WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

QSAR – Based Virtual Screening: Advances and Applications in the Drug Discovery

Shalgaonkar Akshada Pradip; Jare Akash Ambadas

Late Narayandas Bhavandas Chhabada Institute of pharmacy, Raigaon, Satara,

Author for Correspondence: Department of Pharmaceutical Chemistry Pharmacy, Late Narayandas Bhavandas Chhabada Institute of Pharmacy. Raigaon, Satara, Shivaji University, Kolhapur, Maharashtra, India. Gmail: akashjare4711@gmail.com, akshupradipshalgaonkar@gmail.com

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 new hits with desirable features that can then be verified experimentally. The goal of VS, like other computational approaches, is not to replace in vitro or in vivo experiments, but to speed up the discovery process, minimize the number of candidates that need to be tested experimentally, and rationalize their selection. Furthermore, because to its time, cost, resource, and labour savings, VS has become quite popular in pharmaceutical businesses and academic institutions. Because of its high throughput and high hit rate, quantitative structure–activity relationship (QSAR) analysis is the most potent of the VS techniques. As the first stage in developing a QSAR model,

INTRODUCTION

Hanch and Fujita established the quantitative structure–activity relationship (QSAR) analysis method 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/binary (active, inactive, toxic, nontoxic, etc.) biological/toxicological properties using regression and classification techniques (cherakasov et al 2014). QSAR has seen various changes in recent decades, including changes in the dimensionality of molecular descriptors (from 1D to nD) and alternative methodologies for determining a correlation between chemical structures and biological properties. QSAR modelling was originally restricted to a small number of congeneric chemicals and simple regression algorithms. Nowadays,QSAR modelling has increased, varied, and progressed to include the modelling and virtual screening (VS) of very large data sets with thousands of different chemical structures and a range of machine learning algorithms (cherkasov et al. 2014 mitchell). ekins et al., 2014 (Goh et al., 2017)

This review focuses on I a critical study of the benefits and drawbacks of QSAR-based VS in drug discovery; (ii) examples of successful QSAR-based discoveries of compounds with desired features; (iii) best practises for QSAR-based VS; and (iv) future perspectives of this technique.

Validation and Modeling of QSARs: Best Practices

As a result of 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. As obvious as it may appear, numerous inaccuracies in chemical structure and experimental results are regarded as a serious impediment to the development of predictive models.

Taking into account these constraints As a first and required stage in predictive QSAR modelling, rules for chemical and biological data curation were devised. These principles, which have been organised into a solid functional process, enable for the detection, correction, and, if necessary, eradication of structural and biological defects in big data sets. The elimination of organometallics, counterions, mixes, and inorganics, as well as the normalisation of data, are all part of the data curation process.Normalization of certain chemotypes, structural cleaning (e.g., detection of valence violations), tautomeric form standardisation, and ring aromatization are among of the techniques used. To obtain a single bioactivity result, other curation aspects include averaging, aggregating, or removing duplicates. There is a more in-depth description of the aforementioned data curation processes elsewhere.

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 I a clear end point, (ii) an unambiguous algorithm, (iii) a defined domain of applicability, (iv) acceptable goodness-of-fit, robustness, and predictivity metrics, and (v) if practicable, mechanistic interpretation, according to these criteria. . The extra rule requiring extensive data curation as an essential pre-model creation phase, in our opinion, should be placed there.

QSAR'S IMPORTANCE AS A VIRTUAL SCREENING TOOL WILL CONTINUE

The current pipeline for identifying hit compounds in the early phases of drug development is a data-driven approach that relies on bioactivity data from HTS experiments. Given the high expense of getting novel hit compounds in HTS platforms, QSAR modelling has become increasingly important in choosing compounds for synthesis and/or biological investigation. The QSAR models can be used to identify hits as well as optimise hit-to-lead

conversion. Several optimization cycles could be used to obtain a desirable balance between potency, selectivity, and pharmacokinetic and toxicological characteristics, which is essential to develop a new, safe, and effective medicine. Because no substance needs to be manufactured or tested before being evaluated computationally,In companies, universities, and research institutions all over the world, QSAR is frequently used.

