Click to prove you're human

1LabMol – Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, Brazil 2Laboratory of Cheminformatics, Centro Universitário de Anápolis (UniEVANGÉLICA), Anápolis, Brazil 3Laboratory 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 4Department of Chemical Technology, Odessa National Polytechnic University, Odessa, Ukraine

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 nD, 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.

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 nD) 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., 105 to 107 chemical structures) is reduced by QSAR models to a smaller number of compounds, which then will be tested experimentally (i.e., 101 to 103 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 (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. 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-HT1A) 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-HT1A 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-HT1A 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, [(8α)-6-methyl-9,10-didehydroergolin-8-yl]methanol), showed very high binding affinity (Ki) of 2.3 nM against 5-HT1A 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. 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. Acknowledgments The authors would like to thank Brazilian funding agencies, CNPq, CAPES, and FAPEG, for financial support and fellowships.

References 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 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 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar Gomes, M. N. 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar Hansch, C., and Fujita, T. (1964). p -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 CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar Laborda, 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar Mueller, R., Dawson, E. S., Meiler, J., Rodriguez, A. L., Chauder, B. A., Bates, B. S., et al. (2012). Discovery of 2-(2-Benzoxazoyl 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar 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 CrossRef Full Text | Google Scholar 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 PubMed Abstract | CrossRef Full Text | Google Scholar

Keywords: cheminformatics, machine learning, molecular descriptors, computer-assisted drug design, virtual screening 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 Received: 11 August 2018; Accepted: 18 October 2018;Published: 13 November 2018. Edited by: Adriano D. Andricopulo, Universidade de São Paulo, Brazil

*Correspondence: Carolina Horta Andrade, Y2Fyb2xpbmFAdWZnLmJy; Y2Fyb2xoYW5kcmFkZUBnbWFpbC5jb20=

Miller MD, Sheridan RP, Kearsley SK (1999) A program for rapidly producing pharmacophorically relevent molecular superpositions. J Med Chem 42:1505–1514Article CAS Google Scholar Lemmen C, Lengauer T (2000) Computational methods for the structural alignment of molecules. J Comput-Aided Mol Des 14:215–232Article CAS Google Scholar Jain AN (2004) Ligand-based structural hypotheses for virtual screening. J Med Chem 47:947–961Article CAS Google Scholar Hansch C, Leo A (1995) Exploring QSAR: Fundamentals and applications in chemistry and biology. American Chemical Society, Washington, DC Google Scholar Kubinyi H (1997) QSAR and 3D QSAR in drug design. Drug Discov Today 2:457–467Article CAS Google Scholar Cramer RD III, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA) 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967Article CAS Google Scholar Folkers G, Merz A, Rognan D (1993) CoMFA: Scope and limitations. In: Kubinyi H (ed) 3D QSAR in drug design. Theory, methods and applications. ESCOM Science Publishers BV, Leiden Google Scholar Klebe G, Abraham U (1993) On the prediction of binding properties of drug molecules by comparative molecular field analysis. J Med Chem 36:70–80Article CAS Google Scholar Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146Article CAS Google Scholar Cruciani G (ed) (2006) Methods and principles in medicinal chemistry – molecular interaction fields. VCH Publisher, New York Google Scholar Norinder U (1998) Recent progress in CoMFA methodology and related techniques. Perspect Drug Discov Des 12:3–23Article Google Scholar Kim KH, Greco G, Novellino EA (1998) Critical review of recent CoMFA applications. Perspect Drug Discov Des 12:257–315Article Google Scholar Podlogar BL, Ferguson DM (2000) QSAR and CoMFA: A perspective on the practical application to drug discovery. Drug Des Discov 1:4–12 Google Scholar Akamatsu M (2002) Current state and perspectives of 3D QSAR. Curr Top Med Chem 2:1381–1394Article CAS Google Scholar Kubinyi H, Folkers G, Martin YC (eds) (1998) 3D QSAR in drug design. Ligand–protein interactions and molecular similarity. Kluwer/ESCOM, Dodrecht