

Applicability Domain for QSAR Models: Where Theory Meets Reality

Domenico Gadaleta, University of Bari, Bari, Italy

Giuseppe Felice Mangiatordi, University of Bari, Bari, Italy

Marco Catto, University of Bari, Bari, Italy

Angelo Carotti, University of Bari, Bari, Italy

Orazio Nicolotti, University of Bari, Bari, Italy

ABSTRACT

Quantitative Structure-Activity Relationships are widely acknowledged predictive methods employed, for years, in organic and medicinal chemistry. More recently, they have assumed a central role also in the context of the explorative toxicology for the protection of environment and human health. However, their real-life application has not been always enthusiastically welcomed, being often retrospectively used and, thus, of limited importance for prospective goals. The need of making more trustable predictions has thus addressed studies on the so-called Applicability Domain, which represents the chemical space from which a model is derived and where a prediction is considered to be reliable. In the present study, the authors survey a number of approaches used to build the Applicability Domain. In particular, they will focus on strategies based on: a) physico-chemical, b) structural and c) response domains. Moreover, some examples integrating different strategies will be also discussed to meet the needs of both model developers and downstream users.

KEYWORDS

Applicability Domain, Interpolation Space, QSAR, Similarity, Structural Fragments

INTRODUCTION

In the last years, great concerns were raised by numerous troubles in designing and developing new drugs. Failures can occur at different stages of the drug pipeline, from lead/hit discovery to pre-clinical testing to clinical trials (Hay, Thomas, Craighead, Economides, & Rosenthal, 2014). However, it was observed that a great percentage of these flops (about the 60% of all the drugs fail) is due to the occurrence of inadequate ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties, not thoroughly evaluated at the very early phases of drug discovery (Weaver, 2008). Needless to say that the need of making trustable predictions is still extremely important and, thus, the role of *in silico* methods has never been in question although greater attention should be paid in their application and in the interpretation of results (Kubinyi, 1997a; Kubinyi, 1997b; Pisani et al., 2014).

As well-known, Quantitative Structure-Activity Relationships (QSARs) are *in silico* approaches developed to predict a certain response (i.e., endpoint: pharmacological or toxic effect or physico-chemical property) of chemicals. Normally, a QSAR model consists in a mathematical equation relating an endpoint to a series of meaningful molecular descriptors, numerical attributes featuring the chemicals, which are object of investigation (Nicolotti & Carotti, 2006).

DOI: 10.4018/IJQSPR.2016010102

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Besides rational drug design (Kubinyi, 1997b), QSARs have more recently proved to be also a valuable tool in the regulatory context for the evaluation and prediction of risks posed by chemicals on the environment and on human health. In December 2006, the European Community (EC) introduced a new regulation concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) (EC, 2006). The aim of REACH is to ensure a high level of protection of environment and human health by gathering and analyzing a great number of toxicological data for thousands of chemical substances that are on the European market (Nicolotti et al., 2014). Importantly, REACH Article 1 explicitly encourages the use of alternative methods, such as QSARs, for obtaining such information.

The usage of QSAR for regulatory purposes must fulfill a number of important requisites itemized by the Organization for Economic Co-operation and Development (OECD) as follows: 1) a well-defined endpoint; 2) an unambiguous algorithm for model derivation; 3) a clearly defined Applicability Domain (AD); 4) appropriate measures of goodness-of-fit, robustness, and predictivity; and 5) a mechanistic interpretation, if possible. (OECD, 2007).

The present review focuses on the third point, which explicitly claims the need to report, along with a QSAR model, its AD or, in other terms, the areas of the physico-chemical, structural, or biological space where the model is expected to be exploitable and the predictions are assumed to be trustable. Indeed, a QSAR model cannot reliably predict any chemical but only those having physico-chemical, structural and biological features similar to those of Training Set (TS) compounds, used to create a causative relationship with a given endpoint. In this respect, TS compounds define the boundaries of the chemical space where predictions could be considered as the result of data interpolation. On the other hand, predictions made outside such space are the result of data extrapolation and, thus, meaningless, chancy and likely wrong.

