Chapter 10
On Applications of Macromolecular QSAR Theory

Pablo R. Duchowicz
INIFTA, CCT La Plata-CONICET, Argentina

Eduardo A. Castro
INIFTA, CCT La Plata-CONICET, Argentina

ABSTRACT
Present chapter reviews the application of Quantitative Structure-Activity Relationships for the treatment of molecules involving thousands of atoms, such as proteins, nucleic acids (DNA, RNA), or polysaccharides. This is a new developing area of interest in Chemoinformatics, and it is expected to have a growing number of applications during the forthcoming years. Among the several points to be addressed during the modeling of macromolecules, the most important one appears to be the accurate representation of the chemical structure through numerical descriptors. It has to be noticed that descriptors based on optimized three-dimensional geometry are difficult to specify, and it is also a drawback the fact that the experimental geometry is not available. However, different experts in the field have been generalizing the employment of classical types of topological descriptors in macromolecular systems.

INTRODUCTION
During last decades, the Quantitative Structure-Activity Relationships (QSAR) (Hansch, 1995; Kubinyi, 2008) Theory has played an important role in many research areas, such as Medicinal Chemistry, enabling to prevent time consuming and costs associated to experiments. Since the pioneer studies performed by Hansch and Fujita in 1964 (Hansch, 1964), the QSAR formalism has been extensively applied to the study of different biological activities of interest, so the development of the theory is encouraged (Duchowicz, 2008; Duchowicz, 2009a; Duchowicz, 2009b; Goodarzi, 2009; Puzyn, 2009; Selassie, 2002).

DOI: 10.4018/978-1-60960-860-6.ch010
The basis of QSAR relies on the main hypothesis that the biological activity manifested by a chemical compound completely results from its own molecular structure. It is an approach that has a thermodynamical resemblance, in the sense that QSAR is only interested on the initial and final states (molecular structure and final activity, respectively), but does not offer specific details on the usually complex mechanism/path of action involved. However, it is possible to get insight on the underlying mechanism by means of the predicted activity.

In the realms of the theory, the molecular structure is translated into the so-called molecular descriptors, describing some relevant feature of the compounds, with mathematical formulae obtained from Chemical Graph Theory, Information Theory, Quantum Mechanics, Markov Chains Theory, etc. There exist more than a thousand available descriptors in the literature (Diudea, 2001; Katritzky, 1995; Todeschini, 2009; Trinajstic, 1992), and many of these molecular descriptors are topological indices (TIs) or invariants obtained from the molecular graph, whose vertices are atoms weighted with different physicochemical properties (mass, polarity, electronegativity, charge) (Katritzky, 1993).

Even though the relationship between the structure and the activity remains unknown for a given dataset, the QSAR technique has been based on statistically determined linear or nonlinear models relating the chemical behavior of compounds with descriptors, in order to find out useful parallelisms. Altogether, QSAR studies are affected by various factors from which the most important are: (a) the selection of molecular descriptors that should include maximum information of structures and minimum colinearity between them; (b) the use of suitable modeling methods; (c) the number of descriptors to be included in the model; (d) the composition of the training and test sets; and (e) the employment of validation techniques to verify the predictive performance of the developed models. Feature (descriptor) selection has been an active research area in Pattern Recognition, Statistics and Data Mining communities, and its main objective is to select a subset of input variables by eliminating features with little or no predictive information (Guyon, 2003).

In the early 1990’s there has been a great explosion in the proposal of new TIs, and a huge number of TIs have been derived for small-sized molecules that many of them result redundant and their structural information content overlaps with that of others (González-Díaz, 2007d). In addition, most TIs are obtained via vector-matrix-vector procedures (Estrada, 2001). However, the perspective is different for Macromolecular Science, as classical TIs and the more elaborated topographic indices (TPGIs) have never-explored applications on this field. Present chapter reviews such applications performed by different authors during last years.

BACKGROUND

There is a constant need to have an accurate description of the molecular structure, as the characterization of the underlying connectivity of a system enables one to introduce mathematical tools for studying the properties of macromolecules through the QSAR Theory (González-Díaz, 2008b; González-Díaz, 2009). It is possible to extend the classical QSAR applications beyond the study of small-sized molecules. In a chemical graph, each node represents one part of a complex system and the edges represent geometrical or functional relationships between these parts. At the macromolecular scale, such node may be an aminoacid or a protein. A molecular graph can be numerically described using various classic TIs, such as the Wiener index, Zagreb indices, Harary numbers, Randic connectivity indices, valence connectivity indices, Marrero-Ponce quadratic or linear forms, Balaban index, Broto-Moreau autocorrelation, and graph TIs for Markov matrices reported by González-Díaz et al. (2005a).
Related Content

Hydrogen-Bonded Interactions in the Systems L-Cysteine - H2SeO3 and L-Cysteine -H2SeO4: A DFT Study
www.igi-global.com/chapter/hydrogen-bonded-interactions-systems-cysteine/77076?camid=4v1a

A Hybrid Approach Based on Self-Organizing Neural Networks and the K-Nearest Neighbors Method to Study Molecular Similarity
Abdelmalek Amine, Zakaria Elberrichi, Michel Simonet and Ali Rahmouni (2013). Methodologies and Applications for Chemoinformatics and Chemical Engineering (pp. 74-95).
www.igi-global.com/chapter/hybrid-approach-based-self-organizing/77070?camid=4v1a

Graph Mining in Chemoinformatics
www.igi-global.com/chapter/graph-mining-chemoinformatics/45467?camid=4v1a

Bundlet Model for Single-Wall Carbon Nanotubes, Nanocones and Nanohorns
Francisco Torrens and Gloria Castellano (2013). Methodologies and Applications for Chemoinformatics and Chemical Engineering (pp. 228-284).
www.igi-global.com/chapter/bundlet-model-single-wall-carbon/77081?camid=4v1a