
A Concise Review on the Significance of QSAR in Drug Design

Hiteshi Tandon¹, Tanmoy Chakraborty^{2,*}, Vandana Suhag³

¹Department of Chemistry, Manipal University Jaipur, Jaipur, India

²Department of Chemistry, School of Engineering, Presidency University, Bengaluru, India

³Department of Applied Sciences, BML Munjal University, Gurgaon, India

Email address:

tanmoychem@gmail.com (T. Chakraborty), tanmoychakraborty@presidencyuniversity.in (T. Chakraborty)

*Corresponding author

To cite this article:

Hiteshi Tandon, Tanmoy Chakraborty, Vandana Suhag. A Concise Review on the Significance of QSAR in Drug Design. *Chemical and Biomolecular Engineering*. Vol. 4, No. 4, 2019, pp. 45-51. doi: 10.11648/j.cbe.20190404.11

Received: December 2, 2019; **Accepted:** December 16, 2019; **Published:** December 27, 2019

Abstract: Drug designing is a crucial step in the exploration of novel drugs which requires potent methodologies. One of such methodologies is Quantitative Structure Activity Relationship (QSAR) which is a widely used statistical tool that correlates the structure of a molecule to a biological activity as a function of molecular descriptors, thereby, playing an essential role in the drug designing. QSAR utilizes Density Functional Theory (DFT) based chemical descriptors for this purpose. The selection of such significant molecular descriptors from various available descriptors is the foremost challenge in a QSAR as structural descriptors are representative of the molecular characteristics accountable for the relevant activity. Recently, new QSAR approaches have been introduced which further enhance the study of the activities. Further, the constructed QSAR models also need to be tested and validated for their efficiency and practical usage. As the QSAR models are structure specific, they may not be universally applicable. However, because of their high precision and efficacy, they have a promising future in the world of drug design. This review briefly summarizes the role of descriptor based QSAR in drug design in conjunction with existing QSAR approaches and also the utility as well as constraints of this approach in drug design.

Keywords: QSAR, Density Functional Theory (DFT), Quantum Chemical Descriptors, Drug Design

1. Introduction

Drug designing plays a critical role in medicinal chemistry when it is performed with planning and forethought. As per the basic hypothesis of the rational drug design approach, the valuable effects of drugs are generated because of the molecular recognition as well as binding of ligands to the active site of explicit targets, like nucleic acids, receptors and enzymes. Inhibition or promotion of signal transduction can be the consequence of binding for enzymatic activity or molecular transport. Conventionally, medicinal chemistry employed small chemical compounds which already existed in nature and their activity was revealed through serendipity and observation. Exploitation of animal models, designed for human diseases, for methodical exploration of compounds possessing biological activity was a new characterization in the history of drug design. Subsequently with the

introduction of computational chemistry along with advances in molecular biology, superior testing schemes and high output screening became accessible. In tandem, the variety and sophistication of the systems also amplified.

The progressive endeavour of the majority of organic, inorganic, organometallic, and pharmaceutical chemists is the application of theoretical techniques for determining the physicochemical and molecular properties of different drug compounds. Quantitative structure activity relationships (QSARs) as well as Quantitative structure property relationships (QSPRs) are indisputably of immense significance in present chemistry and biochemistry. The primary predicament in chemistry and biochemistry is the relationship of biological or chemical activity with molecular structure and its properties, that is, QSAR (quantitative structure activity relationship), or SAR (structure activity relationship), or QSPR (quantitative structure property relationship). QSAR has emerged as a method that is both

comprehensive as well as competent to be utilized for pharmaceutical problems. QSAR exploits Density functional theory (DFT) based descriptors for predicting the relationship between the molecular properties of the target compound and its various biological phenomena [1]. DFT is a well-accepted method for precisely describing biologically important molecular systems at an effective computational cost. These molecular descriptors are universal variables utilized for QSAR-based activity prediction modeling [2].

