WORLD JOURNAL OF PHARMACY AND PHARMACEUTICAL SCIENCES
ISSN 2278 – 4357 Impact Factor 8.025
Volume 14, Issue 12, 440-452 Review Article Coden USA: WJPPAJ

AI-ENABLED DRUG DISCOVERY: BRIDGING COMPUTATIONAL INNOVATION WITH PHARMACEUTICAL SCIENCES

Gudladona Raghava Ravali[1]*, Purvini K[2], Swathi Kencha[3], Savin Sreenath T P[4]

[1,2,3,4]Department of Pharmaceutical Chemistry, Sri Raghavendra College of Pharmacy, Bangalore.

Article Received on 23 October 2025,
Article Revised on 13 Nov. 2025,
Article Published on 01 Dec. 2025,

https://doi.org/10.5281/zenodo.17735103

***Corresponding Author**

Gudladona Raghava Ravali

Department of Pharmaceutical Chemistry, Sri Raghavendra College of Pharmacy, Bangalore.

How to cite this Article: Gudladona Raghava Ravali1*, Purvini K2, Swathi Kencha3, Savin Sreenath T P4. (2025). AI-Enabled Drug Discovery: Bridging Computational Innovation With Pharmaceutical Sciences. World Journal of Pharmacy and Pharmaceutical Sciences, 14(12), 440–452.

ABSTRACT

Finding a new medicine has always been challenging. It often requires many years of work, major financial investment, and numerous laboratory experiments before a drug finally reaches patients. Because of these demands, the process can feel slow and uncertain. In recent years, however, artificial intelligence has begun to noticeably improve this situation. By examining large collections of chemical, biological, and medical information, artificial intelligence helps researchers understand diseases more clearly and identify strong drug candidates much earlier. This review describes how artificial intelligence is now used across the drug discovery process. It explains how computer-based methods analyze complex data, predict the shapes and behavior of important proteins, and help scientists pinpoint where potential medicines may bind. It also highlights how artificial intelligence tools screen vast numbers of chemical compounds, track how molecules might move inside the body, and estimate safety concerns long before testing begins. Newer approaches—such as generative models that propose new chemical structures and tools that reveal how artificial intelligence makes decisions—are also discussed. Although issues like data quality and prediction reliability still need attention, artificial intelligence is already making drug discovery faster, more organized, and less dependent on trial-and-error. Overall, these technologies are helping researchers work more efficiently and move closer to developing safe and effective medicines in a shorter time.

KEYWORDS: Artificial Intelligence, Drug discovery, Machine Learning, Protein Structure Prediction, Virtual Screening, ADMET Modeling.

1. INTRODUCTION

Drug discovery has long been recognized as one of the most demanding and unpredictable areas of biomedical science. Developing a single therapeutic agent can take more than a decade and often requires enormous financial investment. Despite this effort, the majority of candidates—nearly ninety percent—do not reach regulatory approval, usually because of safety concerns, poor pharmacokinetic performance, or lack of therapeutic effect.[1] These high failure rates reflect the complexity of human biology and the constraints of traditional experimental approaches, which can only examine a fraction of the biological and chemical space relevant to disease.

In recent years, artificial intelligence has begun to reshape this landscape. By learning from vast chemical and biological datasets, artificial intelligence systems can identify hidden patterns, forecast molecular behavior, and support decision-making at every stage of the discovery pipeline. Machine learning and deep learning models have already shown value in tasks such as identifying drug targets from multi-omics data, predicting toxicity, ranking ligands, and proposing new chemical structures.[2,3] Architectures such as graph neural networks, convolutional neural networks, and transformer models have significantly improved the accuracy of protein structure prediction, virtual screening, and property optimization.[4]

This review summarizes these developments in a structured and accessible manner. It expands on the foundations of machine learning, supervised and unsupervised learning, deep learning, and data curation, and provides detailed explanations of protein structure prediction, quantitative structure–activity relationships, drug–target interaction modelling, and broader applications in drug design.

2. FOUNDATIONS OF ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY

Artificial intelligence relies on computational methods capable of recognizing patterns, linking structural features to biological outcomes, and predicting how molecules or proteins will behave in different environments.[5] These methods are particularly useful in drug discovery, where datasets are large, complex, and often nonlinear.

