EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 9
Issue: 2
296-298
Year: 2022

QSAR: A USEFUL TOOL FOR DESIGNING NEW DRUG AND PREDICTING THEIR BIOLOGICAL ACTIVITIES

***Dr. Subhashini Sharma**

Associate Professor, Dept. of Chemistry, Mmh College Ghaziabad, U.P. India.

***Corresponding Author: Dr. Subhashini Sharma**
Associate Professor, Dept. of Chemistry, Mmh College Ghaziabad, U.P. India.

Article Received on 22/12/2021 **Article Revised on 12/01/2022** **Article Accepted on 02/02/2022**

ABSTRACT
It has been seen that discovery of a new drug is such a challenging process that out of hundreds compounds synthesized in research laboratories only one or sometimes none reaches the market ultimate as a' drug'. It has been estimated that traditional process of synthesis of compounds is based on trial and error method and random screening of the compound to test its activity is very expensive and time consuming. The recent advanced technologies in the field of pure and applied science, analytical chemistry and computer science have proved to be useful tools for designing new chemicals and predicting their biological activities before synthesis. QSAR is a quantitative structure activity relationship by which a chemist is enabled to describe and elucidate the effect of various physiochemical properties on drug potency and to determine biological activities of new compounds.

KEYWORDS: QSAR, new drug, biological activity, physiochemical properties, elucidation.

INTRODUCTION
Quantitative structure-activity relationship (QSAR) is a computational modeling method for revealing relationships between structural properties of chemical compounds and biological activities. QSAR modeling is essential for drug discovery, but it has many constraints.

Virtual screening (VS) has emerged in drug discovery as a powerful computational approach to screen large libraries of small molecules for new hits with desired properties that can then be tested experimentally. Similar to other computational approaches, VS intention is not to replace *in vitro* or *in vivo* assays, but to speed up the discovery process, to reduce the number of candidates to be tested experimentally, and to rationalize their choice. Moreover, VS has become very popular in pharmaceutical companies and academic organizations due to its time-, cost-, resources-, and labor-saving. Among the VS approaches, quantitative structure–activity relationship (QSAR) analysis is the most powerful method due to its high and fast throughput and good hit rate.

For the development and designing of a rational drug, various approaches are attempted which can be listed as:
1. Hanch approach
2. Additive model
3. Principal components analysis
4. Molecular orbital method
5. Discriminant Analysis
6. Molecular modeling etc.

As the first preliminary step of a QSAR model development, relevant chemogenomics data are collected from databases and the literature. Then, chemical descriptors are calculated on different levels of representation of molecular structure, ranging from 1D to *n*D, and then correlated with the biological property using machine learning techniques. Once developed and validated, QSAR models are applied to predict the biological property of novel compounds.

Application of QSAR
Quantitative structure–activity relationship (QSAR) analysis is a ligand-based drug design method developed more than 50 years ago by Hansch and Fujita (1964). ... Initially, QSAR modeling was limited to small series of congeneric compounds and simple regression methods. Quantitative structure–activity relationship (QSAR) analysis is a ligand-based drug design method developed more than 50 years ago by Hansch and Fujita (1964). Since then and until now, QSAR remains an efficient method for building mathematical models, which attempts to find a statistically significant correlation between the chemical structure and continuous (pIC50, pEC50, Ki, etc.) or categorical/binary (active, inactive, toxic, nontoxic, etc.) biological/toxicological property using regression and classification techniques, respectively (Cherkasov et al., 2014). In the last decades, QSAR has undergone several transformations, ranging from the dimensionality of the molecular descriptors (from 1D to *n*D) and different methods for finding a correlation between the chemical structures and the

biological property. Initially, QSAR modeling was limited to small series of congeneric compounds and simple regression methods. Nowadays, QSAR modeling has grown, diversified, and evolved to the modeling and virtual screening (VS) of very large data sets comprising thousands of diverse chemical structures and using a wide variety of machine learning techniques (Cherkasov et al., 2014; Mitchell, 2014; Ekins et al., 2015; Goh et al., 2017).

