



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

Bruno J. Neves^{1,2}, Rodolpho C. Braga¹, Cleber C. Melo-Filho¹, José Teófilo Moreira-Filho¹, Eugene N. Muratov^{3,4} and Carolina Horta Andrade^{1*}

¹ LabMol – Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, Brazil, ² Laboratory of Cheminformatics, Centro Universitário de Anápolis (UniEVANGÉLICA), Anápolis, Brazil, ³ Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ⁴ Department of Chemical Technology, Odessa National Polytechnic University, Odessa, Ukraine

OPEN ACCESS

Edited by:

Adriano D. Andricopulo,
Universidade de São Paulo, Brazil

Reviewed by:

Marcus Scotti,
Federal University of Paraíba, Brazil

Nellima Correia Romeiro,
Universidade Federal do Rio
de Janeiro, Brazil

Ana Carolina Rennó Sodero,
Universidade Federal do Rio
de Janeiro, Brazil

*Correspondence:

Carolina Horta Andrade
carolina@ufg.br;
carolhandrade@gmail.com

Specialty section:

This article was submitted to
Experimental Pharmacology
and Drug Discovery,
a section of the journal
Frontiers in Pharmacology

Received: 11 August 2018

Accepted: 18 October 2018

Published: 13 November 2018

Citation:

Neves BJ, Braga RC,
Melo-Filho CC, Moreira-Filho JT,
Muratov EN and Andrade CH (2018)
QSAR-Based Virtual Screening:
Advances and Applications in Drug
Discovery. *Front. Pharmacol.* 9:1275.
doi: 10.3389/fphar.2018.01275

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. 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. Although the experimental testing of computational hits is not an inherent part of QSAR methodology, it is highly desired and should be performed as an ultimate validation of developed models. In this mini-review, we summarize and critically analyze the recent trends of QSAR-based VS in drug discovery and demonstrate successful applications in identifying perspective compounds with desired properties. Moreover, we provide some recommendations about the best practices for QSAR-based VS along with the future perspectives of this approach.

Keywords: cheminformatics, machine learning, molecular descriptors, computer-assisted drug design, virtual screening

INTRODUCTION

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 (pIC_{50} , pEC_{50} , K_i , 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).

This review is devoted to (i) critical analysis of advantages and disadvantages of QSAR-based VS in drug discovery; (ii) demonstration of several successful QSAR-based discoveries of compounds with desired properties; (iii) description of best practices for the QSAR-based VS; and (iv) discussion of future perspectives of this approach.

BEST PRACTICES IN QSAR MODELING AND VALIDATION

High-throughput screening (HTS) technologies resulted in the explosion of amount of data suitable for QSAR modeling. As a result, data quality problem became one of the fundamental questions in cheminformatics. As obvious as it seems, various errors in both chemical structure and experimental results are considered as major obstacle to building predictive models (Young et al., 2008; Southan et al., 2009; Williams and Ekins, 2011).

Considering these limitations, Fourches et al. (2010; 2015; 2016) developed the guidelines for chemical and biological data curation as a first and mandatory step of the predictive QSAR modeling. Organized into a solid functional process, these guidelines allow the identification, correction, or, if needed, removal of structural and biological errors in large data sets. Data curation procedures include the removal of organometallics, counterions, mixtures, and inorganics, as well as the normalization of specific chemotypes, 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. Detailed discussion of aforementioned data curation procedures can be found elsewhere (Fourches et al., 2010, 2015, 2016).

The Organization for Economic Cooperation and Development (OECD) developed a set of guidelines that the researchers should follow to achieve the regulatory acceptance of QSAR models. According to these principles, QSAR models should be associated with (i) defined end point, (ii) unambiguous algorithm, (iii) defined domain of applicability, (iv) appropriate measures of goodness-of-fit, robustness, and predictivity, and (v) if possible, mechanistic interpretation (OECD, 2004). In our

opinion, the additional rule requesting thorough data curation as a mandatory preliminary step to model development should be added there.

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 (Nantasenamat and Prachayasittikul, 2015). 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/or 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. As no compound need to be synthesized or tested before computational evaluation, QSAR represents a labor-, time-, and cost-effective method to obtain compounds with desired biological properties. Consequently, QSAR is widely practiced in industries, universities, and research centers around the world (Cherkasov et al., 2014).

