Chapter 39
Importance of Applicability Domain of QSAR Models

Kunal Roy
Jadavpur University, India

Supratik Kar
Jadavpur University, India

ABSTRACT

Quantitative Structure-Activity Relationship (QSAR) models have manifold applications in drug discovery, environmental fate modeling, risk assessment, and property prediction of chemicals and pharmaceuticals. One of the principles recommended by the Organization of Economic Co-operation and Development (OECD) for model validation requires defining the Applicability Domain (AD) for QSAR models, which allows one to estimate the uncertainty in the prediction of a compound based on how similar it is to the training compounds, which are used in the model development. The AD is a significant tool to build a reliable QSAR model, which is generally limited in use to query chemicals structurally similar to the training compounds. Thus, characterization of interpolation space is significant in defining the AD. An attempt is made in this chapter to address the important concepts and methodology of the AD as well as criteria for estimating AD through training set interpolation in the descriptor space.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) modelling has become an important tool in the new drug candidate design, environmental fate modeling, toxicity and property prediction of chemicals and pharmaceuticals since they offer an economical and time-effective alternative to the medium throughput *in vitro* and low throughput *in vivo* assays (Perkins *et al*., 2003; Selassie, 2003; Walker et al., 2003). A QSAR model is a simple mathematical equation that is evaluated from a set of molecules with known activities/properties/toxicities using computational approaches. Thereafter, the developed predictive QSAR models are also applied by regulatory agencies to evaluate physical, chemical, and biological properties of individual chemical entities using applications specific for decision-making frameworks in risk and safety assessments (Kar & Roy, 2010). The QSAR modeling hypothesis also supports the 3Rs

DOI: 10.4018/978-1-5225-1762-7.ch039

(replacement, refinement and reduction in animals in research) paradigm due to an increased pressure from social and economic background to trim down the use of animal testing as an important alternative method for future prediction of untested chemical entities (Benigni & Giuliani, 2003).

Robust validation of QSAR models plays a key step for the selection of a predictive model that may be considered for future prediction of new molecules. As numerous numbers of researches have been directed to the design of new molecules with the utilization of QSAR technique, validation of a QSAR model has been certified as the most considerable stride (Carlsen *et al*., 2009) for assessing the quality of data, applicability and mechanistic interpretability of the developed model. Thus, a huge number of investigations are currently directed towards the introduction of more appropriate validation approaches for more accurate and predictive QSAR model development. One important objective of QSAR modeling is to predict activity/property/toxicity of new chemical entities falling within the applicability domain of the developed models. The reliability of any QSAR model depends on the confident predictions of these new compounds based on applicability domain (AD) of the modeled compounds and therein arises the importance of the AD study (Ferreira, 2001).

In order to establish the scientific validity of a QSAR model and to facilitate its acceptance for regulatory purposes, the Organization for Economic Cooperation and Development (OECD) in its joint meeting (OECD, 2007) has agreed to five principles that should be followed during the construction of QSAR models. The OECD Principle 3 which defines the need of an AD expresses the fact that QSARs are unavoidably associated with restrictions in terms of the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions. The applicability domain of a QSAR model has been defined as the response and chemical structure space, characterized by the properties of the molecules in the training set. The developed QSAR model can predict a new compound truly only when it falls within the applicability domain of the developed model (Netzeva *et al*., 2005). Thus, to identify the interpolation (true prediction) or extrapolation (less reliable prediction) of query compounds is an important task for a QSAR model developer using the information of applicability domain (Netzeva *et al*., 2005).

Viewing the importance of AD in QSAR model validation, we herein focus to get an overview of different traditional as well as relatively new AD approaches used to judge the quality of the QSAR models. This book chapter will be helpful for the QSAR learners in order to have a clear idea on the principles of available AD approaches useful for judging the predictive quality of QSAR models.

BACKGROUND

A QSAR model is essentially valued in terms of its predictability, indicating how well it is able to predict the endpoint values of the compounds which are not used to develop the correlation. The models that have been suitably validated internally and externally can be considered reliable for both scientific and regulatory purposes (Golbraikh & Tropsha, 2002; Tong *et al*., 2004]. QSAR models should be validated according to the OECD principles for reliable prediction. A meeting of QSAR experts held in Setúbal, Portugal in March 2002 formulated guidelines for the validation of QSAR models, in particular for regulatory purposes (Jaworska *et al*., 2003). These principles were agreed by OECD member countries, QSAR and regulatory communities at the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. These principles are best pos-