Acta Scientific Pharmacology

Volume 3 Issue 6 June 2022

Mini Review

QSAR-based Virtual Screening: Advances and Applications in Drug Discovery

Bruno J Neves[1]*, Rodolpho C Braga[2], Cleber C Melo-Filho[3], José Teofilo Moreira-Filho[4], Eugene N Muratov[5], Carolina Horta Andrade[5] and PGR Achary[5]

[1]Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Federal University of Goias, Goiania, Brazil

[2]Laboratory of Cheminformatics, Centro University of Anapolis, Brazil

[3]Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, University of North Carolina at Chapel Hill, United States

[4]Department of Chemical Technology, Odessa National Polytechnic University, Ukraine

[5]Department of Chemistry, Siksha O Anusandhan, India

***Corresponding Author:** Bruno J Neves, Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Federal University of Goias, Goiania, Brazil.

Received: December 27, 2021

Published: May 27, 2022

Abstract

In the field of drug development, virtual screening (VS) has developed as a potent computer tool for screening huge libraries of small compounds for novel hits with desirable features that can subsequently be verified experimentally. Like that of other computational techniques, isn't to replace in vitro or in vivo experiments, but to speed up the discovery process minimize the number of candidates that must be examined experimentally, and justify their selection. Furthermore, because of its time, cost, resource, and labour savings, Virtual Screening has become quite popular in pharmaceutical businesses. Because of its high throughput and high hit rate, quantitative structure–activity relationship (QSAR) analysis is the most potent of the Virtual Screening techniques. Relevant chemogenomics data is acquired from databases and the literature as the initial stage in developing a QSAR model. In this mini review We outline and critically examine current advances in QSAR-based Virtual Screening in drug discovery and show how it may be used to find potential molecules with desired features [1]. Every day, scientists and researchers all around the world create massive amounts of data; for example, more than 74 million molecules have been recorded in Chemical Abstract Services. According to a recent research, there are currently roughly 1060 compounds classed as novel drug-like molecules. The collection of such molecules is now referred to as "dark chemical space" or "dark chemistry." To examine such hidden molecules scientifically, a large number of live and updated databases (protein, cell, tissues, structure, medicines, and so on) are now accessible. The convergence of three distinct sciences: 'genomics,' 'proteomics,' and 'in-silico simulation,' will transform the drug development process. The screening of a large number of drug-like compounds is a difficult task that must be handled efficiently. Virtual screening (VS) is an essential computational technique in the drug discovery process; nevertheless, experimental drug verification is also vital in the drug development process. One machine learning approach that is often employed in VS techniques is quantitative structure-activity relationship (QSAR) analysis. The current mini-review focuses on the following topics: web-based machine learning tools, the role of QSAR in VS approaches, successful implementations of QSAR-based VS leading to drug discovery, and the benefits and limitations of QSAR [2].

Keywords: Computer-assisted Drug Design; Machine Learning; Cheminformatics; Molecular Descriptors; Virtual Screening

Abbreviations

VS: Virtual Screening; QSAR: Quantitative Structural Activity Relationship; SAR: Structural Activity Relationship

Introduction

Hansch and Fujita created the quantitative structure–activity relationship (QSAR) analysis approach for ligand-based drug discovery more than 50 years ago (1964). QSAR has remained an effective method for developing mathematical models since then, attempting to find a statistically significant correlation between the chemical structure and continuous (pIC50, pEC50, Ki, etc.) or categorical(active, inactive, toxic, nontoxic, etc.) biological/toxicological properties using regression and classification techniques. Respectively QSAR has seen various changes in recent decades, including changes in the dimensionality of molecular descriptors and alternative methodologies for determining a link between chemical structures and biological properties. QSAR modelling was originally restricted to a small number of congeneric chemicals and basic regression algorithms. Nowadays, QSAR modelling has expanded, varied, and developed to include the modelling and virtual screening of very large data sets containing hundreds of different chemical structures, as well as the use of a wide range of machine learning approaches.

