

From Origin to Current Methods: An Overview of Molecular Modeling Applied to Medicinal Chemistry in the Last 30 Years

Vinícius N. Rocha^a and Carlos M. R. Sant'Anna^b*,^{a,b}

^aPrograma de Pós-Graduação em Química, Instituto de Química, Universidade Federal Rural do Rio de Janeiro, 23897-000 Seropédica-RJ, Brazil

^bDepartamento de Química Fundamental, Instituto de Química, Universidade Federal Rural do Rio de Janeiro, 23897-000 Seropédica-RJ, Brazil

Molecular modeling soon attracted the attention of Medicinal Chemistry researchers, given the importance of molecular structure for understanding the mode of action and for designing bioactive compounds. Computer-assisted drug design (CADD) has become widespread and today big pharmaceutical companies routinely use it as a support tool in the search for new drugs. Here, it will be addressed the most relevant topics about CADD from the last 30 years, when research groups in Medicinal Chemistry began to explore molecular modeling in Brazil. This history can be described through phases in which some methods emerged and became predominant, in a continuous evolution, passing, for example, from the basic empirical field and quantum mechanics molecular modeling procedures, through quantitative structure-activity relationship (QSAR) methods, molecular docking and, more recently, virtual screening. Since the mid-2000s, machine learning methods have been increasingly applied to the solution of problems in the context of Medicinal Chemistry, such as the determination of protein 3D structure and the characterization of relationships between chemical structures and their biological activities. Far from being complete, this history continues its evolution, bringing significant contributions to the drug design, either by reducing the time and cost of research, or by enabling and accelerating the finding for new bioactive compounds.

Keywords: molecular modeling, CADD, LBDD, SBDD, machine learning

1. Introduction

The use of molecular models has a long history in chemistry. The difficulty to access molecular structures experimentally was one of the main reasons for this, but even when experimental data on molecular structures began to become available in greater quantities, the interpretation of these data frequently needed molecular models. This was the case when, based on crystallography results obtained by Rosalind Franklin, Watson and Crick,¹ in 1953, proposed a model for deoxyribonucleic acid (DNA) for which they, together with Wilkins, received the Nobel Prize in 1962.¹ It was in fact a physical model constructed manually by them, which obviously is not a practical way to routinely

explore molecular models. It was the introduction of computational methods based on sound theoretical models, molecular graphics, and computers capable of exploring them efficiently that finally determined the popularization of molecular modeling as a research tool.

One of the fundamental concepts of Medicinal Chemistry is that the activity of drugs is intrinsically related to their molecular structures. Since the original proposal made in 1894 by Fischer² that enzymes and their substrates present structural complementarity, in what became known as the lock-and-key model,² the importance of the molecular structure of small molecules (ligands) for their activity on biomacromolecules (and the possible modulation of them by the construction of specific ligands) was implicit, together with all potential pharmacological implications. The structure-activity relationship became a paradigm in Medicinal Chemistry and the quantitative structure-activity relationship (QSAR) models are mathematical translations of this concept. In fact, back in 1993, the current Medicinal Chemistry Division (MED) of the Brazilian Chemical

*e-mail: santanna@ufrj.br



Editor handled this article: Albertina Moglioni (Associate)
This review is dedicated to the memories of Prof Eliezer J. Barreiro and Prof Carlos Alberto Manssour Fraga, who with their passion for Medicinal Chemistry inspired generations of scientists in Brazil.



Society (SBQ) was born as the Quantitative Structure-Activity Relationship (SA) section of SBQ, with the initiative of Prof A. T. Amaral, Prof E. J. Barreiro and Prof R. A. Yunes.³

Because of the central role molecular structure plays in the activity of a bioactive compound, medicinal chemists soon became aware of the importance of structural determination as a tool to aid the design of bioactive molecule - and of the computational methods available for it. The Computer-Aided Drug Design (CADD) emerged and there was an explosion of interest in the potential for CADD in the pharmaceutical industry in the early 1980s; as proposed by Van Drie, this could be traced back at least in part to a 1981 cover article in *Fortune* magazine entitled “The Next Industrial Revolution: designing drugs by computer at Merck”,⁴ which was preceded by a scientific overview in *Science* of the initial evolution of “computer-assisted molecular modeling” at Merck.⁵ A lot of time has passed since then, and today CADD is routinely used in the pharmaceutical industry with significant results; a recent review⁶ showed that between 1981 and 2019 the discovery process of more than 70 commercialized drugs included some kind of computational technique significant enough for being mentioned in literature.

As is the case with most new technologies, CADD took some time to arrive in Brazil, and the first research using CADD here did not take place in industry, but in academia, as far as we know. The research groups of R. B. Alencastro and E. J. Barreiro at Universidade Federal do Rio de Janeiro (UFRJ) are probably the first ones in Brazil to explore molecular modeling methods in the study

of bioactive compounds, as reported in papers published in the 1990s.⁷⁻¹⁰ They were soon followed by many other excellent researchers, and, today, groups that use CADD at some stages of their research work in drug design are quite common in Brazil, including in the pharmaceutical industry.

The aim of this review is to present an overview of the evolution of CADD in the last 30 years (Figure 1); it is not our proposal to present an exhaustive description of each method, but to highlight some of the most significant moments of this already long, successful, and, undoubtedly, fascinating story.

2. Molecular Modeling in CADD: Basics

CADD methods can be conceptually divided into two main groups: Ligand-Based Drug Design (LBDD) and Structure-Based Drug Design (SBDD), where “structure” refers specifically to the target structure, generally a protein. All these methods rely at some point in models that are able to predict the molecular structures of ligands and/or macromolecular targets with enough accuracy. Empirical force field and quantum mechanics are the main theoretical approaches used for this aim until today, at least when stationary points on the potential energy surface for ligand structures are to be obtained. For biomacromolecules to be modelled, other approaches were developed, and will be discussed later.

The theories underlying these methods were developed many years before the advent of CADD, and what occurred during the last decades was in great part the emergence of improvements based on the same approaches.

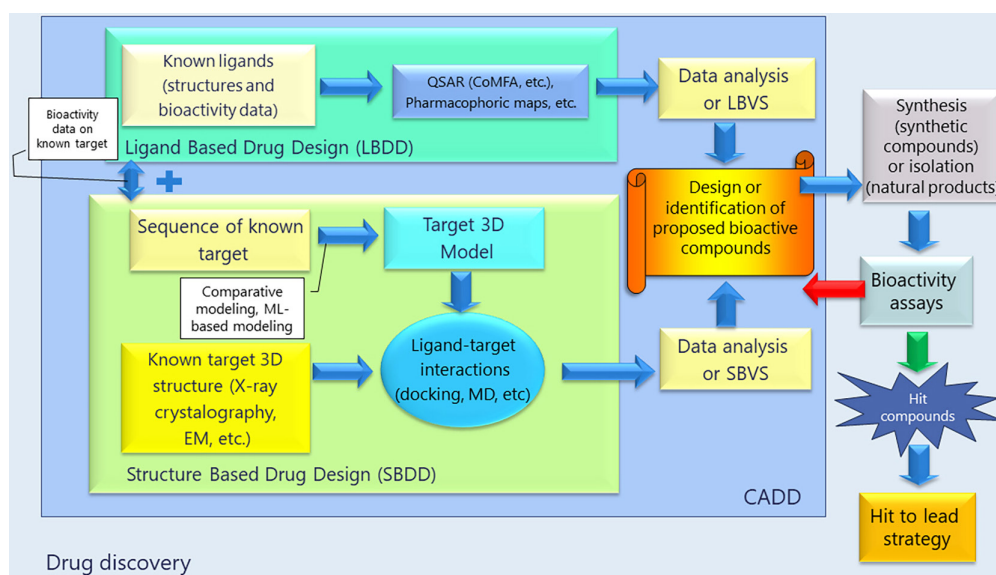


Figure 1. Simplified schematic description of how CADD methods were integrated to the drug discovery process in the last 30 years, which will be discussed in the text. QSAR: quantitative structure activity relationship; MD: molecular dynamics; LBVS: ligand based virtual screening; SBVS: structure based virtual screening.

Molecular mechanics force field methods remain a usual choice when large molecular systems are to be evaluated, since they can produce results of acceptable quality quite quickly. Assisted Model Building with Energy Refinement (AMBER),¹¹ Optimized Potentials for Liquid Simulations (OPLS),¹² Chemistry at Harvard Macromolecular Mechanics (CHARMM),¹³ and Groningen Molecular Simulation (GROMOS),¹⁴ are just a few examples of classical force fields still in use today in routine CADD. Classical forcefields are those composed by the addition of several simple terms describing separate effects of the structure on molecular energy, containing some fixed parameters, such as default geometry parameters (bond distances, bond angles, dihedrals, etc.) and atomic parameters, such as partial atomic charges.

