

Evolution of computational chemistry applied to virtual drug screening: a mini-review

Maria Emanuelle de Oliveira Queirós, Alanna Coutinho Brasileiro Vaz, Lucas Henrique de Oliveira Rodrigues, Maria Eduarda Pinto Cabral, Alicia Giovanna Rodrigues Pereira, Bruna Leitao Matias Ribeiro, Rhayssa Karla Cavalcante de Sena, Beatriz Tamiris da Silva André, Eloisa de Oliveira Cavalcanti, Mariana Alves da Silva, Otávio Henrique do Vale Rodrigues, Arthur Van Lauter Albuquerque Pereira, Diego Santa Clara Marques +, Iranildo José da Cruz Filho, Maria do Carmo Alves de Lima

Universidade Federal de Pernambuco

Abstract. Computational chemistry has become a strategic component of drug design by enabling the evaluation of large chemical libraries, the investigation of molecular interactions, and the prioritization of candidates before experimental testing. This mini-review examines the evolution of computational chemistry applied to virtual screening, highlighting its main approaches, the physicochemical filters used for compound selection, and the incorporation of artificial intelligence-based methods. A narrative review of the literature was conducted using Scopus, Web of Science, PubMed, SciELO, and ScienceDirect, supplemented by Google Scholar. Publications from 2015 to 2026 were prioritized, along with seminal studies on the Lipinski, Ghose, Egan, Muegge, and Veber filters. The literature indicates that structure-based and ligand-based drug design provide complementary strategies for the identification and optimization of bioactive molecules. Methods such as molecular docking, molecular dynamics, QSAR, pharmacophore modeling, and ADMET prediction improve selection efficiency, whereas descriptors such as molecular weight, lipophilicity, solubility, polar surface area, ionization, flexibility, and the fraction of sp^3 carbons (F_{sp^3}) aid in assessing pharmacokinetic potential. Machine learning and deep learning have expanded the ability to explore vast chemical spaces and predict molecular properties. However, the reliability of computational results depends on the quality of the structures, data, and models employed. In conclusion, integrating computational tools, physicochemical criteria, and experimental validation can reduce costs, support decision-making, and increase the efficiency of therapeutic candidate discovery.

Keywords: computational chemistry, virtual screening, drug design, physicochemical properties, artificial intelligence.

Introduction

Drug design and development are multidisciplinary processes that integrate medicinal chemistry, biochemistry, pharmacology, structural biology, and computational sciences. These fields converge to identify and optimize molecules capable of selectively interacting with biological targets while exhibiting adequate potency, selectivity, safety, and pharmacokinetic properties. However, the vast number of potentially available chemical structures makes the experimental evaluation of all candidates impractical, highlighting the need for strategies to guide candidate selection (Carlsson et al., 2024). In this context, computational chemistry has evolved from a tool used primarily for the representation and interpretation of molecular structures into a strategic component of drug discovery. Methods such as

molecular modeling, molecular docking, molecular dynamics, quantitative structure–activity relationship (QSAR) analysis, and virtual screening enable the investigation of ligand–target interactions, the exploration of large chemical libraries, and the prioritization of compounds for synthesis and experimental evaluation. More recently, the incorporation of artificial intelligence and machine learning has expanded the ability to predict biological activity, physicochemical properties, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles (Schneider et al., 2020; Gentile et al., 2022; Sadybekov & Katritch, 2023; Swanson et al., 2024).

Despite these advances, computational results depend on the quality of the structures, data, and models used and should therefore be

interpreted in conjunction with experimental evidence. Accordingly, this mini-review examines the evolution of computational chemistry applied to virtual screening in drug discovery, highlighting its main approaches, the physicochemical filters used for candidate selection, and the integration of these tools with contemporary artificial intelligence methods.

Review methodology

This mini-review was conducted as a narrative literature review, with searches performed in Scopus, Web of Science, PubMed, SciELO, and ScienceDirect. Google Scholar was additionally used to identify studies not retrieved through the primary searches and to screen the reference lists of the selected articles. The search covered publications available up to August 2026 and employed combinations of the following English-language search terms: “computational chemistry,” “computer-aided drug design,” “CADD,” “virtual screening,” “structure-based drug design,” “ligand-based drug design,” “molecular docking,” “molecular dynamics,” “QSAR,” “pharmacophore modeling,” “physicochemical filters,” “drug-likeness,” “ADMET prediction,” “artificial intelligence,” “machine learning,” and “deep learning.” The terms were combined using the Boolean operators AND and OR, as appropriate for each database.

Original research articles and review articles published in Portuguese, English, or Spanish that addressed the evolution and applications of computational chemistry in virtual screening and drug design were included. No initial publication-year restriction was applied, allowing the inclusion of

seminal studies on the Lipinski, Ghose, Egan, Muegge, and Veber filters. However, priority was given to publications from 2015 to 2026. Duplicate records, studies lacking sufficient methodological information, and publications not directly related to the scope of this review were excluded.

Contextualization and analysis

Evolution of computational chemistry

Advances in computing, quantum chemistry methods, and molecular mechanics enabled the three-dimensional representation of molecules and the estimation of their electronic, conformational, and energetic properties. Subsequently, the expansion of chemical and structural databases enabled these tools to be applied to the identification and optimization of bioactive compounds (Figure 1) (Sadybekov & Katritch, 2023).⁷

Early computational methods relied on simplified calculations and were constrained by limited processing power. With the advent of microprocessors, personal computers, high-performance computing, and graphics processing, it became possible to analyze libraries containing thousands or millions of chemical structures (Pandey et al., 2022).

Currently, computational chemistry incorporates artificial intelligence, machine learning, and deep learning, expanding the ability to predict biological activity, ADMET properties, biomolecular complex structures, and novel chemical structures with therapeutic potential (Schneider et al., 2020; Gentile et al., 2022; Abramson et al., 2024; Swanson et al., 2024).

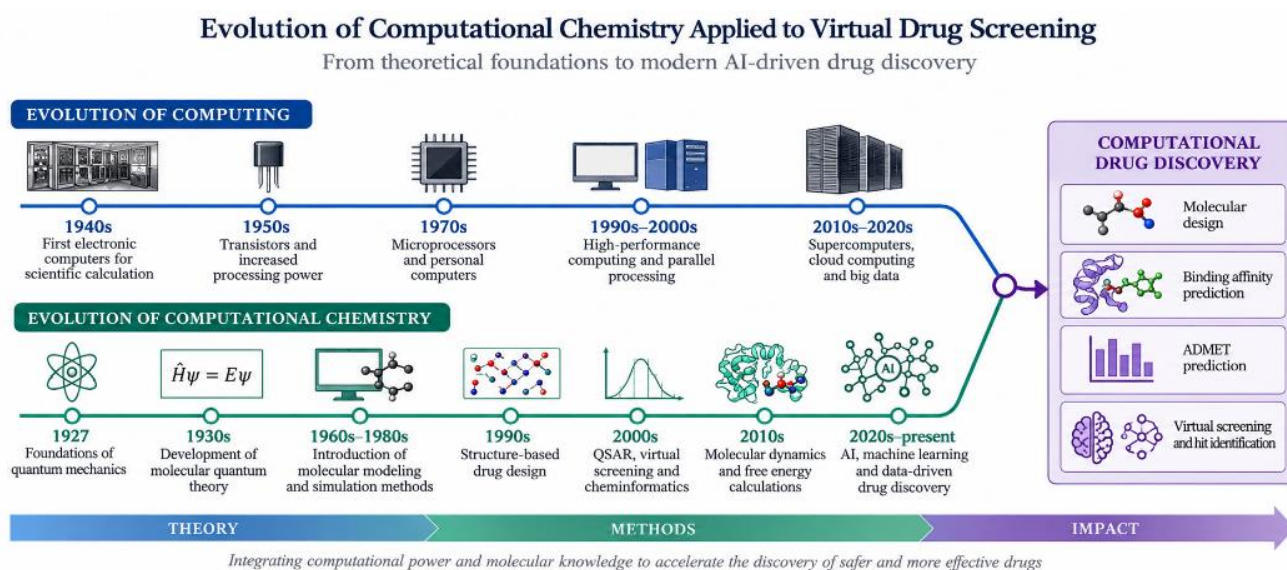


Figure 1. Historical evolution of computing and computational chemistry applied to virtual screening in drug discovery.

The application of these computational approaches to drug design has given rise to two major strategies. Structure-based drug design (SBDD) uses the three-dimensional structures of proteins, enzymes, receptors, or nucleic acids to investigate interactions between ligands and their binding sites. This strategy encompasses molecular docking, structure-based virtual screening, molecular dynamics, and free-energy calculations (Sadybekov & Katritch, 2023; Paggi et al., 2024).

Ligand-based drug design (LBDD) is primarily used when the three-dimensional structure of the target is unavailable. This approach uses information from known bioactive compounds to develop pharmacophore models, quantitative structure–activity relationship (QSAR) models, and machine-learning algorithms (Schneider et al., 2020).

Virtual Screening of Drug Candidates

Virtual screening involves the computational evaluation of molecular libraries to identify candidates with the greatest likelihood of exhibiting biological activity and suitable pharmaceutical properties. This strategy reduces the number of molecules advanced to synthesis and experimental evaluation, thereby reducing time, costs, and reagent consumption (Sadybekov & Katritch, 2023; Gentile et al., 2022).

In structure-based virtual screening, molecules are positioned within the target binding site using molecular docking programs. Docking algorithms explore different ligand orientations and conformations and calculate scores reflecting the geometric, electrostatic, and energetic complementarity of the ligand–target complex (Paggi et al., 2024).

However, docking scores should not be interpreted in isolation as definitive measures of

binding affinity. The results should be evaluated in conjunction with molecular interactions, the conservation of active-site residues, conformational stability, and available experimental evidence (Ferreira et al., 2015; Paggi et al., 2024).

In ligand-based virtual screening, molecules are prioritized based on structural similarity, molecular properties, chemical descriptors, pharmacophore models, or QSAR models. The integration of structure-based and ligand-based approaches increases the robustness of candidate selection, particularly when experimental data on known ligands and a reliable three-dimensional structure of the target are available (Sadybekov & Katritch, 2023).

Physicochemical Filters for Oral Administration

High affinity for a biological target does not necessarily ensure that a molecule can be successfully developed as a drug. A compound must also exhibit solubility, permeability, metabolic stability, and bioavailability compatible with the intended route of administration. For this reason, physicochemical filters are applied during the early stages of virtual screening (Figure 2) (Pennington et al., 2021; Swanson et al., 2024).

Lipinski's Rule of Five proposes that candidates for oral administration are more likely to exhibit adequate absorption when they have a molecular weight below 500 Da, a $\log P \leq 5$, no more than five hydrogen-bond donors, and no more than ten hydrogen-bond acceptors. These criteria should not be regarded as absolute determinants of biological activity but rather as guidelines related to permeability and oral bioavailability (Lipinski et al., 2001).