Google Scholar Kubinyi H, Folkers G, Martin YC (eds) (1998) 3D QSAR in drug design. Recent advances. Kluwer/ESCOM, Dodrecht Google Scholar Masukawa KM, Kollman PA, Kuntz ID (2003) Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. J Med Chem 46:5628–5637Article CAS Google Scholar Tame JR (2005) Scoring functions: The first 100 years. J Comput Aided Mol Des 19:441–451Article CAS Google Scholar Kitchen DB, Decornez H, Furr JR et al. (2004) Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat Rev Drug Discov 11:935–949Article CAS Google Scholar Morris GM, Goodsell DS, Huey R et al. (1994) Distributed automatic docking of flexible ligands to proteins. J Comput Aided Mol Des 8:243–256Article Google Scholar Verdonk ML, Cole JC, Hartshorn M et al. (2003) Improved protein–ligand docking using GOLD. Proteins 52:609–623Article CAS Google Scholar Meng E, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comp Chem 13:505–524Article CAS Google Scholar Kontoyianni M, McClellan LM, Sokol GS (2004) Evaluation of docking performance: Comparative data on docking algorithms. J Med Chem 47:558–565Article CAS Google Scholar Böhm HJ (1998) Prediction of binding constants of protein–ligands: A fast method for the prioritisation of hits obtained from de-novo design or 3D database search programs. J Comput Aided Mol Des 12:309–323Article Google Scholar Tame JRH (1999) Scoring functions: A view from the bench. J Comput Aided Mol Des 13:99–108Article CAS Google Scholar Wang R, Lu Y, Fang X et al. (2004) An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes. J Chem Inf Comput Sci 44:2114–2125CAS Google Scholar Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 56:235–249Article CAS Google Scholar Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small molecule ligands to macromolecular receptors. Angew Chem Int Ed 41:2644–2676Article CAS Google Scholar Huang D, Caflisch A (2004) Efficient evaluation of binding free energy using continuum electrostatics solvation. J Med Chem 47:5791–5797Article CAS Google Scholar Waller CL, Oprea TI, Giolitti A et al. (1993) Three-dimensional QSAR of human immunodeficiency virus (I) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules. J Med Chem 36:4152–4160Article CAS Google Scholar Cho SJ, Garsia ML, Bier J et al. (1996) Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors. J Med Chem 39:5064–5071Article CAS Google Scholar Vaz RJ, McLEan LR, Pelton JT (1998) Evaluation of proposed modes of binding of (2S)-2-[4-[[(3S)-1-acetimidoyl-3-pyrrolidinyl]oxyl]phenyl]-3-(7-amidino-2-naphtyl)-propanoic acid hydrochloride and some analogs to factor Xa using a comparative molecular field analysis. J Comput Aided Mol Des 12:99–110Article CAS Google Scholar Sippl W, Contreras JM, Rival YM et al. (1998) Comparative molecular field analysis of aminopyridazine acetylcholinesterase inhibitors. In: Gundertofte K (ed) Proceedings of the 12th European symposium on QSAR – molecular modelling and predicting of bioactivity. Plenum Press, Copenhagen Google Scholar Sippl W (2000) Receptor-based 3D quantitative structure–activity relationships of estrogen receptor ligands. J Comput Aided Mol Des 14:559–572Article CAS Google Scholar Sippl W (2002) Binding affinity prediction of novel estrogen receptor ligands using receptor-based 3D QSAR methods. Bioorg Med Chem 10:3741–3755Article CAS Google Scholar Ortiz AR, Pisabarro MT, Gago F et al. (1995) Prediction of drug binding affinities by comparative binding energy analysis. J Med Chem 38:2681–2691Article CAS Google Scholar Lozano JJ, Pastor M, Cruciani G et al. (2000) 3D QSAR methods on the basis of ligand-receptor complexes. Application of combine and GRID/GOLPE methodologies to a series of CYP1A2 inhibitors. J Comput Aided Mol Des 13:341–353Article Google Scholar Reynolds CA, Wade RC, Goodford PJ (1989) Identifying targets for bioreductive agents: Using GRID to predict selective binding regions of proteins. J Mol Graph 7:103–108Article CAS Google Scholar Cruciani G, Watson K (1994) Comparative molecular field analysis using GRID force field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. J Med Chem 37:2589–2601Article CAS Google Scholar Cruciani G, Crivori P, Carrupt P-A et al. (2000) Molecular fields in quantitative structure–permeation relationships. J Mol Struct 503:17–30CAS Google Scholar Duca S, Hopfinger AJ (2001) Estimation of molecular similarity based on 4D-QSAR analysis: Formalism and validation. J Chem Inf Comput Sci 41:1367–1387CAS Google Scholar Vedani A, Dobler M (2002) 5D-QSAR: The key for simulating induced fit? J Med Chem 45:2139–2149Article CAS Google Scholar Vedani A, Dobler M, Lill MA (2005) Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. J Med Chem 48:3700–3703Article CAS Google Scholar Gohlke H, Klebe G (2002) Drug score meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. J