The general scheme of QSAR-based VS approach is shown in **Figure 1**. Initially, the data sets collected from external sources are curated and integrated to remove or correct inconsistent data. Using these data, QSAR models are developed and validated following OECD guidelines and best practices of modeling. Then, QSAR models are used to identify chemical compounds predicted to be active against selected endpoints from large chemical libraries (Cherkasov et al., 2014). In principle, VS is often compared to a funnel, where a large chemical library (i.e., 10^5 to 10^7 chemical structures) is reduced by QSAR models to a smaller number of compounds, which then will be tested experimentally (i.e., 10^1 to 10^3 chemical structures) (Kar and Roy, 2013; Tanrikulu et al., 2013). However, it is important to mention that modern VS workflows incorporate additional filtering steps, including: (i) sets of empirical rules [e.g., Lipinski's (Lipinski et al., 1997) rules], (ii) chemical similarity cutoffs, (iii) other QSAR-based filters (e.g., toxicological and pharmacokinetic endpoints), and (iv) chemical feasibility and/or purchasability (Cherkasov et al., 2014). Although the experimental validation of computational hits does not represent part of the QSAR methodology, this should be performed as the final important step. After experimental validation, a multi-parameter optimization (MPO) with QSAR predictions of potency, selectivity, and pharmacokinetic parameters can be conducted. This information will be crucial during hit-to lead and lead optimization design of the compound series, to find the properties balance (potency, selectivity, and PK) related with the effect of different decoration patterns to establish a new series of target compounds for *in vivo* evaluation.

