices (Kier & Hall, 1986), Zagreb index by Trinajstić and co-workers (Gutman & Trinajstić, 1972), Harary index (Plavšić et al., 1993), Verloop's STERIMOL parameter (Verloop, 1987), etc. Graph theoretic descriptors are purely based on a two dimensional basis, whereas concepts of three dimension was also under parallel research and various predictor variables were also derived employing three dimensional basis.