Review Article **ISSN: 2349 – 7106**

Asian Journal of Research in Chemistry and Pharmaceutical Sciences

Journal home page: www.ajrcps.com

QSAR MODELS AND GUIDANCE: A REVIEW

Shirish Ambulgekar*[1]

*[1]School of Chemical Sciences, Swami Raman and Teerth Marathwada University, Vishnupuri, Nanded-431 606, Maharashtra, India.

ABSTRACT

Quantitative structure-activity relationship (QSAR) modeling pertains to the construction of predictive models of biological activities as a function of structural and molecular information of a compound library. The concept of QSAR has typically been used for drug discovery and development and has gained wide applicability for correlating molecular information with not only biological activities but also with other physicochemical properties, which has therefore been termed quantitative structure-property relationship (QSPR). Typical molecular parameters that are used to account for electronic properties, hydrophobicity, steric effects, and topology can be determined empirically through experimentation or theoretically via computational chemistry. A given compilation of data sets is then subjected to data pre-processing and data modeling through the use of statistical and/or machine learning techniques. This review aims to cover the essential concepts and techniques that are relevant for performing QSAR/QSPR studies through the use of selected examples from our previous work.

KEYWORDS

Quantitative structure-activity relationship, QSAR, Quantitative structure-property relationship and Multivariate analysis.

Author for Correspondence:

Shirish Ambulgekar,
School of Chemical Sciences,
Swami Raman and Teerth Marathwada University,
Vishnupuri, Nanded-431 606, Maharashtra, India.

Email: shirishambulgekar@yahoo.com

INTRODUCTION

Quantitative structure-activity and relationships, often simply known as QSAR, is an analytical application that can be used to interpret the quantitative relationship between the biological activities of a particular molecule and its structure[1]. It is considered a major method of chemical researching all over the world today and is frequently used in agricultural, biological, environmental, medicinal, and physical organic studies[2-6].

Quantitative structure-activity and relationships techniques have been used throughout the past century. It was initially used by a scientist from the University of Strasbourg in the 1860's who surprisingly noticed how the toxicity of alcohols in mammalians had improved when a decrease in solubility of water took place[7-9]. The technique of QSAR was later implemented by two other scientists, Crum-Brown and Fraser, and they proposed a correlation between the physiological activities and chemical structures existed for a series of compounds. Even though QSAR has been around, it was not regularly used or found significantly until after scientist, Corwin Hansch began doing his pioneer work in pharmaceutical research[10-14].

Usage and function[14-20]

The main objective of QSAR is to observe the biological responses of a set of molecules, measure it, and statistically relate the measured activity to some molecular structure on their surface. The product of QSAR will then produce useful equations, images or models in either 2D or 3D form that would relate their biological responses or physical properties to their molecular structure.

A General representation of the QSAR equation[21]

Biological Activity = co + cd1 + $(c2d1)^2$ + c3d2 + $c4d2^2$ + ….

*di = the value of the descript or for each molecule in the series

*ci = represents a coefficient calculated by fitting variations in the data by regression analysis

Simplified Overview of the Stages of QSAR Analysis

1. Compile molecular descript or for compounds that are expected to be successful products in a reaction.
   Examples of traditional QSAR descriptors include pKa, Es, log P, pi
2. Express the biological property as a function of the molecular descriptors in a plausible equation, such as the equation listed above.
3. Estimate the activity of a drug candidate based on the molecular descriptors and the QSAR equation by estimating the value based on two known values or by inferring from values within an already observed interval of the data.

Related Application[22-28]

Drug Discovery

Influenza has been known as one of the major causes of death worldwide due to its highly contagious nature and susceptibility in the aging population. Vaccines have been the primary prevention for influenza, but because of its limited effectiveness in patients, an improved alternative such as antiviral drugs have been considered. In today's pharmaceutical market, four drugs have been made available to the general public to treat and/or prevent influenza. They are amantadine, osetlamivir, rimantadine, and zanamivir. These drugs were all analyzed using the unique conceptual modeling device of QSAR, where models of compounds for influenza inhibition were developed and it was from these enhancing models, researchers were able to suggest the inhibitor of influenza to be dependent on its hydrophobicity.