Figure 1, depicts the general scheme of the QSAR-based VS method. The data sets gathered from external sources are first filtered and integrated to remove or repair any inconsistencies. QSAR models are constructed and validated using these data, following OECD rules and modelling best practises. Then, from a large chemical library, QSAR models are utilised to find chemical compounds that are anticipated to be active against specific endpoints. VS is frequently compared to a funnel, in which QSAR models compress a huge chemical library (i.e., 105 to 107 chemical structures) to a smaller number of compounds, which are then tested experimentally. experimentally (i.e., chemical structures 101 to 103). Modern VS workflows, however, include additional filtering phases such as I sets of empirical rules [e.g., Lipinski's rules], (ii) chemical similarity cutoffs, (iii) other QSAR-based filters (e.g., toxicological and pharmacokinetic endpoints), and (iv) chemical feasibility and/or purchasability. Although experimental validation of computational hits is not part of the QSAR technique, it is a vital step that should be completed. A multi-parameter optimization (MPO) with QSAR predictions of potency, selectivity, and pharmacokinetic parameters can be performed after experimental validation. This information will be critical throughout compound series hit-to-lead and lead optimization design to find the characteristics balance (potency, selectivity, and PK) relates to the effect of different decorating patterns on the in vivo evaluation of a new series of target chemicals.

QSAR-BASED VIRTUAL SCREENING VS. HIGH-THROUGHPUT SCREENING

Using automated plate-based experimental tests, high-throughput screening can quickly find significant subsets of molecules with desirable activity from huge screening collections of compounds (105–106 compounds). HTS, on the other hand, has a hit rate of 0.01 percent to 0.1 percent, highlighting the common constraint that most screened compounds are routinely reported as inactive toward the required bioactivity. As a result, the cost of drug discovery rises in proportion to the number of compounds investigated. On the other hand, hit rates from a proven VS approach, including QSAR-based methods, typically vary from 1% to 40%. As a result, VS campaigns are reported to produce more physiologically active molecules at a cheaper cost than HTS programmes.In this perspective, we show that QSAR-based VS could be used to enrich hit rates of HTS campaigns. For example, 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. First, the HTS of approximately 144,000 compounds resulted in a total of 1,356 hits, with a hit rate of 0.94%. Then, this dataset was used to build continuous 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 obtained for biological testing, with 232 of them proving to be active (a hit rate of 28.2%). In another study, researchers screened over 160,000 chemicals to find 624 mGlu5 antagonists. These data were also utilised to create QSAR models, which were then used to screen almost 700,000 molecules from the ChemDiv database. The HTS had a success rate of 0.2 percent, while 88 of the obtained compounds were active, resulting in a hit rate of 3.6 percent.

PRACTICAL APPLICATIONS OF QSAR-BASED VIRTUAL SCREENING

Despite its obvious benefits, QSAR modelling as a VS tool is underappreciated. 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 perspective. Despite the fact that there are just a few VS applications in the literature, the majority of them resulted in the finding of promising 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 different 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 continuing 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. According to a study published by. QSAR models were developed using a data collection of 3,133 chemicals that were reported as active or inactive against P. falciparum chloroquine susceptible strain (3D7). 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 VS of the ChemBridge database using a consensus approach. Following VS, 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 low cytotoxicity in mammalian cells. Experiments confirmed that all 42 substances predicted as inert by the models were inactive.

SCHISTOSOMIASIS

Schistosomiasis is a parasitic infection caused by flatworms of the Schistosoma genus that affects 206 million people globally. The current reliance on praziquantel as the only anti-schistosomal medicine necessitates the rapid development of new anti-schistosomal therapies. Our lab created binary QSAR models for Schistosoma mansoni thioredoxin glutathione reductase (SmTGR), a validated target for schistosomiasis, with the goal of discovering novel medicines. to discover novel compounds with anti-schistosomal action that are structurally different. To accomplish this, we devised a study that included the following steps: I curation of the broadest feasible data set of SmTGR inhibitors, (ii) construction of rigorously verified and mechanistically interpretable models, and (iii) application of created models to the ChemBridge library's VS. We prioritised 29 chemicals using QSAR models. for the purpose of additional testing As a result, 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 new chemical scaffolds, both have efficacy against schistosomula and adult worms at low micromolar concentrations, and hence represent promising antischistosomal hits for further hit-to-lead optimization. We constructed continuous QSAR models for a

data set of oxadiazoles inhibitors of smTGR in another work. 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 life stages of the parasit

TUBERCULOSIS

Every year, Mycobacterium tuberculosis, the bacteria that causes tuberculosis (TB), kills roughly 1.6 million people. The current treatment for this disease takes around 9 months, which leads to noncompliance and, as a result, the development of multidrug-resistant bacteria. Our group employed QSAR models to build novel series of chalcone (1,3-diaryl-2-propen-1-ones) derivatives with the goal of developing new anti-TB medicines. 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 utilising a variety of machine learning algorithms and molecular fingerprints. The derived models' high predictive potential was proven by a fivefold external cross-validation method. We prioritised a series of chalcone derivatives for synthesis and biological evaluation using these models. As a result, 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 found to have little cytotoxicity when tested on commensal bacteria and mammalian cells. These findings point to designed heteroaryl chalcones, which were discovered using QSAR models, as viable anti-TB lead candidates.