QSAR-Based Virtual Screening vs. High-Throughput Screening

High-throughput screening can rapidly identify large subsets of molecules with desired activity from large screening collections of compounds (105–106 compounds) using automated plate-based experimental assays (Mueller et al., 2012). However, the hit rate of HTS ranges between 0.01% and 0.1% and this highlights the frequently encountered limitation that most of the screened compounds are routinely reported as inactive toward the desired bioactivity (Thorne et al., 2010). Consequently, the drug discovery cost increases according to the number of tested compounds (Butkiewicz et al., 2013). On the other hand, typical hit rates from a validated VS method, including QSAR-based, typically range between 1% and 40%. Thus, VS campaigns are found to have a higher rate of biologically active compounds and at a lower cost than HTS.

In this perspective, we show that QSAR-based VS could be used to enrich hit rates of HTS campaigns. For example, Mueller et al. (2010) employed both HTS and QSAR models to search novel positive allosteric modulators for mGlu5, a G-protein coupled receptor involved in disorders like schizophrenia and Parkinson's disease. QSAR models (combining physicochemical descriptors and neural networks), which were subsequently applied to screen a database of approximately 450,000 compounds. Finally, 824 compounds were acquired for biological testing and 232 were confirmed as active (hit rate of 28.2%) (Mueller et al., 2010). In another study, Rodriguez et al. (2010) screened approximately 160,000 compounds to identify 624 antagonists of mGlu5. Further, these data were used to develop QSAR models and, then, applied to screen near 700,000 compounds from ChemDiv database. Among them, 88 of acquired compounds were active, corresponding to a hit rate of 3.6% while the HTS had a hit rate of 0.2% (Mueller et al., 2012).

Tuberculosis

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), kills about 1.6 million people every year (WHO, 2018e). The current treatment of this disease takes approximately 9 months, which normally leads to noncompliance and, hence, the emergence of multidrug-resistant bacteria (AlMatar et al., 2017). Aiming the design of new anti-TB agents, our group used QSAR models to design new series of chalcone (1,3-diaryl-2-propen-1-ones) derivatives. Initially, we retrieved from the literature all chalcone compounds with *in vitro* inhibition data against *M. tuberculosis* H37Rv strain. After rigorous data curation, these chalcones were subject to structure–activity relationships (SAR) analysis. Based on SAR rules, bioisosteric replacements were employed to design new chalcone derivatives with optimized anti-TB activity.

In parallel, binary QSAR models were generated using several machine learning methods and molecular fingerprints. The fivefold external cross-validation procedure confirmed the high predictive power of the developed models. Using these models, we prioritized series of chalcone derivatives for synthesis and biological evaluation (Gomes et al., 2017). As a result, five 5-nitro-substituted heteroaryl chalcones were found to exhibit MICs at nanomolar concentrations against replicating mycobacteria, as well as low micromolar activity against nonreplicating bacteria. In addition, four of these compounds were more potent than standard drug isoniazid. The series also showed low cytotoxicity against commensal bacteria and mammalian cells. These results suggest that designed heteroaryl chalcones, identified with the help of QSAR models, are promising anti-TB lead candidates (Gomes et al., 2017).

Viral Infections

Yearly, influenza epidemics can seriously affect all populations in the world. These annual epidemics are estimated to result in about 5 million cases and 650,000 deaths (WHO, 2018b). Influenza virus is mutating constantly, resulting in novel resistant strains, and hence, the development of new anti-influenza drugs active against these new strains is important to prevent pandemics (Laborda et al., 2016). Aiming the discovery of new anti-influenza A drugs, Lian et al. (2015) built binary QSAR models, using SVM and Naïve Bayesian methods, to predict neuraminidase inhibition, a validated protein target for influenza. Then, four different combinations of machine learning methods and molecular descriptors were applied to screen 15,600 compounds from an in-house database, among which 60 compounds were selected to experimental evaluation on neuraminidase activity. Nine inhibitors were identified, five of which were oseltamivir derivatives exhibiting potent neuraminidase inhibition at nanomolar concentrations. Other four active compounds belonged to novel scaffolds, with potent inhibition at low micromolar concentrations (Lian et al., 2015).