This review focuses on a critical examination of the benefits and drawbacks of QSAR-based Virtual Screening in drug discovery [1].

Materials and Methods

Best practices in QSAR modeling and validation

High-throughput screening (HTS) technology, the amount of data appropriate for QSAR modelling has exploded. As a result, the problem of data quality has become one of the most important issues in cheminformatics. Fourches et al. (2010; 2015; 2016) presented guidelines for chemical and biological data curation as a first and required step in predictive QSAR modelling in light of these constraints. These principles, which have been organized into a solid functional procedure, enable for the detection, correction, and, if necessary, eradication of structural and biological mistakes in big data sets. Data curation procedures include the removal of organometallics, counterions, mixtures, and inorganics, as well as the normalization of specific chemo types, structural cleaning (e.g., detection of valence violations), standardization of tautomeric forms, and ring aromatization. Additional curation elements include averaging, aggregating, or removal of duplicates to produce a single bioactivity result. The Organization for Economic Cooperation and Development (OECD) created a set of standards for researchers to follow in order for QSAR models to be accepted by regulators. QSAR models should be connected with a clear end point, an unambiguous method, a defined sphere of applicability, acceptable goodness-of-fit, robustness, and predictivity metrics, and if practicable, mechanistic interpretation, according to these criteria (OECD, 2004) [1].

Continuing importance of QSAR as virtual screening tool

The current pipeline to discover hit compounds in early stages of drug discovery is a data-driven process, which relies on bioactivity data obtained from HTS campaigns. Since the cost of obtaining new hit compounds in HTS platforms is rather high, QSAR modeling has been playing a pivotal role in prioritizing compounds for synthesis and biological evaluation. The QSAR models can be used for both hits identification and hit-to-lead optimization. In the latter, a favorable balance between potency, selectivity, and pharmacokinetic and toxicological parameters, which is required to develop a new, safe, and effective drug, could be achieved through several optimization cycles. QSAR is a labor, time and money-saving strategy for obtaining molecules with desired biological characteristics. As a result, QSAR is widely used in companies, universities, and research institutions all over the world [1].

Practical applications of QSAR-based virtual screening

Despite its obvious benefits, QSAR modelling as a Virtual Screening tool is unappreciated. Unfortunately, QSAR is still viewed as an add-on to synthesis and biological assessment research, and it is frequently used in studies without any reason or new viewpoint. Despite the fact that there are just a few Virtual Screening applications in the literature, the majority of them resulted in the finding of potential hits and lead candidates. We'll go over some of the successful QSAR-based VS applications for fresh hit discovery and hit-to-lead optimization in the sections below.

Malaria

Malaria is a parasitic disease caused by five distinct species of Plasmodium parasites that is spread to humans by the bite of infected female Anopheles mosquitoes. P. falciparum is the most dangerous species, causing serious disease and death. Malaria is

a widespread illness with current transmission in 91 countries and territories. Malaria caused over 216 million infections and 445,000 deaths in 2016, according to the World Health Organization (WHO) Furthermore, antimalarial medication resistance is a prevalent and developing problem that poses a serious concern to those living in endemic areas. A data collection of 3,133 chemicals reported as active or inert against P. falciparum chloroquine sensitive strain (3D7) was utilize. Dragon descriptors (0D, 1D, and 2D), ISIDA-2D fragment descriptors, and the support vector machines (SVM) approach were used to create the models. The data set was randomly partitioned into modelling and external assessment sets during QSAR modelling and validation. The Sphere Exclusion technique was also used to separate the modelling set into training and test sets many times. The QSAR models were then applied to Virtual Screening of the ChemBridge database using a consensus technique. Following Virtual Screening, 176 candidate antimalarial compounds were discovered and tested in the lab, along with 42 putatively inactive compounds that served as negative controls. In P. falciparum growth inhibition experiments, twenty-five drugs showed antimalarial efficacy and minimal cytotoxicity in mammalian cells. Experiments revealed that all 42 substances predicted as inert by the models were inactive [1]. The verified experimental hits against P. falciparum revealed new chemical scaffolds that might be viable starting points for the creation of new antimalarial medicines.