2.1 Machine Learning in Drug Discovery

Machine learning has become a central force in modern pharmaceutical research. Rather than depending solely on manual interpretation or predefined rules, machine learning enables computers to learn directly from experimental data. This is important because drug discovery involves complicated biological systems, high-dimensional omics data, and subtle structure–activity relationships that are difficult to interpret with conventional tools.

Machine learning models can analyze millions of molecular structures, biological sequences, assay outputs, and pharmacokinetic measurements. Their predictions guide chemists and biologists toward more promising compounds, identify potential toxicity risks, and detect patterns that would likely remain unnoticed through manual analysis.[6] Today, machine learning contributes to nearly every stage of early and late-stage development: forecasting biological activity, identifying structure–activity relationships, classifying toxic or inactive compounds, optimizing lead structures, screening large chemical libraries, predicting pharmacokinetic properties, and supporting drug repurposing.[7] Because these models improve as more data become available, their performance tends to increase over time.

2.2 Supervised Learning

Supervised learning remains the most widely used machine-learning approach in drug discovery because it learns from datasets where the correct output is already known. These labeled datasets may contain values such as IC_{50}, EC_{50}, inhibition constants, active/inactive classifications, solubility data, or permeability measurements.

The typical workflow includes:

- representing molecules through descriptors, fingerprints, or graph structures
- associating each input with a corresponding experimental measurement
- training a model to map features to biological outcomes
- evaluating how well the model predicts properties of new molecules

Supervised learning plays an essential role in predicting binding affinity, toxicity, absorption and distribution properties, drug–target interactions, and multi-target activity (polypharmacology).[8]

2.3 Common Algorithms

A variety of algorithms support AI-driven drug discovery, each suited to different data types and complexity levels.

- Support Vector Machines perform well when data are limited or moderately complex.
- Random Forests handle noise effectively and can process thousands of descriptors.
- Gradient-boosted methods such as XGBoost and LightGBM are powerful tools for predicting pharmacokinetic and toxicity properties.
- Neural networks are preferred for highly nonlinear data and large-scale modelling tasks.

These methods serve as the computational backbone for many predictive tasks across medicinal chemistry and computational biology.

2.4 Unsupervised Learning

Unsupervised learning explores data without relying on predefined labels, making it especially useful in early discovery where outcomes are not yet known.[9]

Common approaches include:

- K-means clustering, which groups molecules or biological samples based on similarity
- Hierarchical clustering, widely used in transcriptomics and proteomics visualization
- Principal component analysis, which reduces the dimensionality of large datasets
- t-SNE and UMAP, which offer detailed visualization of complex biological or chemical spaces

These methods help researchers explore chemical diversity, identify patterns among screening hits, detect disease subtypes, recognize toxicophores, and support the generation of initial hypotheses for supervised models.

2.5 Deep Learning

Deep learning has become a cornerstone of computational drug discovery because of its ability to learn multi-level features directly from raw data.[10] Unlike earlier models that rely on handcrafted descriptors, deep neural networks automatically extract meaningful representations from molecular graphs, protein sequences, and structural data.

Different architectures support different tasks:

- Convolutional neural networks analyze protein pockets and ligand interactions in three dimensions.
- Recurrent neural networks and long short-term memory networks interpret SMILES strings and biological sequences.
- Transformer models, using self-attention mechanisms, excel in predicting protein folding, molecular properties, and drug–target interactions.
- Graph neural networks represent molecules as graphs and capture atomic-level interactions with high precision.Deep learning has significantly advanced structure prediction, virtual screening, toxicity modelling, and de novo molecule design.