Quantitative structure activity relationship (QSAR) models are a statistical solution to the problem of directly calculating physical and biological properties of molecules from their physical structure. The direct prediction of properties is in general not feasible either owing to lack of computing resources or lack of knowledge about the relationship between structure and property. The goal of a QSAR model is to extract information from a set of numerical descriptors characterizing molecular structure and use this information to develop inductively a relationship between structure and property.

Two important questions arise during the modeling process. First, are the data used to build the model representative of the whole dataset and can the model be extended to predict properties for new molecules?. Second, given that a model encodes information about the structures of molecules and relates this to their properties, can we extract and interpret the encoded information?.

Until recently advances in medicinal and pharmaceutical chemistry depended on a trial and error process aided by intuition. Though the properties that would indicate a certain molecule as a drug candidate were known, it was not really feasible to investigate large numbers of molecules for these types of properties. Of course, the nature of these properties

would be represented by structural features of a molecule and thus examination of certain motifs provided a direction for experimental investigations. The problem with this approach is that it does not always lead to an understanding of why a molecule behaves as a drug against its target or why it does so. Furthermore, given a series of compounds it is not always feasible to investigate experimentally which members of the series would be more potent or less toxic. As a result, though medicinal chemistry has resulted in a series of life saving drugs, the process has traditionally been slow and tedious, and in many cases advances have been due to serendipity rather than scientifically guided investigation. In an ideal world one would be able to take a 3-D molecular structure and calculate the required properties. This utopian goal has a number of problems associated with it.

First, what types of properties are to be calculated? Certain intrinsic physical properties can be calculated using ab-initio quantum mechanical computation techniques. Examples include dipole moments, charges and heats of formation. Though these are certainly useful, they do not provide much insight into drug-like properties such as potency and bio availability. In addition, for large collections of molecules, ab-initio techniques become very time consuming. Semi-empirical quantum mechanical methods alleviate the intensive nature of these calculations, but we are still faced with the restriction on the types of properties that can be calculated. Second, the drug-like activity of a molecule is intimately related to the target it is supposed to interact with. Targets generally involve some type of protein to which the putative drug will bind. Thus when considering the activity of a drug, we cannot simply consider the properties of the drug molecule itself. That is, the nature of the interaction between the drug and target must be investigated to understand fully the activity of a drug. However, ab-initio and semi-empirical techniques have traditionally not been suited for the modeling of large protein systems. Though recent advances in linear scaling and hybrid techniques have expanded the purview of quantum mechanical methods to systems containing tens of thousands of molecules, these methods are still not efficient enough to model thousands or millions of molecular structures, and their associated targets, at a time. Third, though the interactions of a drug with its target are certainly important, the drug must be absorbed by cells and the also metabolized and excreted from the body. Thus absorption properties, the nature of the metabolites and other characteristics must also be considered. Clearly, these are very complex properties that involve interactions with a large number of cellular processes. Modeling this quantum mechanically is nearly impossible.

The above discussion illustrates two fundamental problems. It is not feasible to calculate from theory all the properties of a drug molecule that would help us understand its activity and its utility, and we want to be able to analyze large sets of molecules for these properties. Why do we need to analyze large sets of molecules? The reason for this is closely tied to the nature of drug discovery in recent years. The drug discovery process is time consuming and expensive. Often it can take 10 to 15 years for a drug to reach the market from the laboratory. Given this situation, it is important that a company select the proper compound for study. Combined with the results from high throughput screens and in-house libraries, this can mean having to select tens or hundreds of compounds from a collection of millions. Furthermore, the ability to generate an arbitrary number of unique chemical structures in silico, to create virtual libraries, supplants the actual compounds that a company might have synthesized in its physical collection. Clearly, testing each compound libraries (virtual or real) for drug-like properties is out of the question. As we have seen above, calculating properties for collections of this size is either not feasible or impossible. The question thus comes down to this: how can we calculate arbitrary properties of hundreds of thousands of molecules rapidly and accurately? The short answer is that we avoid the calculation step completely and instead predict a property of a set molecules based on a model derived from the measured values of that property for a small subset. In a QSAR analysis, the central task is to find a regression function that predicts the activity of the molecule in high accuracy. Hence, the present study is aimed at to establish the QSAR between experimental antiplasmodial activity and structure