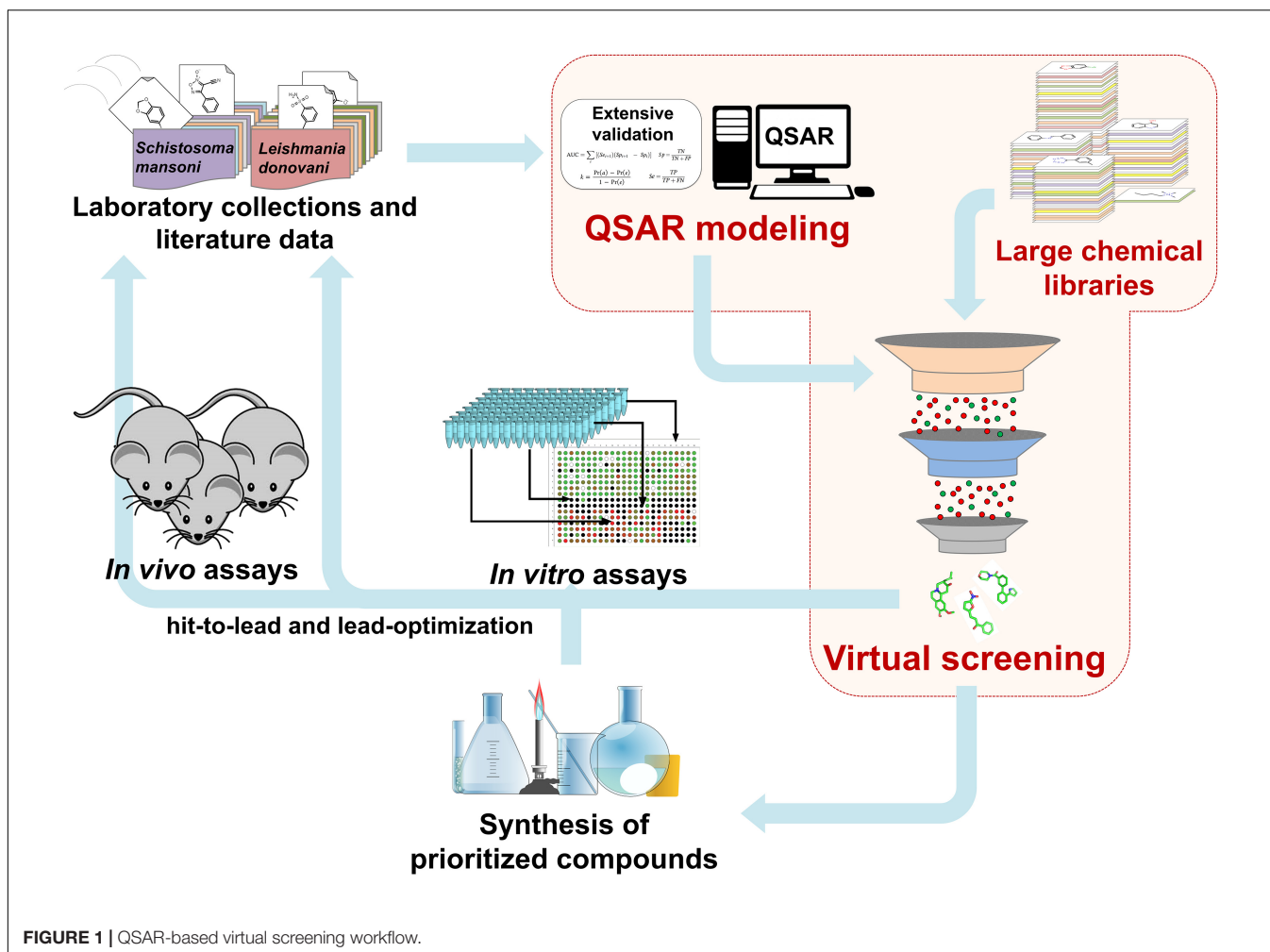


FIGURE 1 | QSAR-based virtual screening workflow.

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 (10^5 – 10^6 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. 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 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).

PRACTICAL APPLICATIONS OF QSAR-BASED VIRTUAL SCREENING

Despite its obvious advantages, QSAR modeling remains underestimated as a VS tool. Unfortunately, QSAR is still seen

as a complementary analysis to studies of synthesis and biological evaluation, often introduced in the study without any justification or additional perspective. Despite the small number of VS applications available in the literature, most of them led to the discovery of promising hits and lead candidates. Below, we discuss some successful applications of QSAR-based VS for the discovery of new hits and hit-to-lead optimization.

Malaria

Malaria is an infectious disease caused by five different species of *Plasmodium* parasites and transmitted to humans through the bite of infected female mosquitoes of the genus *Anopheles*. The most lethal species is *P. falciparum*, which can lead to severe illness and death (Phillips et al., 2017). Malaria is a widespread disease; 91 countries and areas have ongoing transmission. According to World Health Organization (WHO), about 216 million cases and 445,000 deaths from malaria were reported in 2016 (WHO, 2018c). Furthermore, the resistance to antimalarial drugs is a common and growing issue and constitutes a substantial threat for populations in endemic regions (Gorobets et al., 2017; Menard and Dondorp, 2017). In a study reported by Zhang et al. (2013), a data set of 3,133 compounds reported as active or inactive against *P. falciparum* chloroquine susceptible strain (3D7) was used to develop QSAR models. The models were built using Dragon descriptors (0D, 1D, and 2D), ISIDA-2D fragments descriptors and support vector machines (SVM) method. During QSAR modeling and validation, the data set was randomly divided into modeling and external evaluation set. Additionally, the modeling set was divided multiple times in training and test sets using the Sphere Exclusion algorithm. Then, by using a consensus approach, the QSAR models were applied for VS of the ChemBridge database. After VS, 176 potential antimalarial compounds were identified and submitted to experimental validation along with 42 putative inactive compounds, used as negative controls. Twenty-five compounds presented antimalarial activity in *P. falciparum* growth inhibition assays and low cytotoxicity in mammalian cells. All 42 compounds predicted as inactives by the models were confirmed experimentally (Zhang et al., 2013). The confirmed experimental hits presented new chemical scaffolds against *P. falciparum* and could be promising starting points for the development of new optimized antimalarial agents.

Schistosomiasis

Schistosomiasis is a disease caused by flatworms of the genus *Schistosoma* that affects 206 million of people worldwide (WHO, 2018d). The current reliance on only one drug, praziquantel, for treatment and control of this disease calls for the urgent discovery of novel anti-schistosomal drugs (Colley et al., 2014). Aiming at discovering new drugs, our group developed binary QSAR models for *Schistosoma mansoni* thioredoxin glutathione reductase (*SmTGR*), a validated target for schistosomiasis (Kuntz et al., 2007), to find new structurally dissimilar compounds with antischistosomal activity (Neves et al., 2016). To achieve this goal, we designed a study with the following steps: (i) curation of the largest possible data set of *SmTGR* inhibitors, (ii) development of rigorously validated and mechanistically interpretable models,