Schistosomiasis

Schistosomiasis is a parasitic infection caused by flatworms of the Schistosoma genus that affects 206 million people globally WHO, 2018. The present reliance on praziquantel as the only anti-schistosomal medicine necessitates the rapid development of new anti-schistosomal therapies [1]. To investigate new structurally different molecules with antischistosomal action, our group built binary QSAR models for Schistosoma mansoni thioredoxin glutathione reductase (SmTGR), a validated target for schistosomiasis [1]. To do this, we devised a research that included the following steps: (I) curation of the greatest data collection of SmTGR inhibitors conceivable, (ii) construction of rigorously verified and mechanistically interpretable models, and (iii) use of produced models for VS of ChemBridge library. We prioritized 29 substances for further experimental investigation using the QSAR models. As a consequence, we discovered that QSAR models were effective in identifying six new hit compounds active against schistosomula and three new hits active against adult worms (hit rate of 20.6 percent). 2-[2-(3-methyl-4-nitro-5-isoxazolyl)vinyl]pyridine and 2-(benzylsulfonyl)-1,3-benzothiazole, two novel chemical scaffolds, both exhibit efficacy against schistosomula and adult worms at low micromolar concentrations, and hence offer prospective antischistosomal hits for further hit-to-lead optimization [1]. We constructed continuous QSAR models for a data set of oxadiazoles inhibitors of smTGR in another work [1]. We created a consensus model by merging the predictions of individual 2D- and 3D-QSAR models using a combi-QSAR technique. The model was then applied to the ChemBridge database's VS, and the top 10 compounds were tested in vitro against schistosomula and adult worms. In addition, five highly predictive in-house QSAR models were used to forecast key pharmacokinetics and toxicity aspects of the new hits. At low micromolar concentrations, 4-nitro-3,5-bis(1-nitro-1H-pyrazol-4-yl)-1H-pyrazole (LabMol-17) and 3-nitro-4- [(4-nitro-1,2,5-oxadiazol-3-yl)oxy]methyl-1,2,5-oxadiazole (LabMol-19), two compounds containing new chemical scaffolds (hit rate of 20.6 percent), were highly active in both parasite life stages [1].

Tuberculosis

Every year, Mycobacterium tuberculosis, the bacteria that causes tuberculosis (TB), kills roughly 1.6 million people WHO, 2018.The present therapy for this condition takes around 9 months, which frequently results in noncompliance and, as a result, the formation of multidrug-resistant bacteria [1]. With the goal of developing new anti-TB drugs, To create a new set of chalcone (1,3-diaryl-2-propen-1-ones) derivatives, our group employed QSAR models. To begin, we gathered all chalcone compounds with in vitro inhibitory evidence against the M. tuberculosis H37Rv strain from the literature. These chalcones were subjected to structure–activity relationships (SAR) research after thorough data curation. Bioisosteric substitutions were used to develop novel chalcone compounds with optimal anti-TB activity using SAR criteria. Binary QSAR models were created in parallel utilizing a variety of machine learning algorithms and chemical fingerprints. The resulting models' strong predictive potential was proven by a fivefold external cross-validation method. We selected a series of chalcone derivatives for synthesis and biological assessment using these models [1]. As a consequence, five 5-nitro-substituted heteroaryl chalcones were discovered to have MICs against replicating mycobacteria at nanomolar concentrations and low micromolar action against nonreplicating microorganisms. Furthermore, four of these

compounds outperformed the conventional medication isoniazid. The series was also shown to have negligible cytotoxicity when tested on commensal bacteria and mammalian cells. These findings show that QSAR-identified designed heteroaryl chalcones might be promising anti-TB lead candidates [1].