Med Chem 45:4153–4170Article CAS Google Scholar Silber K, Heidler P, Kurz T et al. (2005) AFMoC enhances predictivity of 3D QSAR: A case study with DOXP-reductoisomerase. J Med Chem 48:3547–3563.Article CAS Google Scholar Hillebrecht A, Supuran CT, Klebe G (2006) Integrated approach using protein and ligand information to analyze selectivity- and affinity-determining features of carbonic anhydrase isozymes. ChemMedChem 1:839–853Article CAS Google Scholar Oprea TI, Garcia AE (1996) Three-dimensional quantitative structure–activity relationships of steroid aromatase inhibitors. J Comput Aided Mol Des 10:186–200Article CAS Google Scholar Golbraikh A, Trophsa A (2002) Beware of q2! J Mol Graph Model 20:269–276Article CAS Google Scholar Kubinyi H, Hamprecht FA, Mietzner T (1998) Three-dimensional quantitative similarity–activity relationships (3D QSiAR) from SEAL similarity matrices. J Med Chem 41:2553–2564Article CAS Google Scholar Golbraikh A, Shen M, Xiao Z et al. (2003) Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17:241–253Article CAS Google Scholar Norinder U (1996) Single and domain made variable selection in 3D QSAR applications. J Chemomet 10:95–105Article CAS Google Scholar Doweyko AM (2004) 3D QSAR illusions. J Comput Aided Mol Des 18:587–596Article CAS Google Scholar van Drie JH (2004) Pharmacophore discovery: A critical review. In: Bultinck P, De Winter H, Langenaeker W et al., (eds) Computational medicinal chemistry for drug discovery. Marcel Dekker, New York Google Scholar Head RD, Smythe ML, Oprea TI et al. (1996) VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. J Am Chem Soc 118:3959–3969Article CAS Google Scholar Di Santo R, Costi R, Artico M et al. (2002) Design, synthesis and QSAR studies on N-aryl heteroarylisopropanolamines, a new class of non-peptidic HIV-1 protease inhibitors. Bioorg Med Chem 10:2511–2526Article CAS Google Scholar Ortiz AR, Pisabarro MT, Gago F et al. (1995) Prediction of drug binding affinities by comparative binding energy analysis. J Med Chem 38:2681–2691Article CAS Google Scholar Martin-Santamaria S, Munoz-Muriedas J, Luque FJ et al. (2004) Modulation of binding strength in several classes of active site inhibitors of acetylcholinesterase studied by comparative binding energy analysis. J Med Chem 47:4471–4482Article CAS Google Scholar Murcia M, Ortiz AR (2004) Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. J Med Chem 47:805–820Article CAS Google Scholar Wang T, Wade RC (2002) Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. J Med Chem 45:4828–4837Article CAS Google Scholar Wang T, Wade RC (2001) Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. J Med Chem 44:961–997Article CAS Google Scholar Kim HJ, Chae CH, Yi KY et al. (2004) Computational studies of COX-2 inhibitors: 3D QSAR and docking. Bioorg Med Chem 12:1629–1641Article CAS Google Scholar Lushington GH, Guo JX, Wang JL (2007) Whither combine? New opportunities for receptor-based QSAR. Curr Med Chem 14:1863–1877Article CAS Google Scholar Waller CL, Oprea TI, Giolitti A et al. (1993) Three-dimensional QSAR of human immunodeficiency virus (I) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules. J Med Chem 36:4152–4160Article CAS Google Scholar Cho SJ, Garsia ML, Bier J et al. (1996) Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors. J Med Chem 39:5064–5071Article CAS Google Scholar Muegge I, Podlogary BL (2001) 3D-quantitative structure–activity relationships of biphenyl carboxylic acid MMP-3 inhibitors: Exploring automated docking as alignment method. Quant Struct-Act Relat 20:215–223Article CAS Google Scholar Costantino G, Macchiarulo A, Camaioni E et al. (2001) Modeling of poly(ADP-ribose)polymerase (PARP) inhibitors. Docking of ligands and quantitative structure–activity relationship analysis. J Med Chem 44:3786–3794Article CAS Google Scholar Matter H, Defossa E, Heinelt U et al. (2002) Design and quantitative structure–activity relationship of 3-amidinobenzyl-1H-indole-2-carboxamides as potent, nonchiral, and selective inhibitors of blood coagulation factor Xa. J Med Chem 45:2749–2769Article CAS Google Scholar Matter H, Schudok M, Schwab W et al. (2002) Tetrahydroisoquinoline-3-carboxylate based matrix-metalloproteinase inhibitors: Design, synthesis and structure–activity relationship. Bioorg Med Chem 10:3529–3544Article CAS Google Scholar Tervo AJ, Nyroenen TH, Ronkko T et al. (2003) A structure–activity relationship study of catechol-O-methyltransferase inhibitors combining molecular docking and 3D QSAR methods. J Comput Aided Mol Des 17:797–810Article CAS Google Scholar Tervo AJ, Nyroenen TH, Ronkko T et al. (2004) Comparing the quality and predictiveness between 3D QSAR models obtained from manual and automated alignment. J Chem Inf Comput Sci 44:807–816CAS Google Scholar Egea