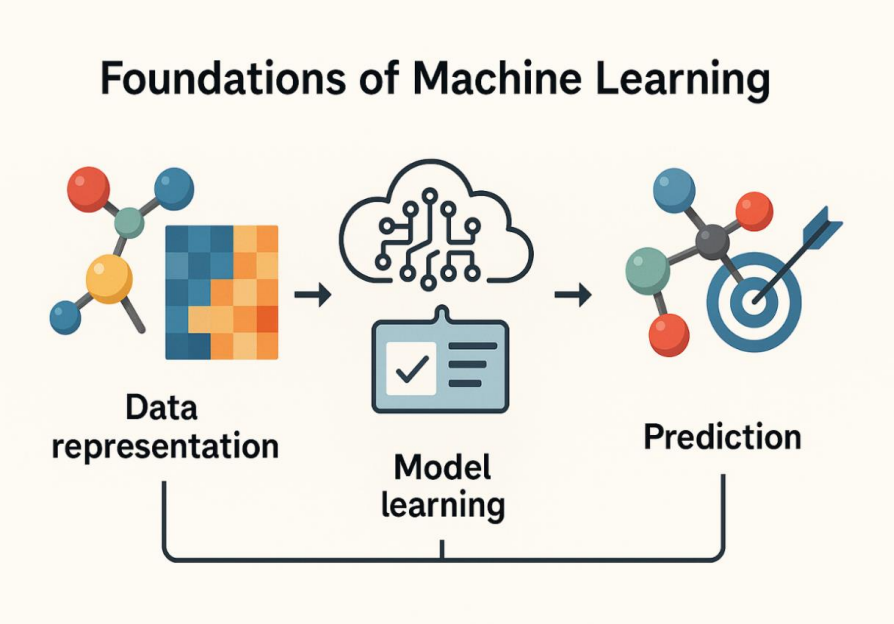

Figure-1

2.6 Data Curation

Data curation is fundamental to building trustworthy AI models. Even the most advanced algorithms cannot compensate for poorly prepared, inconsistent, or noisy data. Chemical datasets must be standardized by removing duplicates, correcting valence issues, ensuring proper stereochemistry, and resolving tautomers.[11] Similarly, biological assay results require harmonized units, removal of outliers, reconciliation of conflicting data, and careful handling of replicates. Omics datasets also require normalization to reduce technical variation. Balanced datasets are essential to avoid bias toward inactive molecules, which commonly dominate screening results. Good curation ensures that models learn meaningful biological patterns instead of artifacts.

2.7 Model Evaluation

Evaluating model performance is crucial to ensure that predictions are reliable and transferable. Classification models are assessed through metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, while regression models are evaluated using error-based metrics including RMSE, MAE, and R^2. Cross-validation techniques—such as k-fold validation and stratified sampling—help prevent overfitting. For tasks like virtual screening, specialized metrics such as enrichment factor and BEDROC measure early recognition of active compounds.[12] Well-validated models support confident decision-making in drug discovery.

3. Systems Biology, Omics, and AI-Based Target Identification

Systems biology and multi-omics technologies have significantly changed how researchers identify therapeutic targets. Rather than focusing on single genes or proteins, modern strategies examine the broader networks that drive disease. Genomics, transcriptomics, proteomics, metabolomics, and epigenomics each provide essential pieces of information, such as mutations, altered gene expression, changes in protein levels, metabolic disturbances, and epigenetic modifications.[13]

Artificial intelligence helps integrate these diverse datasets, revealing relationships and regulatory interactions that are difficult to detect manually. It can rank potential targets based on their biological role, network importance, druggability, and relevance to disease pathways. Approaches such as knowledge graphs, CRISPR screen analysis, phenotypic profiling, and graph neural networks strengthen confidence in target selection. By capturing the complexity of disease biology, AI-driven methods improve the likelihood that early-stage targets will translate successfully into therapies.

4. Protein Structure Prediction and Binding Site Analysis

Recent advances in artificial intelligence have transformed protein structure prediction into a computationally accessible task. Traditional methods such as X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy remain valuable but are limited by experimental constraints, cost, and time demands.[14]

AI-based models, including AlphaFold2 and ESMFold, now predict three-dimensional structures directly from amino acid sequences, often with accuracy close to experimental models. Once high-quality structures are available, identifying potential binding sites is

essential for rational drug design. Classical algorithms such as Fpocket and CASTp detect cavities and pockets based on geometry. More advanced AI-driven tools—including DeepSite, Kalasanty, and P2Rank—can recognize subtle features, detect transient or allosteric sites, and assess druggability with greater precision.[15] Together, these developments enable more effective docking, virtual screening, fragment-based design, and optimization of lead compounds.