and (iii) application of generated models for VS of ChemBridge library. Using the QSAR models, we prioritized 29 compounds for further experimental evaluation. As a result, we found that the QSAR models were efficient for discovery of six novel hit compounds active against schistosomula and three hits active against adult worms (hit rate of 20.6%). Among them, 2-[2-(3-methyl-4-nitro-5-isoxazolyl)vinyl]pyridine and 2-(benzylsulfonyl)-1,3-benzothiazole, two compounds representing new chemical scaffolds have activity against schistosomula and adult worms at low micromolar concentrations and therefore represent promising antischistosomal hits for further hit-to-lead optimization (Neves et al., 2016).

In another study, we developed continuous QSAR models for a data set of oxadiazoles inhibitors of *smTGR* (Melo-Filho et al., 2016). Using a combi-QSAR approach, we built a consensus model combining the predictions of individual 2D- and 3D-QSAR models. Then, the model was used for VS of ChemBridge database and the 10 top ranked compounds were further evaluated *in vitro* against schistosomula and adult worms. Additionally, we applied five highly predictive in-house QSAR models for prediction of important pharmacokinetics and toxicity properties of the new hits. The experimental results showed that 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%), were highly active in both life stages of the parasite at low micromolar concentrations (Melo-Filho et al., 2016).

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.

Mood and Anxiety Disorders

The 5-hydroxytryptamine 1A (5-HT_{1A}) serotonin receptor has been an attractive target for treating mood and anxiety disorders such as schizophrenia (Nichols and Nichols, 2008; Lacivita et al., 2012). However, the currently marketed drugs targeting 5-HT_{1A} receptor possess severe side effects. To address this, Luo et al. (2014) developed a QSAR-based VS workflow to find new hit compounds targeting 5-HT_{1A} receptor. First, binary QSAR models were generated using Dragon descriptors and several machine learning methods. Then, developed QSAR models were rigorously validated and applied in consensus for VS four commercial chemical databases. Fifteen compounds were selected for experimental testing, and nine of them have proven to be active at low nanomolar concentrations. One

of the confirmed hits, [(8 α)-6-methyl-9,10-didehydroergolin-8-yl]methanol, showed very high binding affinity (K_i) of 2.3 nM against 5-HT_{1A} receptor.

Future Directions and 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.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was partially funded by the Grant No. 1U01CA207160 from NIH and Grant No. 400760/2014-2 from CNPq. CHA is Research Fellow in productivity of CNPq.