PF, Klahoz BP, Moras D (2000) Ligand–protein interactions in nuclear receptors of hormones. FEBS Lett 476:62–67Article CAS Google Scholar Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases: 1. Evaluation of different docking/scoring combinations. J Med Chem 43:4759–4767Article CAS Google Scholar Chen YZ, Zhi DG (2001) Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins 43:217–226Article CAS Google Scholar Wolohan P, Reichert DE (2003) CoMFA and docking study of novel estrogen receptor subtype selective ligands. J Comput Aided Mol Des 17:313–328Article CAS Google Scholar Soderholm AA, Lehtovuori PT, Nyronen TH (2005) Three-dimensional structure–activity relationships of nonsteroidal ligands in complex with androgen receptor ligand-binding domain. J Med Chem 48:917–925Article CAS Google Scholar Ai N, DeLisle RK, Yu SJ et al. (2003) Computational models for predicting the binding affinities of ligands for the wild-type androgen receptor and a mutated variant associated with human prostate cancer. Chem Res Toxicol 16:1652–1660Article CAS Google Scholar Moro S, Braiuca P, Deflorian F et al. (2005) Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: Pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study. J Med Chem 48:152–162Article CAS Google Scholar Medina-Franco JL, Rodrýguez-Morales S, Juarez-Gordiano CA et al. (2004) Docking-based CoMFA and CoMSIA studies of non-nucleoside reverse transcriptase inhibitors of the pyridinone derivative type. J Comput Aided Mol Des 18:345–360Article CAS Google Scholar Thaimattam R, Daga P, Rajjak SA et al. (2004) 3D QSAR CoMFA, CoMSIA studies on substituted areas as Raf-1 kinase inhibitors and its confirmation with structure-based studies. Bioorg Med Chem 12:6415–6425Article CAS Google Scholar Sun WS, Park YS, Yoo J et al. (2003) Rational design of an indolebutanoic acid derivative as a novel aldose reductase inhibitor based on docking and 3D QSAR studies of phenethylamine derivatives. J Med Chem 46:5619–5627Article CAS Google Scholar Kim HJ, Chae CH, Yi KY et al. (2004) Computational studies of COX-2 inhibitors: 3D-QSAR and docking. Bioorg Med Chem 12:1629–1641Article CAS Google Scholar Ragno R, Artico M, De Martino G et al. (2005) Docking and 3D QSAR studies on indolyl aryl sulfones. Binding mode exploration at the HIV-1 reverse transcriptase non-nucleoside binding site and design of highly active N-(2-hydroxyethyl)carboxamide and N-(2-hydroxyethyl)carbohydrazide derivatives. J Med Chem 48:213–223Article CAS Google Scholar Assefa H, Kamath S, Buolamwini JK (2003) 3D-QSAR and docking studies on 4-anilinoquinazoline and 4-anilinoquinoline epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors. J Comput Aided Mol Des 17:475–493Article CAS Google Scholar Hu X, Stebbins CE (2005) Molecular docking and 3D QSAR studies of yersinia protein tyrosine phosphatase YopH inhibitors. Bioorg Med Chem 13:1101–1109Article CAS Google Scholar Kuo CL, Assefa H, Kamath S et al. (2004) Application of CoMFA and CoMSIA 3D QSAR and docking studies in optimization of mercaptobenzenesulfonamides as HIV-1 integrase inhibitors. J Med Chem 47:385–399Article CAS Google Scholar Zhou Z, Madura JD (2004) 3D QSAR analysis of HIV-1 RT non-nucleoside inhibitors, TIBO derivatives based on docking conformation and alignment. J Chem Inf Comput Sci 44:2167–2178CAS Google Scholar Sutherland J, Weaver DF (2004) Three-dimensional quantitative structure–activity and structure–selectivity relationships of dihydrofolate reductase inhibitors. J Comput Aided Mol Des 18:309–331Article CAS Google Scholar Carrieri A, Carotti A, Barreca ML et al. (2002) Binding models of reversible inhibitors to type-B monoamine oxidase. J Comput Aided Mol Des 16:769–778Article CAS Google Scholar Sippl W, Höltje H-D (2000) Structure-based 3D-QSAR – merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods. J Mol Struct (Theochem) 503:31–50Article CAS Google Scholar Cinone N, Höltje H-D, Carotti A (2000) Development of a unique 3D interaction model of endogenous and synthetic peripheral benzodiazepine receptor ligands. J Comput Aided Mol Des 14:753–768Article CAS Google Scholar Hammer S, Spika L, Sippl W et al. (2003) Glucocorticoid receptor interactions with glucocorticoids: Evaluation by molecular modeling and functional analysis of glucocorticoid receptor mutants. Steroids 68:329–339Article CAS Google Scholar Classen-Houben D, Sippl W, Höltje HD (2002) Molecular modeling on ligand-receptor complexes of protein-tyrosine-phosphatase 1B. In: Ford M, Livingstone D, Dearden J et al. (eds). EuroQSAR 2002 designing drugs and crop protectants: Processes, problems and solutions. Blackwell Publishing, Bournemouth Google Scholar Broer BM, Gurrath M, Höltje H-D (2003) Molecular modelling studies on the ORL1-receptor and ORL1-agonists. J Comput Aided Mol Des 17:739–754Article