Building on these evidences, the definition of AD is an essential step for both scientific and regulatory purposes as it increases the confidence in predictions and allows a practical use of QSAR models. In the Setubal Workshop (Jaworska, Comber, Van Leeuwen, & Auer, 2003), the AD of a QSAR model was defined as “the physico-chemical, structural, or biological space, knowledge or information on which the TS of the model has been developed, and for which it is applicable to make predictions for new compounds[...]. Ideally, the QSAR should only be used to make predictions within that domain by interpolation not extrapolation”. A broader definition was proposed in the report of the 52nd ECVAM (European Centre for the Validation of Alternative Methods) Workshop (Netzeva et al., 2005, p. 3): “the AD of a QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability.” In the present review, the authors survey the most important approaches so far employed for defining the AD as summarized in Table 1. The ultimate aim is that this paper could represent a valuable yardstick for both QSAR practitioners and downstream users encouraging a greater consciousness in predicting new data and decision-making. Such methods take inspiration from the three main domains of the QSAR space, explicitly cited in both the above reported definitions: a) the physico-chemical, b) the structural, and c) the biological (or response) domain. Additionally, a series of combined approaches described in literature is reported. It is intuitive that the application of strategies encompassing different approaches for characterizing the model’s space has better chance to return a complete and effective definition of AD (Kühne, Ebert, & Schüürmann, 2009).

EVALUATION OF PHYSICO-CHEMICAL DOMAIN

Physico-chemical descriptors represent the coordinates to allocate into a multi-dimensional space the TS compounds from which deriving through interpolation an area that can be employed as AD. In this respect, the use of physico-chemical descriptors has led to develop four major approaches (Netzeva et al., 2005; Jaworska, Nikolova-Jeliazkova, & Aldenberg, 2005; Roy & Kar, 2015) that are: range-, distance-, geometrical-, and probability density distribution-based methods.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/applicability-domain-for-qsar-models/144689

Related Content

A Quantitative Structure-Property Relationship Study of the Adsorption of Amino Acids on Kaolinite Surfaces

Yu Heng Ou, Len Chang and Chia Ming Chang (2018). *International Journal of Quantitative Structure-Property Relationships* (pp. 21-35).

www.irma-international.org/article/a-quantitative-structure-property-relationship-study-of-the-adsorption-of-amino-acids-on-kaolinite-surfaces/204890

HyLaw or Hydrogen Law: A Regulation for Removal of Legal Barriers to the Deployment of Fuel Cells and Hydrogen Applications

Nima Norouzi (2022). *Green Chemistry for the Development of Eco-Friendly Products* (pp. 24-42).

www.irma-international.org/chapter/hylaw-or-hydrogen-law/307045

Computational and Metabolic Studies on a Set of N-Myristoyltransferase Inhibitors Against Trypanosoma Brucei

Luciana Scotti, Hamilton Ishiki, Francisco Jaime Bezerra Mendonça Junior, Frederico Fávoro Ribeiro, Nagendra Sastry Yarla, Marcelo Sobral da Silva, José Maria Barbosa Filho and Marcus Tullius Scotti (2018). *International Journal of Quantitative Structure-Property Relationships* (pp. 80-94).

www.irma-international.org/article/computational-and-metabolic-studies-on-a-set-of-n-myristoyltransferase-inhibitors-against-trypanosoma-brucei/204894

Synthesis and Structure of the Protein

(2024). *Biochemistry in the Space of the Highest Dimension* (pp. 194-227).

www.irma-international.org/chapter/synthesis-and-structure-of-the-protein/345672

Isoxazole: A Bioactive Five-Membered Heterocycle With Diverse Applications

Popat Mohite, Savita Tauro and Aarati Pawar (2025). *Five Membered Bioactive N and O-Heterocycles: Models and Medical Applications* (pp. 103-128).

www.irma-international.org/chapter/isoxazole/370713