REFERENCES

- AlMatar, M., AlMandea, 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
- 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
- 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
- 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
- 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
- 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
- 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
- Ekins, S., Lage de Siqueira-Neto, J., McCall, L.-I., Sarker, M., Yadav, M., Ponder, E. L., et al. (2015). Machine learning models and pathway genome data base for *Trypanosoma cruzi* drug discovery. *PLoS Negl. Trop. Dis.* 9:e0003878. doi: 10.1371/journal.pntd.0003878
- Fourches, D., Muratov, E., and Tropsha, A. (2010). Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* 50, 1189–1204. doi: 10.1021/ci100176x
- Fourches, D., Muratov, E., and Tropsha, A. (2015). Curation of chemogenomics data. *Nat. Chem. Biol.* 11, 535–535. doi: 10.1038/nchembio.1881
- Fourches, D., Muratov, E., and Tropsha, A. (2016). Trust, but verify II: a practical guide to chemogenomics data curation. *J. Chem. Inf. Model.* 56, 1243–1252. doi: 10.1021/acs.jcim.6b00129
- Garbelli, A., Riva, V., Crespan, E., and Maga, G. (2017). How to win the HIV-1 drug resistance hurdle race: running faster or jumping higher? *Biochem. J.* 474, 1559–1577. doi: 10.1042/BCJ20160772
- Goh, G. B., Hodas, N. O., and Vishnu, A. (2017). Deep learning for computational chemistry. *J. Comput. Chem.* 38, 1291–1307. doi: 10.1002/jcc.24764
- Gomes, M. N. M. N., Braga, R. C. R. C., Grzelak, E. M. E. M., Neves, B. J. B. J., Muratov, E., Ma, R., et al. (2017). QSAR-driven design, synthesis and discovery of potent chalcone derivatives with antitubercular activity. *Eur. J. Med. Chem.* 137, 126–138. doi: 10.1016/j.ejmech.2017.05.026
- Gorobets, N. Y., Sedash, Y. V., Singh, B. K., Poonam, A., and Rathi, B. (2017). An overview of currently available antimalarials. *Curr. Top. Med. Chem.* 17, 2143–2157. doi: 10.2174/1568026617666170130123520
- Hansch, C., and Fujita, T. (1964). ρ - σ - π analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86, 1616–1626. doi: 10.1021/ja01062a035
- Kar, S., and Roy, K. (2013). How far can virtual screening take us in drug discovery? *Expert Opin. Drug Discov.* 8, 245–261. doi: 10.1517/17460441.2013.761204
- Kuntz, A. N., Davioud-Charvet, E., Sayed, A. A., Califf, L. L., Dessolin, J., Arnér, E. S. J., et al. (2007). Thioredoxin glutathione reductase from *Schistosoma mansoni*: an essential parasite enzyme and a key drug target. *PLoS Med.* 4:e206. doi: 10.1371/journal.pmed.0040206
- Kurczyk, A., Warszycki, D., Musiol, R., Kafel, R., Bojarski, A. J., and Polanski, J. (2015). Ligand-based virtual screening in a search for novel anti-HIV-1 chemotypes. *J. Chem. Inf. Model.* 55, 2168–2177. doi: 10.1021/acs.jcim.5b00295
- Laborada, P., Wang, S. Y., and Voglmeir, J. (2016). Influenza neuraminidase inhibitors: synthetic approaches, derivatives and biological activity. *Molecules* 21, 1–40. doi: 10.3390/molecules21111513
- Lacivita, E., Di Pilato, P., De Giorgio, P., Colabufo, N. A., Berardi, F., Perrone, R., et al. (2012). The therapeutic potential of 5-HT1A receptors: a patent review. *Expert Opin. Ther. Pat.* 22, 887–902. doi: 10.1517/13543776.2012.703654
- Lian, W., Fang, J., Li, C., Pang, X., Liu, A.-L., and Du, G.-H. (2015). Discovery of influenza A virus neuraminidase inhibitors using support vector machine and Naïve Bayesian models. *Mol. Divers.* 20, 439–451. doi: 10.1007/s11030-015-9641-z
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25. doi: 10.1016/S0169-409X(96)00423-1
- Luo, M., Wang, X. S., Roth, B. L., Golbraikh, A., and Tropsha, A. (2014). Application of quantitative structure-activity relationship models of 5-HT1A receptor binding to virtual screening identifies novel and potent 5-HT1A ligands. *J. Chem. Inf. Model.* 54, 634–647. doi: 10.1021/ci400460q
- Melo-Filho, C. C., Dantas, R. F., Braga, R. C., Neves, B. J., Senger, M. R., Valente, W. C. G., et al. (2016). QSAR-driven discovery of novel chemical scaffolds active against *Schistosoma mansoni*. *J. Chem. Inf. Model.* 56, 1357–1372. doi: 10.1021/acs.jcim.6b00055
- Menard, D., and Dondorp, A. (2017). Antimalarial drug resistance: a threat to malaria elimination. *Cold Spring Harb. Perspect. Med.* 7:a025619. doi: 10.1101/cshperspect.a025619
- Mitchell, J. B. O. (2014). Machine learning methods in cheminformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 468–481. doi: 10.1002/wcms.1183
- Mueller, R., Dawson, E. S., Meiler, J., Rodriguez, A. L., Chauder, B. A., Bates, B. S., et al. (2012). Discovery of 2-(2-Benzoxazolyl amino)-4-Aryl-5-cyanopyrimidine as negative allosteric modulators (NAMs) of metabotropic glutamate receptor5 (mGlu 5): from an artificial neural network virtual screen to an in vivo tool compound. *ChemMedChem* 7, 406–414. doi: 10.1002/cmdc.201100510
- Mueller, R., Rodriguez, A. L., Dawson, E. S., Butkiewicz, M., Nguyen, T. T., Oleszkiewicz, S., et al. (2010). Identification of metabotropic glutamate receptor subtype 5 potentiators using virtual high-throughput screening. *ACS Chem. Neurosci.* 1, 288–305. doi: 10.1021/cn9000389
- Muratov, E. N., Varlamova, E. V., Artemenko, A. G., Polishchuk, P. G., and Kuz'min, V. E. (2012). Existing and developing approaches for QSAR analysis of mixtures. *Mol. Inform.* 31, 202–221. doi: 10.1002/minf.201100129
- Nantasenamat, C., and Prachayasittikul, V. (2015). Maximizing computational tools for successful drug discovery. *Expert Opin. Drug Discov.* 10, 321–329. doi: 10.1517/17460441.2015.1016497
- Neves, B. J., Dantas, R. F., Senger, M. R., Melo-Filho, C. C., Valente, W. C. G., de Almeida, A. C. M., et al. (2016). Discovery of new anti-schistosomal hits by integration of QSAR-based virtual screening and high content screening. *J. Med. Chem.* 59, 7075–7088. doi: 10.1021/acs.jmedchem.5b02038
- Nichols, D. E., and Nichols, C. D. (2008). Serotonin receptors. *Chem. Rev.* 108, 1614–1641. doi: 10.1021/cr078224o
- OECD (2004). *OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models*. Available at: <https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf> [accessed September 20, 2018]
- Phillips, M. A., Burrows, J. N., Manyando, C., van Huijsduijnen, R. H., Van Voorhis, W. C., and Wells, T. N. C. (2017). Malaria. *Nat. Rev. Dis. Prim.* 3:17050. doi: 10.1038/nrdp.2017.50
- Rodriguez, A. L., Grier, M. D., Jones, C. K., Herman, E. J., Kane, A. S., Smith, R. L., et al. (2010). Discovery of novel allosteric modulators of metabotropic glutamate receptor subtype 5 reveals chemical and functional diversity and in vivo activity in rat behavioral models of anxiolytic and antipsychotic activity. *Mol. Pharmacol.* 78, 1105–1123. doi: 10.1124/mol.110.067207