electronic descriptors which may focus on the molecular structures of the compounds. In last decades, QSAR have been applied in many areas enabling to prevent time consuming and cost during the analysis of biological activities of interest. The main hypothesis involved in any QSAR is the assumption that the variation of the behavior of chemical compounds, as expressed by any experimentally measured biological property, can be correlated with numerical entities related to some aspect of the chemical structure termed molecular descriptors. Descriptors are generally used to describe different characteristics/attributes of the chemical structure in order to yield information about the activity/property being studied. In general, QSAR studies are affected by various factors from which the most relevant are: (a) the selection of the best molecular descriptors that should include maximum information of molecular structures and a minimum overlap between them; (b) the optimal number of descriptors to be included in the model; (c) the use of suitable modeling methods; (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.

We consider that the linear methodology is the statistical technique for analyzing present dataset of benzothiazoles derivatives series, as few experimental observations are available on it and thus it is necessary to employ the lowest number of optimized parameters during the model development. In this way, we resort to the Replacement Method (RM) as variable subset selection approach applied on a pool containing more than a thousand of descriptors, as this technique has been successful for selecting relevant structural descriptors. Finally, another main interest of present research is to apply the so derived QSAR models for estimating the antiplasmodial potency on some new structures, for which there still are no experimental activities.

The construction of QSAR/QSPR model typically comprises of two main steps:

(i) Description of molecular structure and

(ii) Multivariate analysis for correlating molecular descriptors with observed activities/properties.

An essential preliminary step in model development is data understanding. Intermediate steps that are also crucial for successful development of such QSAR/QSPR models include data preprocessing and statistical evaluation. A schematic representation of the QSAR process is illustrated in Figure No.2. Data understanding Data understanding is a crucial step that one should not overlook as it helps the researcher to become familiar with the nature of the data prior to actual QSAR/QSPR model construction thereby reducing unnecessary errors or labors that would otherwise occur. An added benefit is that such preliminary observations can often lead to the identification of interesting associations or relationships to study. However, before exploring the data it is essential that thorough literature search on relevant background information pertaining to the biological or chemical system of interest is performed. This can be achieved through what is known as exploratory data analysis which often starts with simple observation of the data matrix particularly the variables (also known as attributes or fields), its corresponding data types, and the data samples (also called records).

Molecular descriptors[28-36]

Molecular descriptors can be defined as the essential information of a molecule in terms of its physicochemical properties such as constitutional, electronic, geometrical, hydrophobic, lipophilicity, solubility, steric, quantum chemical, and topological descriptors. A more in-depth explanation of molecular descriptors can be found in the literature (Helguera *et al*., 2008; Karelson *et al*., 1996; Katritzky and Gordeeva, 1993; Labute, 2000; Randic, 1990; Randic and Razinger, 1997; Xue and Bajorath, 2000) and a more extensive treatment in the encyclopedic Handbook of Molecular Descriptors (Todeschini and Consonni, 2000). From a practical viewpoint, molecular descriptors are chemical information that is encoded within the molecular structures for which numerous sets of algorithms are available for such transformation. Such descriptors could be calculated

using general quantum chemical software such as Gaussian (Frisch *et al*., 2004), Spartan (Wave function, 2004), GAMESS (Gordon and Schmidt, 2005; Schmidt *et al*., 1993), NWChem (Kendall *et al*., 2000), Jaguar (Schrödinger, 2008), MOLCAS (Karlström *et al*., 2003), Q-Chem (Shao *et al*., 2006), Dalton (Angeli *et al*., 2005), and MOPAC (Stewart, 2009) or specialized software such as DRAGON (Taletesrl, 2007; Tetko *et al*., 2005), CODESSA (Katritzky *et al*., 2005), ADRIANA. Code (Molecular Networks GmbH Computer chemie, 2008), and RECON (Sukumar and Breneman, 2002). Once the molecular descriptors have been calculated it will serve as independent variables for further construction of the QSAR model.

Modeled activities/properties[37-43]

The activities and properties that can be modeled by QSAR/QSPR are dependent variables of the QSAR model. These dependent variables are assumed to be influenced by the independent variables which are the molecular descriptors. A variety of biological and chemical properties have successfully been modeled using the QSAR approach.