Among the main improvements that have been introduced later on were the expansion of chemical space coverage of ligand-like molecules among major popular force fields, the inclusion of new charge models for better accuracy and transferability, and new automated parameterization toolkits including machine learning (ML) approaches.¹⁵ Despite the improvements presented by force field methods, when the explicit effect of electrons on a system needs to be considered, such as in bond-making processes in enzymatic reactions and covalent inhibition, redox processes, or for the high-quality description of molecular interactions, quantum methods must be used. It deserves to be highlighted, as far as *ab initio* quantum-mechanics methods are concerned, the growing trend for the use of density functional theory (DFT) methods.^{16,17}

Hartree-Fock (HF) methods, once dominant in literature citations, are based on wave functions, which depend on the $3n$ coordinates for an n -electron system, whereas DFT methods depend on how the electron density varies in the system, so the original formulation of DFT involves only three dimensions for closed-shell systems.¹⁸ In principle, this would result in faster calculations with DFT; however, little or no difference in computational time is in fact observed with the most popular quantum-chemistry based programs. This is probably a consequence of the fact that the computational implementations of the HF and DFT methods have in common the iterative solution of a set of one-particle eigenequations, which in DFT are the Kohn-Sham equations necessary for the calculation of the electron density and the corresponding energy.^{19,20}

Most DFT methods, including the B3LYP hybrid functional²¹ (very popular in CADD literature) include in their formulation a correlation-exchange term that allows to at least partially consider the effect of electron correlation, which is absent in the HF methods of the same level.¹⁸ However, the effects on molecular geometry optimization

are generally small and, specifically when it comes to CADD studies, where the experimental variables to be confronted with the model results present a considerable level of uncertainty, it is not clear whether the present preference for DFT is in fact justifiable.

However, there were some situations where DFT was demonstrated in fact to outperform a number of alternative methods, such as in the evaluation of the thermochemistry of systems containing transition metals.^{22,23} The evaluation was based on two well-designed benchmark sets for reaction energies and barrier heights: MOR41, which covers chemically relevant reactions of closed-shell complexes,²² and ROST61, which contains reactions of open-shell single-reference transition metal complexes.²³ For both benchmark sets, double-hybrid functionals²⁴ surpassed all other assessed methods, PWPB95-D3 for the first one and PWPB95-D4 for the second, in this case approaching the estimated accuracy of the reference method, CCSD(T). Such reaction energetics can have some importance in CADD studies involving metalloenzymes or metal-containing drug candidates, for example.

On the other hand, post-HF methods, such as Coupled-Cluster²⁵ and Configuration Interaction,²⁶ produce results with higher accuracy than HF and DFT methods in general, but computational costs can limit their applications in CADD studies, where large collections of ligands and/or large molecules are frequently involved.

For very large systems, quantum calculations can be done at the semi-empirical level, which can be applied with a low computational cost and, consequently, relatively fast calculation speed, to systems with thousands of atoms, including complete proteins and their complexes with ligand molecules. Almost 30 years from now, Dewar's group launched AM1 (Austin Model 1), a landmark in the semiempirical methods history, but with parameters still limited to a few elements, most of them related to organic molecules.²⁷ Semi-empirical quantum mechanical methods capable of modeling transition metal-containing systems, and also containing enhancements in the description of molecular interactions, were released in the last decades, including PM6,²⁸ PM7,²⁹ and PM6-DnHmX methods;³⁰⁻³² recently, a new one was presented, PM6-ORG, with a reparameterization claimed to improve protein modeling.³³ It must be remembered that since semi-empirical methods partially depend on parameters obtained from experimental information, the quality of the results is expected to be influenced by the extent of available data. Therefore, some variation in the accuracy of results is expected, especially when atoms and bonds with less available data are used for parameterization, such as for transition metal complexes.³⁴ For such systems, a careful preliminary step

based on comparison with structures obtained by high-accuracy DFT or wave function-based calculations and experimental procedures is recommended in order to verify the performance of the chosen semi-empirical method with similar systems.

An interesting alternative approach for the inclusion of quantum calculations in the modeling of large molecular systems is the combined Quantum Mechanics/Molecular Mechanics approach (QM/MM). It was first implemented in 1976 by Warshel and Levitt³⁵ for the investigation of an enzymatic reaction. The proposal of QM/MM is to provide reliable chemical accuracy using a quantum method for the treatment of the most important region of the system, such as the active site of an enzyme, while the remaining system is modeled with the faster molecular mechanics method. In this way, the field effect of the surroundings on the quantum-mechanical calculation can be included.

An important step for the effective use of the QM/MM approach is the appropriate delimitation of the QM region, which can be sometimes challenging. Based on a set of model clusters with well-defined configurations to mimic the basic types of non-covalent interactions in proteins, Kollar and Frece³⁶ using the DFT-B3LYP/6-31G**//OPLS-2005 QM/MM approach, recently proposed recommendations for the addition of chemical groups or protein residues into the QM region for leading to a more realistic description of ligand-protein interactions.

With respect to specific applications in the drug design process, Kar³⁷ presented an interesting review where he highlights the methodologies through which the QM/MM approach proved critical for understanding the drug-target interaction, together with a discussion of some recent hybrid QM/MM approach results applied to FDA-approved drugs.

A natural evolution of the QM/MM concept that is being explored in recent years is the combination of QM/MM with molecular dynamics simulations, the QM/MM/MD approach, a powerful and promising tool for the investigation of chemical reactions in complex biochemical systems.³⁸

3. The Evolution of Ligand-Based Models in CADD

Undoubtedly, QSAR is the most prominent and explored method in the LBDD branch of CADD. QSAR is an extension of the original proposal of Hammett^{39,40} presented in the 1930s of a correlation between the effect of the addition of substituents on benzoic acid with pK_a ; Hammett^{39,40} presented parameters called electronic

σ - ρ constants and established the linear free-energy relationship (LFER) principle. Later, in the 1950's, Taft⁴¹ proposed the first parameters related to steric effects. Then, in the next decade, the seminal works of Hansch, Fujita, Free and Wilson⁴²⁻⁴⁵ laid the foundations of quantitative structure-activity relationships applied to compounds with biologic activity.

The first electronic and steric parameters are examples of what were later called descriptors, which may be any quantity related to the molecular structure as a whole, such as clogP, or to some part of it, such as the electron density at a given atom.⁴⁶ QSAR are functions that allow, for a group of bioactive molecules, a quantitative relationship between some selected descriptors and the measured activity data, such as half-maximal inhibitory concentration (IC_{50}), the dissociation constant related to the inhibitor-enzyme binding (K_i), etc., to be established. Depending on the type of descriptors, we can classify QSAR models according to their dimensionality; when the descriptors are somehow related to the three-dimensional structure of the ligands, for example, the method is called 3D-QSAR. The 1980s were the moment for the birth of 3D-QSAR, as obtaining the 3D structures for large groups of ligands, necessary to build robust QSAR models, was becoming easy with the availability of molecular modeling software, together with faster and more accessible computers.^{47,48}

Cramer *et al.*⁴⁹ introduced a new kind of descriptors, based on the concept of molecular fields, which are built on two premises: (i) the observed biological effect is usually resulting from non-covalent interactions; and (ii) most of these interactions are mediated by forces that could be acceptably evaluated with the Van der Waals (the 6-12 Lennard-Jones potential) and Coulomb terms of molecular mechanics force fields. These terms are calculated on the nodes of a 3D grid involving the energy-minimized structures of the active compounds, which need to be aligned according to some criteria. The method, called Comparative Molecular Field Analysis (CoMFA), was made effective by a number of advances in molecular graphics and by the emergence of a new method of data analysis in the 1980s, Partial Least Squares (PLS);⁵⁰ applied to QSAR, PLS was able to derive robust linear equations from tables having many more descriptors than compounds.

In the next decade, another 3D-QSAR method was presented, the Comparative Molecular Similarity Indices Analysis (CoMSIA), which was developed to overcome certain limitations of CoMFA; in place of the molecular mechanics fields of CoMFA, in CoMSIA similarity fields were employed as descriptors to describe steric, electrostatic, hydrophobic and hydrogen bonding properties.⁵¹ Both of these molecular field methods remain

significantly in use today; a direct search with the words “CoMFA” and “CoMSIA” on Google Scholar returned 3,730 and 2,070 citations, respectively, in the period 2020-2023.

As observed by Hopfinger *et al.*,⁵² the 3D-QSAR had three inherent problems: the identification of the bioactive conformations of flexible compounds; the molecular alignment; and the partitioning of each molecule in the training set with respect to interactions with the receptor. To solve these problems, they developed 4D-QSAR, in which the fourth dimension is the “dimension” of an ensemble sampling generated by a molecular dynamics (MD) approach. The method includes the conformational flexibility and the freedom of alignment by ensemble averaging of the descriptors, which are the grid cell occupancy descriptors (GCOD), generated for a number of different atom types, called interaction pharmacophore elements (IPEs). These IPEs are defined as “any type”, “nonpolar”, “polar-positive charge”, “polar-negative charge”, “hydrogen bond acceptor”, “hydrogen bond donor”, and “aromatic”, which are defined onto a grid around the conformational ensemble.⁵²

It should be mentioned that shortly after the emergence of 4D-QSAR, the first example of the use of 4D-QSAR by Brazilian research groups was published at the end of the 1990s, a paper resulting from the collaboration of Albuquerque, Alencastro and Barreiro with Hopfinger,⁵³ creator of the method. In the late 2000s, Ferreira and co-workers,⁵⁴ at Universidade Estadual de Campinas (Unicamp), presented LQTA-QSAR, a free package that explores jointly the main features of CoMFA and 4D-QSAR paradigms to develop 4D-QSAR models, for which the ligand conformations are obtained with GROMACS, the popular free and open-source software suite for high-performance MD and output analysis launched in the 1990s.^{55,56}

An interesting improvement of the method emerged also in the 2000s, the RD-4D-QSAR, with RD meaning receptor dependent. This approach can be applied when the target structure is available, as the main feature of RD-4D-QSAR is that the resultant pharmacophore sites generated in the analysis are explicitly dependent upon the combined geometries of the ligand bound to the receptor, which considerably improves the overall quality of the models. In the broad sense, RD-4D-QSAR can be considered a SBDD method, since it depends on the target structure.⁵⁷ However, this story is not over; far from being exhausted, 4D-QSAR appears to be experiencing a revival in recent years, and an interesting review on this was recently presented.⁵⁸

Although the original proposal of QSAR was the development of statistically robust activity prediction models, they can also be applied to virtual screening

procedures, which will be discussed later. But before that, it is time to discuss the evolution of some of the most employed methods when the target 3D structure is known or can be modelled.