- Southan, C., Várkonyi, P., and Muresan, S. (2009). Quantitative assessment of the expanding complementarity between public and commercial databases of bioactive compounds. *J. Cheminform.* 1:10. doi: 10.1186/1758-2946-1-10
- Tanrikulu, Y., Krüger, B., and Proschak, E. (2013). The holistic integration of virtual screening in drug discovery. *Drug Discov. Today* 18, 358–364. doi: 10.1016/j.drudis.2013.01.007
- Thorne, N., Auld, D. S., and Inglese, J. (2010). Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr. Opin. Chem. Biol.* 14, 315–324. doi: 10.1016/j.cbpa.2010.03.020
- Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* 29, 476–488. doi: 10.1002/minf.201000061
- WHO (2018a). *HIV/AIDS*. Available at: <http://www.who.int/news-room/fact-sheets/detail/hiv-aids> [accessed September 20, 2018].
- WHO (2018b). *Influenza (Seasonal)*. Available at: [http://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)) [accessed September 20, 2018].
- WHO (2018c). *Malaria*. Available at: <http://www.who.int/news-room/fact-sheets/detail/malaria> [accessed September 20, 2018].
- WHO (2018d). *Schistosomiasis*. Available at: <http://www.who.int/news-room/fact-sheets/detail/schistosomiasis> [accessed September 20, 2018].
- WHO (2018e). *Tuberculosis*. Available at: <http://www.who.int/news-room/fact-sheets/detail/tuberculosis> [accessed September 20, 2018].
- Williams, A. J., and Ekins, S. (2011). A quality alert and call for improved curation of public chemistry databases. *Drug Discov. Today* 16, 747–750. doi: 10.1016/j.drudis.2011.07.007
- Young, D., Martin, T., Venkatapathy, R., and Harten, P. (2008). Are the chemical structures in your QSAR correct? *QSAR Comb. Sci.* 27, 1337–1345. doi: 10.1002/qsar.200810084
- Zhang, L., Fourches, D., Sedykh, A., Zhu, H., Golbraikh, A., Ekins, S., et al. (2013). Discovery of novel antimalarial compounds enabled by QSAR-based virtual screening. *J. Chem. Inf. Model.* 53, 475–492. doi: 10.1021/ci300421n

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Neves, Braga, Melo-Filho, Moreira-Filho, Muratov and Andrade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.