Data pre-processing[44-46]

Data pre-processing can be considered to be one of the most important phase of data mining as it helps to ensure the integrity of the data set before proceeding further with data mining analysis. Essentially, the quality of a data mining analysis is a function of the quality of the data to be analyzed. This is often summarized by the "garbage in-garbage out" rule. Therefore, to obtain reliable QSAR models it is important to handle the data with great care.

Data cleaning[47-53]

The preliminary steps in data preprocessing typically requires *data cleaning* as raw data often contain anomalies, errors, or inconsistencies such as missing data, incomplete data, and invalid character values which may cause trouble for data mining software if left untreated. This matter is made complicated when information are consolidated from various sources as such data would need to be prepared to conform to designated criteria and redundant information would also need to be eliminated.

Data transformation[54-65]

There exists a great deal of variability in the range and distribution of each variable in the data set. However, this may pose a problem for data mining algorithms such as neural network which involves distance measurements in the learning step. Such situation is handled by applying statistical techniques such as min-max normalization or z-score standardization. In min-max normalization, the minimum and maximum value of each variable is adjusted to a uniform range between 0 and 1.

Multivariate analysis[66-74]

Multivariate analysis is essentially an approach to quantitatively discern relationships between the independent variables (e.g. molecular descriptors) and the dependent variables (e.g. biological/chemical properties of interest). The classical approach is a linear regression technique typically involving the establishment of a linear mathematical equation.

CONCLUSION

The past few decades have witnessed many advances in the development of computational models for the prediction of a wide span of biological and chemical activities that are beneficial for screening promising compounds with robust properties. In this review article, we have provided a brief introduction to the concepts of QSAR along with examples from our previous investigation son diverse biological and chemical systems. It should be noted that the applicability of QSAR models are only useful in the domains that they were trained and validated. As such, QSAR models spanning wider domains of molecular diversity have the benefit of being valid for wider spans of molecules. It is also interesting to note that there are many paths for researchers in the field of QSAR/QSPR in their quest of establishing relationships between structure and activities/properties. Such abstract nature holds the beauty of the field as there are endless possibilities in reaching the same destination of designing novel molecules with desirable properties.

ACKNOWLEDGEMENT

I would like to express profound gratitude to ***SRTM University, Nanded*** for the invaluable support, encouragement, supervision and useful suggestions. It was a privilege and honor working under ***Dr. Zubaidha P. K.*** for this work. Her deep insight in to the subject and clear thoughts were always the guiding light for the completion of this task.

Moreover, my sincere thanks to my friends and staff of department of School of Chemical sciences, family members, relatives, senior professionals, and colleague for their support and encouragement.

I would like express my deep gratefulness to ***Lord Shri Shri Laxmi - Vekateshwara*** and ***Shri Hanumana***, I remain always thankful for His blessings.

CONFLICT OF INTEREST

We declare that we have no conflict of interest.