VIRAL INFECTIONS

Every year, influenza epidemics 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. Because the influenza virus is continually changing, resulting in novel resistant strains, it is critical to discover new anti-influenza medications that are effective against these new strains in order to avoid pandemics. To forecast neuraminidase inhibition, a proven protein target for influenza, researchers created binary QSAR models using SVM and Nave Bayesian approaches, with the goal of discovering new anti-influenza A medicines. The researchers then used four distinct combinations of machine learning approaches and molecular descriptors to screen 15,600 molecules from an internal database, of which 60 were chosen for experimental neuraminidase activity testing. There were nine inhibitors found, five of which were oseltamivir derivatives that inhibited neuraminidase at nanomolar doses. The other four active compounds were novel scaffolds that inhibited effectively at low micromolar doses.

According to the World Health Organization, roughly 35 million people are HIV-positive. 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 formation of resistance and lack of tolerability, there is a strong demand for the development of new anti-HIV medications. Developed a two-step VS strategy to prioritise compounds against HIV integrase, a critical target in the viral replication cycle, with the goal of developing novel anti-HIV-1 medicines. The first stage used binary QSAR models, while 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 moderate anti-HIV-1 potencies were discovered among them, and so constitute potential 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 therapeutic target for 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. To address this, researchers created a QSAR-based VS approach to uncover new 5-HT1A receptor-targeting hit molecules. First, utilizing 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 databases 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, had a very high binding affinity (Ki) of 2.3 nM for the 5-HT1A receptor.

FUTURE DIRECTIONS AND CONCLUSION

To summarise, QSAR modelling is a time-, labor-, and cost-effective technique for identifying hit compounds and lead candidates in the early stages of the drug development process. When looking at the examples 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 projects fail at the model construction stage, despite the success tales. This is due to a lack of awareness that QSAR is a highly interdisciplinary and application-oriented topic, as well as a general lack of knowledge of industry best practises. We previously characterised this as a result of an unacceptably high number of "button pushers," or researchers that do modelling without first evaluating and assessing the data and the modelling procedure itself This was further explained by the enigmatic simplicity of obtaining a computational model and performing even complex computations without knowing the approach's logic and constraints. Furthermore, many even seasoned researchers focus their efforts on a "vicious statistical cycle," with the primary goal of validating models using as many indicators as feasible. The QSAR modelling in this example is limited to a single straightforward question: "What are the appropriate metrics or statistical methods?" Although we understand that the correct statistical methodology and, in particular, rigorous external validation are crucial and important steps in any computer-aided drug discovery project, we wish to emphasise that QSAR modelling is only applicable in a limited number of situations. if it is used to solve a specific problem and results in the creation of new compounds with the necessary qualities. In terms of future directions, we'd want to emphasise that the big data era has only just begun, and we're still in the chemical/biological data accumulation stage. To avoid a situation where the quantity of tested compounds

available in the literature exceeds the modelling capability, new machine learning algorithms and data curation procedures capable of processing millions of compounds must be developed and implemented as soon as possible. Finally, the overall success of any QSAR-based VS project is determined by a scientist's capacity to think critically and pick the most promising hits based on his previous experiences. Furthermore, the success rate of collaborative drug development efforts, in which the final selection of computational hits is done by both a modeller and an expert in a specific field, is relatively high. The success rate of projects driven primarily by computational or experimental scientists is substantially lower.

REFERENCES

1. 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

2. publish Abstract | CrossRef Full Text | Google Scholar

3. Bajorath, J. (2012). Computational chemistry in pharmaceutical research: at the crossroads. J. Comput. Aided.

4. Mol. Des. 26, 11–12. doi: 10.1007/s10822-011-9488-z

5. 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

7. PubMed Abstract | CrossRef Full Text | Google Scholar

8. Butkiewicz, M., Lowe, E. W., Mueller, R., Mendenhall, J. L., Teixeira, P. L., Weaver, C. D., et al. (2013). Benchmarking ligand-based virtual high-throughput screening with the pubchem database. Molecules 18, 735–756. doi: 10.3390/molecules18010735

9. PubMed Abstract | CrossRef Full Text | Google Scholar

10. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., et al. (2014). QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57, 4977–5010. doi: 10.1021/jm4004285

11. PubMed Abstract | CrossRef Full Text | Google Scholar

12. Cihlar, T., and Fordyce, M. (2016). Current status and prospects of HIV treatment. Curr. Opin. Virol. 18, 50–56. doi: 10.1016/j.coviro.2016.03.004

13. PubMed Abstract | CrossRef Full Text | Google Scholar

14. Colley, D. G., Bustinduy, A. L., Secor, W. E., and King, C. H. (2014). Human schistosomiasis. Lancet 383, 2253–2264. doi: 10.1016/S0140-6736(13)61949-2