4. The Evolution of Target-Based Models in CADD

After a target associated to a disease is validated, the determination of its 3D structure is a necessary step if SBDD is the strategy to be employed. Undoubtedly, most of targets explored in drug design are proteins, and the main resource of experimental 3D protein structures has been the open access Protein Data Bank (PDB), established in 1971 at Brookhaven National Laboratory under the leadership of W. Hamilton and originally containing only 7 structures.⁵⁹ If we go back 30 years from now, the number of 3D structures deposited in the PDB in 1994 were less than 10% (1,289 structures) from what was deposited during the last year (13,585 structures); the total number of deposited structures jumped from 2,871 in 1994 to 213,221 in 2023.⁶⁰

This number alone show that the use of SBDD today has many more possibilities than 30 years ago, but the use of the SBDD strategy is not limited by the availability of structures deposited in the PDB. Protein 3D modeling by the methods traditionally used for small molecules is not a practical approach due to the complexity of the conformational space of these biomacromolecules, but there is a quite effective alternative technique for protein modeling: homology (comparative) modeling, which has evolved rapidly since the first studies. The year of 1993 can be considered a milestone in comparative protein modeling, since Peitsch and Jongeneel⁶¹ presented a paper where they described the procedure to build a 3D protein model that became the basis of the SWISS-MODEL server for comparative automated modeling of 3D protein structures.⁶² With this tool, 3D structures could be modeled just from their amino acid sequence, as long as a suitable 3D template with at least 30% identity with the target protein was also available.⁶² Amino acid sequences can be easily retrieved from databases such as the open access UniProtKB, which presently contains 248,805,733 entries (248,234,451 from TrEMBL and 571,282 from Swiss-Prot).⁶³

Coincidentally, in the same year of 1993, Sali and Blundell⁶⁴ presented MODELLER, a comparative protein modeling software designed to find the most probable conformation for a protein sequence given its alignment with available related 3D structures, which also became quite popular over the years.⁶⁴ Many other tools for homology modeling are now available and additional information, including the limitations of the method, can

be obtained in reviews such as the work from Muhammed and Aki-Yalcin.⁶⁵

Another major leap in the modeling of protein 3D structures occurred in 2021 with the emergence of AlphaFold, developed by DeepMind and EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute).⁶⁶ AlphaFold was the first machine learning (ML)-based computational method that could regularly predict 3D structures of proteins with atomic precision, even in cases where no similar structure to be used as a model was known. The AlphaFold Protein Structure Database now provides open access to over 200 million protein 3D structure predictions.⁶⁷

But having access to a quality 3D structure for the target protein, whether experimentally or through modeling methods, is just the beginning of an SBDD project - the next big challenge is defining how candidate ligands would bind to it. This is a very difficult task because of the number of degrees of freedom associated to the formation of a ligand/protein complex. The easiest way to generate such structures was, as a first approach, to simplify the problem by neglecting some of these degrees during the search of ligand/protein geometries, which is a common assumption in one of the most popular methods nowadays in SBDD projects: molecular docking.

Some kind of molecular docking, although not using this name, was already present in the literature of the 1970s. An early example is the work of Levinthal *et al.*⁶⁸ about a computerized molecular model that was used to deduce the arrangement of sickle cell hemoglobin molecules in tubular fibers. In the 1980s, some important improvements were introduced in the procedure, such as the use of “hard sphere” repulsion and hydrogen bonding terms to describe protein-ligand interactions,⁶⁹ and the word “docking” began to be used.⁷⁰ These interaction terms are examples of what is known in modern molecular docking protocols as scoring functions, which are used to quantify the interaction of the ligand-protein system; the aim of the scoring function is to classify the different interaction modes (or “poses”), allowing the ranking of the poses that were generated by some specific docking algorithm.⁷¹

Scoring functions are a continuously evolving aspect in molecular docking. If, on the one hand, speed in evaluating poses is important in high throughput docking, on the other hand the demands on the quality of the results are becoming increasingly greater, especially when trying to correlate them with bioactivity data. In fact, the main objective of scoring functions is the prediction of the ligand-protein interaction mode and this can be reasonably achieved by traditional force field-based, knowledge-based or empirical scoring functions, with the main scoring

functions leading to hit rates in redocking assays greater than 70%. However, many other factors that are generally not included in common scoring functions are important for binding affinity, such as desolvation.⁷²⁻⁷⁴ Therefore, it is not surprising that the search for correlations between docking scores obtained with the most commonly used scoring functions and binding affinity data often fails.

Developing scoring functions for specific target classes is a strategy that has proven to work better in some cases,^{75,76} but this strategy has obvious limitations. In the search of general use scoring functions capable of predicting bioactivity, ML methods are undoubtedly those with the most promising results, as demonstrated in the work of Guedes *et al.*⁷⁷ presenting a new ML-derived scoring function, DockTScore; multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) algorithms were used to derive scoring functions involving force-field terms, solvation and lipophilic interactions terms, and an improved term accounting for ligand torsional entropy contribution to ligand binding. DockTScore presented a Pearson's correlation coefficient of 0.705 for binding affinity prediction from diverse protein targets available on the PDBbind v2013 set,⁷⁸ being competitive with the best-evaluated scoring functions. It was only supplanted by BT-Score, an ensemble ML scoring function of boosted decision trees and thousands of predictive descriptors to estimate binding affinity,⁷⁹ which reproduced data of out-of-sample test complexes with a correlation coefficient of 0.825, and RF-Score:VinaElem,⁸⁰ with a correlation coefficient of 0.752. Affinity predictions with DockTScore can be made at the DockThor portal,⁸¹ a free protein-ligand docking server developed in Brazil at the National Laboratory for Scientific Computing (LNCC/MCTIC).

In addition to the scoring function, the other central element of a molecular docking method is the availability of an efficient algorithm for automated docking; one of the first examples was the incremental construction presented in the work of Rarey *et al.*;⁸² in this approach, the ligand is fragmented and the fragments are introduced step by step in the binding cavity until the ligand structure is complete. Another category that was explored was that of the stochastic methods, such as the genetic algorithm (GA), where the ligand structure is randomly introduced in the binding cavity. The binding modes (or poses), after translation to a kind of “genetic code” describing their main features, such as their torsion angles, are progressively changed by a series of genetic operations, such as mutation and reproduction (i.e., combination of “genetic codes”) between previous “parent” poses, leading to a succession of better generations of poses.^{83,84}

Evidently, considering Koshland's induced fit proposal,⁸⁵ it may seem that assuming a rigid structure for the target is too drastic an approach for molecular docking, especially when proteins with highly flexible binding sites are involved. As always is the case when working with models, some critical sense is necessary in the use of molecular docking; for example, trying to dock candidate ligands to a crystallographic structure of a ligand-free protein (apo conformation) is frequently a frustrating experience, since the binding site may be too constricted to adequately accommodate a ligand inside, so working with structures containing cocrystallized ligands is generally a clever choice.

On the other hand, although most of molecular docking programs adopt the rigid-body approach for the target structure, some allow a partial movement in its structure; an example of such approach is available in GOLD (Genetic Optimisation for Ligand Docking), a GA-based program launched in the 1990s, in which some amino acid side chains or even limited peptide backbone sections can be chosen to adopt different conformations during the docking run.^{84,86}

An interesting approach to simulate the effect of target flexibility during molecular docking is ensemble docking, which was introduced in 1999.⁸⁷ The idea is to implement docking runs on different conformations of the same target that, in the original paper, were obtained by MD, which was applied in order to define the conformation of a loop in human immunodeficiency virus (HIV) integrase that was ill-defined by crystallography. In this way, although the target remains rigid during the docking run, the effect of its different conformations on the definition of the ligand poses is at least partially included. The greater the number of MD snapshots included in the ensemble, the greater the simulation of target flexibility, but the computational cost for the complete ensemble docking will increase proportionally. A comprehensive review about the topic was presented by Amaro *et al.*⁸⁸ 2018. Naturally, another way for generating the ensemble is by selection of different target conformations obtained from its crystallographic structures containing different cocrystallized ligands.

Molecular docking algorithms were originally developed to identify the interaction mode of a ligand to a selected protein binding site. In some cases, however, the exact location of the binding site at which ligand interaction results in the observed bioactivity is unclear; some compounds may present their bioactivity as consequence of interactions with an allosteric binding site, for example. This situation presents a much greater challenge than docking at a predefined binding site and different methods have been developed to find candidate binding sites in an entire protein and perform docking runs on them.⁸⁹ This

process was defined as blind docking and can be executed by some programs or web servers: in QuickVina-W^{90,91} and SwissDock,^{92,93} the docking box covers all cavities found on the entire protein surface; in SITEHOUND-web, the location of potential binding sites is performed firstly and is followed by multiple independent docking runs on smaller boxes centered on the predicted binding sites;⁹⁴ in FRAG, a blind docking protocol based on Autodock-Vina,⁹⁵ binding site identification and pose prediction are accomplished at the same time by a systematic exploration of the protein volume performed with several preliminary docking calculations.⁹⁶

An entirely different approach that can be applied to the ligand-protein complex problem is MD, from which a trajectory described by the system during a specific time interval can be completely evaluated. MD calculations can be used for the direct evaluation of the binding affinity; by the principles of statistical mechanics, the Gibbs free energy change (ΔG) between two states can be obtained as the expectation value by integration over a representative fraction of the full conformational space of the system.⁹⁷ Different from molecular docking, the process to generate different conformations in MD is deterministic—each structure is generated from a previous one by applying classical physics equations to the molecular system over time.