Google Scholar Sippl W, Contreras JM, Parrot I et al. (2001) Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. J Comput Aided Mol Des 15:395–410Article CAS Google Scholar Wichapong K, Lindner M, Pianwanita S et al. (2008) Receptor-based 3D-QSAR studies of checkpoint wee1 kinase inhibitors. Eur J Med Chem 44:1383–1395Article CAS Google Scholar Barner EL, Gray SL (1998) Donepezil in alzheimer disease. Ann Pharmacother 32:70–77Article CAS Google Scholar Raves ML, Harel M, Pang YP et al. (1997) Structure of acetylcholinesterase complexed with the nootropic alkaloid huperzine A. Nat Struct Biol 4:57–63Article CAS Google Scholar Contreras JM, Rival Y, Chayer S et al. (1999) Aminopyridazines as acetylcholinesterase inhibitors. J Med Chem 42:730–741Article CAS Google Scholar Kryger G, Silman I, Sussman JL (1999) Structure of acetylcholinesterase complexed with E2020 (aricept): Implications for the design of new anti-alzheimer drugs. Structure Fold Des 15:297–307Article Google Scholar San Juan AA (2008) 3D-QSAR models on clinically relevant K103N mutant HIV-1 reverse transcriptase obtained from two strategic considerations. Bioorg Med Chem Lett 18:1181–1194Article CAS Google Scholar Pately PD, Pately MR, Kaushik-Basu N et al. (2008) 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors. J Chem Inf Model 48:42–51Article CAS Google Scholar Murumkar PR, Giridhar R, Yadav MR (2008) 3D-quantitative structure–activity relationship studies on benzothiadiazepine hydroxamates as inhibitors of tumor necrosis factor-α converting enzyme. Chem Biol Drug Des 71:363–373Article CAS Google Scholar Muddassar M, Pasha FA., Chung HW et al. (2008) Receptor guided 3D-QSAR: A useful approach for designing of IGF-1R inhibitors. J Biomed Biotech 2008:837653Article CAS Google Scholar San Juan AA (2008) Towards predictive inhibitor design for the EGFR autophosphorylation activity. Eur J Med Chem 43:781–791Article CAS Google Scholar Fischer B, Fukuzawa K, Wenzel W (2008) Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in silico drug discovery. Proteins 70:1264–1273Article CAS Google Scholar Roncaglioni A, Benfenati E (2008) In silico-aided prediction of biological properties of chemicals: Oestrogen receptor-mediated effects. Chem Soc Rev 37:441–450Article CAS Google Scholar Sadeghian H, Seyedi SM, Saberi MR et al. (2008) Design and synthesis of eugenol derivatives, as potent 15-lipoxygenase inhibitors. Bioorg Med Chem 16:890–901Article CAS Google Scholar Holder S, Lilly M, Brown ML (2007) Comparative molecular field analysis of flavonoid inhibitors of the PIM-1 kinase. Bioorg Med Chem 15:6463–6473Article CAS Google Scholar Abu Hammad AM, Afifi FU, Taha MO (2007) Combining docking, scoring and molecular field analyses to probe influenza neuraminidase-ligand interactions. J Mol Graph Model 26:443–456Article CAS Google Scholar Dezi C, Brea J, Alvarado M et al. (2007) Multistructure 3D-QSAR studies on a series of conformationally constrained butyrophenones docked into a new homology model of the 5-HT2A receptor. J Med Chem 50:3242–3255Article CAS Google Scholar Korhonen S-P, Tuppurainen K, Asikainen A et al. (2007) SOMFA on large diverse xenoestrogen dataset: The effect of superposition algorithms and external regression tools. QSAR Comb Sci 26:809–819Article CAS Google Scholar Du L, Li M, You Q et al. (2007) A novel structure-based virtual screening model for the hERG channel blockers. Biochem Biophys Res Commun 355:889–894Article CAS Google Scholar Tuccinardi T, Ortore G, Rossello A et al. (2007) Homology modeling and receptor-based 3D-QSAR study of carbonic anhydrase IX. J Chem Inf Model 47:2253–2262Article CAS Google Scholar Pissurlenkar RRS, Shaikh MS, Coutinho EC (2007) 3D-QSAR studies of dipeptidyl peptidase IV inhibitors using a docking based alignment. J Mol Mod 13:1047–1071Article CAS Google Scholar Tuccinardi T, Nuti E, Ortore G. et al. (2007) Analysis of human carbonic anhydrase II: Docking reliability and receptor-based 3D-QSAR study. J Chem Inf Model 47:515–525Article CAS Google Scholar Huang H, Pan X, Tan N et al. (2007) 3D-QSAR study of sulfonamide inhibitors of human carbonic anhydrase II. Eur J Med Chem 42:365–372Article CAS Google Scholar Jójárt B, Márki A (2007) Receptor-based QSAR studies of non-peptide human oxytocin receptor antagonists. J Mol Graph Model 25:711–720Article CAS Google Scholar Jacobsson M, Liden P, Stjernschantz E et al. (2003) Improving structure-based virtual screening by multivariate analysis of scoring data. J Med Chem 46:5781–5789Article CAS Google Scholar

