# Chapter 37
# On Extended Topochemical Atom (ETA) Indices for QSPR Studies

**Kunal Roy**
*Jadavpur University, India*

**Rudra Narayan Das**
*Jadavpur University, India*

## ABSTRACT

*Development of predictive models has been accepted as an important strategy to aid in toxicity screening of chemicals, determination of physicochemical as well as other biological activity of new molecules, and also in the generation and optimization of lead compounds in rational drug discovery process. The journey of quantitative structure-property relationship (QSPR) modeling started with the development of various property-based and two-dimensional descriptors to model various physicochemical and biological properties (including toxicity). Topological descriptors contain significant information encoded in the molecular structure. Extended topochemical atom (ETA) indices, a relatively new class of topological descriptors, are the focus point in this chapter. ETA indices contain important information regarding the nature of the atoms, bonds, atomic electronic environment and consider the contribution of different functional groups, molecular fragments, and branching to the response as evidenced by different reports showing their successful application in modeling different endpoints including toxicity, drug activity, and physicochemical properties. Extensive research is still going on for the refinement of the ETA indices by the incorporation of some novel parameters, and future reports on ETA indices will include these new indices.*

## 1. INTRODUCTION

Use of chemicals for various purposes has been an inseparable event to human life since time immemorial. Starting from the ancient times to the modern world, the use of chemicals has increased to a great extent to fulfill different demands of mankind. Millions of chemicals are synthesized and released into the market now-a-days to meet the need of the modern industrialization. About 56 million organic and inorganic substances are reported to be present in the CAS registry and there is an addition of around 12,000 new

DOI: 10.4018/978-1-4666-5125-8.ch037

numbers daily (http://en.wikipedia.org/wiki/CAS_registry_number). After synthesis and release of these chemicals, determination of their various physicochemical properties as well as biological activities (including toxicities) becomes essential for various needs. Now, individual assessment of this large number of chemicals for a particular property or a number of properties thereof is really troublesome and impracticable. Appropriate use of computational technique can help in this context to predict the desired property or activity of a large number of chemicals using the experimental values of comparatively lesser number of chemicals. In drug discovery and development, the ultimate main aim of research is to synthesize novel therapuetically and pharmacologically active molecules with minimum side effects and toxicity profile. Because of involvement of huge time and cost in drug discovery process, it is not possible to rely upon trial and error method. Suitable computational technique is very useful in guiding medicinal chemists for the proper selection of compounds. Use of quantitative structure–property/activity/toxicity relationship (QSPR/QSAR/QSTR) analysis makes the drug discovery and toxicological screening process more successful and economic as well (Choplin, 1990; Franke, 1984). Apart from this, development and use of computationally derived predictive models has also got other applications, namely in agriculture, forensic science etc. (Roy, 2004). Regarding toxicity assessment of chemicals, the European Union has implemented Registration, Evaluation and Authorization of Chemicals (REACH) program which requires human and environmental hazard assessment of all chemicals produced or imported >1 ton per annum (tpa) in the European Union (Aptula & Roberts, 2006). In order to avoid toxicity testing in animals, use of alternative tools for the determination of potential hazardous effects of chemicals was prescribed in the European Union White Paper concerning a future of chemical policy (Christensen *et al.*, 2003). QSAR (QSTR) modeling has been proposed by the Office of Toxic Substances of the US Environmental Protection Agency (Auer *et al.*, 1990) and the Agency for Toxic Substances and Disease Registry (ATSDR) (El-Masri *et al.*, 2002) to aid toxicological screening.

Biological activity or toxicity shown by a molecule is a result of its interaction with the biological system or some specific biological receptor using basic physicochemical forces (Andrews & Tintelnot, 1990; Franke, 1984; Martin, 1989). The type and extent to which a compound undergoes such interaction is dependent on the structural properties of the compounds. Proper encoding of these structural features controlling activity and/or property of a molecule in the form of numerical quantities called descriptors forms the basis for the development of predictive models in the form of QSPR/QSAR/QSTR. This study helps in optimizing compounds so as to get better analogues of a series of existing compounds exhibiting a particular activity or toxicity (lead optimization).

Topological indices are simple two dimensional descriptors which can be easily calculated and interpreted. These indices have been widely used by the QSAR community to represent chemical structures numerically (Jurs *et al.*, 1995). A large amount of novel research has been carried out on model development with topological descriptors for different endpoints from the past, and such approaches have been found to be fruitful in many instances. This chapter gives an overview of history of evolution of QSAR techniques and glimpses of different two-dimensional descriptors for effective model development and then specifically focuses on the generation, use and applicability of extended topochemical atom (ETA) indices which were originally derived based on refinement of the topochemically arrived unique (TAU) scheme indices developed in the eighties.