Viral infections

Every year, influenza outbreaks can have a devastating impact on the entire world's population. Each year, these epidemics are predicted to cause 5 million illnesses and 650,000 fatalities WHO, 2018 Because the influenza virus is continually changing, resulting in new resistant strains, the discovery of new anti-influenza medications that are effective against these new strains is critical to preventing pandemics [1]. used binary QSAR models to predict neuraminidase inhibition, a confirmed protein target for influenza, with the goal of discovering novel anti-influenza A medicines. They used SVM and Nave Bayesian approaches. The researchers then used four distinct combinations of machine learning approaches and chemical descriptors to screen 15,600 compounds from an internal database, of which 60 were chosen for experimental neuraminidase activity testing. Five of the inhibitors were oseltamivir derivatives that inhibited neuraminidase at nanomolar doses. The other four active compounds were new scaffolds that exhibited significant inhibition even at low micromolar doses [1]. According to the World Health Organization, roughly 35 million individuals are HIV-positive WHO, 2018. Antiretroviral medication is required to treat HIV infections for the rest of one's life, and it targets distinct stages of the HIV replication cycle. As a result of the rise of resistance and lack of tolerability, there is a strong need for innovative anti-HIV medications [1]. To select compounds against HIV integrase, a critical target in the viral replication cycle, researchers devised a two-step VS method. The first stage used binary QSAR models, whereas the second relied on privileged pieces. After that, 1.5 million commercially available compounds were evaluated, and 13 were chosen to be investigated in vitro for HIV-1 replication inhibition. Two novel chemotypes with modest anti-HIV-1 potencies were discovered among them, and so constitute fresh starting points for structural optimization research in the future.

Mood and anxiety disorders

The 5-hydroxytryptamine 1A (5-HT1A) serotonin receptor has long been a promising target for the treatment of mood and anxiety disorders like schizophrenia. However, the medications that target the 5-HT1A receptor that are currently on the market have serious negative effects. A QSAR-based VS methodology to uncover novel 5-HT1A receptor-targeting hit molecules [1]. First, utilising Dragon descriptors and a variety of machine learning algorithms, binary QSAR models were created. The created QSAR models were then carefully tested and applied to VS four commercial chemical datasets in consensus. A total of fifteen compounds were chosen for testing, with nine of them proving to be active at low nanomolar doses. [(8-)-6-methyl-9,10-didehydroergolin-8-yl] methanol, one of the verified hits, has a very high binding affinity (Ki) of 2.3 nM for the 5-HT1A receptor.

Conclusion

To conclude, QSAR modelling is a time-, labor-, and cost-effective technique for identifying hit compounds and lead candidates in the early phases of the drug development process. When looking at the instances of QSAR-based VS in the literature, it's clear that many of them resulted in the discovery of intriguing lead candidates. However, many QSAR initiatives fail at the model construction stage, despite the success tales. This is due to a lack of awareness that QSAR is a highly multidisciplinary and application-oriented discipline, as well as a general lack of knowledge of industry best practices [1]. We have characterized this as a result of an unacceptably large number of "button pushers," or researchers that do modelling without first comprehending and assessing the data and modelling process [1]. This was further explained by the enigmatic simplicity of generating a computational model and doing even sophisticated computations without knowing the approach's logic and constraints [1]. Furthermore, many even seasoned researchers focus their efforts on a "vicious statistical cycle", with the primary objective of validating models using as many indicators as feasible. The QSAR modelling is limited to a single basic query in this case: "What are the most effective metrics or statistical methods?". Although we recognize that the right statistical approach and, in particular, rigorous external validation are critical in any computer-aided drug discovery study, we want to emphasize that QSAR modelling is only useful if it is used to solve a formulated problem and results in the development of new compounds with desired properties.

Conflict of Interest

The authors state that there were no commercial or financial ties that may be considered as a possible conflict of interest during the research.

Bibliography

1. Neves Bruno J., *et al*. "QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery". *Frontiers in Pharmacology* 9 (2018): 1-2.

2. Acharya Patnala R. "Applications of Quantitative Structure-Activity Relationships (QSAR) Based Virtual Screening in Drug Design: A Review". *Mini-Reviews in Medicinal Chemistry* 20.14 (2020): 1375-1388.