BIBLIOGRAPHY

1. Akahori Y, Nakai M, Yakabe Y, Takatsuki M, Mizutani M, Matsuo M and Shimohigashi Y. Two-step models to predict binding affinity of chemicals to the human estrogen receptor by three-dimensional quantitative structure – activity relationships (3D-QSARs) using receptor-ligand docking simulation, *SAR and QSAR in Environmental Research,* 16(4), 2005, 323-337.
2. Anstead G M, Carlson K E and Katzenellenbogen J A. The estradiol pharmacophore: Ligand structure estrogen receptor binding affinity relationships and a model for the receptor binding site, *Steroids*, 62(3), 1997, 268- 303.
3. Arena V C, Sussman N B, Mazumdar S, Yu S and Macina O T. The Utility of Structure -Activity Relationship (SAR) Models for Prediction and Covariate Selection in Developmental Toxicity: Comparative Analysis of Logistic Regression and Decision Tree Models. SAR and QSAR in Environmental Research 15(1), 2004, 1 - 18.
4. Asikainen A H, Ruuskanen J and Tuppurainen K A. Consensus kNN QSAR: a versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands, *Environmental Science and Technology*, 38(24), 2004, 6724-6729.
5. Benfenati E, Roncaglioni A, Boriani E, Porcelli C, Spreafico M and Lo Piparo E. Validation of selected, non-commercial (Q)SAR models for ER and AR binding, Final report of JRC contract CCR.IHCP.C430414.X0. 2005.
6. Brown L P, Lewis D F V, Flint O P, Orton T C and Gibson G G. Teratogenicity of phenylhydantoins in an *in vitro* system: Molecular orbital-generated quantitative structure-toxicity relationships, *Xenobiotica*, 19(12), 1989, 1471-1481.
7. Cramer R D, Patterson D E and Bunce J D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110(18), 2002, 5959-5967.
8. Cronin M T D and Worth A P. (Q)SARs for Predicting Effects Relating to Reproductive Toxicity, *QSAR and Combinatorial Science*, 27(1), 2008, 91-100.
9. Cruciani G, Crivori P, Carrupt P A and Testa B. Molecular fields in quantitative structure-permeation relationships: the Vol Surf approach, *Journal of Molecular Structure: THEOCHEM*, 503(1-2), 2000, 17-30.
10. Dawson D A. Additive incidence of developmental malformation for *Xenopus embryos* exposed to a mixture of ten aliphatic carboxylic acids, *Teratology*, 44(5), 1991, 531-546.
11. Dawson D A and Wilke T S. Initial evaluation of developmental malformation as an end point in mixture toxicity hazard assessment for aquatic vertebrates, *Ecotoxicology and Environmental Safety*, 21(2), 1991, 215-226.
12. Dawson D A, Schultz T W and Hunter R S. Developmental toxicity of carboxylic acids to Xenopus embryos: A quantitative structure-activity relationship and computer-automated structure evaluation, *Teratogenesis, Carcinogenesis, and Mutagenesis*, 16(2), 1996, 109-124.
13. Demyttenaere-Kovatcheva D A, Cronin M T D, Benfenati E, Roncaglioni A Lo Piparo E. Identification of the structural requirements of the receptor-binding affinity of diphenolic azoles to

estrogen receptors alpha and beta by three-dimensional quantitative structure-activity relationship and structure-activity relationship analysis, *Journal of Medicinal Chemistry*, 48(24), 2005, 7628-7636.
14. Devillers J, Chezeau A and Thybaud E. PLS-QSAR of the adult and developmental toxicity of chemicals to *Hydra attenuate*, *SAR and QSAR in Environmental Research*, 13(7), 2002, 705-712.
15. Devillers J, Chezeau A, Thybaud E and Rahmani R. QSAR modeling of the adult and developmental toxicity of glycols, glycol ethers and xylenes to *hydra attenuate*, *SAR and QSAR in Environmental Research*, 13(5), 2002, 555-566.
16. Dix D J, Houck K A, Martin M T, Richard A M, Setzer R W and Kavlock R J. The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. Toxicological Sciences 95(1), 2007, 5-12.
17. Enoch S J, Cronin MTD, Madden J C and Hewitt M. Formation of Structural Categories to Allow for Read Across for Teratogenicity, *QSAR and Combinatorial Science*, 28(6-7), 2009, 696-708.
18. Enslein K, Lander T R and Strange J R. Teratogenesis: a statistical structure-activity model, *Teratogenesis, Carcinogenesis, and Mutagenesis*, 3(3), 1983, 289-309.
19. Fabjan E, Hulzebos E, Mennes W and Piersma A H. A Category Approach for Reproductive Effects of Phthalates, *Critical Reviews in Toxicology*, 36(9), 2006, 695-726.
20. Fang H, Tong W, Shi L M, Blair R, Perkins R, Branham W, Hass B S, Xie Q, Dial S L, Moland C L and Sheehan D M. Structure activity Relationships for a Large Diverse Set of Natural, Synthetic, and Environmental Estrogens, *Chemical Research in Toxicology*, 14(3), 2001, 280-294.
21. Filimonov D and Poroikov V. Probabilistic Approaches in Activity Prediction. In: Chemoinformatics Approaches to Virtual Screening (Alexandre Varnek and Alexander Tropshaeds.), *Cambridge, RSC Publishing*, 2008, 182-216.
22. Filimonov D, Poroikov V, Borodina Yu and Gloriozova T. Chemical Similarity Assessment through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors, *Journal of Chemical Information and Computer Sciences*, 39(4), 1999, 666-670.
23. Gallegos Saliner A, Netzeva T I and Worth A P. Prediction of Estrogenicity: Validation of a Classification Model, *SAR and QSAR in Environmental Research*, 17(1), 2006, 195-223.
24. Ghafourian T and Cronin M T. The impact of variable selection on the modelling of oestrogenicity, *SAR and QSAR in Environmental Research*, 16(1-2), 2005, 171-190.
25. Gombar V K, Enslein K and Blake B W. Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models, *Chemosphere*, 31(1), 1995, 2499-2510.
26. Hewitt M, Madden J C, Rowe P H and Cronin M T D. Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier, *SAR and QSAR in Environmental Research*, 18(1), 2007, 57-76.
27. Hewitt M, Ellison C M, Enoch S J, Madden J C and Cronin M T D. Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity, *Reproductive Toxicology*, 30(1), 2010, 147-160.
28. Hong H, Fang H, Xie Q, Perkins R, Sheehan D M and Tong W. Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor, *SAR and QSAR in Environmental Research*, 14(5-6), 2003, 373-388.
29. Hong H, Tong W, Xie Q, Fang H and Perkins R. An in silico ensemble method for lead discovery: decision forest, *SAR and QSAR in Environmental Research*, 16(4), 2005, 339 - 347.
30. Hulzebos E M and Posthumus R. (Q)SARs: gatekeepers against risk on chemicals?, *SAR and QSAR in Environmental Research*, 14(4), 2003, 285-316.
31. Islam M A, Nagar S, Das S, Mukherjee A and Saha A. Molecular design based on receptor-independent pharmacophore: application to