According to WHO, approximately 35 million people are infected with HIV (WHO, 2018a). The treatment for HIV infections requires a lifelong antiretroviral therapy, targeting different stages of HIV replication cycle. Consequently, because of the emergence of resistance and the lack of tolerability, development of novel anti-HIV drugs is of high demand (Cihlar and Fordyce, 2016; Garbelli et al., 2017). With the purpose of discovering new anti-HIV-1 drugs, Kurczyk et al. (2015) developed a two-step VS approach to prioritize

compounds against HIV integrase, an important target to viral replication cycle. The first step was based on binary QSAR models, and the second on privileged fragments. Then, 1.5 million of commercially available compounds were screened, and 13 compounds were selected to be tested *in vitro* for inhibiting HIV-1 replication. Among them, two novel chemotypes with moderate anti-HIV-1 potencies were identified, and therefore, represent new starting points for prospective structural optimization studies.

CONCLUSION

To summarize, we would like to emphasize that QSAR modeling represents a time-, labor-, and cost-effective tool to discover hit compounds and lead candidates in the early stages of drug discovery process. Analyzing the examples of QSAR-based VS available in the literature, one can see that many of them led to the identification of promising lead candidates. However, along with success stories, many QSAR projects fail on the model building stage. This is caused by the lack of understanding that QSAR is highly interdisciplinary and application field as well as general ignorance of the best practices in the field (Tropsha, 2010; Ban et al., 2017). Earlier, we have explained this by the undesirably high population of "button pushers," that is, researchers who conduct modeling without understanding and analyzing the data and modeling process itself (Muratov et al., 2012). This was also explained by the elusive ease of obtaining computational model and making even advanced calculations without understanding of the sense and limitations of the approach (Bajorath, 2012). In addition to this, a lot of even experienced researchers target their efforts to a "vicious statistical cycle," which main goal is to validate models using as many metrics as possible. In this case, the QSAR modeling is restricted to a single simple question: "What is the best metrics or the best statistical method"? Although we recognize that the right choice of statistical approach and especially rigorous external validation are necessary and represent an essential step in any computer-aided drug discovery study, we want to reinforce that QSAR modeling is useful only if it is applied for the solution of a formulated problem and results in development of new compounds with desired properties.

As future directions, we would like to point out that the era of big data has just started, and it is still in the chemical/biological data accumulation stage. Therefore, to avoid the situation that the number of assayed compounds available on literature exceeds the modeling capability, the development, and implementation of new machine learning algorithms and data curation methods capable of handling millions of compounds are urgently needed. Finally, the overall success of any QSAR-based VS project depends on the ability of a scientist to think critically and prioritize the most promising hits according to his experience. Moreover, the success rate of collaborative drug discovery projects, where the final selection of computational hits is done by both a modeler and an expert in a given field, is much higher than success rate of the projects driven solely by computational or experimental scientists.

REFERENCES

1. CohenJS, Hogan ME, The new genetic medicines. Scientific American (Dec) 1994; 50(entire issue)
2. StromBL, Tugwell P.Pharmacoepidemilogy: current status, prospects and problems. Ann Intern Med, 1990; 113: 179-181.
3. Long JW.Clinical management of prescription drugs. Hagarstown: Harper and Row, 1983.
4. AlMatar, M., AlMandeal, H., Var, I., Kayar, B., and Köksal, F. (2017). New drugs for the treatment of *Mycobacterium tuberculosis* infection. *Biomed. Pharmacother*, 91: 546–558. doi: 10.1016/j.biopha.2017.04.105 PubMed Abstract | CrossRef Full Text | Google Scholar
5. Bajorath, J. (2012). Computational chemistry in pharmaceutical research: at the crossroads. *J. Comput. Aided. Mol. Des.*, 26: 11–12. doi: 10.1007/s10822-011-9488-z PubMed Abstract | CrossRef Full Text | Google Scholar
6. Ban, F., Dalal, K., Li, H., LeBlanc, E., Rennie, P. S., and Cherkasov, A. (2017). Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *J. Chem. Inf. Model.*, 57: 1018–1028. doi: 10.1021/acs.jcim.7b00137 PubMed Abstract | CrossRef Full Text | Google Scholar