Chapter 3
Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease

Matheus P. Freitas
Federal University of Lavras, Brazil

Mariene H. Duarte
Federal University of Lavras, Brazil

ABSTRACT

Multivariate Image Analysis applied in Quantitative Structure-Activity Relationship (MIA-QSAR) is a simple method to achieve, at least in a variety of examples, QSAR models with predictive abilities comparable to those of sophisticated tridimensional methodologies. MIA-QSAR is based on the correlation between properties (e.g. biological activities) and chemical descriptors, which are pixels of images representing chemical structures in a congeneric series of molecules. The MIA-QSAR approach has been improved since its creation, in 2005, both in terms of data analysis and development of more descriptive information. This chapter reports the MIA-QSAR method, including its augmented version, named aug-MIA-QSAR because of the introduction of new dimensions to better encode atomic properties. In addition, the application to a case study illustrates the main practical differences between traditional and augmented MIA-QSAR. The use of a neglected disease as example represents a challenge in QSAR, which is particularly focused on diseases with higher economical appearance.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) methods have shown to be powerful tools to comprehend the action mechanisms and structural profiles required to improve a drug performance, as well as to estimate the bioactivity of a non-existing, proposed drug candidate. QSAR approaches are genuinely structure-based methods, in which biological properties are reflected by structural changes in molecules, despite the receptor information included in the modern multidimensional QSAR techniques. However, traditional QSAR, which is based on descriptors easily accessible or calculated, such as log*P*, connectiv-

DOI: 10.4018/978-1-4666-8136-1.ch003

ity indices and a variety of other parameters derived from the 2D structure of a molecule, is not inferior to methods based on the three-dimensional molecular structure, at least in many practical cases (Brown & Martin, 1997; Estrada, Molina, & Perdomo-López, 2001).

Multivariate image analysis applied in QSAR (MIA-QSAR) appeared to correlate drawings of molecular structures with the corresponding biological activities (Freitas, Brown, & Martins, 2005); therefore, it is essentially a 2D QSAR technique, since descriptors are obtained from the projection of a molecule in the plane. The structural changes in a congeneric series of drug-like compounds explain the variance in the bioactivities block; in MIA-QSAR, structural changes correspond to different coordinates of the pixels composing the molecular drawings. Because the substitution pattern along with the congeneric series of compounds is captured in the calibration step, usually performed using partial least squares (PLS) regression, prediction of the bioactivities of similar compounds are often feasible. However, much chemical information is lost when a given substituent is represented as letters in the drawings, although other molecular properties, such as steric effects (*e.g.* for large side chains) and shape (*e.g.* the hexagonal benzene ring), are appropriately encoded. For instance, the MIA-QSAR model is capable of recognizing that a bromine substituent bonded to a given aromatic carbon causes an effect on the bioactivity of a molecular scaffold, but its description as "Br" in the drawing does not have chemical meaning. Thus, an augmented version for the MIA-QSAR method (aug-MIA-QSAR) was developed by introducing "dimensions" to better encode atomic sizes (using spheres with sizes proportional to the van der Waals radii) and different types of atoms (using spheres with different colors) in a molecule (Nunes & Freitas, 2013). Because aug-MIA-QSAR has been recently implemented, there are many challenges to improve its predictive ability, as well as its chemical interpretation; research directions include testing regression and variable selection methods for the aug-MIA descriptors, and also searching for ways to indicate how different atomic sizes and colors impact the trends in bioactivity in a series of drug-like compounds.

QSAR methods have been used in numerous studies in order to find a correlation between chemical structures and biological activities related to profitable diseases, such as obesity, sexual dysfunction and hypertension. Nevertheless, little attention has been devoted to neglected diseases, which mostly affect poor people from the third world, being forgotten by the big pharmaceutical companies (Ramalho, Freitas, & da Cunha, 2012). This chapter describes the development of a variety of QSAR methods during decades, focusing on the MIA-based approaches, which is demonstrated here to a set of thiosemicarbazones as anti-*Trypanosoma cruzi* agents (Garkani-Nejad & Ahmadi-Roudi, 2010). The efficacy of the current chemotherapy against *T. cruzi* is quite variable in different regions of high endemicity, because the high biological, biochemical and genetic diversity of *T. cruzy* strains (Zingales et al., 2009). Thus, the development of drugs based upon the structural optimization of existing compounds, such as the Nifuroxazide, is advantageous, considering time and efforts consumption.

Background

QSAR methods are based on the possibility of the bioactivity (or any other property) in being a function of the molecular structure, that is bioactivity = *f*(structure), in which the structure is represented by molecular parameters called descriptors. This aims at planning new substances with improved therapeutic profile. Therefore, this research field is expected to be of general interest, because the use of QSAR methods avoids exhaustive exploratory syntheses, since these methods enable designing molecules with well-defined properties, reducing time and costs during the drug development.