With the advancement of computational resources, there is a gradual uplifting of the used dimensions of quantitative structure–activity relationship (QSAR) descriptors. The two-dimensional (2D) and lower-dimensional models suffer from various drawbacks that led to the introduction of 3D-QSAR. This approach has been enhanced with significant advancements in order to study multiple three-dimensional (3D) features of chemicals, establishing a correlation between structure and biological activity. The 3D-QSAR techniques are broadly divided into alignment-based methods [comparative molecular field analysis (CoMFA), self-organizing molecular field analysis (SOMFA), comparative molecular similarity indices analysis (CoMSIA), receptor surface analysis (RSA), and molecular shape analysis (MSA)] and alignment-independent methods [comparative molecular moment analysis (CoMMA), weighted holistic invariant molecular (WHIM) descriptor analysis, VolSurf, Compass, comparative spectral analysis (CoSA), grid-independent descriptors (GRIND)]. The fundamental concept, methodology, and limitations of some of the major approaches are discussed in this chapter to give an overview of this topic. 3D-QSAR; comparative molecular field analysis (CoMFA); comparative molecular moment analysis (CoMMA); comparative molecular similarity indices analysis (CoMSIA); comparative spectral analysis (CoSA); molecular shape analysis (MSA); receptor surface analysis (RSA); self-organizing molecular field analysis (SOMFA); weighted holistic invariant molecular (WHIM) Contents The basic principle of a quantitative structure–activity relationship (QSAR) study is that the deviations in biological response among a series of compounds are accountable for the differences in the structural properties. In the classical QSAR studies, biological responses have been correlated with atomic, group, or molecular properties such as lipophilicity, polarizability, electronic, and steric properties (Hansch analysis) or with certain structural features (Free–Wilson analysis). However, in these techniques, one cannot ignore their limited utility for designing diverse functional new molecules due to the lack of consideration of the three-dimensional (3D) structures of the molecules. As a consequence, 3D-QSAR has emerged as a natural extension to the classical Hansch and Free–Wilson approaches that exploits the 3D properties of the ligands to predict their biological response by employing robust chemometric tools. The 3D-QSAR is a broad term encompassing all those QSAR methods that correlate macroscopic target properties with computed atom-based descriptors derived from the spatial representation of the molecular structures. These approaches have served as a valuable predictive tool in the design of pharmaceuticals and agrochemicals [1–3]. The prime goal of any 3D-QSAR method is to establish the relationship between biological activity and spatial properties of chemicals like steric, electrostatic, and lipophilic ones. The 3D-QSAR methodology is computationally more exhaustive and complex than 2D-QSAR approaches. Normally, it consists of several steps to acquire numerical descriptors from the compound structures: It is interesting to point out that some methods, independent of the alignment strategy, have also been developed with the progress of 3D-QSAR approaches [4]. One has to understand that the QSAR model is not a substitute for the experimental assays, although experimental techniques are also not free of inaccuracies. However, QSAR researchers are trying to develop a model that is as close as possible to the real one, and for this purpose, the 3D-QSAR techniques have to rely on some basic assumptions, which are illustrated here: • Binding of a drug molecule or ligand with the receptor is considered directly related to the biological response. Effects on second messengers or other signaling effects between receptor binding and experimentally observed response are not normally considered. • Molecular properties (physical, chemical, and biological) are encoded with a set of numbers or descriptors. • It is believed in general that compounds with common structures have comparable properties, and thus they have similar binding modes and accordingly equivalent biological activities and vice versa. • Structural properties leading to a biological response are usually determined by nonbonding forces, mainly steric and electrostatic ones. • Another important assumption is that the biological response is shown by the ligand itself, not by its metabolite product. • The lowest-energy conformation of the ligand is its bioactive conformation, which exerts binding effects. • The geometry of the receptor binding site is considered rigid, though there are a few exceptions. • The loss of translational and rotational degrees of freedom (entropy) upon binding is believed to follow a similar pattern for all these compounds. • The protein binding site is assumed to be the same for all of the studied ligands. • The major factors that contribute to the overall free energy of binding, like desolvation