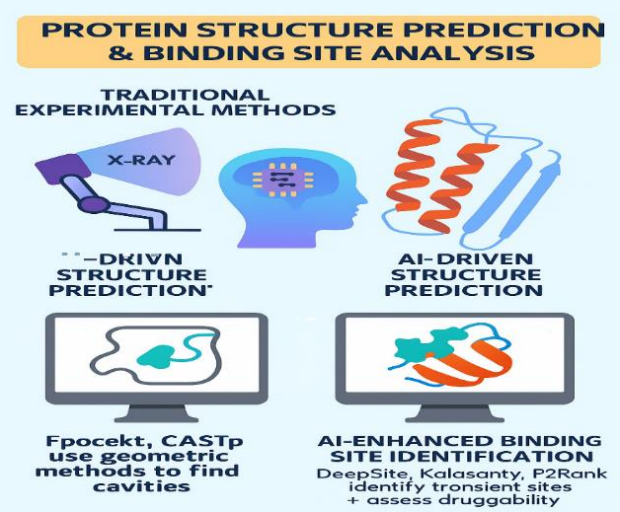

Figure 2

5. Virtual Screening and Molecular Docking

Virtual screening and molecular docking have become essential components of computational drug discovery because they allow researchers to examine large chemical libraries quickly and identify molecules that are likely to bind to a target protein with meaningful affinity. Structure-based virtual screening relies on the three-dimensional structure of a protein and uses docking algorithms to position candidate molecules within the binding pocket while scoring functions estimate how strongly each molecule may bind.[16] Widely used tools such as AutoDock Vina, Glide, and GOLD have played a major role in hit discovery, although they can be limited by their handling of protein flexibility, water interactions, and dynamic pocket changes.

Recent advances in artificial intelligence have further strengthened virtual screening by enabling “deep docking,” where machine-learning models predict docking scores for millions or even billions of molecules. This dramatically reduces computational time by prioritizing only the most promising candidates for full docking calculations. Ligand-based methods, including similarity searches, shape-matching approaches, and pharmacophore models,

identify compounds that share essential features with known active molecules. Approaches such as PyRMD, which applies random matrix theory to classify actives and enrich hit lists, further expand the utility of ligand-based screening.[17] Together, structure-based tools and AI-driven screening significantly enhance the efficiency and accuracy of early-stage hit identification.

6. QSAR and Drug–Target Interaction Prediction

Quantitative structure–activity relationship (QSAR) modelling aims to link chemical structure with biological activity, providing a predictive framework for evaluating new molecules. Earlier QSAR methods used handcrafted descriptors and linear models, but modern versions rely on more flexible learning systems such as deep neural networks, graph-based models, and transformer architectures. These advanced models can capture nonlinear relationships and physicochemical interactions directly from molecular structures, allowing for more accurate activity prediction.[18]

Drug–target interaction (DTI) prediction has also benefited from recent advances in artificial intelligence. Instead of requiring structural data, DTI models such as DeepDTA, GraphDTA, and transformer-based approaches learn from molecular sequences and protein sequences to estimate binding affinity. These tools help researchers predict multi-target effects, identify repurposing opportunities, and prioritize hits earlier in the discovery process. Collectively, QSAR and DTI models form a critical part of the modern computational toolkit for molecular prioritization.

7. Molecular Dynamics

Molecular dynamics (MD) has become indispensable for understanding how proteins and ligands behave over time at the atomic level. While docking provides a static representation of how a molecule might fit within a binding site, MD captures the continuous motion, flexibility, and induced-fit effects that occur under physiological conditions.[19] Proteins naturally undergo dynamic processes—loop movements, side-chain rotations, shifts in pocket geometry—that can influence ligand binding.

MD simulations use established force fields such as AMBER and CHARMM to simulate atomic interactions. The workflow generally includes system preparation, energy minimization, equilibration phases under constant temperature and pressure, and a production stage where trajectories are generated. These trajectories provide detailed information on

system stability and interactions through metrics such as RMSD, RMSF, radius of gyration, solvent-accessible surface area, hydrogen bonding patterns, and non-bonded interaction energies.

Binding affinity can be estimated more precisely through post-processing methods such as MM/PBSA, MM/GBSA, free energy perturbation, and thermodynamic integration. Enhanced sampling techniques—including metadynamics, accelerated molecular dynamics, and replica-exchange molecular dynamics—allow simulations to explore rare events like ligand unbinding or hidden pocket formation. With AI-based clustering and conformational prediction tools now accelerating analysis, MD has evolved into a powerful complement to docking and virtual screening.[20]

8. Artificial Intelligence for ADMET and Toxicology Modelling

Artificial intelligence has had a transformative impact on predicting absorption, distribution, metabolism, excretion, and toxicity, all of which determine whether a drug candidate can safely progress to clinical trials. Machine-learning and deep-learning models can now estimate solubility, permeability, bioavailability, plasma-protein binding, blood–brain barrier penetration, metabolic stability, clearance, and half-life with strong reliability.[21] These models also help anticipate interactions with cytochrome P450 enzymes, predict metabolic hot spots, and flag potential drug–drug interactions at an early stage.

AI-based toxicology tools identify structural motifs associated with liver injury, cardiac toxicity, kidney damage, and genetic toxicity. Specialized predictors, including graph neural networks for hepatotoxicity and transformer models for cardiotoxicity, contribute valuable insights into safety profiles. Predictions of hERG inhibition—critical for assessing cardiac risk—have also become more accurate through deep-learning approaches. Multi-task learning frameworks enable simultaneous prediction of multiple ADMET endpoints, while physiologically based pharmacokinetic (PBPK) modelling integrates biological system knowledge to simulate patient-specific outcomes. By reducing late-stage failures, AI-driven ADMET modelling helps streamline the drug development pipeline.

9. Emerging Technologies: Generative Artificial Intelligence, Pharmacophores, Quantum Machine Learning, and Explainable Artificial Intelligence

Emerging artificial intelligence technologies are redefining what is possible in drug design. Generative models—including variational autoencoders, generative adversarial networks,

reinforcement-learning frameworks, and diffusion models—enable the creation of novel molecular structures optimized for potency, selectivity, and pharmacokinetic properties.[22] These methods represent a shift from screening existing molecules to designing entirely new ones.

Pharmacophore modelling remains valuable by identifying the essential spatial arrangements of chemical features responsible for activity. It allows researchers to search for new ligands even when complete protein structures are unavailable.

Quantum machine learning is an emerging frontier that combines quantum-mechanical accuracy with scalable learning algorithms. It has the potential to improve predictions of molecular energetics, electronic structure, and conformational dynamics.

Explainable artificial intelligence has become increasingly important for regulatory and scientific acceptance. Methods such as SHAP values, LIME, attention-weight visualization, and graph explainers help researchers and regulators understand why a model makes certain predictions, increasing interpretability and trust in AI-guided decisions.

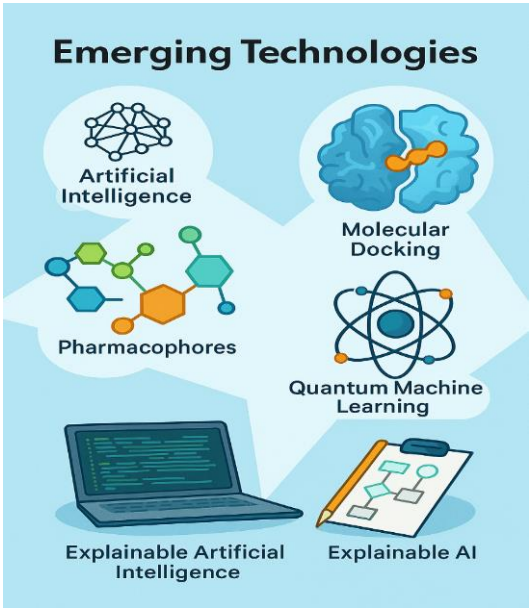

Figure: 3.

10. Artificial Intelligence in Clinical Development and Regulatory Science

Artificial intelligence is reshaping clinical development by helping researchers identify the right patient groups, predict clinical outcomes, determine appropriate dosing, and design adaptive trial strategies.[23] Machine-learning models analyze electronic health records,

imaging data, biomarkers, and genomic information to identify patients who are most likely to respond to a treatment. Predictive tools help forecast disease progression and detect safety issues earlier in clinical testing.

Regulatory agencies increasingly emphasize transparency and fairness in AI models. Explainable artificial intelligence tools help justify predictions, while evolving guidelines address AI-generated drug candidates, digital biomarkers, and predictive toxicology models. In addition, AI enhances pharmacovigilance by monitoring large datasets for rare adverse events, contributing to improved patient safety after approval.

11. DISCUSSION

The integration of artificial intelligence across drug discovery and development has led to significant changes in how therapeutic candidates are identified, optimized, and advanced. Multi-omics methods enable more accurate identification of biologically relevant targets. Protein structure prediction supports rational, structure-based design. Virtual screening and docking accelerate hit discovery, while QSAR and DTI models enhance prioritization. Generative models contribute to novel compound design, and molecular dynamics simulations deepen the understanding of biomolecular behavior.[24]

Despite these advancements, challenges remain. Issues such as inconsistent data quality, limited model interpretability, integration barriers between software platforms, biological complexity, and evolving regulatory expectations continue to slow progress. Future research will focus on integrating multi-modal data, improving transparency, and leveraging quantum-enhanced computational tools.

12. CONCLUSION

Artificial intelligence is fundamentally reshaping how new medicines are discovered. By replacing much of the slow trial-and-error approach with fast, data-guided decision-making, AI allows researchers to analyze large-scale biological and chemical information, uncover meaningful patterns, and choose promising candidates earlier in the development pipeline. This shift is making drug discovery more efficient, targeted, and innovative.

AI’s contributions span the entire process—from understanding disease biology and identifying therapeutic targets to predicting protein structures, evaluating chemical libraries, modelling molecular interactions, and estimating safety profiles. Advances in virtual

screening, docking, QSAR modelling, drug–target interaction prediction, and molecular dynamics have greatly improved scientific accuracy. AI-driven ADMET prediction now helps reduce late-stage failures, while generative AI expands exploration into new chemical space. With the emergence of quantum-based tools, explainable AI, and autonomous laboratories, the future of drug discovery appears increasingly integrated and automated.

In summary, artificial intelligence is not only improving pharmaceutical research—it is redefining it. Over the next decade, deeply integrated, AI-driven pipelines are expected to accelerate the development of safe, effective, and personalized therapies at a pace that was previously unimaginable.

13. REFERENCES

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov., 2010; 9(3): 203–214.
2. Schneider G. Automating drug discovery. Nat Rev Drug Discov., 2018; 17(2): 97–113.
3. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov., 2019; 18: 463–477.
4. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature., 2021; 596: 583–589.
5. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. Mol Pharm., 2016; 13: 1445–1454.
6. Patel L, Shukla T, Huang X, et al. Machine learning methods in drug discovery. Molecules., 2020; 25: 527.
7. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today, 2018; 23: 1241–1250.
8. Sheridan RP. Using random forest to model the domain applicability of another random forest model. J Chem Inf Model., 2013; 53: 2837–2850.
9. Ding M, Weng Z, Chen Y. Unsupervised learning for molecular embedding. J Cheminform., 2021; 13: 1–15.
10. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol., 2019; 37: 1038–1040.
11. Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics. J Chem Inf Model., 2010; 50(7): 1189–1204.

12. Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the early recognition problem. J Chem Inf Model., 2007; 47: 488–508.
13. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol., 2017; 18: 83.
14. Callaway E. The revolution in protein structure prediction. Nature., 2020; 588: 203–204.
15. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model., 2018; 58: 287–296.
16. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking. J Comput Chem., 2010; 31: 455–461.
17. Brereton RG. Chemometrics: data analysis for the laboratory and chemical plant. Wiley, 2003.
18. Baskin II, Winkler DA, Tetko IV. A renaissance of neural networks in QSAR modeling. J Chem Inf Model., 2016; 56: 249–262.
19. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron., 2018; 99: 1129–1143.
20. Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. Science, 2019; 365: eaaw1147.
21. Lagorce D, Bouslama L, Becot J, et al. ADMET prediction using machine learning. J Med Chem., 2020; 63: 6–27.
22. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning. Science, 2018; 361: 360–365.
23. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng., 2018; 2: 719–731.
24. Fleming N. How artificial intelligence is changing drug discovery. Nature, 2018; 557: S55–S57.