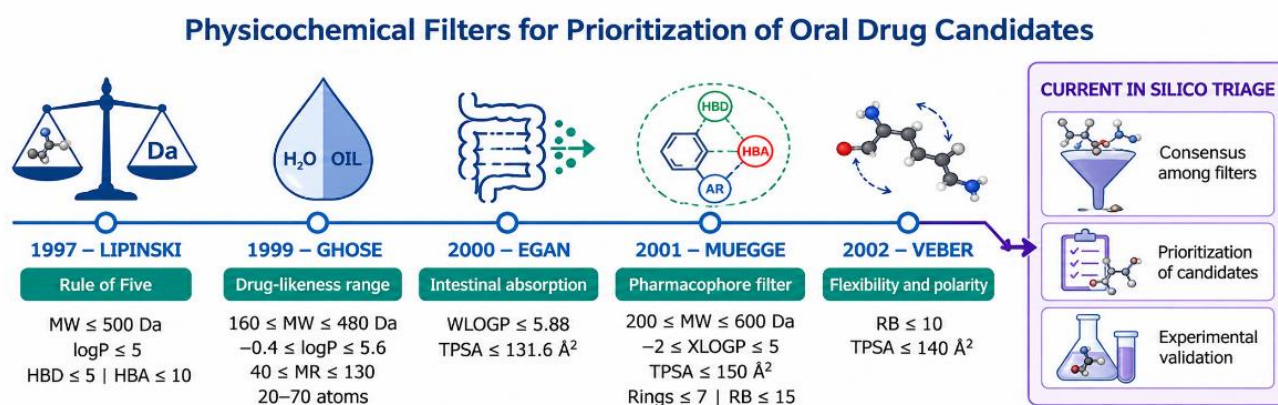


Figure 2. Evolution of the main physicochemical filters used to prioritize candidates for oral administration, including the Lipinski, Ghose, Egan, Muegge, and Veber criteria.

Other models have complemented this assessment. The Ghose filter considers molecular weight, lipophilicity, molar refractivity, and atom count (Ghose et al., 1999). The Egan model relates intestinal absorption to lipophilicity and topological

polar surface area (TPSA) (Egan et al., 2000). The Muegge filter combines molecular properties and functional groups to identify structures consistent with drug-like chemical space (Muegge et al., 2001).

Veber et al. (2002), in turn, highlighted the importance of molecular flexibility and polarity, proposing no more than ten rotatable bonds and a polar surface area not exceeding approximately 140 Å² as favorable parameters.

The combined application of these filters enables a more comprehensive assessment of drug candidates. A compound may satisfy Lipinski's Rule of Five and still exhibit poor aqueous solubility, excessive flexibility, aggregation, metabolic instability, or potentially toxic functional groups. Therefore, candidate prioritization should consider physicochemical, pharmacokinetic, and structural properties collectively rather than relying on compliance with any single filter (Pennington et al., 2021).

Key physicochemical properties

Physicochemical properties play a central role in determining the behavior of drug candidates, influencing their solubility, membrane permeability, absorption, distribution, bioavailability, and overall pharmacokinetic profile (Wu et al., 2024). Their assessment during virtual screening helps prioritize compounds with physicochemical profiles compatible with further development. Key parameters include molecular weight, lipophilicity, aqueous solubility, polar surface area, hydrogen-bonding capacity, molecular flexibility, saturation, three-dimensional character, and ionization state (Kralj et al., 2023) (Figure 3).

Molecular Weight

Molecular weight influences the diffusion, permeability, and distribution of a compound. Excessive molecular weight can hinder membrane permeation and limit absorption. However, exceptions exist, particularly for natural products, peptides, and molecules transported by specific mechanisms (Lipinski et al., 2001; Pennington et al., 2021).

Lipophilicity

Lipophilicity, commonly expressed as logP, reflects the affinity of the nonionized form of a compound for lipid environments. High values may favor membrane permeation but are also associated with poor aqueous solubility, high plasma protein binding, metabolism, and toxicity (Lipinski et al., 2001; Pennington et al., 2021).

LogD is more informative for ionizable molecules because it accounts for the distribution of a compound at a given pH.

Aqueous Solubility

Aqueous solubility, commonly expressed as logS, determines the concentration of a compound available for absorption and interaction with its target. Highly lipophilic or excessively planar and aromatic compounds often exhibit poor solubility because of strong crystal lattice interactions and low affinity for water (Lipinski et al., 2001; Pennington et al., 2021).