energy, temperature, diffusion, transport, pH, salt concentration, and plasma protein binding, are difficult to identify and thus are generally ignored. The 3D-QSAR methods can be classified based on a variety of criteria, as given in Table 8.1. Most commonly and successfully employed 3D-QSAR methods are discussed in the following sections of this chapter. Table 8.1 Categorization of 3D-QSAR techniques Basis of classification Type Examples of techniques Based on employed chemometric techniques Linear CoMFA, CoMSIA, AFMoC, GERM, CoMMA, SoMFA Nonlinear Compass Based on the alignment criterion Alignment-dependent CoMFA, CoMSIA, MSA, RSA, GERM, AFMoC, HIFA, VFA, MQSM Alignment-independent Compass, CoMMA, HQSAR, WHIM, GRIND, VolSurf, CoSA Based on intermolecular modeling or the information employed to develop QSAR Ligand-based CoMFA, CoMSIA, MSA, RSA, Compass, GERM, CoMMA, SoMFA Receptor-based AFMoC, HIFA Comparative molecular field analysis (CoMFA) is a molecular field–based, alignment-dependent, ligand-based method developed by Cramer et al. [5], which helps in building the quantitative relationship of molecular structures and its response property. The method mostly focuses on ligand properties like steric and electrostatic ones, and the resulting favorable and unfavorable receptor–ligand interactions. As CoMFA is an alignment-dependent, descriptor-based method, all aligned ligands are placed in an energy grid, and by placing an appropriate probe at each lattice point, energy is calculated. The resultant energy calculated at each unit fraction corresponds to electrostatic (Coulombic) and steric (van der Waals) properties. These computed values serve as descriptors for model development. These descriptor values are then correlated with biological responses employing a robust linear regression method like partial least squares (PLS). The PLS results serve as an important signal to identify the favorable and unfavorable electrostatic and steric potential and also correlate it with biological responses. The formalism of the CoMFA methodology is described next: a. Structures of all molecules are drawn using any structure-drawing software. b. The bioactive conformation of each molecule is generated and energy minimization is carried out. c. All the molecules are superimposed or aligned using either manual or automated methods employed in the working software, in a manner defined by the supposed mode of interaction with the receptor. d. Thereafter, the overlaid compounds are positioned in the center of a lattice grid with a spacing of 2 Å. e. In the 3D space, the steric and electrostatic fields are calculated around the molecules with different probe groups positioned at all intersections of the lattice. Computation of the steric field uses the Lennard–Jones equation as follows: (8.1) In Eq. (8.1), ε is the depth of the potential well, σ is the finite distance at which the interparticle potential is zero, r is the distance between the particles, and rm is the distance at which the potential reaches its minimum. At rm, the potential function has the value −ε. The distances are given as rm=21/6σ.Again, computation of electrostatic field follows the Coulombic interaction equation as follows: (8.2) where q1 and q2 denote point charges, r is the distance between charges, and ε is the dielectric constant of the medium. f. The interaction energy or field values forming a pool of the descriptor/variable matrix are correlated with the biological response data employing the PLS technique, which identifies and extracts the quantitative influence of specific features of molecules on their activity. g. The results may be expressed as correlation equations with the number of latent variable terms, each of which is a linear combination of original independent lattice descriptors. h. For visual interpretation, the PLS output is illustrated in the form of interactive graphics consisting of colored contour plots of coefficients of the corresponding field variables at each lattice intersection, and showing the imperative favorable and unfavorable regions in the 3D space, which are closely associated with the biological activity. The CoMFA formalism is schematically illustrated in Figure 8.1. There are diverse factors that can control the complete performance of the constructed CoMFA model. These are described in the next sections. Like any 2D-QSAR method, one has to use precise activity data in order to create a good 3D-QSAR model. The following conditions should be fulfilled for maintaining the accuracy and appropriateness of the biological response data [3,6]: Representing the initial molecular structure is an important issue in 3D-QSAR analysis. This can be done by both experimental and computational approaches. A huge number of experimentally determined crystal structures are