The details determined by the descriptors rely on the dimensionality of the molecule in addition to the algorithm employed in computing the descriptors [3]. One-dimensional descriptors refer to the quantities which encode chemico-physical properties and constitutional parameters. These descriptors are not sensitive to molecular topology and are liable to be identical for different compounds. Consequently, they are frequently utilized jointly with other descriptors. For chemical space description, often two-dimensional descriptors employed. These kinds of descriptors aren't only molecular conformation independent but graph invariant as well. Three-dimensional descriptors offer a rigorous description of molecular structures. These entail conformational probing and can classify isomers, although the process is computationally expensive. Four-dimensional descriptors resemble three-dimensional descriptors and assess several structural conformations at the same time. Another category of molecular descriptors comprises of fingerprints [3–5]. Generally used fingerprints are the extended-connectivity fingerprints (ECFP) [6], PubChem [7] and Molecular ACCess System (MACCS) [8] substructure fingerprints.

The QSAR was initially detected by Hammett in the 1930s and developed by Hansch and Fujita [9] later in the mid-1960s. A number of QSAR methodologies were proposed by numerous authors following 70 years advancement in the field like 2DQSAR [9–11], 3DQSAR which corresponds to CoMSIA [12, 13] and CoMFA [14, 15] in addition to 4DQSAR [16] and HQSAR (Hologram quantitative structure-activity relationship) [17]. Until now QSAR is a well recognized method and has been used more or less in every branch of chemistry, including medicinal chemistry [18–20], agricultural chemistry [9, 18], environmental chemistry [18, 21] and toxicology [22]. Particularly in pharmaceutical chemistry QSAR has turned out to be a model means for drug innovation and is an essential part of all industrial drug design software packages. In fact, QSAR is one of the most commonly used tools in drug design for more than 50 years. QSAR technique relies on the physicochemical and structural information of molecules and thus can offer further structural and physical insights.

The idea of QSAR is to alter pursuit for compounds with anticipated properties via chemical insight and practice into a computerized and mathematically quantified model. As soon as a relationship amidst structure and activity is established, a number of compounds, in addition to those which are not synthesized so far, can be effortlessly displayed on the computer to facilitate selection of structures with required

properties. Consequently, it becomes possible to opt for potential compounds to synthesize and analyze in the laboratory. Hence, the QSAR technique conserves resources and hastens the course of development of novel molecules for use as drugs, additives, materials, or various other purposes. Although it is complicated to determine efficacious structure activity correlations, the recent rampant growth in the quantity of papers concerning with QSAR analysis evidently reveals the fast advancement in this realm. To achieve an effective and valuable correlation, it is essential that suitable descriptors are utilized, regardless of them being theoretical, empirical, semi-empirical or derived from an easily accessible experimental structural characteristics.

Numerous descriptors reveal elementary molecular properties and therefore can offer an understanding of the physicochemical character of the activity under concern. Current developments in computational hardware and the progress in the construction of powerful algorithms have supported the custom progress of molecular quantum mechanical computations. In a reasonably small computational timespan, latest semi-empirical methodologies deliver practical quantum-chemical molecular quantities. Therefore, quantum chemical estimations are an enticing resource of new molecular descriptors, theoretically which suggest every single electronic and geometric property of molecules and corresponding interactions. In fact, a lot of latest QSAR studies have engaged quantum chemical descriptors solely or in unification with other common descriptors. Rather than empirical methods, quantum chemistry presents a more precise and exhaustive account of electronic effects [23]. By directly deriving electronic descriptors from the molecular wave function, quantum chemical methods can be employed to QSARs. In lots of cases it has been known that inaccuracy owing to the approximate nature of quantum-chemical procedure and ignorance of solvation effects can be transferred to a great extent within structurally associated series. As a result, despite the fact that the absolute values of computed descriptors are not directly related, the relative values can still be significant [24]. In addition, on the basis of atoms or groups, molecular wave function derived electronic descriptors can also be divided, permitting the explanation of different molecular areas individually. Largely, work utilizing quantum chemical descriptors has been done in the area of QSAR rather than QSPR, i.e. correlation of the descriptors with biological activities such as hallucinogenic activity, enzyme inhibition efficiency, etc. have been established [25–28]. This happened to a certain extent since the chase for quantitative relationships with chemical structure, in the past, begun with the foundation of scheme for theoretical drug design. Quantum-chemical descriptors also correlate different physical properties of molecules with their reactivity [29, 30]. The current review article scrutinizes the relevance of quantum chemical descriptors in the progress of QSAR/QSPR concerning chemical, biochemical, physical, and pharmacological properties of compounds.

2. Quantum Chemical Descriptors

For an entire molecule as well as for molecular components and substituents, reactivity, shape and binding properties are characterized by a large number of molecular and local quantities which are defined via quantum chemical methods. The most often used quantum-chemistry based descriptors are atomic charges, HOMO and LUMO energies, orbital electron densities, superdelocalizabilities, atom-atom polarizabilities, molecular polarizabilities, dipole moments, polarity indices, electronegativities, global electrophilicities, global nucleophilicities and lately global compressibilities [1, 31–39]. As a huge amount of precise physical data is encrypted in a number of theoretical descriptors, two key benefits appear due to their employment in the design of a training batch in a QSAR analysis. First, on grounds of their molecular structure, various chemical species along with their different parts and substituents can be directly identified. Second, the suggested mechanism of action can be directly described with respect to the chemical reactivity of the chemical species under analysis [40]. The obtained QSAR models, as a result, comprise of the information which is related to the nature of the intermolecular forces of compounds under consideration which helps in ascertaining their biological and/or other activity. However, quantum-chemically derived descriptors have some characteristics in common with experimentally measured quantities, but basically they are different from each other. There exists no statistical error in calculations done quantum chemically which is unlikely in the case of experimental measurements. Though there is innate error linked with the suppositions necessary to assist the computations. For the most part only the direction of the error is recognized and not the magnitude [41]. The computational error is believed to be roughly constant all through the series of interrelated compounds while applying quantum chemical descriptors. A primary limitation of quantum chemistry based descriptors is the deficiency to directly illustrate bulk effects [42].

Ligand-based drug design (LBDD) approach is a strategy which does not call for the biological target structure [43]. QSAR is a LBDD scheme and it has been extensively exploited in drug design, primarily intended to anticipate the biological activity of a compound set contrary to a definite target to adjust the binding affinity [44, 45]. QSAR models offer exact estimations of measured endpoints rather than an independent classification of biological activity. These quantitative strategies have also been exploited in other projects, for instance toxicity profile, optimization of pharmacokinetics [46–48] and virtual screening [49–52]. Many significant QSAR studies are present in literature, which contain explanation of potent computational techniques and algorithms [53, 54], validation procedures [55], practical uses [56, 57] in addition to different issues and concerns and how they have been dealt with [58-63].

The foremost challenge in the drug design, regardless the extensive use of QSAR approaches, is the realization of such models because it considerably relies on the accessibility of an enormous amount of data, which prevails as a challenge in

drug design. This glitch is largely associated with the complication involving the quality of accumulated public data, including vague illustration of chemical structures and imprecise activity details [64]. Moreover, the nature of miscellaneous experimental procedures can generally head towards the data belonging to various probability distributions, which makes the application of these methods impractical. Mostly the relationship between chemical variations of two data sets is tough to be studied as biological activities of those sets are calculated under dissimilar experimental settings [65]. As stated in a review on quantitative structure activity relationship, in 2014, the transfer of one QSAR models to another is one of the issues in QSAR modeling [66], as the conventional schemes have been characteristically designed for each target property separately. Apart from this, the quantum chemical descriptors cannot theoretically provide an explanation for temperature and entropic effects since a quantum chemical computation is carried out for a single structure at minimum energy. As a result, it relates to the hypothetical physical condition of the gas at infinitely low pressure and temperature (0 K), in addition to ignoring the zero-point vibrations of the molecule. For a specified property or course of action, when such factors are intervening, quantum-chemical descriptors are inadequate for their explanation and if any correlation is achieved using them then it can be considered as accidental. Nevertheless, nearly every standard quantum-chemical application packages such as AMPAC 9.2, Gaussian 16, MOPAC 2016 have an option to determine the rotational, vibrational, and translational partition functions of the molecule at the specified temperature and their relevant contributions to the thermodynamic functions such as molecular entropy, molecular enthalpy and so on.

3. Descriptor Selection and QSAR Model Development

A model drug candidate is required possess definite properties, i.e., chemical properties, solubility, enzymatic stability, permeation across biological membranes, low clearance by the liver or kidney, potency, and safety. Out of numerous available descriptors, selection of the fundamental molecular descriptors is the most important challenge in a QSAR. Hence, to understand the QSAR model, to decrease over-fitting, accelerate training and to enhance the overall model predictability, choice of appropriate and interpretable descriptors to set up QSAR models is an extremely crucial step. Nevertheless, this is a challenging and complicated step as well. The principal supposition of QSAR methodology is that the discrepancy observed in the biological activity is correlated to molecular structure [67, 68]. Thus, using mathematical expression, biological activity of corresponding molecular frameworks can be expressed as a function of definite structural molecular characteristics, that is, descriptors, through regression analysis to assess the relative significance of those characteristics responsible for the

biological effect. The traditional QSAR methodologies [69, 70] involve molecular descriptors with global molecular properties of ligands. QSAR models based on distinct and suitable descriptors are used from linear or multilinear regression analysis (MLR) to present complex multivariate analysis [71–79]. However, non-linear techniques should be preferred over others as a modeling approach due to the sophistication of drug metabolism.

4. QSAR Approaches

The QSAR model construction is characterized by having specific schemes for computing and choosing the molecular descriptors and explicit statistical procedures for formulating the resulting models. Characterization of the QSAR group is done by designing the model system in the lack of a definite structure intended for the molecular target. Reformulating the conventional QSAR methods with refined mathematical tools and well-designed theoretical models, recently, three modern QSAR methods were introduced [80, 81]. First one being FB-QSAR (Fragment-Based two dimensional QSAR), where as per the substitute being examined, the molecular structures in a series of drug candidates are segregated into a number of fragments. The physicochemical characteristics of molecular fragments are compared with the bioactivities of drug candidates with the help of two sets of coefficients out of which one is for the molecular fragments while the other for the physicochemical characteristics. The second type is known as MF-3D-QSAR (Multiple Field Three Dimensional QSAR). In this QSAR type, additional molecular potential field (thermodynamic and non thermodynamic) is incorporated in CoMFA [15] (Comparative Molecular Field Analysis) using two sets of coefficients, one for the Cartesian three dimensional space position and the other for potential field. CoMFA (Comparative Molecular Fields Analysis) [14] was formulated in 1988 which is a milestone in QSAR because it was for the first time that 3D structure of the ligands was used to describe structure-activity relationships. Interaction of chemical probes with ligands is plotted on a plane or lattice encompassing a set of compounds in CoMFA which is superimposed in 3D space. This plane or lattice corresponds to a substitute of the binding site of the real biological receptor. The third category includes AABPP (Amino Acid-Based Peptide or Protein Prediction). In this, the QSAR procedure is employed for the peptide and protein activity evaluation. Using two sets of coefficients, one for the physicochemical properties of amino acids whilst the other for the residues in the peptide chain, the bioactivities of peptides and proteins are linked with the physicochemical properties of each or some amino acid residues in the chain. All of the three QSAR approaches mentioned above are characterized by a set of three dimensional concurrent equations that enclose two sets of undetermined coefficients. In the meantime, an iterative double least square (IDLS) scheme is developed alternatively and persistently for determining the two sets of coefficients in a research dataset. In relation to traditional QSAR approaches, the recent QSAR methodologies can

amazingly augment the predictive power of QSAR and offer further information on the molecular structure.

5. Conclusions

This review presents the advancement of Quantitative Structure Activity Relationship (QSAR) in the field of drug designing using quantum chemically derived molecular descriptors. Through utilizing QSAR correlations based on quantum chemical descriptors, one can develop hypothetical structures which either don't exist or have never been synthesized. The review also draws attention towards some facets which should be stressed while utilizing QSAR approach for drug designing like the diversity and depiction of the data sets, i.e., training and test sets, variable assortment and promising application of the new statistical schemes. Although, it should be noted that these descriptors are not fully universal and, may have severe downsides as per the nature of the chemical processes or structures involved. New QSAR approaches have also been introduced recently which can remarkably enhance the predictive power of QSAR and present supplementary information related to molecular structure. Although it is apparent that quantum chemical based descriptors have immense applicability and potentiality in QSAR studies in the field of drug design given that their usage is crucially scrutinized and validated for a certain property or phenomenon, there are still a lot of elementary aspects to be further probed with QSAR methods. Therefore, it is perceptible that existing and future advances will continue to facilitate and broaden the use of descriptor based QSAR technique in the studies of novel drugs candidates as an indispensable part of drug design, and it is probable to remain as such in the near future.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors are thankful to Manipal University Jaipur, Jaipur; Presidency University, Bengaluru and BML Munjal University, Gurgaon for providing research facility.

References

- [1] Tandon, H., Chakraborty, T., Suhag, V. (2019) A New Model of Atomic Nucleophilicity Index and Its Application in the Field of QSAR. *International Journal of Quantitative Structure-Property Relationships* 4 (3), 99-117.
- [2] Yang, H., Sun, L., Li, W., Liu, G., Tang, Y. (2018) In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front. Chem.* 6, 30.
- [3] Khan, A. U. (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov. Today* 21 (8), 1291-1302.

- [4] Leach, A. R., Gillet, V. J. (2007) An introduction to cheminformatics. Springer, Dordrecht.
- [5] Todeschini, R., Consonni, V. (2000) Handbook of molecular descriptors. Wiley-VCH, Weinheim.
- [6] Rogers, D., Hahn, M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model* 50 (5), 742-754.
- [7] National Institutes of Health (2009) PubChem substructure fingerprint. ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt. Accessed 13 Dec 2019.
- [8] Duan, J., Dixon, S. L., Lowrie, J. F., Sherman, W. (2010) Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* 29 (2), 157-170.
- [9] Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., et al (1963) The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J. Am. Chem. Soc.* 85 (18), 2817-2824.
- [10] Hansch, C., Leo, A. J. (1979) *Substitute Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York.
- [11] Du, Q., Mezey, P. G., Chou, K. C. (2005) Heuristic molecular lipophilicity potential (HMLP): A 2D-QSAR study to LADH of molecular family pyrazole and derivatives. *J. Comput. Chem.* 26 (5), 461-470.
- [12] Klebe, G., Mietzner, T. (1994) A fast and efficient method to generate biologically relevant conformations. *J. Comput Aided Mol. Des.* 8 (5), 583-606.
- [13] Klebe, G., Abraham, U. (1999) Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput. Aided Mol. Des.* 13 (1), 1-10.
- [14] Marshall, G. R., Cramer III, R. D. (1988) Three-dimensional structure-activity relationships. *Trends pharmacol. Sci.* 9 (8), 285-289.
- [15] Cramer III, R. D., Patterson, D. E., Bunce, J. D. (1989) Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Boil. Res.* 291, 161.
- [16] Hopfinger, A. J., Wang, S., Tokarski, J. S., Jin, B., Albuquerque, M., et al (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* 119 (43), 10509-10524.
- [17] Tong, W., Lewis, D. R., Perkins, R., Chen, Y., Welsh, et al (1998) Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.* 38 (4), 669-677.
- [18] Roy, K. (2017) *Advances in QSAR Modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences*. Springer, Cham.
- [19] Rose, K., Hall, L. H., Kier, L. B. (2002) Modeling blood-brain barrier partitioning using the electrotopological state. *J. Chem. Inf. Comput. Sci.* 42 (3), 651-666.
- [20] Tandon, H., Chakraborty, T., Suhag, V. (2019) A Brief Review on Importance of DFT in Drug Design. *Res. Med. Eng. Sci.* 7 (4), RMES. 000668 (791-795).
- [21] Kier, L. B., Hall, L. H. (2001) Database organization and searching with E-state indices. *SAR QSAR Environ. Res.* 12 (1-2), 55-74.
- [22] Hall, L. H., Vaughn, T. A. (1997) QSAR of phenol toxicity using electrotopological state and kappa shape indices. *Med. Chem. Res.* 7 (6-7), 407-416.
- [23] Cartier, A., Rivail, J. L. (1987) Electronic descriptors in quantitative structure-activity relationships. *Chemom. Intell. Lab. Sys.* 1 (4), 335-347.
- [24] Redl, G., Berkoff, C. E. (1974) Quantitative drug design. *Chem. Soc. Rev.* 3 (3), 273-292.
- [25] Franke, R. (1984) *Theoretical Drug Design Methods (Pharmacochimistry Library) Vol. 7*, Elsevier, Amsterdam.
- [26] Gupta, S. P., Singh, P., Bindal, M. C. (1983) QSAR studies on hallucinogens. *Chem. Rev.* 83 (6), 633-649.
- [27] Gupta, S. P. (1987) QSAR studies on enzyme inhibitors. *Chem. Rev.* 87 (5), 1183-1253.
- [28] Gupta, S. P. (1991) QSAR (quantitative structure-activity relationship) studies on local anesthetics. *Chem. Rev.* 91 (6), 1109-1119.
- [29] Brown, R. E., Simas, A. M. (1982) On the applicability of CNDO indices for the prediction of chemical reactivity. *Theor. Chim. Acta.* 62 (1), 1-16.
- [30] Murugan, R., Grendze, M. P., Toomey, J. E. Jr., Katritzky, A. R., Karelson, M. et al (1994) Predicting Physical Properties from Molecular Structure. *Chemtech* 24, 17-23.
- [31] Bonin, D. K., Kresin, V. V. (1997) *Electric-Dipole Polarizabilities of Atoms, Molecules and Clusters*, World Scientific, Singapore.
- [32] Miller, K. J. (1990) Additivity methods in molecular polarizability. *J. Am. Chem. Soc.* 112 (23), 8533-8542.
- [33] Pauling, L. (1932) The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* 54 (9), 3570-3582.
- [34] Pauling, L. (1960) *The nature of the chemical bond*, 3rd Edn., Cornell University Press, Ithaca, New York.
- [35] Parr, R. G., Szentpály, L. v., Liu, S. (1999) Electrophilicity Index. *J. Am. Chem. Soc.* 121 (9), 1922-1924.
- [36] Tandon, H., Chakraborty, T., Suhag, V. (2019) A New Scale of the Electrophilicity Index Invoking the Force Concept and its Application in Computing the Internuclear Bond Distance. *J. Struct. Chem.* 60 (11), 1725-1734.
- [37] Shalini, A., Tandon, H., Chakraborty, T. (2017) Molecular Electrophilicity Index – A Promising Descriptor for Predicting Toxicological Property. *J. Bioequiv. Availab.* 9 (6), 518-527.
- [38] Tandon, H., Chakraborty, T., Suhag, V. (2019) A new scale of atomic static dipole polarizability invoking other periodic descriptors. *J. Math. Chem.* 57 (9), 2142-2153.
- [39] Tandon, H., Chakraborty, T., Suhag, V. (2019) A model of atomic compressibility and its application in QSAR domain for toxicological property prediction. *J. Mol. Model.* 25 (10), 303.

- [40] Cocchi, M., Menziani, M. C., De Benedetti, P. G., Cruciani, G. (1992) Theoretical versus empirical molecular descriptors in monosubstituted benzenes: A chemometric study. *Chemom. Intell. Lab. Sys.* 14 (1-3), 209-224.
- [41] Magee, P. S., King, J. W. (1989) Correlations and mechanisms of chemical toxicity in animals. In ACS symposium series, Oxford University Press 413, 390-399.
- [42] Turki, T., Wei, Z., Wang, J. T. (2017) Transfer learning approaches to improve drug sensitivity prediction in multiple Myeloma patients. *IEEE Access* 5, 7381-7393.
- [43] Du, Q. S., Huang, R. B., Chou, K. C. (2008) Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Protein Pept. Sci.* 9 (3), 248-259.
- [44] Gertrudes, J. C., Maltarollo, V. G., Silva, R. A., Oliveira, P. R., Honorio, K. M., et al (2012) Machine learning techniques and drug design. *Curr. Med. Chem.* 19 (25), 4289-4297.
- [45] Maltarollo, V. G., Gertrudes, J. C., Oliveira, P. R., Honorio, K. M. (2015) Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opin. Drug Metab. Toxicol.* 11 (2), 259-271.
- [46] Egeghy, P. P., Sheldon, L. S., Isaacs, K. K., Özkaynak, H., Goldsmith et al (2016) Computational exposure science: an emerging discipline to support 21st-century risk assessment. *Environ. Health Perspect.* 124 (6), 697-702.
- [47] Chemi, G., Gemma, S., Campiani, G., Brogi, S., Butini, S. et al (2017) Computational tool for fast in silico evaluation of hERG K⁺ channel affinity. *Front. Chem.* 5, 7.
- [48] Brogi, S., Papazafiri, P., Roussis, V., Tafí, A. (2013) 3D-QSAR using pharmacophore-based alignment and virtual screening for discovery of novel MCF-7 cell line inhibitors. *Eur. J. Med. Chem.* 67, 344-351.
- [49] Melo-Filho, C. C., Dantas, R. F., Braga, R. C., Neves, B. J., Senger et al (2016) QSAR-driven discovery of novel chemical scaffolds active against *Schistosoma mansoni*. *J. Chem. Inf. Model.* 56 (7), 1357-1372.
- [50] Neves, B. J., Dantas, R. F., Senger, M. R., Melo-Filho, C. C., Valente, W. C. et al (2016) Discovery of new anti-schistosomal hits by integration of QSAR-Based virtual screening and high content screening. *J. Med. Chem.* 59 (15), 7075-7088.
- [51] Zaccagnini, L., Brogi, S., Brindisi, M., Gemma, S., Chemi et al (2017) Identification of novel fluorescent probes preventing PrPSc replication in prion diseases. *Eur. J. Med. Chem.* 127, 859-873.
- [52] Sliwoski, G., Kothiwale, S. K., Meiler, J., Lowe, E. W., Barker, E. L. (2014) Computational methods in drug discovery. *Pharmacol. Rev.* 66 (1), 334-395.
- [53] Raies, A. B., Bajic, V. B. (2016) In silico toxicology: computational methods for the prediction of chemical toxicity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 6 (2), 147-172.
- [54] Gramatica, P., Sangion, A. (2016) A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. *J. Chem. Inf. Model.* 56 (6), 1127-1131.
- [55] Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin et al (2014) QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57 (12), 4977-5010.
- [56] Fang, C., Xiao, Z. (2016) Receptor-based 3D-QSAR in drug design: methods and applications in kinase studies. *Curr. Top. Med. Chem.* 16 (13), 1463-1477.
- [57] Cronin, M. T. D., Schultz, T. W. (2003) Pitfalls in QSAR. *J. Mol. Struct. Theochem* 622 (1-2), 39-51.
- [58] Arthur, M. D. (2008) QSAR: dead or alive? *J. Comput. Aided Mol. Des.* 22, 81-89.
- [59] Dearden, J. C., Cronin, M. T. D., Kaiser, K. L. E. (2009) How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* 20 (3-4), 241-266.
- [60] Scior, T., Medina-Franco, J. L., Do, Q. T., Martínez-Mayorga, K., Yunes Rojas, J. A. et al (2009) How to recognize and workaround pitfalls in QSAR studies: a critical review. *Curr. Med. Chem.* 16 (32), 4297-4313.
- [61] Wang, T., Wu, M. B., Lin, J. P., Yang, L. R. (2015) Quantitative structureactivity relationship: promising advances in drug discovery platforms. *Expert Opin. Drug Discov.* 10 (12), 1283-1300.
- [62] Ponzoni, I., Sebastián-Pérez, V., Requena-Triguero, C., Roca, C., Martínez, M. J. et al (2017) Hybridizing feature selection and feature learning approaches in QSAR modeling for drug discovery. *Sci. Rep.* 7 (1), 2403.
- [63] Zhao, L., Wang, W., Sedykh, A., Zhu, H. (2017) Experimental errors in QSAR modeling sets: What we can do and what we cannot do. *ACS Omega* 2 (6), 2805-2812.
- [64] Richter, L., Ecker, G. F. (2015) Medicinal chemistry in the era of big data. *Drug Discov. Today* 14, 37-41.
- [65] Gadaleta, D., Mangiatordi, G. F., Catto, M., Carotti, A., Nicolotti, O. (2016) Applicability domain for QSAR models: where theory meets reality. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, 1 (1), 45-63.
- [66] Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., Hopfinger, A. J. (2010) 4D-QSAR: perspectives in drug design. *Molecules* 15 (5), 3281-3294.
- [67] Cohen, N. C. (Ed.) (1996) *Guidebook on molecular modeling in drug design*. Academic Press: London.
- [68] Hansch, C.; Fujita, T. (1964) Rho-Sigma-Pi Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86 (8), 1616-1626.
- [69] Kubinyi, H. (1993) QSAR: Hansch Analysis and Related Approaches In *Methods and Principles in Medicinal Chemistry*. Vol. 1, Mannhold, R.; Kroegsgard-Larsen, P.; Timmerman, H. (Eds.) Wiley-VCH, Weinheim, pp 240.
- [70] Dunn Iii, W. J., Scott, D. R., Glen, W. G. (1989) Principal components analysis and partial least squares regression. *Tetrahedron Comput. Methodol.* 2 (6), 349-376.
- [71] de Julian-Ortiz, J. V., Garcia-Domenech, R., Galvez, J., Pogliani, L. (2005) Predictability and prediction of lowest observed adverse effect levels in a structurally heterogeneous set of chemicals. *SAR QSAR Environ. Res.* 16 (3), 263-272.

- [72] Prado-Prado, F. J., Gonzalez-Diaz, H., de la Vega, O. M., Ubeira, F. M., Chou, K. C. (2008) Unified QSAR approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg. Med. Chem.* 16 (11), 5871-5880.
- [73] Hemmateenejad, B. (2005) Correlation ranking procedure for factor selection in PC-ANN modeling and application to ADMETox evaluation. *Chemom. Intell. Lab. Syst.* 75 (2), 231-245.
- [74] Rogers, D., Hopfinger, A. J. (1994) Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* 34 (4), 854-866.
- [75] Goulon, A., Picot, T., Duprat, A., Dreyfus, G. (2007) Predicting activities without computing descriptors: graph machines for QSAR. *SAR QSAR Environ. Res.* 18 (1), 141-153.
- [76] Yuan, H., Wang, Y., Cheng, Y. (2007) Local and global quantitative structure-activity relationship modeling and prediction for the baseline toxicity. *J. Chem. Inf. Model* 47 (1), 159-169.
- [77] Fernandez, M., Caballero, J. (2006) Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor. *J. Mol. Graphics Modell.* 25 (4), 410-422.
- [78] Ren, S., Kim, H. (2003) Comparative assessment of multiresponse regression methods for predicting the mechanisms of toxic action of phenols. *J. Chem. Inf. Comput. Sci.* 43 (6), 2106-2110.
- [79] Du, Q. S., Huang, R. B., Wei, Y. T., Wang, C. H., Chou, K. C. (2007) Peptide reagent design based on physical and chemical properties of amino acid residues. *J. Comput. Chem.* 28 (12), 2043-2050.
- [80] Du, Q. S., Huang, R. B., Wei, Y. T., Du, L. Q., Chou, K. C. (2007) Multiple field three dimensional quantitative structure-activity relationship (MF-3D-QSAR). *J. Comput. Chem.* 29 (2), 211-219.
- [81] Du, Q. S., Wei, Y. T., Pang, Z. W., Chou, K. C. and Huang, R. B. (2007) Predicting the affinity of epitope-peptides with class I MHC molecule HLA-A*0201: an application of amino acid-based peptide prediction. *Prot. Eng. Des. Select.* 20 (9), 417-423.