In MD, the interaction and motion of atoms are described by Newton's physics: from an initial molecular geometry, the forces between interacting atoms can be obtained as derivatives of their energies, calculated by means of a classical force field, which then are used for the calculation of accelerations, velocities, and the resulting new positions of the atoms after a small timestep. This new molecular geometry presents a new energy, which, in turn, allows the process to be repeated; the result of this process is the definition of a trajectory described by the system during a sufficiently long time interval. Typically, the fastest events of biochemical interest take place on timescales of nanoseconds (10^{-9} s), but since the timesteps in a MD simulation must be in the order of femtoseconds (10^{-15} s) to assure numerical stability, at least millions of calculations must be done. Because kinetic energy is available, potential energy barriers along the molecular trajectory can be overcome.⁹⁸

However, solving this sophisticated set of equations is only part of the problem, since the calculations must be done for each atom of a system usually composed of protein, ligand, a surrounding box of solvent molecules, and sometimes a fragment of the cell membrane, in order to describe the real ligand-protein environment as best as possible. The large number of atoms, combined with the complexity of the calculations to describe the system's trajectory, results in high computational cost and time.

The foundations of MD were presented in the 1950s,⁹⁹ but its use in studying biochemical systems was only stimulated in the 1970s by the pioneering studies by Karplus and co-workers¹⁰⁰ and by Levitt and Warshel,¹⁰¹ which used MD simulations to obtain different conformations of proteins and nucleic acids. A milestone in the use of MD simulations in Brazil was undoubtedly the THOR program, developed in the late 1990s through the collaboration of different research groups located at UFRJ, Universidade Federal da Bahia (UFBA), Centro Brasileiro de Pesquisas Físicas (CBPF) and Universidade de São Paulo (USP).¹⁰² It inspired a generation of computationally oriented scientists in South America (indeed, the previously referred DockThor was named after THOR).

The evolution of computational power over the years have allowed a continuous increase in the size of molecular systems that could be evaluated. As soon as in the middle of the 2000s, all-atom MD calculations of a complete virus were achieved, encompassing up to 1 million atoms for over 50 ns.¹⁰³ This was a time when performing MD simulations required a supercomputer, but the evolution of computer hardware in the 2010s, particularly the use of graphics processing units (GPUs), allowed MD simulations on large systems to be run on relatively inexpensive machines with reasonable computation times.¹⁰⁴

In addition to hardware evolution, many MD-related theoretical achievements were produced over time, including enhanced sampling methods, focusing on free-energy perturbation, metadynamics, steered MD, and other methods most consistently used to study drug-target binding. A discussion about this theoretical evolution and the impact of MD on drug discovery can be found, for example, in the works of De Vivo *et al.*¹⁰⁵ and of Hollingsworth and Dror.¹⁰⁶ Powerful boosting methodologies to introduce quantum mechanics in MD calculations are being developed; an interesting example is multiscale simulations in computational chemistry (MiMiC), a framework designed to facilitate massively parallel multiscale MD. The main goal of a multiscale approach to MD is to alleviate the computational cost of high-level methods, including quantum mechanics approaches, while maintaining their flexibility and accuracy.¹⁰⁷

5. Virtual Screening

Virtual screening (VS) is undoubtedly one of the most explored topics in CADD in recent years. Many excellent reviews¹⁰⁸⁻¹¹⁴ on VS have been published and interested readers are encouraged to consult them to access more information. This story begins with a failure: in the final

1980s the pharmaceutical industry began to invest heavily on experimental high-throughput screening (HTS) and combinatorial chemistry in a tentative to overcome the lead discovery bottleneck at the time, but the results were far from expected, with significant costs resulting in only low hit rates.¹¹⁵

In the search for more efficient ways to discover lead compounds, computational alternatives began to be tested and, according to Klebe,¹¹⁴ the term “virtual screening” appeared for the first time in the late 1990s. However, as soon as 1982, there was some initial trials to identify ligands for the HIV protease, using as candidates some rigid entries from the Cambridge Crystallographic Database and an initial molecular docking method.⁶⁹ This work, although not employing the term “virtual screening”, contains the key characteristics of a VS study, since the main proposal in VS is to search chemical structures databases to find out lead candidates by computational means, in this case an initial version of molecular docking.⁶⁹

Since then, molecular docking remains as one of the most explored methods in VS studies. Its use in VS campaigns is part of what can be classified as target structure-based VS (SBVS), in contrast to ligand-based VS (LBVS) (Figure 2).¹⁰⁸ The first approach requires the availability of a 3D structure of a validated target, which can be obtained as discussed previously, to which the interaction of candidate ligands must be somehow assessed and quantified; molecular docking is an obvious choice for this purpose, mainly because of its fast execution, an important characteristic when databases composed of millions of structures are sometimes evaluated.

For this purpose, it is worth mentioning an interesting work of Taranto and co-workers¹¹⁷ focusing on the development of a rather complete workflow, called MolAr (Molecular Architect), which was designed to integrate diverse free programs to run many of the necessary steps to implement VS campaigns, interconnected with a friendly-user interface.¹¹⁸ MolAr includes protein preparation, combining comparative modeling with MODELLER,⁶⁴ and definition of protonation states with PROPKA,¹¹⁹⁻¹²¹ and a VS procedure through AutoDock Vina,⁹⁵ DOCK 6,^{122,123} or a consensus of both.

On the other hand, the techniques employed in the LBVS approach generally use molecules with known biological activity as patterns for the screening in virtual libraries of new chemical structures in the search of candidate structures that share some level of similarity with them.¹⁰⁸ According to the pattern generation procedure, the LBVS methods can be divided into three main types: similarity screening,¹²⁴ screening by a reference pharmacophore,¹²⁵ and QSAR-based approaches.¹²⁶

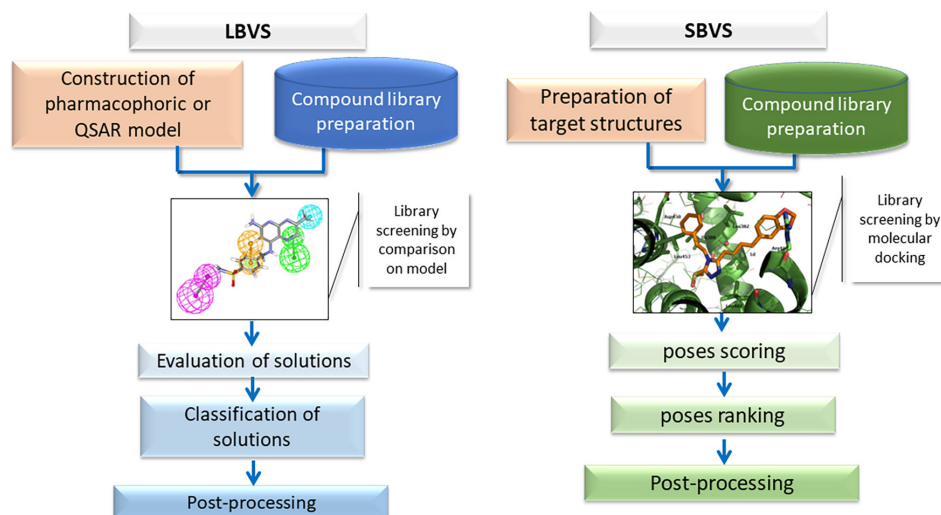


Figure 2. Comparison of the LBVS and SBVS approaches. Preparation of compound libraries may involve many steps, including filtering procedures (Lipinski's Rule of Five,¹¹⁶ ADME-Tox, etc.). In LBVS, the screening process may involve comparison with a pharmacophore map, a QSAR model or simply a similarity analysis (Tanimoto coefficient, etc.). Both procedures can be combined in the search of a consensus selection or even coupled, using selected compounds in a faster LBVS procedure as entries of the SBVS procedure.

LBVS and SBVS methods all depend on the availability of chemical databases. Nowadays there are many commercially and free databases available online.¹⁰⁸ Undoubtedly, one of the most extensive and explored databases is ZINC, which was launched in 2005.¹²⁷ The present version, ZINC20,¹²⁸ is a free database of commercially available compounds for VS, containing over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC maintains an overwhelming preference over the remaining large chemical databases, with an average use between 2015-2020 of 31%, a number that is more than double that of the second place.¹²⁹

There are some examples of Brazilian chemical databases that were developed in recent years, such as the Laboratory for the Evaluation and Synthesis of Bioactive Substances (LASSBio) Chemical Library,^{130,131} from UFRJ, whose compound collection has its major part released for public access, and the Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products Database (NuBBE_{DB}), from Universidade Estadual de São Paulo (UNESP), a free database of chemical and biological information from Brazilian biodiversity.¹³²

6. Machine Learning in CADD

As is the case in probably any area of science, machine learning (ML) is now being largely explored in Medicinal Chemistry. The main purpose of ML is to understand patterns in data and be able to predict these patterns.¹³³⁻¹³⁵ ML algorithms are able to “learn” by adjusting the distance between the original data and the predicted data, minimizing the error function (known as loss function) of

their predictions.¹³³ The precondition for the method to be applied is the hypothesis of the existence of a correlation between any intended characteristic of the real system (dependent variable, normally called labels or Y) and other intrinsic or extrinsic characteristics (independent variables, normally called as features or X).

These characteristics must be firstly measured and will consist of the data that will feed the method. Experimental data will be used to train the model, from where patterns can be learned, representing the training data, a collection of experimentally measured X and Y. If a well-established theoretical method is available, it is possible to use theoretical data as well. Generally, Y as experimental data is preferred as it is closer to the ground truth of the analyzed phenomena and better suited to model validation. The successfully trained model is then fed with test data, consisting of another set of X, but with a corresponding set of Y not available to the model, but known by the modeler.

If the assumptions of the model are satisfactory, the ML model is validated and ready to predict the learned relationships in systems where Y of interest is not experimentally measured. Most of the modeling process depends on the data treatment, allowing coherent inferences and, consequently, predictions. The better and accessible the data are, more robust and interpretable a ML model can be. This is especially important as the non-linearity and dimensionality grows, making interpretation of predictions less feasible, known as the “black-box effect” of any ML model.¹³⁶

ML algorithms that use labeled data (X and Y are discriminated) are quoted as supervised ML algorithms. For scenarios where data is not labeled, which means data

are not described as features nor labels from a system, for example in chemical similarity and clustering studies, ML algorithms are quoted as unsupervised. The complexity of biological and chemical systems frequently impels the use of hybrid models in a multi-process fashion, to classify, quantify, standardize, transform and inverse-transform information. That is the case of semi-supervised ML models, for example some ensemble models that use meta-learners trained on a series of predictions or deep learning (DL) schemes.¹³⁵

In the context of Medicinal Chemistry, data are known for their natural complexity, high variance, little or no standardization, high dimensionality and, commonly, high experimental cost, reflecting narrow population samples.^{137,138} In order for chemical data to be computed as data points for ML methods, the molecular information must be encoded. Simplified molecular input line entry simplification (SMILES),¹³⁹ 2D/3D molecular graphs,¹⁴⁰ and undirected graphs (UGs) representations¹⁴¹ were introduced as the standard encoding procedures to represent spatial aspects of chemical compounds as atoms, connectivity, ramifications, aromaticity and non-aromaticity, isomerism, fragments, and chirality. A number of CADD applications benefited from it as a practical and fast way of sampling an enormous portion of chemical space and describe it.^{134,136,137,141-143}

Although allusion to ML application in CADD is growing fast in the recent literature, it is not in fact a recent topic; a notorious example was published by Cramer *et al.*¹⁴⁴ as soon as 1974, although primitive and not entirely algorithmic as nowadays, as a result of the Hansch and Free-Wilson⁴⁴ linear regression methodology limitations with growing SAR complexity. In this work, a substructure-activity relationship prediction model was presented as a novel approach to drug design, an excellent example of chemical data management. A set of 850 compounds, examined for their antiarthritic immunoregulatory effects in a rat-model, was used and further pruned for bias reduction to 770 compounds.¹⁴⁴

Since this scientific milestone, a massive development in the field of computational and cheminformatic sciences arose and new algorithms were built to address every kind of data. Some are artificial neural networks (ANNs) variations (radial basis function (RBF), Kohonen's self-organizing maps, DL, convoluted neural networks, recurrent neural networks), ensemble methods (gradient boosting, random forest and XGBoost as well meta-learners that uses blending and stacking methodology), SVM with a variety of kernels (a class of algorithms for pattern analysis) for example, linear, polynomial and Tanimoto kernels.^{133-136,143,145} These methods can be applied by programming language packages,

such as R¹⁴⁶ and Python¹⁴⁷ with practical implementation, for example by Google Colaboratory.¹⁴⁸

The use of ML in CADD projects increased in the 2000s, and a few illustrations will be presented here. In 2001, a quantitative and classification screening tool for large chemical libraries ANNs and the heuristics of k-nearest neighbors (k-NNs) were presented; the focus was the screening of cyclooxygenase-2 (COX-2) inhibitors. Using experimental IC₅₀ data and feature selection, the model was capable to discriminate between active and inactive compounds with 83.3% of accuracy; its methodological superiority compared to multiple linear regression models (MLR) was proposed in the case of noisy and non-linear SAR.¹⁴⁹

Another example was a QSAR model, using ANNs based on RBF with GA for prediction of glycine/NMDA receptor antagonist inhibition.¹⁵⁰ Doniger *et al.*¹⁵¹ compared the performance of SVMs with multi-layer perceptron (MLP) neural networks to predict the blood-brain barrier permeability of different classes of molecules, to develop a method to predict the ability of drug compounds to penetrate the central nervous system. SVMs showed an average 81.5% accuracy against 75.5% of MLP over 30 test sets.

RF algorithm was introduced later as a powerful classification and quantification tool for QSAR modeling, capable of achieving high prediction accuracies, robust enough to handle high-dimensionality, introducing auto-feature selection. This algorithm is from a class of ensemble methods and went well compared to ANNs and SVM models.¹⁵²

A new classification metamodel (a model that consists of statements about models) for predicting protein secondary structures was published in 2006, based on two cascaded models, each with an ensemble of three SVM binary classifiers, employing one-versus-rest learning approach. The first model was trained on amino acid sequences to assign weights to the three-state protein structure (helix, sheet and coil) from 3 different public datasets. The second model was trained on the previous predictions to further classify the secondary structure, and an accuracy of 79.34% was achieved.¹⁵³

As the 2010 decade came to an end, excellent reviews were published, highlighting the plethora of ML applications in various subfields of Medicinal Chemistry at that time.¹⁵⁴ SVM and RF methods became the dominant methods, giving excellent performances and generalization capacities over shallow neural networks models, and versatility to organize chemical information with diverse kernels to suit the research peculiarities. They continue to appear in recent publications, with modern methodologies,

which is remarkable for methods from the pre-deep learning (DL) era.¹⁵⁵ Cascaded models are good examples of the transition from shallow ML models to DL models.¹⁵³

In 2013, Lusci *et al.*¹⁴¹ presented a DL regression model for water solubility prediction as a good example of how the length of datasets can be relevant for DL methods. Undirect graph-recursive neural networks (UG-RNNs) were built utilizing public benchmark datasets used for solubility predictions in literature, two containing thousands of molecules, and the rest with less than 200 molecules. Undirect graph encoding of molecules is a way of describing chemical information representing an atom or arbitrary molecule fragment as interconnected nodes (vectors) and directing them towards one specific node (root). For the two bigger datasets, UG-RNN topped the coefficients of determination R^2 (values ranging 0.90 to 0.91) showing an outstanding performance. However, when comes to narrow datasets, the predictive performance was not consistent enough. Insights were giving to address the performances like the noisy and fewer observations from where no generalization could be captured against overfit.¹⁴¹

Xu *et al.*,¹³⁶ in 2015, developed a deep neural network (DNN) model to predict drug-induced liver's injury (DILI), an unwished collateral effect associated with many drugs in literature. DILI is not easy of detection by *in vivo* protocols, encouraging *in silico* models to take the task. The model was based on UG-RNN structure commented above. All models outperformed or performed closer (accuracy scores ranging from 0.60 to 0.76) to original classification models from literature, associated with some of public datasets. The best model was trained and tested with the combination of two datasets, with 86% of classes correctly predicted, proving again how sample population can positively impact the performance of the model. To readers interested in details of how data management can impact performance of DL models in multiple levels, this work is highly suggested.¹³⁶

Synthesis is a fundamental topic in Medicinal Chemistry and finding ways to improve it is a desire of every synthetic chemist. The main idea of retrosynthesis is to decompose the chemical product from an unknown reaction, by formal organic rules and expertise, into its synthesis path. In 2017, a retrosynthetic encoder-decoder DL model was developed to predict the likely reactants to react in a specific reaction type to form products; in the model, chemical reactions are represented as sequence-to-sequence data, and two concatenated RNNs were constructed, based on SMILES representation of molecules and a database of 50,000 reaction experiments from USA patent literature, categorized in 10 different classes of reactions. It was shown that the model performed better

than rule-based expert models and ML models that use rule-based expert methodology, even surpassing some of its known limitations giving a notably upper hand to challenges within retrosynthesis computational analysis.¹⁴³

Another state-of-art model for retrosynthesis was built by Segler *et al.*¹⁵⁶ in 2018, based on DNNs and Monte Carlo tree; in quotation to authors, "These deep neural networks were trained on essentially all reactions ever published in organic chemistry".¹⁵⁶

From a different perspective, chemical data mining is becoming increasingly important in the Medicinal Chemistry field. This is because the need to amplify, standardize and filter chemical information from all kinds of scientific documents, considering the endless applications inside the drug development scope. In 2019, Staker *et al.*¹⁵⁷ presented a paper stating another cascaded DL model, with capabilities to extract chemical information from PDF inputs, capitalizing from the pages containing 2D standard representations of molecules their SMILES codes. For three datasets used to test the model, the accuracy of SMILES generation ranged from 0.77 to 0.83.

As a final example, let us remember that not distantly the world faced the coronavirus virus disease 2019 (COVID-19) pandemics, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) coronavirus. In front of that reality, Wang *et al.*¹⁵⁸ built a DNNs classification model to screen for inhibitors of the SARS-CoV-2 chymotrypsin-like protease (3CL^{pro}), responsible for the replication and transcription of the virus. Many inhibitors are known to covalently bind to a 3CL^{pro} active site cysteine residue (Cys145). The main assumption was a sequence similarity up to 96% between SARS-CoV 3CL^{pro} and SARS-CoV-2 3CL^{pro}. Two DNN classifiers were trained separately on two sets of know inhibitors, to take acquaintance of covalent and non-covalent bindings from literature and performed well on test sets. Then, these classifiers were used to screen a library containing 39,000 compounds; the screened compounds were experimentally tested, and from 32 top-ranked tested compounds, 6 showed low micromolar range IC_{50} activity, 1.4 μ M, at best.

For more complete information compiling the advances of ML applications on Medicinal Chemistry, some excellent reviews on the topic are indicated to the interested reader.^{135,138,159-161}

7. Conclusion

Although the theoretical basis for many of the methods used in molecular modeling had been laid decades earlier, it was in the 1980s that they began to be explored in

CADD by the pharmaceutical industry, probably because of the increasing of computational power of machines at a progressively accessible cost observed at that time. Since then, its influence in supporting the design of drug candidates has increased and today can be considered consolidated, as can be seen by the number of bioactive compounds whose design benefited from one or more of the CADD methods and which effectively became marketed drugs. In Brazil, CADD arrived in the 1990s through the pioneering work of some academic research groups and, since then, it has become an integral part of several projects that have resulted in the training of postgraduate students and researchers and in the publication of articles and patent filings for drug candidates.

The methodologies in the SBDD and LBDD areas that make up CADD have evolved in several ways from those early years to the present day (Figure 3). Many methods emerged, became dominant for some time, and were then replaced by others as the main interest in the search for drug candidates.

Empirical field and quantum mechanical-based models remain the workhorses behind generating ligand structures necessary for developing CADD projects. Multidimensional QSAR, although not so present as before, continues to be an important topic in CADD, especially when considering its application as a VS technique, where it now shares space with other LBDD approaches, such as pharmacophore models. VS remains at the forefront of CADD efforts, where molecular docking is the main SBDD method in use today. The future here is on the improvement of binding affinity prediction, a limitation for older linear fitness score functions. ML-based functions are proving the best way to follow in this aspect, although the enhancement in the physical representation of the binding process, including MD and even the partial inclusion of quantum mechanics effects, such as in QM/

MM/MD, is an exciting emerging frontier in CADD. The computational cost is high, but the development of new programs, capable of extract the maximum performance of massively parallel multiscale MD, may soon improve the access of such techniques to the CADD arsenal.

But it is unquestionable, as chemical nonlinearity began to be addressed, confronted and dealt with, that several state-of-the-art studies have expanded the applications of ML in CADD and other areas of Medicinal Chemistry. While more traditional methods will continue to play their role in CADD for many years, a new chapter in the history of CADD is now being written with ML methods covering different subjects such as 3D protein structure prediction, generation of LBDD models, binding affinity prediction for SBDD models, ADME-Tox predictions, chemical data mining, and many others involved in the development of new drugs. Let us look forward to the promising next 30 years to come.

Acknowledgments

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Pesquisa-CNPq (grant 315948/2021-3), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (Faperj) and INCT-INOFAR.



Vinicius Nunes da Rocha has a degree in Industrial Chemistry (UFRRJ, 2018) and a Master's degree in Theoretical Chemistry (UFRRJ, 2021). He is now developing a PhD thesis in Chemistry at UFRRJ, in the Medicinal Chemistry area. He

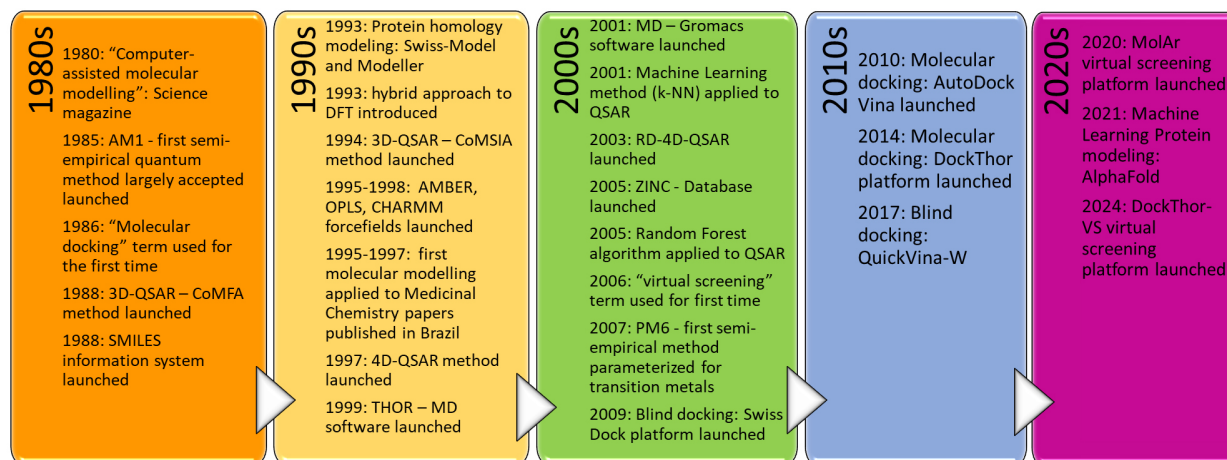


Figure 3. Timeline describing the evolution of molecular modelling methods applied to Medicinal Chemistry projects. This timeline is not intended to be exhaustive, since only key moments that were discussed in the text are highlighted.

has experience in the field of chemistry, with emphasis on computational chemical kinetics and in silico drug design. Currently working on drug development for the treatment of bipolar mood disorder, through molecular modeling combined with machine learning and data analysis.



Carlos Mauricio R. Sant'Anna has a degree in Industrial Chemistry (UFF, 1986), a Master's degree in Polymer Science and Technology (UFRJ, 1990) and a PhD in Organic Chemistry (UFRJ, 1997). He joined as a Professor at the Federal Rural University of Rio de Janeiro in 1998,

where he has been Full Professor since 2015. He has experience in chemistry, with an emphasis on molecular modeling, working mainly in the design and development of bioactive compounds, having published more than 100 articles involving these themes. He has been a collaborating Professor at the Laboratory for the Evaluation and Synthesis of Bioactive Substances (LASSBio-UFRJ) since 1997.

References

- Watson, J.; Crick, F.; *Nature* **1953**, *171*, 737. [Crossref]
- Fischer, E.; *Ber. Ges. Dtsch. Chem.* **1894**, *27*, 2985. [Crossref]
- do Amaral, A. T.; Andrade, C. H.; Kümmerle, A. E.; Guido, R. V. C.; *Quim. Nova* **2017**, *40*, 694. [Crossref]
- Van Drie, J. H.; *J. Comput.-Aided Mol. Des.* **2007**, *21*, 591. [Crossref]
- Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, G. M.; *Science* **1980**, *208*, 1425. [Crossref]
- Sabe, V. T.; Ntombela, T.; Jhamba, L. A.; Maguire, G. E. M.; Govender, T.; Naicker, T.; Kruger, H. G.; *Eur. J. Med. Chem.* **2021**, *224*, 113705. [Crossref]
- Albuquerque, M. G.; Rodrigues, C. R.; Alencastro, R. B.; Barreiro, E. B.; *Int. J. Quantum Chem., Quantum Biol. Symp.* **1995**, *22*, 181. [Crossref]
- Barreiro, E. J.; Rodrigues, C. R.; Albuquerque, M. G.; de Sant'Anna, C. M. R.; de Alencastro, R. B.; *Quim. Nova* **1997**, *20*, 694. [Crossref]
- de Sant'Anna, C. M. R.; de Alencastro, R. B.; Barreiro, E. J.; Fraga, C. A. M.; *J. Mol. Struct.* **1995**, *340*, 193. [Crossref]
- De Sant'Anna, C. M. R.; de Alencastro, R. B.; Fraga, C. A. M.; Barreiro, E. J.; Motta Neto, J. D.; *Int. J. Quantum Chem.* **1996**, *60*, 1069. [Crossref]
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A.; *J. Am. Chem. Soc.* **1995**, *117*, 5179. [Crossref]
- Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J.; *J. Am. Chem. Soc.* **1996**, *118*, 11225. [Crossref]
- MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, Jr. R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M.; *J. Phys. Chem. B* **1998**, *102*, 3586. [Crossref]
- Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W.; *J. Comput. Chem.* **2004**, *25*, 1656. [Crossref]
- He, X.; Walker, B.; Man, V. H.; Ren, P.; Wang, J.; *Curr. Opin. Struct. Biol.* **2022**, *72*, 187. [Crossref]
- van Mourik, T.; Bühl, M.; Gageot, M. P.; *Philos. Trans. R. Soc., A* **2014**, *372*, 20120488. [Crossref]
- Verma, P.; Truhlar, G. G.; *Trends Chem.* **2020**, *2*, 302. [Crossref]
- Burke, K.; Wagner, L. O.; *Int. J. Quant. Chem.* **2014**, *113*, 96. [Crossref]
- Kohn, W.; Sham, L.; *J. Phys. Rev.* **1965**, *140*, A1133. [Crossref]
- Politzer, P.; Abu-Awwad, F.; *Theor. Chem. Acc.* **1998**, *99*, 83. [Crossref]
- Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J.; *J. Phys. Chem.* **1994**, *98*, 11623. [Crossref]
- Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P.; *J. Chem. Theory Comput.* **2018**, *14*, 2596. [Crossref]
- Maurer, L. R.; Bursch, M.; Grimme, S.; Hansen, A.; *J. Chem. Theory Comput.* **2021**, *17*, 6134. [Crossref]
- Grimme, S.; *J. Chem. Phys.* **2006**, *124*, 034108. [Crossref]
- Purvis, G. D. III; Bartlett, R. J.; *J. Chem. Phys.* **1982**, *76*, 1910. [Crossref]
- Pople, J. A.; Seeger, R.; Krishnan, R.; *Int. J. Quantum Chem.* **1977**, *12*, 149. [Crossref]
- Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P.; *J. Am. Chem. Soc.* **1985**, *107*, 3902. [Crossref]
- Stewart, J. J. P.; *J. Mol. Model.* **2007**, *13*, 1173. [Crossref]
- Stewart, J. J. P.; *J. Mol. Model.* **2013**, *19*, 1. [Crossref]
- Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P.; *J. Chem. Theory Comput.* **2009**, *5*, 1749. [Crossref]
- Korth, M.; Pitonák, M. J.; Řezáč, M. J.; Hobza, P. A.; *J. Chem. Theory Comp.* **2010**, *6*, 344. [Crossref]
- Řezáč, J.; Hobza, P.; *Chem. Phys. Letters* **2011**, *506*, 286. [Crossref]
- Stewart, J. J. P.; Stewart, A. C.; *J. Mol. Model.* **2023**, *29*, 9. [Crossref]
- Minenkov, Y.; Sharapa, D. I.; Cavallo, L.; *J. Chem. Theory Comput.* **2018**, *14*, 3428. [Crossref]
- Warshel, A.; Levitt, M.; *J. Mol. Biol.* **1976**, *103*, 227. [Crossref]
- Kollar, J.; Freceer, V.; *J. Mol. Model.* **2018**, *24*, 11. [Crossref]
- Kar, R. K.; *Drug Discovery Today* **2023**, *28*, 103374. [Crossref]
- Tzeliou, C. E.; Mermigki, M. A.; Tzeli, D.; *Molecules* **2022**, *27*, 2660. [Crossref]

39. Hammett, L. P.; *J. Am. Chem. Soc.* **1937**, *59*, 96. [Crossref]
40. Hammett, L. P.; *Chem. Rev.* **1935**, *17*, 125. [Crossref]
41. Taft, R. W.; *J. Am. Chem. Soc.* **1952**, *74*, 3120. [Crossref]
42. Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M.; *Nature* **1962**, *194*, 178. [Crossref]
43. Free Jr., S. M.; Wilson, J. W.; *J. Med. Chem.* **1964**, *7*, 395. [Crossref]
44. Hansch, C.; Fujita, T.; *J. Am. Chem. Soc.* **1964**, *86*, 1616. [Crossref]
45. Hansch, C.; *Acc. Chem. Res.* **1969**, *2*, 232. [Crossref]
46. Verma, J.; Khedkar, V. M.; Coutinho, E. C.; *Curr. Top. Med. Chem.* **2010**, *10*, 95. [Crossref]
47. Hopfinger, A. J.; Tokarski, J. S.; *Three-Dimensional Quantitative Structure-Activity Relationship Analysis. In: Practical Application of Computer-Aided Drug Design*; Charifson, P. S., ed.; Marcel Dekker, Inc.: New York, USA, 1997, p. 105-164.
48. Martin, Y. C.; *3D QSAR: Current State, Scope, and Limitations. In: 3D QSAR in Drug Design - Recent Advances*, vol. 3; Kubinyi, H.; Folkers, G.; Martin, Y. C., eds.; Kluwer Academic Publishers: New York, USA, 1998, p. 3-23.
49. Cramer, R. D.; Patterson, D. E.; Bunce, J. D.; *J. Am. Chem. Soc.* **1988**, *110*, 5959. [Crossref]
50. Wold, S.; Ruhe, A.; Wold, H.; Dunn III, W. J.; *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 135. [Crossref]
51. Klebe, G.; Abraham, U.; Mietzner, T.; *J. Med. Chem.* **1994**, *37*, 4130. [Crossref]
52. Hopfinger, A.; Wang, S.; Tokarski, J.; Jin, B.; Albuquerque, M.; Madhav, P.; Duraiswami, C.; *J. Am. Chem. Soc.* **1997**, *119*, 10509. [Crossref]
53. Albuquerque, M. G.; Hopfinger, A. J.; Barreiro, E. J.; de Alencastro, R. B.; *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 925. [Crossref]
54. Martins, J. P. A.; Barbosa, E. G.; Pasqualoto, K. F. M.; Ferreira, M. M. C.; *J. Chem. Inf. Model.* **2009**, *49*, 1428. [Crossref]
55. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R.; *GROMACS*, version 1.0; The University of Groningen, NL, 1995.
56. Lindahl, E.; Hess, B.; van der Spoel, D.; *J. Mol. Model.* **2001**, *7*, 306. [Crossref]
57. Pan, D.; Tseng, Y.; Hopfinger, A. J.; *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1591. [Crossref]
58. Bak, A.; *Int. J. Mol. Sci.* **2021**, *22*, 5212. [Crossref]
59. PDB History, <https://www.rcsb.org/pages/about-us/history>, accessed in May 2024.
60. PDB Statistics, <https://www.rcsb.org/stats>, accessed in May 2024.
61. Peitsch, M. C.; Jongeneel, C. V.; *Int. Immunol.* **1993**, *5*, 233. [Crossref]
62. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C.; *Nucleic Acids Res.* **2003**, *31*, 3381. [Crossref]
63. UniProtKB, <https://www.uniprot.org/>, accessed in May 2024.
64. Sali, A.; Blundell, T. L.; *J. Mol. Biol.* **1993**, *234*, 779. [Crossref]
65. Muhammed, M. T.; Aki-Yalcin, E.; *Chem. Biol. Drug Des.* **2019**, *93*, 12. [Crossref]
66. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Michalina Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; *Nature* **2021**, *596*, 583. [Crossref]
67. The AlphaFold Protein Structure Database, <https://alphafold.ebi.ac.uk/>, accessed in May 2024.
68. Levinthal, C.; Wodak, S. J.; Kahn, P.; Dadivanian, A. K.; *Proc. Natl. Acad. Sci. USA.* **1975**, *72*, 1330. [Crossref]
69. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E.; *J. Mol. Biol.* **1982**, *161*, 269. [Crossref]
70. Desjarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R.; *J. Med. Chem.* **1986**, *29*, 2149. [Crossref]
71. Pagadala, N. S.; Syed, K.; Tuszyński, J.; *Biophys. Rev.* **2017**, *9*, 91. [Crossref]
72. Guedes, I. A.; Pereira, F. S. S.; Dardenne, L. E.; *Front. Pharmacol.* **2018**, *9*, 1089. [Crossref]
73. Huang, S.-Y.; Zou, X.; *J. Chem. Inf. Model.* **2010**, *50*, 262. [Crossref]
74. Kar, P.; Lipowsky, R.; Knecht, V.; *J. Phys. Chem. B* **2013**, *117*, 5793. [Crossref]
75. Seifert, M. H. J.; *J. Comput. Aided Mol. Des.* **2009**, *23*, 633. [Crossref]
76. Politi, R.; Convertino, M.; Popov, K.; Dokholyan, N. V.; Tropsha, A.; *J. Chem. Inf. Model.* **2016**, *56*, 1032. [Crossref]
77. Guedes, I. A.; Barreto, A. M. S.; Marinho, D.; Krempser, E.; Kuenemann, M. A.; Sperandio, O.; Dardenne, L. E.; Miteva, M. A.; *Sci. Rep.* **2021**, *11*, 3198. [Crossref]
78. Li, Y.; Liu, Z. H.; Li, J.; Han, L.; Liu, J.; Zhao, Z. X.; Wang, R. X.; *J. Chem. Inf. Model.* **2014**, *54*, 1700. [Crossref]
79. Ashtawy, H. M.; Mahapatra, N. R.; *J. Chem. Inf. Model.* **2018**, *58*, 119. [Crossref]
80. Li, H.; Leung, K.-S.; Wong, M.-H.; Ballester, P.; *Molecules* **2015**, *20*, 10947. [Crossref]
81. DockThor, <https://dockthor.incc.br/v2/>, accessed in May 2024.
82. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G.; *J. Mol. Biol.* **1996**, *261*, 470. [Crossref]
83. Jones, G.; Willett, P.; Glen, R. C.; *J. Mol. Biol.* **1995**, *245*, 43. [Crossref]
84. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R.; *J. Mol. Biol.* **1997**, *267*, 727. [Crossref]
85. Koshland Jr., D. E.; *J. Cell Comp. Physiol.* **1959**, *54*, 245. [Crossref]

86. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R.; *GOLD*, Cambridge Crystallographic Data Centre, UK, 1995.
87. Carlson, H. A.; Masukawa, K. M.; McCammon, J. A.; *J. Phys. Chem. A* **1999**, *103*, 10213. [Crossref]
88. Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C.; *Biophys. J.* **2018**, *114*, 2271. [Crossref]
89. Ghersi, D.; Sanchez, R.; *Proteins* **2009**, *74*, 417. [Crossref]
90. Hassan, N. M.; Alhossary, A. A.; Mu, Y.; Kwok, C. K.; *QuickVina-W*, Nanyang Technological University School of Computer Engineering, Singapore, 2015.
91. Hassan, N. M.; Alhossary, A. A.; Mu, Y.; Kwok, C. K.; *Sci. Rep.* **2017**, *7*, 15451. [Crossref]
92. Grosdidier, A.; Zoete, V.; Michielin, O.; *SwissDock*; Swiss Institute of Bioinformatics, Switzerland, 2011.
93. Grosdidier, A.; Zoete, V.; Michielin, O.; *Nucleic Acids Res.* **2011**, *39*, W270. [Crossref]
94. Hernandez, M.; Ghersi, D.; Sanchez, R.; *Nucleic Acids Res.* **2009**, *37*, W413. [Crossref]
95. Trott, O.; Olson, A. J.; *J. Comp. Chem.* **2010**, *31*, 455. [Crossref]
96. Grasso, G.; Di Gregorio, A.; Mavkov, B.; Piga, D.; Labate, G. F. D.; Danani, A.; Deriu, M. A.; *J. Biomol. Struct. Dyn.* **2022**, *40*, 13472. [Crossref]
97. Coveney, P. V.; Wan, S.; *Phys. Chem. Chem. Phys.* **2016**, *18*, 30236. [Crossref]
98. Frenkel, D.; Smit, B.; *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press, Inc.: San Diego, CA, 2001.
99. Alder, B. J.; Wainwright, T. E.; *J. Chem. Phys.* **1957**, *27*, 1208. [Crossref]
100. McCammon, J.; Gelin, B.; Karplus, M.; *Nature* **1977**, *267*, 585. [Crossref]
101. Levitt, M.; Warshel, A.; *Nature* **1975**, *253*, 694. [Crossref]
102. Moret, M. A.; Pascutti, P. G.; Bisch, P. M.; Mundim, K. C.; *J. Comput. Chem.* **1998**, *19*, 647. [Crossref]
103. Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K.; *Structure* **2006**, *14*, 437. [Crossref]
104. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C.; *J. Chem. Theory Comput.* **2013**, *9*, 3878. [Crossref]
105. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A.; *J. Med. Chem.* **2016**, *59*, 4035. [Crossref]
106. Hollingsworth, S. A.; Dror, R. O.; *Neuron* **2018**, *99*, 1129. [Crossref]
107. Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Bircher, M. P.; Ippoliti, E.; Carloni, P.; Rothlisberger, U.; *J. Chem. Theory Comput.* **2019**, *15*, 5601. [Crossref]
108. Rocha, S. F. L. S.; Olanda, C. G.; Fokoue, H. H.; Sant'Anna, C. M. R.; *Curr. Top. Med. Chem.* **2019**, *19*, 1751. [Crossref]
109. Fradera, X.; Babaoğlu, K.; *Curr. Protoc. Chem. Biol.* **2017**, *9*, 196. [Crossref]
110. Irwin, J. J.; Shoichet, B. K.; *J. Med. Chem.* **2016**, *59*, 4103. [Crossref]
111. Ferreira, L. G.; Santos, R. N.; Oliva, G.; Andricopulo, A. D.; *Molecules* **2015**, *20*, 13384. [Crossref]
112. Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z.; *Curr. Top. Med. Chem.* **2014**, *14*, 1923. [Crossref]
113. Seifert, M. H. J.; Lang, M.; *Mini Rev. Med. Chem.* **2008**, *8*, 63. [Crossref]
114. Klebe, G.; *Drug Discovery Today* **2006**, *11*, 580. [Crossref]
115. Lahana, R.; *Drug Discovery Today* **1999**, *4*, 447. [Crossref]
116. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.; *Adv. Drug Delivery Rev.* **1997**, *23*, 3. [Crossref]
117. Maia, E. H. B.; Medaglia, L. R.; Silva, A. M.; Taranto, A. G.; *MolAr*; Laboratory of Bioinformatics and Drug Design, UFSJ, Brazil, 2020.
118. Maia, E. H. B.; Medaglia, L. R.; Silva, A. M.; Taranto, A. G.; *ACS Omega* **2020**, *5*, 6628. [Crossref]
119. PropKa Online, <https://www.ddl.unimi.it/vegaol/propka.htm>, accessed in May 2024.
120. Li, H.; Robertson, A. D.; Jensen, J. H.; *Proteins* **2005**, *61*, 704. [Crossref]
121. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A.; *Nucleic Acids Res.* **2007**, *35*, W522. [Crossref]
122. Lang, P. T.; Moustakas, D.; Brozell, S.; Carrascal, N.; Mukherjee, S.; Prentis, L.; Singleton, C.; Zhou, Y.; Fochtman, B.; Balias, T.; McGee Jr., T. D.; Allen, W. J.; Bickel, J.; Matos, G. D. R.; Pak, S.; Corbo, C.; Boysan, B.; Holden, P.; Pegg, S.; Raha, K.; Shivakumar, D.; Rizzo, R.; Case, D.; Shoichet, B.; Kuntz, I.; *DOCK6*, version 6.0; University of California, USA, 2009.
123. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D.; *RNA* **2009**, *15*, 1219. [Crossref]
124. Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G.; *Methods* **2015**, *71*, 58. [Crossref]
125. Chen, Z.; Tian, G.; Wang, Z.; Jiang, H.; Shen, J.; Zhu, W.; *J. Chem. Inf. Model.* **2010**, *50*, 615. [Crossref]
126. Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H.; *Front. Pharmacol.* **2018**, *9*, 1275. [Crossref]
127. Irwin, J. J.; Shoichet, B. K.; *J. Chem. Inf. Model.* **2005**, *45*, 177. [Crossref]
128. ZINC20, <https://zinc.docking.org/>, accessed in May 2024.
129. Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A.; *J. Chem. Inf. Model.* **2020**, *60*, 6065. [Crossref]
130. LASSBio Chemical Library, <https://lassbiochemicallib.wixsite.com/home>, accessed in May 2024.
131. Colodette, N. M.; Franco, L. S.; Maia, R. C.; Fokoue, H. H.; Sant'Anna, C. M. R.; Barreiro, E. J.; *J. Comput. Aided Mol. Des.* **2020**, *34*, 1091. [Crossref]

132. Pilon, A. C.; Valli, M.; Dametto, A. C.; Pinto, M. E. F.; Freire, R. T.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S.; *Sci. Rep.* **2017**, *7*, 7215. [Crossref]
133. Sakiyama, Y.; *Expert Opin. Drug Metab. Toxicol.* **2009**, *5*, 149. [Crossref]
134. Gawehn, E.; Hiss, J. A.; Schneider, G.; *Mol. Inform.* **2015**, *35*, 3. [Crossref]
135. Priya, S.; Tripathi, G.; Singh Bukhsh, D.; Jain, P.; Kumar, A.; *Chem. Biol. Drug Design* **2022**, *100*, 136. [Crossref]
136. Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L.; *J. Chem. Inf. Model.* **2015**, *55*, 2085. [Crossref]
137. Panteleev, J.; Gao, H.; Jia, L.; *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2807. [Crossref]
138. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Zhao, S.; *Nat. Rev. Drug Discovery* **2019**, *18*, 463. [Crossref]
139. Weininger, D.; *J. Chem. Inf. Model.* **1988**, *28*, 31. [Crossref]
140. Bonchev, D.; Rouvray, D., H.; *Chemical Graph Theory: Introduction and Fundamentals*, vol. 1, 1sted.; Taylor & Francis: Oxfordshire, UK, 1991.
141. Lusci, A.; Pollastri, G.; Baldi, P.; *J. Chem. Inf. Model.* **2013**, *53*, 1563. [Crossref]
142. Swamidass, S. J.; Chen, J.; Bruand, J.; Phung, P.; Ralaivola, L.; Baldi, P.; *Bioinformatics* **2005**, *21*, i359. [Crossref]
143. Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V.; *ACS Central Science* **2017**, *3*, 1103. [Crossref]
144. Cramer, R. D.; Redl, G.; Berkoff, C. E.; *J. Med. Chem.* **1974**, *17*, 533. [Crossref]
145. Lavecchia, A.; *Drug Discovery Today* **2015**, *20*, 318. [Crossref]
146. R Core Team; *R, R Foundation for Statistical Computing*, AUST, 2021.
147. van Rossum, G.; *Python tutorial, Technical Report CS-R9526*; Centrum voor Wiskunde en Informatica (CWI), NL, 1995.
148. Google Colab, <https://colab.google>, accessed in May 2024.
149. Kauffman, G. W.; Jurs, P. C.; *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 1553. [Crossref]
150. Patankar, S. J.; Jurs, P. C.; *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1053. [Crossref]
151. Doniger, S.; Hofmann, T.; Yeh, J.; *J. Comp. Biol.* **2002**, *9*, 849. [Crossref]
152. Svetnik, V.; Wang, T.; Tong, C.; Liaw, A.; Sheridan, R. P.; Song, Q.; *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 786. [Crossref]
153. Karypis, G.; *Proteins* **2006**, *64*, 575. [Crossref]
154. Wale, N.; *Drug Dev. Res.* **2011**, *72*, 112. [Crossref]
155. Rodrigues-Pérez, R.; Bajorath, J.; *J. Comput. Aided Mol. Des.* **2022**, *36*, 355. [Crossref]
156. Segler, M. H. S.; Preuss, M.; Waller, M. P.; *Nature* **2018**, *555*, 604. [Crossref]
157. Staker, J.; Marshall, K.; Abel, R.; McQuaw, C. M.; *J. Chem. Inf. Model.* **2019**, *59*, 1017. [Crossref]
158. Wang, L.; Yu, Z.; Wang, S.; Guo, Z.; Sun, Qi, S.; Lai, L.; *Eur. J. Med. Chem.* **2022**, *244*, 114803. [Crossref]
159. Dara, S.; Dhamecherla, S.; Jadav, S. S.; Babu, C. M., Ahsan, M. J.; *Artif. Intell. Rev.* **2022**, *55*, 1947. [Crossref]
160. Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V.; *J. Chem. Inf. Model.* **2015**, *55*, 263. [Crossref]
161. Di Lascio, E.; Gerebtzoff, G.; Rodríguez-Pérez, R.; *Mol. Pharmaceutics* **2023**, *20*, 1758. [Crossref]

Submitted: February 6, 2024

Published online: June 14, 2024