estrogen receptor ligands, *Biological and Pharmaceutical Bulletin,* 31(7), 2008, 1453- 1460.
32. Jacobs M N, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M and Van der Burg B. The use of metabolising systems for in vitro testing of endocrine disruptors, *Current Drug Metabolism,* 9(8), 2008, 796-826.
33. Jensen G E, Niemela J R, Wedebye E B and Nikolov N G. QSAR models for reproductive toxicity and endocrine disruption in regulatory use – a preliminary investigation, *SAR and QSAR in Environmental Research,* 19(7), 2008, 631-641.
34. Julien E, Willhite C C, Richard A M, DeSesso J M and Group I R W. Challenges in constructing statistically based structure-activity relationship models for developmental toxicity, *Birth Defects Research Part A: Clinical and Molecular Teratology,* 70(12), 2004, 902-911.
35. Kavlock R J. Structure-activity relationships in the developmental toxicity of substituted phenols: *in vivo* effects, *Teratology,* 41(1), 1990, 43-59.
36. Knudsen T B, Martin M T, Kavlock R J, Judson R S, Dix D J and Singh A V. Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB, *Reproductive Toxicology,* 28(2), 2009, 209-219.
37. Kramer V J and Giesy J P. Specific binding of hydroxylated polychlorinated biphenyl metabolites and other substances to bovine calf uterine estrogen receptor: structure-binding relationships, *Science of the Total Environment,* 233(1-3), 1999, 141-161.
38. Kurunczi L, Seclaman E, Oprea T I, Crisan L and Simon Z. MTD PLS: A PLS Variant of the Minimal Topologic Difference Method. III. Mapping Interactions between Estradiol Derivatives and the Alpha Estrogenic Receptor, *Journal of Chemical Information and Modeling,* 45(5), 2005, 1275-1281.
39. Lill M A, Dobler M and Vedani A. In silico prediction of receptor-mediated environmental toxic phenomena-application to endocrine disruption, *SAR and QSAR in Environmental Research,* 16(1-2), 2005, 149- 169.
40. Lill M A, Vedani A and Dobler M. Raptor: Combining Dual-Shell Representation, Induced-Fit Simulation, and Hydrophobicity Scoring in Receptor Modeling: Application toward the Simulation of Structurally Diverse Ligand Sets, *Journal of Medicinal Chemistry,* 47(25), 2004, 6174-6186.
41. Marini F, Roncaglioni A and Novic M. Variable Selection and Interpretation in Structure Affinity Correlation Modeling of Estrogen Receptor Binders, *Journal of Chemical Information and Modeling,* 45(6), 2005, 1507-1519.
42. Martin M T, Mendez E, Corum D G, Judson R S, Kavlock R J, Rotroff D M and Dix D J. Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. Toxicological Sciences 110(1), 2009, 181-190.
43. Maslankiewicz L H E, Vermeire T G, Muller JJA, Piersma A H (2005). Can chemical structure predict reproductive toxicity? Available at: http://hdl.handle.net/10029/7374.
44. Matthews E J, Kruhlak N L, Daniel Benz R and Contrera J F. A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database, *Regulatory Toxicology and Pharmacology,* 47(2), 2007a, 115-135.
45. Matthews E J, Kruhlak N L, Daniel Benz R, Ivanov J, Klopman G and Contrera J F. A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals, *Regulatory Toxicology and Pharmacology,* 47(2), 2007b, 136-155.
46. Mekenyan O, Dimitrov S, Schmieder P and Veith G. In silico modelling of hazard endpoints: current problems and perspectives, *SAR and QSAR in Environmental Research,* 14(3), 2003a, 361-371.
47. Mekenyan O, Nikolova N and Schmieder P. Dynamic 3D QSAR techniques: applications in toxicology, *Journal of Molecular Structure: Theochem,* 622(1), 2003b, 147-165.

48. Mekenyan O G, Schultz T W, Veith G D and Kamenska V. 'Dynamic' QSAR for semicarbazide-induced mortality in frog embryos, *Journal of Applied Toxicology*, 16(4), 2003b, 355-363.
49. Mukherjee S, Saha A and Roy K. QSAR of estrogen receptor modulators: exploring selectivity requirements for ER[alpha] versus ER[beta] binding of tetrahydroisoquinoline derivatives using E-state and physicochemical parameters, *Bioorganic and Medicinal Chemistry Letters*, 15(4), 2005, 957-961.
50. Netzeva T I, Gallegos Saliner A and Worth A P. Comparison of the Applicability Domain of a QSAR for Estrogen city with a Large Chemical Inventory, *Environmental Toxicology and Chemistry*, 25(3), 2006, 1223-1230.
51. OECD (2009). (Quantitative) Structure Activity Relationships [(Q)SARs] of Estrogen binding affinity to support prioritization of heterogenous chemicals within defined inventories for screening and testing. Schmieder P., R. Kolanczyk, H. Aladjov, M. Hornung, M. Tapper, J. Denny, B. Sheedy. Organisation for Economic Cooperation and Development.Available at: http://www.olis.oecd.org/olis/2009doc.nsf/LinkTo/NT00004CD6/$FILE/JT03268747.PDF.
52. Pastor M, Cruciani G, McLay I, Pickett S and Clementi S. GRid-IN dependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors, *Journal of Medicinal Chemistry*, 43(17), 2000, 3233-3243.
53. Pearl G M, Livingston-Carr S, Durham S. Integration of computational analysis as a sentinel tool in toxicological assessments, *Current Topics in Medicinal Chemistry*, 1(2), 2001, 247-255.
54. Poroikov V, Filimonov D, Lagunin A, Gloriozova T and Zakharov A. PASS: identification of probable targets and mechanisms of toxicity, *SAR and QSAR in Environmental Research*, 18(1), 2007, 101-110.
55. Pike ACW, Brzozowski A M, Hubbard R E, Bonn T, Thorsell A-G, Engstrom O, Ljunggren J, Gustafsson J-A and Carlquist M. Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, *EMBO Journal*, 18(17), 1999, 4608-4618.
56. Richard A M and Hunter ES III. Quantitative structure-activity relationships for the developmental toxicity of haloacetic acids in mammalian whole embryo culture, *Teratology*, 53(6), 1996, 352-360.
57. Ridings J E and Baldwin J A. A qualitative assessment of developmental toxicity within a series of structurally related dopamine mimetics, *Toxicology*, 76(3), 1992, 197-207.
58. Saliner A G, Netzeva T I and Worth A P. Prediction of estrogenicity: validation of a classification model, *SAR and QSAR in Environmental Research*, 17(2), 2006, 195 - 223.
59. Salumld B, Polikarpov I and Andricopulo A D. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies, *Journal of Molecular Graphics and Modelling*, 26(2), 2007, 434-442.
60. Salumlv B, Polikarpov I and Andricopulo A D. Structure-Based Approach for the Study of Estrogen Receptor Binding Affinity and Subtype Selectivity, *Journal of Chemical Information and Modeling*, 48(11), 2008, 2243-2253.
61. Scialli A R. The challenge of reproductive and developmental toxicology under REACH, *Regulatory Toxicology and Pharmacology* 51(2), 2008, 244-250.
62. Schmieder P, Mekenyan O, Bradbury S &Veith G. QSAR prioritization of chemical inventories for endocrine disruptor testing, *Pure and Applied Chemistry*, 75(2), 2003, 2389-2396.
63. Schultz T W and Dawson D A. Structure-Activity Relationships for Teratogenicity and Developmental Toxicity, in: W. Karcher, J. Devillers (Eds.), Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicity, *European Commission, Brussels*, 1990, 389-409.
64. Serafimova R, Todorov M, Nedelcheva D, Pavlov T, Akahori Y, Nakai M and Mekenyan O. QSAR and mechanistic interpretation of estrogen receptor

binding, *SAR and QSAR in Environmental Research*, 18(3), 2007, 389-421.

65. Soderholm A A, Viiliainen J, Lehtovuori P T, Eskelinen H, Roell D, Baniahmad A and Nyronen T H. Computationally Identified Novel Diphenyl- and Phenylpyridine Androgen Receptor Antagonist Structures, *Journal of Chemical Information and Modeling*, 48(9), 2008, 1882-1890.
66. Sussman N B, Arena V C, Yu S, Mazumdar S and Thampatty B P. Decision tree SAR models for developmental toxicity based on an FDA/TERIS database, *SAR and QSAR in Environmental Research*, 14(2), 2003, 83-96.
67. Taha M O, Tarairah M, Zalloum H and Abu-Sheikha G. Pharmacophore and QSAR modeling of estrogen receptor [beta] ligands and subsequent validation and in silico search for new hits, *Journal of Molecular Graphics and Modelling*, 28(5), 2010, 383-400.
68. Tamura H, Ishimoto Y, Fujikawa T, Aoyama H, Yoshikawa H and Akamatsu M. Structural basis for androgen receptor agonists and antagonists: Interaction of SPEED 98-listed chemicals and related compounds with the androgen receptor based on an in vitro reporter gene assay and 3D-QSAR, *Bioorganic and Medicinal Chemistry*, 14(21), 2006, 7160-7174.
69. Tong W, Xie Q, Hong H, Shi L, Fang H and Perkins R. Assessment of Prediction Confidence and Domain Extrapolation of Two Structure–Activity Relationship Models for Predicting Estrogen Receptor Binding Activity, *Environmental Health Perspectives*, 112(2), 2004, 1249-1254.
70. Vedani A, Dobler M and Lill M A. Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor, *Journal of Medicinal Chemistry*, 48(11), 2005, 3700- 3703.
71. Vedani A and Smiesko M. In silico toxicology in drug discovery - concepts based on three-dimensional models, *Alternatives to Laboratory Animals*, 37(5), 2009, 477-496.
72. Vedani A, Smiesko M, Spreafico M, Peristera O and Dobler M. Virtual ToxLab - in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products, Two years and 2,000 compounds of experience: a progress report, *Altex*, 26(3), 2009, 167-176.
73. Vinggaard A M, Niemela J, Wedebye E B and Jensen G E. Screening of 397 Chemicals and Development of a Quantitative Structure Activity Relationship Model for Androgen Receptor Antagonism, *Chemical Research in Toxicology*, 21(4), 2008, 813-823.
74. Zhao C Y, Zhang R S, Zhang H X, Xue C X, Liu H X, Liu M C, Hu Z D and Fan B T. QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor, *SAR and QSAR in Environmental Research*, 16(4), 2005, 349 - 367.

Please cite this article in press as: Shirish Ambulgekar. QSAR Models and Guidance: A Review, *Asian Journal of Research in Chemistry and Pharmaceutical Sciences*, 4(1), 2016, 1-10.