Polar Surface Area

Topological polar surface area (TPSA) is related to the presence of heteroatoms and polar groups. This property contributes to hydrogen-bond formation and influences both solubility and permeability (Egan et al., 2000; Veber et al., 2002).

High TPSA values can hinder membrane permeation, whereas very low values may increase lipophilicity and reduce solubility. TPSA should therefore be interpreted in conjunction with logP, ionization state, and molecular weight.

• Main Physicochemical Properties

How molecular structure influences the behavior of a drug candidate

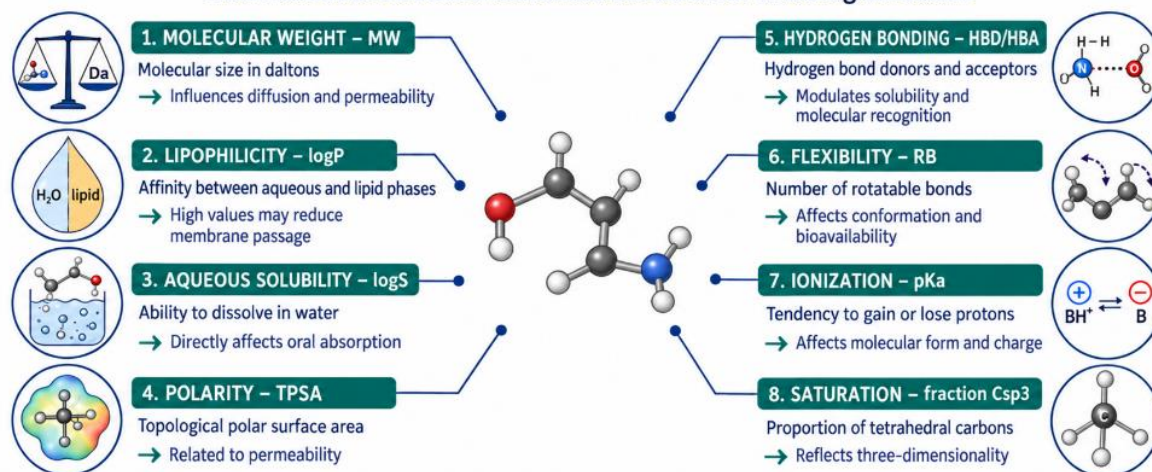


Figure 3. Key physicochemical properties considered in the selection of drug candidates: molecular weight, lipophilicity, aqueous solubility, polarity, hydrogen-bonding capacity, molecular flexibility, ionization, and degree of saturation.

Hydrogen Bonding

Hydrogen-bond donors and acceptors participate in molecular recognition between a ligand and its target. However, an excessive number of these groups increases the energetic cost of desolvation before membrane permeation or binding to the target (Lipinski et al., 2001; Pennington et al., 2021).

Molecular Flexibility

The number of rotatable bonds is a measure of molecular flexibility. Highly flexible molecules can adopt multiple conformations in solution, increasing the entropic cost associated with adopting the bioactive conformation. Conversely, excessively rigid structures may not adapt adequately to the binding site (Veber et al., 2002).

Saturation and Three-Dimensional Character

The fraction of sp^3 carbons (F_{sp^3}) and the degree of saturation are also important molecular features. Increased three-dimensional character can enhance conformational diversity while reducing molecular planarity and crystal packing. However, this feature must be balanced against the need to maintain the appropriate spatial orientation of pharmacophoric groups (Lovering et al., 2009; Pennington et al., 2021).

Ionization

Ionization is primarily determined by the pK_a of a compound and the pH of the surrounding medium. The ionized form generally exhibits greater aqueous solubility, whereas the nonionized form tends to cross membranes more readily by passive diffusion. Thus, the distribution of ionized and nonionized forms throughout the gastrointestinal tract directly influences oral absorption.

Integration of Filters in Candidate Selection

Current virtual screening workflows may begin with the preparation and standardization of chemical structures, followed by the removal of duplicates, reactive compounds, and assay-interfering compounds. Physicochemical filters, pharmacokinetic property models, and toxicity predictions are subsequently applied (Shi et al., 2024; Swanson et al., 2024).

Compounds that pass these filters may then undergo molecular docking, analysis of ligand–target interactions, and, when necessary, molecular dynamics simulations. Agreement among different computational methods increases the reliability of candidate prioritization (Sadybekov & Katritch, 2023; Paggi et al., 2024).

A candidate may be prioritized when it exhibits interactions consistent with the binding site, adequate complex stability, physicochemical properties compatible with the intended route of administration, and no major toxicological alerts.

Nevertheless, computational predictions do not replace experimental assays. Rather, they guide compound selection and support the formulation of

hypotheses that must be validated through biochemical, cellular, pharmacokinetic, and toxicological studies (Sadybekov & Katritch, 2023; Paggi et al., 2024).

Final Considerations

Computational chemistry has established virtual screening as a strategic tool for the selection and optimization of drug candidates. The integration of molecular docking, molecular dynamics, physicochemical filters, ADMET prediction, and artificial intelligence can reduce costs and guide experimental testing. However, these predictions should be evaluated collectively and experimentally validated to ensure greater reliability in the discovery of new therapeutic agents.

Declaration of Artificial Intelligence Contribution

ChatGPT (OpenAI, GPT-5 model) was used exclusively to assist with language editing and editorial organization. The scientific content, reference selection, interpretation of the information, and final version of the manuscript were reviewed and approved by the authors.

References

- ABRAMSON, J.; ADLER, J.; DUNGER, J.; EVANS, R.; GREEN, T.; PRITZEL, A.; RONNEBERGER, O.; WILLMORE, L.; BALLARD, A. J.; BAMBRICK, J.; BODENSTEIN, S. W.; EVANS, D. A.; HUNG, C.-C.; O'NEILL, M.; REIMAN, D.; TUNYASUVUNAKOOL, K.; WU, Z.; ŽEMGULYTĖ, A.; ARVANITI, E.; BEATTIE, C.; BERTOLLI, O.; BRIDGLAND, A.; CHEREPANOV, A.; CONGREVE, M.; COWEN-RIVERS, A. I.; COWIE, A.; FIGURNOV, M.; FUCHS, F. B.; GLADMAN, H.; JAIN, R.; KHAN, Y. A.; LOW, C. M. R.; PERLIN, K.; POTAPENKO, A.; SAVY, P.; SINGH, S.; STECULA, A.; THILLAISUNDARAM, A.; TONG, C.; YAKNEEN, S.; ZHONG, E. D.; ZIELINSKI, M.; ŽÍDEK, A.; BAPST, V.; KOHLI, P.; JADERBERG, M.; HASSABIS, D.; JUMPER, J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, v. 630, p. 493–500, 2024. DOI: 10.1038/s41586-024-07487-w.
- CARLSSON, J.; LUTTENS, A. Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Current Opinion in Structural Biology*, v. 87, p. 102829, 2024. DOI: 10.1016/j.sbi.2024.102829.
- EGAN, W. J.; MERZ, K. M.; BALDWIN, J. J. Prediction of drug absorption using multivariate statistics. *Journal of Medicinal Chemistry*, v. 43, n. 21, p. 3867–3877, 2000. DOI: 10.1021/jm000292e.
- FERREIRA, L. G.; DOS SANTOS, R. N.; OLIVA, G.; ANDRICOPULO, A. D. Molecular docking and structure-based drug design strategies. *Molecules*, v. 20, n. 7, p. 13384–13421, 2015. DOI: 10.3390/molecules200713384.

- GENTILE, F.; YAACOU, J. C.; GLEAVE, J.; FERNANDEZ, M.; TON, A.-T.; BAN, F.; STERN, A.; CHERKASOV, A. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols*, v. 17, p. 672–697, 2022. DOI: 10.1038/s41596-021-00659-2.
- GHOSE, A. K.; VISWANADHAN, V. N.; WENDOLOSKI, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery: 1. A qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, v. 1, n. 1, p. 55–68, 1999. DOI: 10.1021/cc9800071.
- KRALJ, S.; JUKIČ, M.; BREN, U. Molecular filters in medicinal chemistry. *Encyclopedia*, v. 3, n. 2, p. 501–511, 2023. DOI: 10.3390/encyclopedia3020035.
- LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, v. 46, n. 1–3, p. 3–26, 2001. DOI: 10.1016/S0169-409X(00)00129-0.
- LOVERING, F.; BIKKER, J.; HUMBLET, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *Journal of Medicinal Chemistry*, v. 52, n. 21, p. 6752–6756, 2009. DOI: 10.1021/jm901241e.
- MUEGGE, I.; HEALD, S. L.; BRITTELLI, D. Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, v. 44, n. 12, p. 1841–1846, 2001. DOI: 10.1021/jm015507e.
- PAGGI, J. M.; BELLO, M. L.; DROR, R. O. Leveraging AlphaFold in virtual screening: a practical evaluation of the AlphaFold structural database. *Drug Discovery Today*, v. 29, p. 103889, 2024. DOI: 10.1016/j.drudis.2024.103889.
- PANDEY, M.; FERNANDEZ, M.; GENTILE, F.; ISAYEV, O.; TROPSHA, A.; STERN, A. C.; CHERKASOV, A. The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, v. 4, p. 211–221, 2022. DOI: 10.1038/s42256-022-00463-x.
- PENNINGTON, L. D.; HESSE, M. J.; KOESTER, D. C. The role of molecular modeling and structure–activity relationships in medicinal chemistry. *Future Medicinal Chemistry*, v. 13, 2021. DOI: 10.4155/fmc-2020-0305.
- SADYBEKOV, A. V.; KATRITCH, V. Computational approaches streamlining drug discovery. *Nature*, v. 616, p. 673–685, 2023. DOI: 10.1038/s41586-023-05905-z.
- SCHNEIDER, P.; WALTERS, W. P.; PLOWRIGHT, A. T.; SIEROKA, N.; LISTGARTEN, J.; GOODNOW JR., R. A.; FISHER, J.; JANSEN, J. M.; DUCA, J. S.; RUSH, T. S.; ZENTGRAF, M.; HILL, J. E.; KRUTOHLOW, E.; KOHLER, M.; BLANEY, J.; FUNATSU, K.; LUEBKEMANN, C.; SCHNEIDER, G. Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*, v. 19, p. 353–364, 2020. DOI: 10.1038/s41573-019-0050-3.
- SHI, S.; FU, L.; YI, J.; YANG, Z.; ZHANG, X.; DENG, Y.; WANG, W.; WU, C.; ZHAO, W.; HOU, T.; ZENG, X.; LYU, A.; CAO, D.-S. ChemFH: an integrated tool for screening frequent false positives in chemical biology and drug discovery. *Nucleic Acids Research*, v. 52, n. W1, p. W439–W450, 2024. DOI: 10.1093/nar/gkae359.
- SWANSON, K.; WALTHER, P.; LEITZ, J.; MUKHERJEE, S.; WU, J. C.; SHIVNARAIN, R. V.; ZOU, J. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical datasets. *Bioinformatics*, v. 40, n. 7, p. btae416, 2024. DOI: 10.1093/bioinformatics/btae416.
- VEBER, D. F.; JOHNSON, S. R.; CHENG, H.-Y.; SMITH, B. R.; WARD, K. W.; KOPPLE, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, v. 45, n. 12, p. 2615–2623, 2002. DOI: 10.1021/jm020017n.
- WU, K. et al. Overcoming challenges in small-molecule drug bioavailability: a review of key factors and approaches. *International Journal of Molecular Sciences*, v. 25, n. 23, p. 13121, 2024. DOI: 10.3390/ijms252313121.