accessible in databases like the Cambridge Structural Database [7] and the Protein Data Bank [8]. The obtainable crystal structures present the benefit that some conformational information about the flexible molecule is included. Computationally, the 3D structures can be generated by three methods: Once the starting 3D molecular structures are generated, their geometries are refined by minimizing their conformational energies using following structure optimization techniques, including: • Molecular mechanics: It does not explicitly consider the electronic motion, so they are fast, accurate, and can be employed for large molecules like enzymes. • Quantum mechanics or ab initio: It takes into account the 3D distribution of electrons around the nuclei, and thus it is extremely precise. The identified major drawbacks of this methods are that they are time-consuming and computationally intensive, and they cannot handle large molecules. • Semiempirical: Semiempirical quantum chemical methods attempt to address two restrictions—namely, slow speed and low accuracy of quantum mechanical (e.g., Hartree–Fock) calculations by omitting certain integrals based on experimental data, such as ionization energies of atoms or dipole moments of molecules. Thus, semiempirical methods are very fast, applicable to large molecules, and may give precise results when applied to molecules that are similar to the molecules used for parameterization. Molecules to be used for semiempirical calculations may contain hundreds of atoms. Modern semiempirical models are based on the neglect of diatomic differential overlap (NDDO) methods like MNDO, AM1, PM3, and PDDG/PM3. The following conformational search methods can be implemented: • Systematic search (or grid search): It generates all probable conformations by systematically varying each of the torsion angles of a molecule by some increment, keeping the bond lengths and bond angles fixed. • Monte Carlo: It simulates dynamic behavior of a compound and generates the conformations by making random changes in its structure, calculating and comparing its energy with that of the previous conformation, and accepting the result if it is unique. • Random search: It generates a set of conformations by repetitively and arbitrarily changing either the Cartesian (x, y, z) or the internal (bond lengths, bond angles, and torsion/dihedral angles) coordinates of a starting geometry of the molecule under consideration. • Molecular dynamics: It employs Newton’s second law of motion (force=mass×acceleration) to simulate the time-dependent movements and conformational changes in a molecular system, and results in a so-called trajectory showing how the positions and velocities of atoms in the molecular system vary with time. • Simulated annealing: It theoretically heats up the molecular system under consideration to high temperatures to overcome huge energy barriers, and after equilibrating there for some time using molecular dynamics, cools down the system slowly and gradually to obtain low-energy conformations according to the Boltzmann distribution. • Distance geometry algorithm: It generates a random set of coordinates by selecting random distances within each pair of upper and lower bounds to form constraints in a distance matrix, which are employed to create energetically feasible conformations of a set of molecules. • Genetic and evolutionary algorithms: It is based on the concept of biological evolution and initially creates a population of promising solutions to the problem. The solutions with the best fitness scores undergo crossovers and mutations over a time, and proliferate their good distinctiveness down the generations resulting in better solutions in the form of new conformers. The bioactive conformation defines a particular conformation of the molecule in which it is bound to the receptor. The intrinsic forces between the atoms in the molecule, as well as extrinsic forces between the molecule and its surrounding environment, considerably influence the bioactive conformation of the molecule [6]. Bioactive conformations of the compounds can be attained both by experimental and theoretical techniques. Experimental methods for creating bioactive conformations comprise the techniques described. The precise 3D structure of the macromolecules can be obtained by this method. Drug–receptor complexes generated by X-ray crystallography logically offer the exact information, but this method has several disadvantages: The 3D structural data is obtained in the solution and is a method of selection when the molecule cannot be crystallized through experimental ways, as in the case of the membrane-bound receptors or receptors, which have not yet been isolated due to stability, resolution, or other issues. The imperative features of this method are: