Preprints.org

Review

Not peer-reviewed version

Meta-Analysis and Review of In Silico Methods in Drug Discovery – Part 2: Principles of Cellular Complexity, Molecular Similarity and Singularity

Sonia Morya , Debosree Ghosh , Arife Uzundurukan , Mark Nelson , Shahidul Islam , John Victor Christy , Christopher Mattew Teske , Mohamed Elzagheid , Holli-Joi Martin , Eugene Muratov , Domenico Fuoco *

Posted Date: 11 November 2026

Keywords: chemical space; molecular similarity; activity cliffs; virtual screening; artificial intelligence; ADMET; PBPK modeling; drug discovery

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Meta-Analysis and Review of In Silico Methods in Drug Discovery – Part 2: Principles of Cellular Complexity, Molecular Similarity and Singularity

Sonia Morya [1], Debosree Ghosh [2], Arife Uzundurukan [3], Mark Nelson [4], Shahidul Islam [5], John Victor Christy [6], Christopher Matthew Teske [7], Mohamed Elzagheid [8], Holli-Joi Martin [9], Eugene Muratov [10] and Domenico Fuoco [11,*]

[1] Department of Food Technology and Nutrition, School of Agriculture, Lovely Professional University, Phagwara-144411, Punjab, India
[2] Department of Physiology, Government General Degree College, Kharagpur II, P.O - Madpur Dist - Paschim Medinipur, 721149, West Bengal, India
[3] Division of Surgical and Interventional Sciences, Department of Surgery, Faculty of Medicine and Health Sciences, McGill University, Montréal, QC H3G 2M1, Canada
[4] Applied Molecular Design, LLC, 73 Versailles Way, Midway, 84049, Utah, USA
[5] Département de chimie, Université du Québec à Montréal (UQÀM), Montréal, Québec, Canada
[6] Major Surface Combatants (MSC), DGMEPM, Materiel Group, Royal Canadian Navy, Department of National Defence (Contractor), ICE Dragon Corrosion Inc Unit 104, 130 Boundstone Way, Kanata, K2T 0M7, Ontario, Canada
[7] Hepatology and Oncology Research Center, Department of Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, 48103, Michigan, USA
[8] Libyan Authority for Scientific Research, Tripoli, Libya and Center for Strategic and International Research, Tripoli
[9] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, US
[10] UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA
[11] Chemical Engineering, Polytechnique Montréal, Montréal, Québec, Canada
* Correspondence: domenicofuoco@live.ca

Abstract

Scope: This meta-analysis and review examine the conceptual and translational evolution of *in silico* methods in drug discovery, with emphasis on how chemical space, molecular similarity, cellular complexity, and localized singularities shape next-generation computational pharmacology. Building on Part 1, which defined the historical "what" and "when" of computational drug discovery, this Part 2 addresses the "how" and "why" by analyzing the principles that connect molecular descriptors, virtual screening, scaffold hopping, activity cliffs, and AI-driven lead prioritization. **Methods:** A systematic meta-analysis of 97 highly cited studies, selected from an initial pool of 300 top-cited publications, evaluates the ongoing paradigm shift from traditional computer-assisted drug discovery (CADD) toward advanced AI-driven workflows. **Results:** Thus, the current drug discovery practice enables the screening of ultra-large chemical libraries, mapping chemical diversity, generating hit maps, and prioritizing compounds through early ADMET assessment. The review shows that these methods can substantially accelerate early discovery and reduce experimental burden, but their impact remains mainly translational rather than automatically clinical. **Limitations and Uncertainties:** Major bottlenecks persist because model performance is constrained by chemical data quality, biological label uncertainty, missing or biased negative evidence, limited applicability-domain coverage, static representations of protein targets, incomplete treatment of solvent thermodynamics, and fragmented ADMET prediction. **Conclusion:** This work therefore argues that reliable AI-driven drug discovery requires more than faster prediction: it depends on fit-for-purpose

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

datasets, explicit uncertainty handling, dynamic protein and solvent modeling, Systems ADMET, AI-PBPK integration, and closed-loop experimental validation. Overall, chemical space is presented not only as a repository of possible molecules, but as an operational framework for connecting molecular design, biological complexity, and clinically meaningful translation.

Keywords: chemical space; molecular similarity; activity cliffs; virtual screening; artificial intelligence; ADMET; PBPK modeling; drug discovery

Executive Summary

Domain Elements	Core Conceptual Synthesis
Definition	Post-2000 *in silico* methods leveraging **Big Data, Artificial Intelligence (AI), and Chemical Space** metrics have systematically pivoted the early phases of pharmaceutical research and development from a traditional "one drug–one target" paradigm to multi-target network pharmacology. This aim of this review is to provide a "big picture" about the evolution from QSAR (mainly research-based approach of late '80s to current **Pharmacology Network**).
Needs	Evidently, the evolution of digital library from thousands to billions of items useful to research and the high-speed computing systems have provided the technological base to this revolution. Consequently, advanced, multi-dimensional modeling tools capable of simulating **dynamic biological networks, protein conformational flexibility, and localized chemical singularities** (activity cliffs) are considered the new standard in computational tools.
Opportunity	Utilizing high-speed deep learning architectures (e.g., Graph Neural Networks) and Molecular Dynamics trajectories to **screen ultra-large virtual libraries**, retrieve >80–90% of experimentally confirmed hits and drastically compress target-to-Phase I trial timelines. Unsurprisingly, the **2024 Nobel Prize in Chemistry** was awarded to a team of researchers recognizing the revolutionary role of artificial intelligence in mapping protein structures — fully validating this modern trajectory.
Gap	A stark translational barrier marked by **poor data quality/annotation errors** in public databases, a distinct "solvent blind spot" in static docking software, and fragmented, standalone ADMET filters that miss complex bioactivation loops and toxic systemic cross-talk. In Part 1 (already published) of this meta-analysis and review, it was discussed about the What and When. In this Part 2, we are going to discuss the "How and Why".
Recommendation	Transition away from isolated, linear algorithmic filters toward **closed-loop Systems ADMET and dynamic AI-PBPK (Physiologically Based Pharmacokinetic) integrated architectures** powered by mechanistic ordinary differential equations (ODEs) to capture the biological realities of human physiology.

Introduction

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

In the previous Part 1 was established the definition of Chemical Space and its "expanding universe" of billions of molecules. In this Part 2, Chemical Space serves as the multidimensional stage where molecular similarities and cellular complexity perform their balancing act [1–3].

Chemical Space matters conceptually since it serves as both map ("Space" concept = where are we?) and a design canvas (enormous number of dimensional connections to finding new escape-solutions = where should we go next?). Consequently, today medicinal chemists are living in a totally different era than the just-in-time "me too solution" typical of the '80s and '90s [4].

In fact, to understand how to run analysis in this multi-dimensional space is the key factor in accelerating successful drug discovery outcomes. Chemical Space is intended not just a collection of molecules (by definition is the biggest collection ever of **drug candidates**), but the ensemble of the biological functions (**targets**), and the given atomic coordinates (**pharmacophore**) to match each drug candidate with its own target [5–7]. Thus, implicitly, Chemical Space includes also all the physiological systems of the given hosts (**network pharmacology**). Moreover, considering that the biological functions are also the macromolecules that are responsible for the pathology inside the host and that each drug candidate can have several shades of affinity for many of those targets, the results is the rationalization of **multi-targets-drug** [8,9].

Consequently, this AI-driven era provides highly capable cheminformatics platforms to parse and analyze massive datasets, opening unprecedented avenues for **drug repurposing** and the discovery of novel **orphan drugs** [10–13].

Modeling- and Mapping/Visualization- tools

Computer-aided drug design (CADD) visualizations are translated in real time by dedicated software (Schrodinger, MOE, ChemAxon, RDKit, Dragon and Mordred) into a multi-dimensional array of mathematical descriptors, including scalar parameters (ex.: lipophilicity and molecular weight) and structural vectors (ex.: shape index and charge distribution) [14–16]. Which are called molecular descriptors. These critical molecular descriptors bridge the gap between abstract 3D models and predictable physicochemical behaviors, capturing essential parameters such as topological indices, electronic properties, and conformational limitations [17–19]. Together, these molecular descriptors allow dedicated algorithm to calculate mathematical distances within chemical space and to discover new relation (or mechanism of action) within the biological functions [20].

A premier example of molecular descriptor is the **3D Molecular Electrostatic Potential (MEP) Grid Vector**, which maps the magnitude and direction of coulombic forces across a molecule's surface. These spatial vectors do not always shift linearly with structural modification [21]. Instead, they can trigger an **activity cliff** (in order words, a localized **chemical singularity**), where a fractional atomic alteration, such as adding a single hydroxyl group, dramatically warps the 3D electrostatic vector [22–24].

This minor vector transformation can induce a catastrophic, thousand-fold surge or collapse in target protein binding affinity, completely defying standard linear structure-activity assumptions. In other words, a chemical singularity can also be viewed as a specific distance among other chemical referenced points over a multi-dimensional space that is de facto a point of inflection within the surround space in all its dimensional layers [25–27].

Activity cliffs embody the pinnacle of Structure-Activity Relationship (SAR) discontinuity, functioning as localized singularities within chemical space. In these regions, the standard linear assumptions of computational modeling fail. As observed in standard Factor Xa inhibitor series, a fractional modification — such as introducing a single hydroxyl group or inverting a stereocenter — subtly updates a molecule's 2D topology while causing a catastrophic, non-linear shift in its 3D electrostatic potential vector. This small vector shift triggers an all-or-nothing cooperative binding event, forcing a million-fold leap across a vertical landscape cliff [28,29].

Table 1 represents the evolution of the Structure-Activity-Relationship (SAR) since the beginning of modern medicinal chemistry. In fact, the best way to understand the How and Why in medicinal chemistry is to use the evolution of tools used in SAR analysis.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Moreover, at the end of this article are reported three boxes to deep dive in the definitions (BOX 1) and scientistic knowledge (BOX 2 and BOX 3) that are used in the current drug discovery process.

Moving from Computer-Assisted Drug Discover (CADD) to AI-driven drug discovery (AIDD)

The transition from CADD to AIDD can be understood as a shift from linear, filter-based workflows toward integrated, evidence-aware discovery systems [30–32]. Traditional virtual screening first identifies candidate ligands, low-dimensional hit maps then organize enriched chemical regions, tree-map filtering narrows candidates according to structural and property relationships, and ADMET optimization evaluates whether promising molecules can survive the physiological constraints of absorption, distribution, metabolism, excretion, and toxicity. In modern AI-driven workflows, these steps are no longer isolated checkpoints but interconnected layers of a closed discovery loop that links chemical-space navigation to biological validation [33,34].

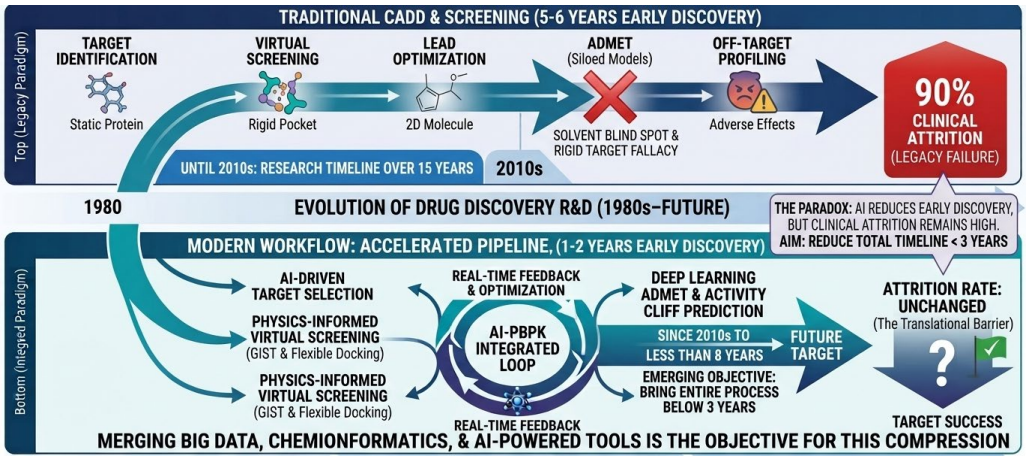

Figure 1. Comparative Timeline Analysis of Drug Discovery R&D Paradigm Shifts (1980s–Future). The upper panel illustrates the traditional, sequential computer-aided drug design (CADD) pipeline prominent until the 2010s. This legacy framework routinely exceeded a 15-year total research timeline, constrained by a static protein target fallacy, solvent ignorance, and siloed, linear ADMET filters that resulted in a staggering 90% clinical attrition rate. The middle/lower panels chart the modern transition toward an accelerated, integrated workflow (2010s–present), which has compressed the timeline to under 8 years but leaves downstream clinical attrition rates largely unchanged due to persistent physiological barriers. The emerging paradigm highlights the core strategic objective of merging Big Data, Cheminformatics, and AI-driven physics-informed architectures (such as GIST and multi-compartment AI-PBPK loops) to simulate non-linear human physiology *in silico*, with the ultimate target of bringing the entire end-to-end R&D pipeline below 3 years while maximizing clinical translation success.

Core in silico methods and workflows

Computational methods and specially AI approaches now touch every stage of drug discovery, from target selection and virtual screening to ADMET prediction and clinical repurposing. The analysis of the literature has provided the consensus that in silico methods consistently reduce time, cost, and experimental burden (in silico medicine tremendous success) [35–37]. Although, the translation of these results to clinically approved drugs remain constrained by data validation issues [38–40].

Screening ultra-large or enriched chemical library has yielded high hit rates, often discovering new chemotypes and sub-micromolar ligands (potent and high potent drug candidates) [41].

Chemical Space-driven virtual screening can retrieve >80-90% of experimentally confirmed hits with substantial reductions in compute cost, effectively enriching for hits compared with brute-force screens [42,43].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Ligand-based 3D shape and chemotype similarity tools (e.g., SHAFTS) display excellent enrichment and scaffold-hopping capabilities. These methods serve as initial high-throughput triage steps, which are subsequently validated using downstream systems pharmacology filters like **AI-PBPK modeling**.

The integration of computer-aided drug design (CADD), machine learning (ML), and generative artificial intelligence (AI) has fundamentally restructured the early phases of pharmaceutical research and development (R&D). Computational platforms can screen billions of virtual molecules in a fraction of the time required by traditional high-throughput screening (HTS) [44–46]. However, a persistent paradox remains despite massive capital investment and the computational generation of thousands of optimized "hits," the structural rate of attrition in clinical pipelines has not materially changed [47–50]. There is a real translational gap that separate *in silico* models from *in vitro*, *in vivo*, and clinical reality and structural misalignments between physical binding affinities and real-world bioactivity. The challenges of dynamic target systems, the limitations of current absorption, distribution, metabolism, excretion, and toxicity (ADMET) modeling, and data fragmentation across the pharmaceutical pipeline still exist [51,52].

While computational frameworks have successfully compressed the timeline required to advance a compound from initial conceptualization to a Phase I trial (frequently down to under 18 months), they have introduced a distinct translational bottleneck [53–55]. The historical probability of success for an oncology or neurology asset transitioning from Phase I to full regulatory approval has remained stubbornly low (**Figure 1**). This indicates that while computational discovery methods excel at finding molecular keys that fit specific, isolated protein locks, they regularly fail to predict how those keys operate within the complex, interconnected networks of human physiology [56].

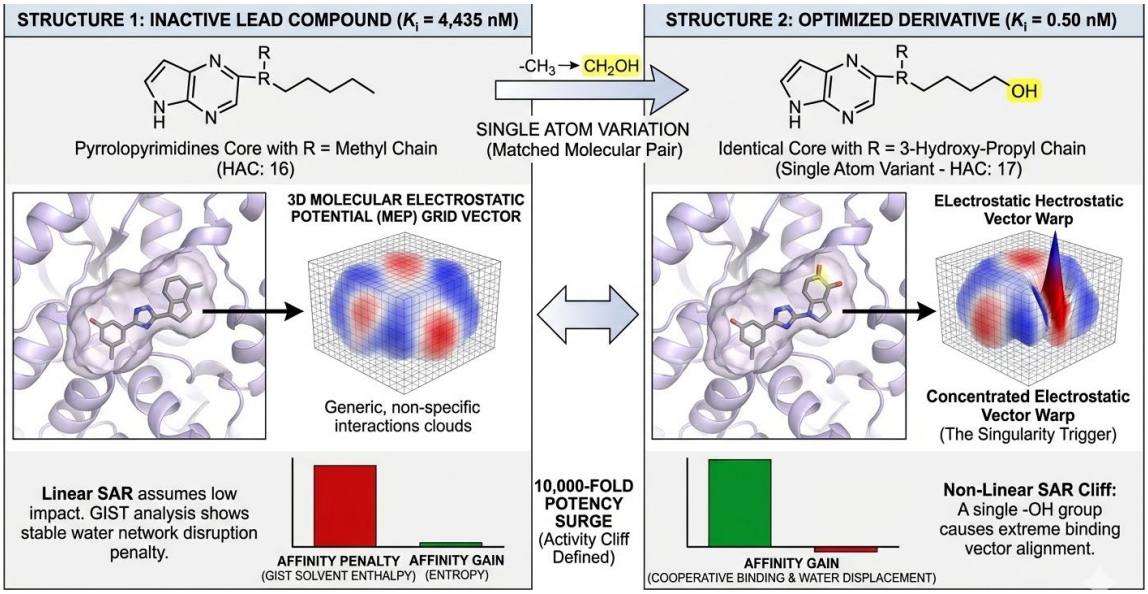

Figure 2. Molecular Topology of an Activity Cliff via a Matched Molecular Pair (MMPA). The upper panel contrasts flat 2D representations of a pyrrolopyrimidine core where a single atom variation (terminal modification changing a neutral methyl group to a 3-hydroxy-propyl chain; HAC16 →17) breaks the core molecular similarity principle. While linear structure-activity relationship (SAR) models predict marginal target potency variations, Grid Inhomogeneous Solvation Theory (GIST) profiling (middle/lower panels) reveals a catastrophic, non-linear warp in the 3D Molecular Electrostatic Potential (MEP) Grid Vector. This highly concentrated electrostatic singularity triggers an all-or-nothing cooperative binding event and optimal pocket water networks displacement, driving a localized, 10,000-fold affinity surge from an inactive lead (K_i=4,435nM) to a hyper-potent derivative (K_i = 0.500 nM).

Mapping Chemical Space for Lead Prioritization

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Modern AI-driven drug discovery increasingly depends on the ability to transform vast chemical libraries into navigable spatial representations [57]. High-speed cheminformatics platforms convert gigascale collections of virtual molecules into searchable descriptor-based spaces, while dimensionality-reduction algorithms generate visual maps that organize molecular diversity across interpretable coordinates [58,59]. These projections allow researchers to move beyond over-explored chemical regions and identify underrepresented areas with potential biological relevance [60]. Within these maps, active-learning and deep-learning workflows can then generate targeted hit maps, isolate enriched chemotypes, and prioritize clusters with predicted binding affinity. Tree-map categorization further refines this process by organizing candidates according to scaffold similarity, drug-likeness, and predicted ADMET behavior, enabling unstable, reactive, or poorly bioavailable compounds to be deprioritized before synthesis. Together, spatial projection, hit mapping, hierarchical visualization, and ADMET filtering convert chemical space from an abstract theoretical concept into a practical decision-making framework for lead identification and optimization [61–63].

Effectiveness, challenges, and translation

In silico and AI-driven methods have already proven highly effective at improving the efficiency of early drug discovery [64]. They can rapidly search ultra-large chemical libraries, enrich virtual screens for experimentally testable hits, support scaffold hopping, identify novel chemotypes, and prioritize compounds with more favorable predicted potency or ADMET profiles [65–68]. In this sense, their strongest contribution is not replacing experimental pharmacology but reducing the size of the experimental problem by converting enormous chemical possibility into smaller, more rationally selected candidate sets [69].

These gains, however, are translational rather than automatically clinical. A compound that appears promising in a computational screen may still fail when exposed to the layered realities of biological activity, cellular context, pharmacokinetics, metabolism, toxicity, formulation, and patient variability. Therefore, the success of CADD and AIDD should be measured not only by the number of generated hits or the speed of target-to-lead progression, but also by the extent to which predictions remain valid across *in vitro*, *in vivo*, and eventually clinical settings [38,39].

This distinction creates the conceptual bridge to the bottlenecks discussed below. Computational discovery is powerful precisely because it accelerates hypothesis generation, but its reliability depends on the quality of input data, the biological meaning of model labels, the coverage of the applicability domain, and the ability of models to represent dynamic protein behavior, solvent effects, systemic ADMET interactions, and closed-loop experimental validation [70,71].

Bottlenecks 1: Data quality, labeling inconsistencies, and limited domain coverage

The performance of any *in silico* method is fundamentally constrained by the quality and representativeness of the data on which it is developed [72]. Increasing dataset size or model complexity cannot compensate for systematically incorrect chemical structures, inconsistent representations, salts and mixtures, duplicates, improperly standardized chemotypes, conflicting annotations, or sparse coverage of relevant chemical and biological space [73]. Early work by Fourches, Muratov, and Tropsha showed that these apparently routine chemical-data problems can materially affect computational analyses and model performance, leading to the broader “trust, but verify” framework in which rigorous chemical and biological curation is treated as an essential component of model development rather than preliminary housekeeping [74,75].

For this reason, best-practice QSAR and machine-learning workflows require data curation, outlier assessment, appropriate dataset construction, external validation, and clear applicability-domain definition before predictions can be considered reliable [76–78]. Poorly curated input data can distort which compounds, targets, or chemical series appear most promising, and errors introduced during curation or annotation may be amplified when modern virtual screening narrows millions of candidates to a small experimental shortlist [79–82].

Biological data introduce an additional layer of uncertainty because model labels are conditional experimental summaries rather than absolute ground truth. An “active” or “inactive” designation can depend on assay format, biological system, concentration, exposure duration, endpoint, target

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

construct, cell line, and the threshold used to convert a continuous measurement into a categorical label [83]. Consequently, measurements assigned the same label are not necessarily biologically equivalent, while apparently conflicting records may reflect true context dependence rather than simple experimental error. Dataset modelability therefore depends not only on the number of available compounds, but also on whether chemically similar compounds display internally consistent biological responses under comparable experimental conditions [84].

Our experience curating antiviral bioactivity data illustrates the practical magnitude of this problem. Development of the Small Molecule Antiviral Compound Collection (SMACC) required extensive manual and automated reconciliation of ChEMBL data for 13 emerging viruses because nominally structured records frequently contained inconsistent ontology assignments, incomplete experimental descriptions, and biologically misleading target annotations [80,81]. Missing cell-type information, imprecise target labels, and uncertain activity calls can directly reshape machine-learning labels, causing models to learn dataset artifacts instead of true biological relationships.

A related problem is the distinction between negative evidence and missing evidence. Public databases often overrepresent positive or biologically interesting findings because successful experiments are more likely to be published and deposited than negative studies, while many inactive observations from screening campaigns remain inaccessible or incompletely described. The resulting datasets can distort class balance and the apparent boundaries of biological activity, causing models to learn what has historically been measured and reported rather than the full landscape of what is biologically possible [80,81].

This limitation connects directly to domain coverage. Computational models are most reliable where high-quality experimental evidence is dense, consistent, and biologically relevant; they may perform well in data-rich benchmarks but become unreliable when applied to underrepresented diseases, targets, mechanisms, assays, or chemical classes [11,20].

In classical QSAR modeling, this is formalized as the applicability-domain problem: a prediction should be considered reliable only when the query compound falls within a region of chemical space adequately represented by the training data. In contemporary AI-driven drug discovery, the same principle must be extended beyond chemistry to include biological targets, disease states, assay conditions, annotation quality, and experimental context. A numerical prediction can still be generated outside these domains, but the existence of an output does not mean that the prediction is supported by relevant evidence [75–79].

The key message is that rigorous data curation, provenance tracking, explicit treatment of conflicting evidence, and quantitative assessment of domain coverage are not optional technical details; they define when an *in silico* prediction can be trusted and when it should be considered unsupported. [20,74]

Ultimately, reliable computational drug discovery depends not simply on having more data, but on establishing fit-for-purpose datasets in which chemical identity, biological meaning, experimental context, uncertainty, and limits of inference are sufficiently well characterized. As computational approaches become more powerful, this requirement becomes more important because increasingly complex models can extract signals without automatically distinguishing biologically meaningful relationships from systematic artifacts [10,20].

Beyond the quality and coverage of the data themselves, a second source of uncertainty emerges from the way computational models simplify biological systems into static, isolated, and mathematically tractable representations [81–84].

Bottlenecks 2: Gaps between computational predictions and biological/clinical validation "Very few drugs are discovered primarily by computations"

Epistemological and Biophysical Gaps

The primary gap between computational output and biological reality stems from fundamental flaws in how multi-scale biophysical systems are translated into mathematical formulas [85,86]. A foundational pillar of *in silico* ligand screening is the optimization of binding affinity scoring functions — such as those utilized in Autodock Vina or Schrodinger's Glide — which treat ligand-

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

target interactions as isolated, equilibrium-driven events. Within these static frameworks, scoring functions prioritize the maximization of target receptor occupancy (θ), classically defined by the Hill-Langmuir equation: [41,87].

$$\theta = \frac{[L]^n}{K_d^n + [L]^n}$$

where [L] represents the free ligand concentration, K_d is the equilibrium dissociation constant, and n is the Hill coefficient.

However, a high-affinity binder (minimizing K_d) does not automatically translate into a downstream therapeutic effect. Computational docking algorithms routinely score molecules based strictly on static, time-independent thermodynamic interactions — such as hydrogen bonding networks, hydrophobic pocket matching, and salt bridges — at a fixed, rigid designated pocket [88].

In living tissue, cellular networks are governed by active homeostasis and non-linear, multi-compartment feedback loops that operate far from thermodynamic equilibrium. If a computational drug candidate achieves mathematically perfect target occupancy ($\theta \rightarrow 1$) and completely blocks a single kinase receptor, the biological system routinely bypasses this localized perturbation by upregulating alternative, parallel signaling pathways [89–91]. This homeostatic resilience can be modeled dynamically through the rate of change of a downstream cellular survival or metabolic effector (Y) using an ordinary differential equation (ODE):

$$\frac{d[Y]}{dt} = \left(\frac{V_{\max,1} \cdot [S_1]}{K_{m,1}\left(1 + \frac{[L]}{K_i}\right) + [S_1]} \right) + \left(\frac{V_{\max,2} \cdot [S_2]}{K_{m,2} + [S_2]} \cdot \alpha_{\text{up}} \right) - k_{\text{deg}}[Y]$$

Here, the targeted pathway (the first term) is effectively suppressed as a function of the ligand concentration and its inhibition constant (K_i). Yet, the absolute signal output of [Y] is maintained due to the secondary term, where (alpha_{up}) represents a dynamic feedback upregulation factor driven by the host network's homeostatic sensor:

$$\alpha_{\text{up}} = \alpha_0 + \frac{K_f^m}{K_f^m + [Y]^m}$$

As [Y] drops initially due to *in silico* optimized inhibition, the fractional feedback term approaches unity, maximizing (alpha_{up}) and driving flux through the alternative pathway substrate ($[S_2]$)). This systemic redirection renders a mathematically flawless ligand biologically inert. Consequently, traditional structure-based discovery frameworks introduce a profound epistemological error by treating a dynamic, adaptive biological network as a series of isolated, predictable protein-ligand locks. This process is summarized in **Figure 3**.

Proteins are not rigid, deterministic three-dimensional geometries; they exist as complex, time-dependent thermodynamic ensembles of moving conformations [92]. While deep learning structural prediction architectures—most notably DeepMind's AlphaFold2 (AF2) and AlphaFold3 (AF3) — have transformed structural biology by generating coordinate files with near-experimental accuracy, they introduce a systemic, architectural error when directly coupled with ligand-docking workflows. Because these models are trained to map sequences onto a singular, low-energy ground state (X0), they fundamentally fail to capture the continuous free-energy landscape (Delta G(X)) sampled by functional proteins in an intracellular environment. Under physiological conditions driven by local pH gradients, macromolecular crowding, and intracellular thermodynamics, the probability P(X) of a protein occupying any given conformational state (X) is governed by the Boltzmann distribution [92–96];

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

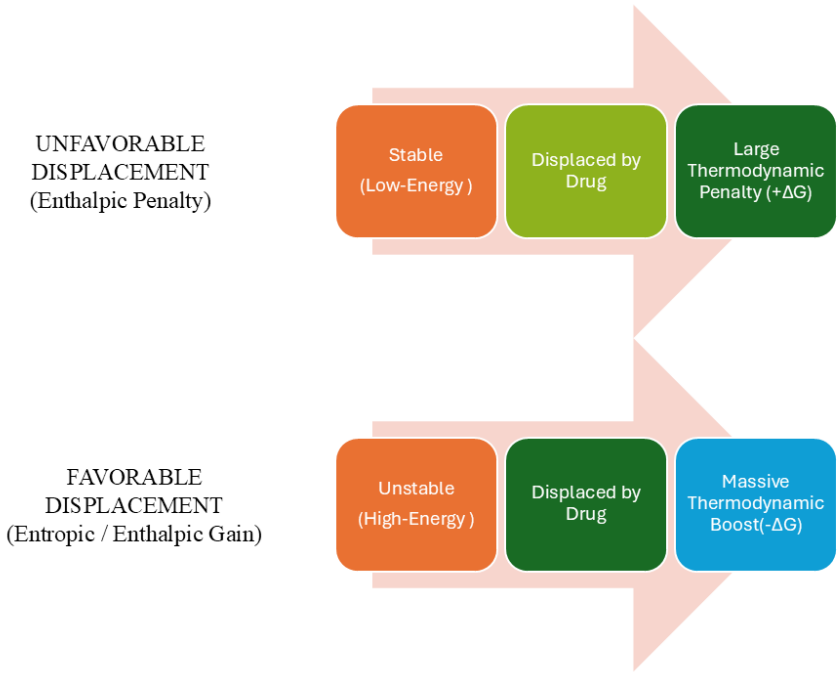

Figure 3. Favorable and unfavorable drug displacement.

$$P(X) = \frac{e^{-\beta E(X)}}{\int e^{-\beta E(X')} dX'}$$

where beta E(X) represents the potential energy of the specific conformation X, and beta is the thermodynamic beta expressed as 1/(kT), where k is the Boltzmann constant and T is absolute temperature. The denominator represents the configuration partition function, which integrates the Boltzmann factors over the entire accessible chemical and geometric coordinate space (X^I) to normalize the system. Mathematically, this distribution reveals why traditional structure-based pipelines are prone to high failure rates. By optimizing virtual screening pipelines against a lone static snapshot (X_0), docking algorithms implicitly treat a highly fluid physiological system as a static target. If the target snapshot X_0 possesses a high probability $P(X_0)$ approaches zero meaning algorithms are effectively searching for a structural "lock" that may only exist transiently or not at all *in vivo*, as reported in Figure 3.

Induced-Fit Docking (IFD) addresses structural rigidity by relaxing, localized side-chain coordinates while keeping the main-chain backbone rigid. However, this local approximation fails to capture large-scale, cooperative conformational transitions and cannot access the true thermodynamic ensemble. To overcome these limitations, modern workflows utilize long-duration Molecular Dynamics (MD) simulations paired with Markov State Models (MSMs) to organically sample natural energy landscapes and map high-probability conformational states. IFD merely perturbs localized amino acid side-chain coordinates (r_{side}) while maintaining a rigid main-chain backbone (R_{back}) [97–99]:

$$\mathbf{r}_{side} \rightarrow \mathbf{r}_{side} + \delta\mathbf{r}, \quad \text{subject to } \mathbf{R}_{back} = \text{constant}$$

This local approximation completely fails to simulate large-scale, cooperative structural transitions such as domain-swapping or loop closures, which regularly mask or completely assemble active binding sub-pockets. To bridge this translation gap and move beyond static constraints, modern discovery architectures run long-duration, all-atom Molecular Dynamics (MD) simulations

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

[94]. By immersing an AlphaFold-derived starting structure into an explicit solvent box containing N water molecules and physiological ions, the trajectory generates tens of thousands of structural snapshots over microsecond timescales via numerical integration of Newton's equations of motion [101,102]. The layout engine completely broke down the mathematical format here. It must be restored to its proper differential notation:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_i V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

These highly dimensional trajectories are subsequently clustered into discrete, functional states using Markov State Models (MSMs). The transitions between macroscopic conformational states are parameterized by a transition probability matrix T(tau), where each element T_{ij} re2presents the conditional probability that a protein in state i will transition to state j within a lag time tau:

$$T_{ij}(\tau) = P\big(x(t+\tau) = j \mid x(t) = i\big)$$

By executing virtual screening ensembles across the full spectrum of high-probability MSM states (sum_j T_{ij} = 1), the ligand is optimized to block the target throughout its entire natural physiological range of motion.

A secondary, yet equally critical blind spot in traditional structure-based screening is the mathematical treatment of the solvent environment within a target protein's binding pocket. Historically, molecular docking scoring functions treated protein cavities as empty space or modeled water as a uniform, continuous background dielectric fluid. In real biological systems, however, water molecules are active, dynamic participants whose local displacement drives the thermodynamics of binding [103–105].

The total standard free energy of binding (Delta G_{bind}) is fundamentally tied to whether an incoming ligand can favorably displace ordered water networks out of the pocket and back into the bulk solvent phase [99]:

$$\Delta G^{\circ}_{\text{bind}} = \Delta H_{\text{bind}} - T\Delta S_{\text{bind}} = (\Delta H_{\text{lig-prot}} + \Delta H_{\text{desolv}}) - T(\Delta S_{\text{lig-prot}} + \Delta S_{\text{desolv}})$$

When a target pocket contains highly unstable, trapped water molecules that are prevented from forming their ideal tetrahedral network of four hydrogen bonds with the surrounding hydration crust, they exist in a high-potential-energy state. To map these localized thermodynamic properties on a discrete voxel grid, advanced profiling tools such as Grid Inhomogeneous Solvation Theory (GIST) break down the binding cavity into a 3D grid of tiny voxels (V). For each voxel, the local water enthalpy (E_V) and translational/rotational entropy (S_V) are calculated relative to bulk solvent values (E_{bulk}, S_{bulk}):

$$\Delta E_V = \int_V \rho(\mathbf{r})\big(\langle E(\mathbf{r})\rangle - E_{\text{bulk}}\big)d\mathbf{r}$$

$$-T\Delta S_V = -T\left(S_{\text{trans},V} + S_{\text{rot},V} - S_{\text{bulk}}\right)$$

where rho(r) represents the local water density profile inside the voxel.

If a computational drug candidate is designed with a hydrophobic group that matches this voxel coordinate, it forces these unstable, high-energy water molecules out into the bulk solvent. Once

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

released, these water molecules rapidly form stable hydrogen bonds with the surrounding fluid, triggering a sharp entropic release (Delta S_{desolv} >> 0). This entropic displacement creates a massive favorable thermodynamic push (-Delta G), allowing the drug to adhere firmly to the pocket [105].

Conversely, if the pocket contains stable water molecules that have already formed strong, low-energy hydrogen bonds with the protein backbone, removing them introduces a severe enthalpic penalty (Delta H_{desolv} >> 0). If a generative AI architecture designs a molecule that forces the displacement of these stable water fluids without perfectly matching or compensating for those exact hydrogen bonds, the drug will fail to attach due to the massive free energy cost (+Delta G).

Because standard generative deep learning platforms struggle to process these dense, shifting 3D voxel data maps, they routinely design molecules that look flawless on simplified protein surfaces but fail completely in the wet lab [106,107].

The Multi-Scale Modeling Crisis in ADMET Predictions

Even if an in silico computational framework successfully generates a molecular entity possessing authentic binding affinity against an isolated biological target, the candidate must navigate highly complex systemic physiological barriers to achieve therapeutic efficacy [18,39,69]

Traditional computer-aided ADMET predictive modeling relies on fragmented, linear pipelines composed of isolated, binary filters. Within this legacy framework, generated chemical structures are evaluated sequentially through distinct, uncoupled algorithmic black boxes — such as isolated passive permeability classifiers, human plasma protein binding regressors, and *hERG* safety profiles. Because these independent models fail to account for co-dependent, non-linear physiological cross-talk, they generate severe structural blind spots that trigger catastrophic translational failures during *in vivo* validation [52,62].

For instance, optimizing a basic amine solely to minimize local cytochrome P450 (CYP) metabolic clearance frequently drives up its lipophilicity. While this shields the molecule in an isolated metabolic filter, it accelerates excessive real-world binding to human serum albumin (HSA), neutralizing the unbound free drug fraction required for therapeutic efficacy [108,109].

Furthermore, static, uncoupled filters are inherently blind to multi-stage biotransformation cycles and micro-environmental variations [110].

They completely fail to predict when benign parent compounds are converted by hepatic enzymes into highly toxic, reactive intermediate metabolites (bioactivation). Similarly, by assuming linear passive diffusion, they miss non-linear transporter saturation kinetics and fail to anticipate the rapid protonation and subsequent lysosomotropic trapping of basic drugs within acidic sub-cellular compartments. Because a living organism operates as a highly dynamic, contemporaneous network of biological interactions, these isolated, linear computational assumptions ultimately shield complex systemic toxicities from early virtual screening detection [111,112].

Table 2. Computational ADMET Prediction Methods and Their Corresponding Biological Obstacles and Translational Failure Mechanisms.

ADMET Phase	Computational Method	Biological Obstacle / Translational Failure Mechanism
Absorption	Caco-2 permeability & logP deep learning classifiers	Fails to predict unexpected active efflux via P-glycoprotein pump mechanisms.
Distribution	Plasma protein binding (PPB) quantitative models	Compounds bind excessively to human serum albumin, dropping free drug levels below therapeutic range.
Metabolism	Cytochrome P450 (CYP) regioselectivity models	Liver enzymes metabolize the compound into unstable, highly toxic structural variants.
Excretion	Renal clearance machine learning models	Fails to simulate organic anion/cation transport mechanics across real kidney membranes.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Toxicity	hERG channel virtual screening & Ames test classifiers	Misses complex systemic toxicities like drug-induced liver injury (DILI) and immune reactions.

These metabolites can be highly toxic, causing drug-induced liver injury (DILI) or severe heart arrhythmias. This is known as bioactivation or biotransformation failure cycle (**Figure 4**). Because isolated computational pipelines fail to model this multi-stage biotransformation before running toxicity checks, these dangerous side effects remain completely invisible during *in silico* screening. Due to these reasons and others, it is extremely important to Track these Metabolites and avoid any misleading information on their toxicity [111–113].

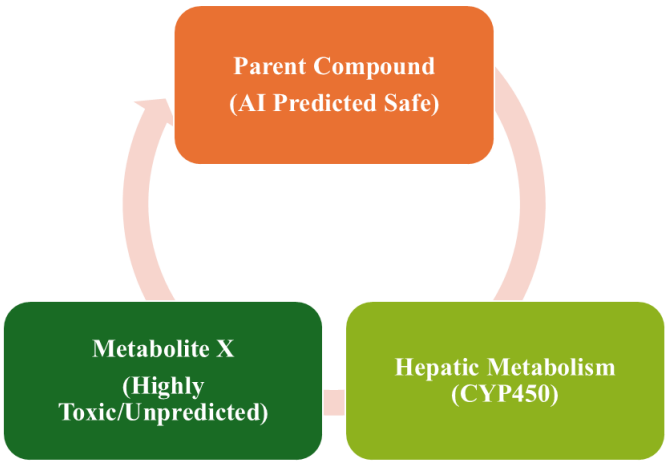

Figure 4. Biotransformation Failure Cycle.

Traditional computational ADMET tools heavily depend on linear assumptions of passive diffusion, relying strictly on molecular descriptors like molecular weight and (log P). By optimizing basic amines solely for neutral extracellular bilayers (pH 7.4), traditional architectures fail to anticipate the rapid protonation and subsequent lysosomotropic trapping (**Figure 5**) that occurs within acidic lysosomal compartments (pH 4.5–5.0), leaving serious intra-cellular toxicity completely invisible during early virtual screenings [62,63].

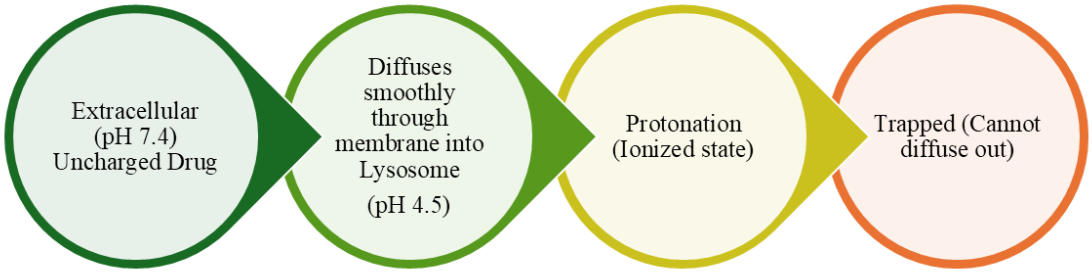

Figure 5. Lysosomotropic Trapping Mechanism.

Physiochemically, this lysosomotropic phenomenon is driven by the severe pH gradient existing between the neutral extracellular fluid (pH 7.4) and the acidic lysosomal lumen (pH 4.5–5.0). For basic compounds possessing a high (pKa), the Henderson-Hasselbalch equilibrium dictates that upon entering the acidic organelle, the molecule undergoes rapid protonation. Because the resulting ionized species cannot passively retro-diffuse across the lipophilic membrane, the drug becomes irreversibly sequestered, driving an intracellular concentration surge that can exceed extracellular

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

levels by orders of magnitude. Traditional, uncoupled ADMET models fail to capture this spatial restriction, mistaking structural permeability for therapeutic availability. This sub-cellular sequestration not only triggers direct vacuolation and cell death but severely depletes the free, unbound drug fraction required to engage downstream therapeutic targets. Mitigating this profound translational blind spot requires moving away from static, binary classifiers toward integrated, multi-compartment simulations capable of modeling dynamic ionization fluxes across complex cellular volumes.

The Technical Architecture of Systems ADMET and AI-PBPK Integration

To address the high failure rates caused by isolated ADMET filters, the pharmaceutical industry is moving toward integrated Systems ADMET architectures driven by Physiologically Based Pharmacokinetic (PBPK) models and machine learning. Rather than treating absorption, distribution, metabolism, and excretion as separate, unrelated endpoints, this approach uses AI-predicted molecular properties as inputs for multi-compartment ordinary differential equations (ODEs) that simulate how a drug candidate moves through a complete, connected organism over time [114–116].

- Stage 1 (Deep Learning Featurization): The molecular input—expressed as a SMILES string or a 3D graph — is processed through end-to-end Multitask Graph Neural Networks (GNNs) or Message Passing Neural Networks (MPNNs). Instead of calculating independent, uncoupled scores, the AI extracts a unified vector of interconnected biophysical and biochemical property predictions directly from the chemical topology.
- Stage 2 (Mechanistic PBPK Simulation Engine): This AI-predicted property vector is plugged directly into a closed mathematical system of multi-compartment ordinary differential equations (ODEs) that define the human body. Here, each major organ is mathematically parameterized by its actual physiological volume, blood perfusion rate, and tissue-to-plasma partition coefficients [115].

Computational drug discovery techniques can enhance early structural optimization and chemical synthesis. It cannot, however, work without complex biology and physical chemistry. Physical and biological validation is critical due to continuous limitations in knowing how molecules operate in living systems, regulating protein mobility, and mimicking systemic toxicity. To narrow this gap, integrated, closed-loop systems that prioritize the dynamic realities of human biology must replace disparate, standalone algorithms.

By simulating these multi-scale parameters simultaneously, the integrated framework continually tracks the dynamic relationship between free, active drug availability and systemic clearance over time. Moving from rigid, isolated filters to dynamic AI-PBPK modeling allows researchers to generate highly accurate virtual plasma concentration-time curves before chemical synthesis ever begins. This systemic integration successfully catches complex distribution, saturation, and safety bottlenecks early, drastically compressing target-to-Phase I trial timelines while mitigating pharmacokinetic attrition in clinical trials [109–115].

Emerging directions

The following emerging directions can be understood as complementary responses to the limitations described above, each aiming to improve reliability by combining better data selection, greater model transparency, stronger physical grounding, and tighter experimental feedback.

Active learning for low-data regimes

Active learning for low-data regimes optimizes computational drug discovery when large-scale training data is unavailable or expensive to generate. By employing automated, query-by-committee (QBC) iterative loops, the framework uses assemblages of machine learning potentials to quantify prediction uncertainty and intelligently select only the most informative molecular candidates for experimental assaying. This selective, high-efficiency data generation loop can fine-tune Chemical Language Models (CLMs) using single orphan compounds as templates. Ultimately, this process eliminates explicit, hand-crafted chemical design rules, enables accurate *de novo* molecular generation with nanomolar potency, and drastically reduces iterative wet-lab experimental cycles [117–119].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Explainable and "veridical" AI, physics-informed models, and tighter integration with experimental cycles

Explainable and "veridical" AI frameworks shift computational discovery away from predictive "black boxes" by integrating domain-specific scientific insights with machine learning models [114,115]. By combining causal reasoning with physics-informed Machine Learning (PIML), these systems move beyond the data-retrieval limits and hallucinations of standard Large Language Models to simulate complex physical systems—like fluid dynamics and thermodynamics — with high data efficiency in low-data regimes [116]. When paired with high-throughput experimentation (HTE) and system dynamics modeling, this tighter integration creates high-impact feedback loops that accelerate target identification and *de novo* molecule design. Ultimately, this approach reduces iterative experimental cycles and planning delays, while providing transparent, risk-aware project validation [123–125].

Disease-specific AI pipelines are revolutionizing precision oncology and cardiology by shifting diagnostics from static risk assessments to dynamic, individualized care [20,76,80]. In complex domains like pancreatic cancer, these frameworks combine physics-informed deep learning with multi-modal imaging (CT, MRI, PET) and radiomics to probe the tumor microenvironment and personalize targeted drug delivery. However, translating these advanced models into highly regulated medical environments creates significant regulatory bottlenecks due to data privacy concerns, a lack of common validation terminology, and poor model interpretability. To overcome these hurdles and achieve U.S. FDA acceptance, the industry must transition toward standardized international policies, contextual validation, and risk-based oversight to ensure true AI "model actionability" at the bedside [126,127].

Despite substantial advances in computational screening and predictive modeling, translation into clinically approved drugs remains constrained by data quality, interpretability limitations, and incomplete experimental validation [123–128].

Conclusion

Chemical space is not only an abstract universe of possible molecular structures, but the operational landscape in which drug candidates, biological targets, pharmacophores, molecular descriptors, and physiological networks are connected. In this review, chemical space is therefore presented as both a map and a design framework: it helps researchers define where known chemistry is concentrated, where unexplored regions may contain new therapeutic opportunity, and how molecular similarity can guide lead identification while still being disrupted by localized singularities such as activity cliffs.

In silico methods have become essential to modern drug discovery because they allow ultra-large virtual screening, scaffold hopping, hit-map generation, early ADMET prioritization, and rapid movement from chemical hypothesis to experimental shortlist. However, the central message of this paper is that computational speed does not automatically translate into biological or clinical reliability. Predictive models remain constrained by the quality, provenance, annotation consistency, and domain coverage of the datasets on which they are trained.

This limitation becomes especially important when AI models are applied beyond well-curated, experimentally dense regions of chemical and biological space. Poorly standardized structures, conflicting activity labels, missing assay metadata, and underrepresented disease or target classes can cause models to learn artifacts rather than mechanisms. For this reason, rigorous data curation, explicit uncertainty handling, applicability-domain assessment, and transparent validation should be considered core components of computational drug-discovery workflows rather than optional preprocessing steps.

The review also emphasizes that molecular recognition cannot be reduced to static ligand-target matching. Protein flexibility, solvent displacement, cellular feedback, metabolic bioactivation, transporter behavior, lysosomotropic trapping, and systemic ADMET interactions create multi-scale biological realities that isolated docking scores or linear filters often fail to capture. Future progress will therefore depend on closed-loop architectures that combine explainable and physics-informed

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

AI with molecular dynamics, solvent thermodynamics, Systems ADMET, PBPK modeling, and iterative experimental validation. In this sense, the next phase of AI-driven drug discovery will be defined not simply by faster prediction, but by more biologically realistic, evidence-aware, and clinically translatable computational systems.

Methods

The literature screening protocol, database retrieval criteria, and quantitative validation benchmarks for this series were initially established in Part 1. Briefly, a systematic multi-repository search was executed across public and proprietary data engines (including **PubMed**, **Scopus**, **IEEE Xplore**, and **ChEMBL**) targeting peer-reviewed publications tracking the mathematical and structural evolution of QSAR, CADD, and AIDD workflows spanning from **1980 to the present** [1].

For this Part 2 review, a subset of **97 studies** was rigorously isolated from an initial evaluation pool of **300 top-cited articles**. Study inclusion criteria were strictly restricted to frameworks presenting explicit quantitative evaluation parameters regarding:

1. **Chemical Singularity Overrides:** Algorithmic validation against localized activity cliffs.
2. **Pocket Solvent Thermodynamics:** Quantitative benchmarking using Grid Inhomogeneous Solvation Theory (GIST) or WaterMap voxel metrics.
3. **Multi-Scale Systems Integration:** Implementations coupling GNN-derived biochemical vectors with ordinary differential equation (ODE) based Physiologically Based Pharmacokinetic (PBPK) architectures.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Box 1: Core Theoretical Concepts and Mathematical Formulations in Computational Pharmacology

1. Molecular Similarity (Pharmacophore and Drug-Likeness Score)

- **Definition:** The operational assumption that structurally, sterically, or electronically similar molecules elicit similar biological responses.
- **Mechanistic Basis:** It shifts traditional pipelines from rigid, static matching toward multidimensional chemotype mapping (e.g., 3D ligand-based alignment metrics like **SHAFTS**). By mapping topological indices, shape features, and pharmacophoric points, models determine a compound's mathematical distance within an operational chemical space to prioritize synthetic candidates.
- **Historical Static Assumption (Target Receptor Occupancy):**

$$\theta = \frac{[L]^n}{K_d^n + [L]^n}$$

Where [L] represents free ligand concentration, K_d is the equilibrium dissociation constant, and n is the Hill coefficient. Legacy workflows rely strictly on maximizing target occupancy ($\theta \to 1$) via isolated, equilibrium-driven thermodynamic interactions, assuming a lower K_d automatically dictates downstream therapeutic success.

2. Chemical Singularity (Activity Cliffs)

- **Definition:** Highly localized regions in chemical space where minor structural modifications yield disproportionate, catastrophic disruptions in biological activity.
- **Mechanistic Basis:** Fractional atomic alterations—such as adding a single hydroxyl group or inverting a stereocenter—subtly alter a molecule's 2D topology but drastically warp its **3D Molecular Electrostatic Potential (MEP) Grid Vector**. This sudden vector transformation triggers an all-or-nothing cooperative binding event, causing a 100-fold to 1,000-fold surge or collapse in target binding affinity, completely defying linear structure-activity relationship (SAR) assumptions.

3. Pocket Solvent Thermodynamics (Grid Inhomogeneous Solvation Theory - GIST)

- **Definition:** The quantitative profiling of localized water networks trapped within a target protein's binding cavity to evaluate the thermodynamic penalties and gains of ligand-induced displacement.
- **Mechanistic Basis and Solvation Energy Voxel Calculations**

$$\Delta E_V = \int_V \rho(\mathbf{r})(\langle E(\mathbf{r})\rangle - E_{\text{bulk}})d\mathbf{r}$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

$$-T\Delta S_V = -T(S_{\text{trans},V} + S_{\text{rot},V} - S_{\text{bulk}})$$

Where $\rho(\mathbf{r})$ represents the localized water density profile inside a discrete 3D voxel coordinate (V), assessing local enthalpy (E_V) and translational/rotational entropy (S_V) relative to bulk solvent values. Displacing high-energy, unstable water molecules into the bulk solvent triggers a massive entropic push ($-\Delta G$). Conversely, removing stable, low-energy water molecules that are hydrogen-bonded to the protein backbone introduces severe enthalpic penalties, causing predicted hits to fail in wet-lab validation.

4. Polyphormacology (The Multi-Target Focus)

- **Definition**: The design or discovery of a single chemical entity (or a combination of therapeutics) that interacts with **multiple distinct biological targets** (receptors, enzymes, or ion channels) to exert a phenotypic therapeutic effect.
- **Core Concept**: Moving away from Paul Ehrlich's reductionist "magic bullet" (one drug, one target) toward intended, *selective promiscuity*. It focuses heavily on the structural capability of the ligand to match multiple pockets.

5. Network Pharmacology (The Systemic Focus)

- **Definition**: A highly integrated approach that combines **systems biology, network analysis, and network medicine** to map how multi-target drugs affect interconnected biological pathways, cellular cascades, and disease phenotypes.
- **Core Concept**: Viewing the organism as an "interactome" or graph theory network where proteins are **nodes** and interactions are **edges**. Instead of analyzing a drug as a static key in a lock, network pharmacology maps how triggering node A ripples through a dynamic pathway to alter complex disease modules without triggering toxicity.

6. Cellular Complexity (Basal and Induced Trans-membrane Communications)

- **Definition:** The multi-scale, dynamic network of biological barriers and homeostatic feedback loops that govern a drug candidate's behavior in living tissue, rendering isolated target inhibition insufficient.
- **Mechanistic Basis:**
    - *Basal State:* The initial, unperturbed physiological barriers and passive baseline properties (e.g., Caco-2 permeability thresholds or baseline metabolic clearance).
    - *Induced State:* The non-equilibrium cellular response triggered by localized drug perturbations. If an in silico optimized candidate successfully blocks a single node, living tissue routinely upregulates parallel, alternative signaling pathways to bypass the bottleneck.
- **Homeostatic Cellular Resilience (Dynamic Effector Efflux Line kinetics):**

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

$$\frac{d[Y]}{dt} = \left(\frac{V_{\max,1} \cdot [S_1]}{K_{m,1}\left(1 + \frac{[L]}{K_i}\right) + [S_1]} \right) + \left(\frac{V_{\max,2} \cdot [S_2]}{K_{m,2} + [S_2]} \cdot \alpha_{\text{up}} \right) - k_{\text{deg}}[Y]$$

$$\alpha_{\text{up}} = \alpha_0 \cdot \frac{K_f^m}{K_f^m + [Y]^m}$$

Where the rate of change of a downstream cellular survival or metabolic effector (Y) is modeled via an ordinary differential equation (ODE). Even if the targeted pathway (first term) is suppressed by ligand concentration and its inhibition constant (K_i), absolute effector output is maintained via alternative substrates (S_2) and a dynamic feedback upregulation factor ($alpha_{up}$), neutralizing a mathematically flawless ligand.

5. The Epistemological Paradox of Static Modeling

- **Definition:** The foundational mismatch introduced by treating dynamic, adaptive, and highly interconnected biological networks as a series of isolated, rigid protein-ligand locks.
- **Mechanistic Core:** While structural prediction models map sequence data onto singular, low-energy ground states (X_0), they fail to capture the continuous, time-dependent thermodynamic ensembles (P(X)) governed by Boltzmann distributions under physiological conditions. Reliable computational translation requires moving beyond standalone linear filters toward closed-loop Systems ADMET and dynamic AI-PBPK architectures capable of simulating non-linear, multi-compartment organismal kinetics.

6. SHAFTS stands for **SHApe-FeaTure Similarity**

It is a highly regarded 3D molecular similarity calculation and cheminformatics program used in computer-aided drug design (CADD). Instead of just analyzing flat 2D structures, SHAFTS adopts a **hybrid target-template overlay metric** that combines two crucial dimensions:

1. **Volumetric Molecular Shape:** It evaluates how well the overall 3D volume of a candidate molecule physically overlays with a known active reference compound (the template).
2. **Pharmacophore Features ("Colored" Chemistry Groups):** It simultaneously checks if the critical chemical groups—like hydrogen bond donors/acceptors, charge centers, and hydrophobic zones—line up precisely in 3D space.
3. **Scaffold Hopping Power:** Because it looks at 3D shape and pharmacophoric features rather than strict atom-to-bond connectivity, it excels at identifying entirely new chemical classes (chemotypes) that can fit into the same target pocket. This makes it a primary tool for discovering novel active matter before moving compounds into heavier, dynamic simulations like Molecular Dynamics or PBPK modeling.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Box 2: Taxonomy of the 20 Most Utilized Molecular Descriptors in Cheminformatics and QSAR Modeling

1. Constitutional Descriptors (Global Composition)

- **Molecular Weight (MW):** The sum of atomic masses; a fundamental metric for scaling transport phenomena, molecular size, and boundary flux.
- **Heavy Atom Count (HAC):** The total number of non-hydrogen atoms; sets the baseline molecular complexity scale for network partitioning.
- **Fraction Csp3 (fCsp3):** The ratio of sp^3-hybridized carbons to total carbon count; quantifies molecular fraction out-of-plane and correlates with solubility and melting point characteristics.

2. Topological & Connectivity Indices (2D Graph Invariants)

- **Wiener Index (W):** The sum of the shortest path distances between all pairs of heavy atoms in a molecular graph; tracks structural branching and compact geometry.
- **Randić Connectivity Index (χ):** Evaluates bond-degree matrices to quantify steric accessibility, surface branching, and structural densification.
- **Balaban's J Index:** A highly discriminating topological index based on distance matrices; minimizes degeneracy across complex congeneric series.
- **Kier & Hall Valence Connectivity Indices (χ^v):** Incorporates core and valence electron configurations into graph paths to capture electronic properties alongside structural connectivity.

3. Physicochemical & Hydrophobic Descriptors (Phase Partitioning)

- **Crippen Wildman LogP (SLogP):** Atom-centered fragmentation values that calculate the octanol-water partition coefficient; a core metric dictating passive bilayer permeability thresholds.
- **Topological Polar Surface Area (TPSA):** Sum of surface contributions from polar atoms (primarily nitrogen, oxygen, and attached hydrogens); heavily used to predict blood-brain barrier transport and cellular absorption.
- **Molar Refractivity (MR):** Combines molecular volume and polarizability metrics via the Lorentz-Lorenz equation to model steric bulk and dispersive target interaction forces.

4. Electrostatic & Charge Descriptors (Electronic Matrices)

- **Partial Charges (Gasteiger-Marsili):** Iterative calculation of (sigma)-electron charge distributions across localized atomic topologies to evaluate early coulombic forces.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

- **Maximum/Minimum Negative Partial Charge (Qmax/Qmin):** Identifies highly localized electron density centers responsible for hydrogen-bonding networks and target polarization.
- **BCUT Descriptors:** Eigenvalues extracted from modified adjacency matrices weighted by atomic charge, polarizability, and mass; essential for high-yield chemical space mapping.

5. Geometrical & 3D Conformation-Dependent Descriptors (Spatial Topology)

- **Asphericity (Ω):** Quantifies a molecule's deviation from spherical symmetry based on its principal moments of inertia; maps structural shape factors.
- **Eccentricity:** Measures the elongation of the molecular boundary ellipsoid, distinguishing linear structures from compact chemotypes.
- **Radius of Gyration (R_g):** Captures the distribution of molecular mass relative to its rotational center of inertia to gauge structural extension and effective bulk.

6. Molecular Fingerprints & Fragment Descriptors (Substructural Environments)

- **Extended-Connectivity Fingerprints (ECFP4 / Morgan Circular):** Iteratively hashes concentric atom environments up to a specified diameter; the standard descriptor for mapping chemical similarity and identifying localized activity cliffs.
- **Path-Based Fingerprints (RDKit / Daylight-style):** Hashes topological linear tracks up to a designated bond length to map structural connectivity profiles across larger chemical libraries.
- **MACCS Keys:** A fixed 166-bit dictionary mapping the presence or absence of predefined structural fragments and chemical functional patterns.
- **BertzCT Index:** A comprehensive graph invariant that quantifies topological complexity by combining structural skeleton branching matrices with heteroatom diversity profiles.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Box 3: Architecture of Dynamic AI-PBPK Integrated Systems and ODE Parameterization

1. Multitask GNN Featurization (The System Input Vectors)

- **Mechanistic Role:** Converts 2D/3D chemical graphs directly into multi-scale biological inputs without relying on isolated, uncoupled filters. End-to-end Multitask Graph Neural Networks (GNNs) or Message Passing Neural Networks (MPNNs) process the molecular topology to extract a unified vector of interconnected biophysical and biochemical property predictions simultaneously.
- **Key AI-Predicted Parameters:**
  - $f_{u,p}$ **and** $f_{u,\text{mic}}$**:** Unbound fraction in plasma and fraction unbound in microsomal/hepatocyte assays, respectively.
  - **CL_{int}:** Intrinsic metabolic clearance mapped against localized Cytochrome P450 (CYP) regioselectivity profiles.
  - K_p**:** Organ-specific tissue-to-plasma partition coefficients, driving the directional flux of the drug asset.

2. Multi-Compartment Physiological Parameterization

To simulate how a molecule moves through a connected organism over time, the structural model maps actual biological dimensions across discrete anatomical compartments (e.g., Blood, Liver, Kidney, Gut, and Lung). Each organ is mathematically parameterized by its explicit physiological constraints:

- **Anatomical Volume (V_i):** The actual physical volume of the specific tissue or organ compartment i.
- **Blood Perfusion Rate (Q_i):** The volumetric flow rate of blood delivering the drug candidate to tissue compartment i.
- **Tissue-to-Plasma Partitioning Matrix ($K_{p,i}$):** The thermodynamic ratio of drug concentration in tissue i versus plasma at equilibrium, establishing whether an asset exhibits a high distribution volume or remains locked in systemic circulation.

3. Mechanistic Ordinary Differential Equations (ODEs)

Rather than evaluating absorption, distribution, metabolism, and excretion as standalone, static endpoints, the system tracks mass-balance kinetics dynamically. The rate of change of drug concentration over time across interconnected compartments is governed by simultaneous, non-linear differential equations:

- **Perfusion-Limited Tissue Compartment (e.g., Kidney, Muscle):**

$$\frac{dC_{\text{tissue}}}{dt} = \frac{Q_{\text{tissue}}}{V_{\text{tissue}}} \cdot \left(C_{\text{arterial}} - \frac{C_{\text{tissue}}}{K_p} \right)$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

Where the rate of drug accumulation is driven strictly by blood perfusion (Q) and restricted by the tissue partition coefficient (Kp).

- **Eliminating Organ Compartment with Non-Linear Kinetics (e.g., Liver Metabolism):**

$$\frac{dC_{\text{liver}}}{dt} = \frac{Q_{\text{liver}}}{V_{\text{liver}}} \cdot \left(C_{\text{arterial}} - C_{\text{venous, liver}}\right) - \frac{V_{\max} \cdot \left(C_{\text{liver}} \cdot f_{u,p}\right)}{K_m + \left(C_{\text{liver}} \cdot f_{u,p}\right)}$$

Where metabolic clearance scales non-linearly via Michaelis-Menten kinetics (V_{max}, K_m) using the unbound free drug fraction ($f_{u,p}$) as the driving force.

4. Clinical Translation and Concentration-Time Profiles

- **Virtual Plasma Curves:** By solving these multi-scale parameters simultaneously using ODE solvers, the dynamic AI-PBPK system generates highly accurate virtual plasma concentration-time profiles before chemical synthesis ever begins.
- **Catching Toxicity Bottlenecks:** This integration captures complex biological phenomena that isolated filters miss entirely — including transporter saturation kinetics, complex multi-stage distribution loops, and toxic systemic cross-talk (e.g., drug-induced liver injury or severe heart arrhythmias).
- **Clinical Window Optimization:** It maps the drug's behavior directly against the maximum tolerated concentration (MTC) and minimum effective concentration (MEC), allowing researchers to optimize the therapeutic window and compress target-to-Phase I trial timelines.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

References

1. Uzundurukan, A., Nelson, M., Teske, C., et al. (2025). Meta-analysis and review of in silico methods in drug discovery–part 1: Technological evolution and trends from big data to chemical space. *The Pharmacogenomics Journal*, *25*(3), Article 8. https://doi.org/10.1038/s41397-025-00368-z
2. David, L., Thakkar, A., Mercado, R., & Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: A review and practical guide. *Journal of Cheminformatics*, *12*, Article 56. https://doi.org/10.1186/s13321-020-00460-5
3. Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. *Chemical Reviews*, *119*(18), 10520–10594. https://doi.org/10.1021/acs.chemrev.8b00728
4. Lill, M. A. (2007). Multi-dimensional QSAR in drug discovery. *Drug Discovery Today*, *12*(23-24), 1013–1017. https://doi.org/10.1016/j.drudis.2007.08.004
5. Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical Reviews*, *96*(3), 1027–1044. https://doi.org/10.1021/cr950202r
6. Oprea, T. I., & Gottfries, J. (2001). Chemography: the art of navigating in chemical space. *Journal of combinatorial chemistry*, *3*(2), 157–166. https://doi.org/10.1021/cc0000388
7. Dobson C. M. (2004). Chemical space and biology. *Nature*, *432*(7019), 824–828. https://doi.org/10.1038/nature03192
8. Cheng, F., et al. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature Communications*, *9*, Article 2691. https://doi.org/10.1038/s41467-018-05116-5
9. Geerts, H., Wikswo, J., van der Graaf, P. H., Bai, J. P. F., Gaiteri, C., Bennett, D., Swalley, S. E., Schuck, E., Kaddurah-Daouk, R., Tsaioun, K., & Pelleymounter, M. (2020). Quantitative systems pharmacology for neuroscience drug discovery and development: Current status, opportunities, and challenges. *CPT: Pharmacometrics & Systems Pharmacology*, *9*(1), 5–20. https://doi.org/10.1002/psp4.12478
10. Lo, Y.-C., Rensi, S. E., Torng, W., & Altman, R. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, *23*(8), 1538–1546. https://doi.org/10.1016/j.drudis.2018.05.010
11. Niazi, S. K., & Mariam, Z. (2023). Recent advances in machine-learning-based chemoinformatics: A comprehensive review. *International Journal of Molecular Sciences*, *24*(14), Article 11488. https://doi.org/10.3390/ijms241411488
12. Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C., & Ahsan, M. J. (2021). Machine learning in drug discovery: A review. *Artificial Intelligence Review*, *55*(3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4
13. Carracedo-Reboredo, P., Liñares-Blanco, J., Rodríguez-González, A., et al. (2021). A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal*, *19*, 4538–4558. https://doi.org/10.1016/j.csbj.2021.08.011
14. Fuoco D. (2012). Classification Framework and Chemical Biology of Tetracycline-Structure-Based Drugs. *Antibiotics (Basel, Switzerland)*, *1*(1), 1–13. https://doi.org/10.3390/antibiotics1010001
15. Carlotti, B., Fuoco, D., & Elisei, F. (2010). Fast and ultrafast spectroscopic investigation of tetracycline derivatives in organic and aqueous media. *Physical chemistry chemical physics : PCCP*, *12*(48), 15580–15591. https://doi.org/10.1039/c0cp00044b
16. Fuoco D. (2012). A New Method for Characterization of Natural Zeolites and Organic Nanostructure Using Atomic Force Microscopy. *Nanomaterials (Basel, Switzerland)*, *2*(1), 79–91. https://doi.org/10.3390/nano2010079
17. Acosta-Murillo, R., Ortíz-Bayliss, J. C., & Zapata-Morín, P. (2026). Enhancing cancer drug discovery: QSAR modeling with machine learning and chemical representations. *PLOS One*, *21*, Article e0343654. https://doi.org/10.1371/journal.pone.0343654
18. Alqahtani, S. (2017). In silico ADME-Tox modeling: Progress and prospects. *Expert Opinion on Drug Metabolism & Toxicology*, *13*(11), 1147–1158. https://doi.org/10.1080/17425255.2017.1389897
19. Altharawi, A., & Alqahtani, S. M. (2026). Integrative computational chemistry approaches in modern drug discovery: Advances in docking, pharmacophore modeling, molecular dynamics, and virtual screening. *Pharmaceutics*, *18*(5), Article 0565. https://doi.org/10.3390/pharmaceutics18050565

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

20. Tropsha, A., Isayev, O., Varnek, A., Schneider, G., & Cherkasov, A. (2023). Integrating QSAR modelling and deep learning in drug discovery: The emergence of deep QSAR. *Nature Reviews Drug Discovery*, *23*(2), 141–155. https://doi.org/10.1038/s41573-023-00832-0
21. Reymond, JL., van Deursen, R., Blum L.C., Ruddigkeit, L. (2010). Chemical space as a source for new drugs. *Med. Chem. Commun.* 2010; 1 (1): 30–38. https://doi.org/10.1039/c0md00020e
22. Ancuceanu, R., Lascu, B. E., Drăgănescu, D., & Dinu, M. (2025). In silico ADME methods used in the evaluation of natural products. *Pharmaceutics*, *17*(8), Article 1002. https://doi.org/10.3390/pharmaceutics17081002
23. Azam, F., & Almahmoud, S. A. (2026). Open-source molecular docking and AI-augmented structure-based drug design: Current workflows, challenges, and opportunities. *International Journal of Molecular Sciences*, *27*(7), Article 3302. https://doi.org/10.3390/ijms27073302
24. Das, U. (2025). Generative AI for drug discovery and protein design: The next frontier in AI-driven molecular science. *Medicine in Drug Discovery*, *25*, Article 100213. https://doi.org/10.1016/j.medidd.2025.100213
25. Hansen, K., Biegler, F., Ramakrishnan, R., Pronobis, W., von Lilienfeld, O. A., Müller, K. R., & Tkatchenko, A. (2015). Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The journal of physical chemistry letters*, *6*(12), 2326–2331. https://doi.org/10.1021/acs.jpclett.5b00831
26. Deng, Y., & Roux, B. (2009). Computations of standard binding free energies with molecular dynamics simulations. *The Journal of Physical Chemistry B*, *113*(16), 5525–5543. https://doi.org/10.1021/jp807701h
27. Gimenez, B., Santos, M. S., Ferrarini, M., & Fernandes, J. (2010). Evaluation of blockbuster drugs under the rule-of-five. *Die Pharmazie*, *65*(2), 148–152. https://doi.org/10.31083/ph.2010.9733
28. Gramatica, P. (2007), Principles of QSAR models validation: internal and external. QSAR Comb. Sci., 26: 694-701. https://doi.org/10.1002/qsar.200610151
29. Tropsha A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular informatics*, *29*(6-7), 476–488. https://doi.org/10.1002/minf.201000061
30. Jorgensen W. L. (2004). The many roles of computation in drug discovery. *Science (New York, N.Y.)*, *303*(5665), 1813–1818. https://doi.org/10.1126/science.1096361
31. Lavecchia A. (2015). Machine-learning approaches in drug discovery: methods and applications. *Drug discovery today*, *20*(3), 318–331. https://doi.org/10.1016/j.drudis.2014.10.012
32. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature reviews. Drug discovery*, *18*(6), 463–477. https://doi.org/10.1038/s41573-019-0024-5
33. Sabe, V. T., Ntombela, T., Jhamba, L. A., Maguire, G. E. M., Govender, T., Naicker, T., & Kruger, H. G. (2021). Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *European Journal of Medicinal Chemistry*, *224*, Article 113705. https://doi.org/10.1016/j.ejmech.2021.113705
34. Ferreira, F. J. N., & Carneiro, A. S. (2025). AI-driven drug discovery: A comprehensive review. *ACS Omega*, *10*(28), 23889–23903. https://doi.org/10.1021/acsomega.5c00549
35. Gogoi, P. P., Singha, B., Longkumer, P., Boruah, N., & Sinha, U. (2026). Classical docking to machine learning based docking: Molecular docking in drug discovery. *Current Topics in Medicinal Chemistry*. https://doi.org/10.2174/0115680266424314251204071847
36. Hou, T., Wang, J., Li, Y., & Wang, W. (2010). Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, *51*(1), 69–82. https://doi.org/10.1021/ci100275a
37. Jabeen, F. (2024). *Application of machine learning and deep learning approaches in cheminformatics for drug discovery* [Doctoral dissertation, Carleton University]. https://doi.org/10.22215/etd/2024-16292
38. Wu, F.-X., Zhou, Y.-Q., Li, L., Shen, X., Chen, G., Wang, X.-Q., Liang, X., Tan, M., & Huang, Z.-N. (2020). Computational approaches in preclinical studies on drug discovery and development. *Frontiers in Chemistry*, *8*, Article 726. https://doi.org/10.3389/fchem.2020.00726

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

39. Ferreira, L. L. G., & Andricopulo, A. (2019). ADMET modeling approaches in drug discovery. *Drug Discovery Today*, *24*(5), 1157–1165. https://doi.org/10.1016/j.drudis.2019.03.015
40. Lusher, S. J., McGuire, R., Van Schaik, R. C., Nicholson, C. D., & De Vlieg, J. (2014). Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*, *19*(7), 859–868. https://doi.org/10.1016/j.drudis.2013.12.004
41. Yang, C., Chen, E., & Zhang, Y. (2022). Protein–ligand docking in the machine-learning era. *Molecules*, *27*(14), Article 4568. https://doi.org/10.3390/molecules27144568
42. Blum, L. C., & Reymond, J. L. (2009). 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, *131*(25), 8732–8733. https://doi.org/10.1021/ja902302h
43. Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, *45*(1), 177–182. https://doi.org/10.1021/ci049714+
44. Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature reviews. Drug discovery*, *4*(8), 649–663. https://doi.org/10.1038/nrd1799
45. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, *23*(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039
46. Jakhar, R., Dangi, M., Khichi, A., & Chhillar, A. K. (2020). Relevance of molecular docking studies in drug designing. *Current Bioinformatics*, *15*(7), 666–678. https://doi.org/10.2174/1574893615666191219094216
47. Kolluri, S., Lin, J., Liu, R., Zhang, Y.-W., & Zhang, W. (2022). Machine learning and artificial intelligence in pharmaceutical research and development: A review. *The AAPS Journal*, *24*(1), Article 34. https://doi.org/10.1208/s12248-021-00644-3
48. Lee, T.-S., Allen, B. K., Giese, T. J., et al. (2020). Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *Journal of Chemical Information and Modeling*, *60*(11), 5595–5623. https://doi.org/10.1021/acs.jcim.0c00613
49. Protti, Í. F., Rodrigues, D., Fonseca, S. K., Alves, R., Oliveira, R. B., & Maltarollo, V. (2021). Do drug-likeness rules apply to oral prodrugs? *ChemMedChem*, *16*(7), 1157–1165. https://doi.org/10.1002/cmdc.202000805
50. Cheng, Y., Gong, Y., Liu, Y., Song, B., & Zou, Q. (2021). Molecular design in drug discovery: A comprehensive review of deep generative models. *Briefings in Bioinformatics*, *22*(6), Article bbab344. https://doi.org/10.1093/bib/bbab344
51. Segall, M. D., & Barber, C. (2014). Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discovery Today*, *19*(5), 688–693. https://doi.org/10.1016/j.drudis.2014.01.006
52. Tong, X., Liu, X.-H., Tan, X., et al. (2021). Generative models for de novo drug design. *Journal of Medicinal Chemistry*, *64*(19), 14015–14047. https://doi.org/10.1021/acs.jmedchem.1c00927
53. Zeng, X., Wang, F., Luo, Y.-Z., et al. (2022). Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*, *3*(10), Article 100794. https://doi.org/10.1016/j.xcrm.2022.100794
54. Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: Progress in machine intelligence for rational drug discovery. *Drug Discovery Today*, *22*(11), 1680–1685. https://doi.org/10.1016/j.drudis.2017.08.010
55. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, *40*(Database issue), D1100–D1107. https://doi.org/10.1093/nar/gkr777
56. Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, *145*, 22–29. https://doi.org/10.1016/j.chemolab.2015.04.013
57. Fuoco D (2025) The stealth effect from a medicinal chemist perspective: definition and updates. Front. Drug Deliv. 5:1564120. https://doi.org/10.3389/fddev.2025.1564120
58. Song, Y.-X., Fürtös, A., Fuoco, D., Boumghar, Y., & Patience, G. S. (2023). Meta-analysis and review of cannabinoids extraction and purification techniques. *Canadian Journal of Chemical Engineering*, *101*(6), 3108-3131. https://doi.org/10.1002/cjce.24786
59. Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, *9*, 71. https://doi.org/10.1186/1741-7007-9-71

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

60. Zhang, Y., & Lee, A. (2019). Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning. *Chemical Science*, *10*(35), 8154–8163. https://doi.org/10.1039/c9sc00616h
61. Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods*, *44*(1), 235–249. https://doi.org/10.1016/S1056-8719(00)00107-6
62. Lipinski, C. A. (2004). Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies*, *1*(4), 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007
63. Pracht, P., , Bohle, F., , & Grimme, S., (2020). Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical chemistry chemical physics : PCCP*, *22*(14), 7169–7192. https://doi.org/10.1039/c9cp06869d
64. Ekins, S., Ring, B. J., Grace, J., McRobie-Belle, D. J., & Wrighton, S. A. (2000). Present and future in vitro approaches for drug metabolism. *Journal of pharmacological and toxicological methods*, *44*(1), 313–324. https://doi.org/10.1016/s1056-8719(00)00110-6
65. Egan, W. J., Merz, K. M., Jr, & Baldwin, J. J. (2000). Prediction of drug absorption using multivariate statistics. *Journal of medicinal chemistry*, *43*(21), 3867–3877. https://doi.org/10.1021/jm000292e
66. Boudovitch, D., Sakaya, A., Uzundurukan, A., Leroux, J., & Fuoco, D. (2023). Review of commercially available nano-drugs and nano-delivery systems: Challenges and perspectives. *Preprints*, Article 2023121480. https://doi.org/10.20944/preprints202312.2023121480
67. Fuoco, D., di Tomasso, J., Boulos, C., Kilgour, R. D., Morais, J. A., Borod, M., & Vigano, A. (2015). Identifying nutritional, functional, and quality of life correlates with male hypogonadism in advanced cancer patients. *Ecancermedicalscience*, *9*, 561. https://doi.org/10.3332/ecancer.2015.561
68. Xiong, G., Wu, Z., Yi, J., et al. (2021). ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research*, *49*(W1), W5–W14. https://doi.org/10.1093/nar/gkab255
69. Butina, D., Segall, M. D., & Frankcombe, K. (2002). Predicting ADME properties in silico: methods and models. *Drug discovery today*, *7*(11), S83–S88. https://doi.org/10.1016/s1359-6446(02)02288-2
70. Dearden J. C. (2003). In silico prediction of drug toxicity. *Journal of computer-aided molecular design*, *17*(2-4), 119–127. https://doi.org/10.1023/a:1025361621494
71. Ouma, R. B. O., Ngari, S. M., & Kibet, J. K. (2024). A review of the current trends in computational approaches in drug design and metabolism. *Discover Public Health*, *21*, Article 108. https://doi.org/10.1186/s12982-024-00229-3
72. Tutone, M., & Almerico, A. M. (2026). Computational strategies reshaping modern drug discovery. *Molecules*, *31*(2), Article 200. https://doi.org/10.3390/molecules31020200
73. Koirala, M., Yan, L., Mohamed, Z., & DiPaola, M. (2025). AI-integrated QSAR modeling for enhanced drug discovery: From classical approaches to deep learning and structural insight. *International Journal of Molecular Sciences*, *26*(19), Article 9384. https://doi.org/10.3390/ijms26199384
74. Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, *50*(7), 1189–1204. https://doi.org/10.1021/ci100176x
75. Fourches, D., Muratov, E., & Tropsha, A. (2016). Trust, but verify II: A practical guide to chemogenomics data curation. *Journal of Chemical Information and Modeling*, *56*(7), 1243–1252. https://doi.org/10.1021/acs.jcim.6b00129
76. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., & Tropsha, A. (2014). QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry*, *57*(12), 4977–5010. https://doi.org/10.1021/jm4004285
77. Golbraikh, A., Muratov, E., Fourches, D., & Tropsha, A. (2014). Data set modelability by QSAR. *Journal of Chemical Information and Modeling*, *54*(1), 1–4. https://doi.org/10.1021/ci400572x
78. Alves, V. M., Auerbach, S. S., Kleinstreuer, N., Rooney, J. P., Muratov, E. N., Rusyn, I., Tropsha, A., & Schmitt, C. (2021). Curated data in—trustworthy in silico models out: The impact of data quality on the

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

reliability of artificial intelligence models as alternatives to animal testing. *Alternatives to Laboratory Animals*, *49*(3), 73–82. https://doi.org/10.1177/02611929211029635

79. Martin, H. J., Melo-Filho, C. C., Korn, D., Eastman, R. T., Rai, G., Simeonov, A., Zakharov, A. V., Muratov, E., & Tropsha, A. (2023). Small molecule antiviral compound collection (SMACC): A comprehensive, highly curated database to support the discovery of broad-spectrum antiviral drug molecules. *Antiviral Research*, *217*, Article 105620. https://doi.org/10.1016/j.antiviral.2023.105620
80. Martin, H.-J., Melo-Filho, C. C., Zakharov, A. V., Muratov, E., & Tropsha, A. (2025). On the importance of data curation for knowledge mining in antiviral research. *Science Progress*, *108*(1), Article 00368504241301535. https://doi.org/10.1177/00368504241301535
81. Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947–1958. https://doi.org/10.1021/ci034160g
82. Alamo, D. D., Sala, D., Mchaourab, H. S., & Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, *11*, Article e75751. https://doi.org/10.7554/eLife.75751
83. Mitjavila-Domènech, N., Díaz-Holguín, A., Hu, H., Kahlous, N. A., Cabeza de Vaca, I., Wallner, B., & Carlsson, J. (2026). Improving AlphaFold2 Performance in Virtual Screens Targeting GPCRs by Enhancing Binding-Site Conformational Sampling. *Journal of chemical information and modeling*, *66*(10), 6045–6056. https://doi.org/10.1021/acs.jcim.6c00034
84. Mitjavila-Domènech, N., Díaz-Holguín, A., Hu, H., Kahlous, N. A., Cabeza de Vaca, I., Wallner, B., & Carlsson, J. (2026). Improving AlphaFold2 performance in virtual screens targeting GPCRs by enhancing binding-site conformational sampling. *Journal of Chemical Information and Modeling*, *66*(10), 6045–6056. https://doi.org/10.1021/acs.jcim.6c00034
85. Ma, Z., Ajibade, A., & Zou, X. (2024). Docking strategies for predicting protein-ligand interactions and their application to structure-based drug design. *Communications in Information and Systems*, *24*(3), 199–230. https://doi.org/10.4310/CIS.241021221101
86. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O., & Roitberg, A. E. (2018). Less is more: Sampling chemical space with active learning. *The Journal of chemical physics*, *148*(24), 241733. https://doi.org/10.1063/1.5023802
87. Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: From docking to molecular dynamics. *Current Opinion in Structural Biology*, *48*, 93–102. https://doi.org/10.1016/j.sbi.2017.10.010
88. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry*, *47*(7), 1750–1759. https://doi.org/10.1021/jm030644s
89. Moret, M., Friedrich, L., Grisoni, F. *et al.* Generative molecular design in low data regimes. *Nat Mach Intell* **2**, 171–180 (2020). https://doi.org/10.1038/s42256-020-0160-y
90. Ballarotto, M., Willems, S., Stiller, T., Nawa, F., Marschner, J. A., Grisoni, F., & Merk, D. (2023). *De Novo* Design of Nurr1 Agonists *via* Fragment-Augmented Generative Deep Learning in Low-Data Regime. *Journal of medicinal chemistry*, *66*(12), 8170–8177. https://doi.org/10.1021/acs.jmedchem.3c00485
91. Gupta, T., Gong, W., Ma, C., Pawlowski, N., Hilmkil, A., Scetbon, M., Rigter, M., Famoti, A., Llorens, A. J., Gao, J., Bauer, S., Kragic, D., Schölkopf, B., & Zhang, C. (2024). The essential role of causality in foundation world models for embodied AI. *arXiv*. https://doi.org/10.48550/arXiv.2402.06665
92. Ma, C., Zhang, H., Rao, Y., Jiang, X., Liu, B., Sun, Z., Song, Z., Gao, Y., Cui, Y., Liu, X., & Li, Z. (2025). AI-driven virtual cell models in preclinical research: technical pathways, validation mechanisms, and clinical translation potential. *NPJ digital medicine*, *9*(1), 25. https://doi.org/10.1038/s41746-025-02198-6
93. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., ... & Jumper, J. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*(7873), 590–596. https://doi.org/10.1038/s41586-021-03828-1
94. Schauperl, M., & Denny, R. (2022). AI-based protein structure prediction in drug discovery: Impacts and challenges. *Journal of Chemical Information and Modeling*, *62*(11), 2642–2651. https://doi.org/10.1021/acs.jcim.2c00026

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

95. Lyu, J., Kapolka, N. J., Gumpper, R., Alon, A., Wang, L., DiBerto, J. F., ... & Shoichet, B. K. (2024). AlphaFold2 structures guide prospective ligand discovery. *Science*, *384*(6697), Article eadn6354. https://doi.org/10.1126/science.adn6354
96. Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W., ... & Zhavoronkov, A. (2023). AlphaFold accelerates artificial intelligence powered drug discovery: Efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical Science*, *14*(6), 1443–1452. https://doi.org/10.1039/d2sc05709c
97. Lin, J.-H. (2011). Accommodating protein flexibility for structure-based drug design. *Current Topics in Medicinal Chemistry*, *11*(2), 171–178. https://doi.org/10.2174/156802611794863580
98. Campbell, A. J., Lamb, M., & Joseph-McCarthy, D. (2014). Ensemble-based docking using biased molecular dynamics. *Journal of Chemical Information and Modeling*, *54*(7), 2127–2138. https://doi.org/10.1021/ci400729j
99. Hidayat, A. F. A., & Mohamad, S. B. (2025). Incorporating target protein structure flexibility and dynamics in computational drug discovery using ensemble-based docking analysis. *Journal of Visualized Experiments*, *(220)*, Article e67174. https://doi.org/10.3791/67174
100. Fux, A., Rozenes, S., & Cohen, Y. (2026). The Impact of AI Integration on Project Lifecycle Dynamics. *Applied Sciences*, *16*(6), 2893. https://doi.org/10.3390/app16062893
101. Gapsys, V., Hahn, D. F., Tresadern, G., Mobley, D. L., Rampp, M., & de Groot, B. L. (2022). Pre-exascale computing of protein–ligand binding free energies with open source software for drug design. *Journal of Chemical Information and Modeling*, *62*(5), 1172–1177. https://doi.org/10.1021/acs.jcim.1c01445
102. Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J., & Hou, T. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chemical Reviews*, *119*(16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055
103. Ramsey, S., Nguyen, C., Salomon-Ferrer, R., Walker, R. C., Gilson, M. K., & Kurtzman, T. (2016). Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *Journal of Computational Chemistry*, *37*(21), 2029–2037. https://doi.org/10.1002/jcc.24417
104. Nguyen, C. N., Young, T. K., & Gilson, M. K. (2012). Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *The Journal of Chemical Physics*, *137*(4), Article 044101. https://doi.org/10.1063/1.4733951
105. Szalai, T. V., Bajusz, D., Börzsei, R., Zsidó, B. Z., Ilaš, J., Ferenczy, G. G., Hetényi, C., & Keserű, G. M. (2024). Effect of water networks on ligand binding: Computational predictions vs experiments. *Journal of Chemical Information and Modeling*, *64*(23), 8980–8998. https://doi.org/10.1021/acs.jcim.4c01291
106. Nizamuddin, S. F. S., Ikramuddin, M., & Dhore, N. S. (2025). Artificial intelligence in pharmaceutical sciences: Transforming drug discovery, formulation, and manufacturing. *Periodikos*, *2*(1), 145–158. https://doi.org/10.5281/zenodo.17945149
107. Narigina, M., Vindecs, A., Bošković, D., Merkuryev, Y., & Romanovs, A. (2025). AI-Powered Stroke Diagnosis System: Methodological Framework and Implementation. *Future Internet*, *17*(5), 204. https://doi.org/10.3390/fi17050204
108. Li, J., Fu, J., Geng, X., & Wang, H. (2026). Foundation models in clinical oncology: Progresses and perspectives. *Cancer letters*, *639*, 218220. https://doi.org/10.1016/j.canlet.2025.218220
109. Subhan, A., Manoharan, G. (2025). Advancing cancer care through artificial intelligence: from innovative models to clinical decision-making and regulatory integration. *Clin. Cancer Bull.* **4**, 23. https://doi.org/10.1007/s44272-025-00052-0
110. Fuoco, D., Patience, G. S., & Shokouhi, S. (2026). Perspectives in biomedicine: Superfoods and antioxidant power classification. *The Canadian Journal of Chemical Engineering*. https://doi.org/10.1002/cjce.70554
111. Driscoll, J. P., Sadlowski, C. M., Shah, N. R., & Feula, A. (2020). Metabolism and bioactivation: It's time to expect the unexpected. *Journal of Medicinal Chemistry*, *63*(12), 6303–6314. https://doi.org/10.1021/acs.jmedchem.0c00026
112. Wang, S., Ballard, T. E., Christopher, L. J., Foti, R. S., Gu, C., et al. (2023). The importance of tracking "missing" metabolites: How and why? *Journal of Medicinal Chemistry*, *66*(23), 15586–15612. https://doi.org/10.1021/acs.jmedchem.3c01293

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 11 November 2026 https://doi.org/10.20944/preprints202611.1119.v1

113. Giacomini, K. M., Huang, S. M., Tweedie, D. J., Benet, L. Z., Brouwer, K. L., Chu, X., ... & International Transporter Consortium. (2010). Membrane transporters in drug development. *Nature Reviews Drug Discovery*, *9*(3), 215–236. https://doi.org/10.1038/nrd3028
114. Zhang, J., Li, H., Zhang, Y., Huang, J., Ren, L., Zhang, C., Zou, Q., & Zhang, Y. (2025). Computational toxicology in drug discovery: Applications of artificial intelligence in ADMET and toxicity prediction. *Briefings in Bioinformatics*, *26*(1), Article bbaf533. https://doi.org/10.1093/bib/bbaf533
115. Chen, X., & Lin, Z. (2026). Integration of machine learning with PBPK and QSAR modeling approaches to facilitate drug discovery and development. *CPT: Pharmacometrics & Systems Pharmacology*, *15*(2), Article e70228. https://doi.org/10.1002/psp4.70228
116. Tanihata, S., & Iwata, H. (2026). Next-generation artificial intelligence for ADME prediction in drug discovery: From small molecules to biologics. *Yonago Acta Medica*, *69*(1), 1–13. https://doi.org/10.33160/yam.2026.02.001
117. Wang, L., Zhou, Z., Yang, X., Shi, S., Zeng, X., & Cao, D. (2024). The present state and challenges of active learning in drug discovery. *Drug Discovery Today*, Article 103985. https://doi.org/10.1016/j.drudis.2024.103985
118. Masood, M. A., Kaski, S., & Cui, T. (2025). Molecular property prediction using pretrained-BERT and Bayesian active learning: A data-efficient approach to drug design. *Journal of Cheminformatics*, *17*, Article 42. https://doi.org/10.1186/s13321-025-00986-6
119. Williams, H. J., Pickett, S. D., Baxter, A., & Palmer, D. S. (2026). Query Matters: How Selection Strategies Influence Active Learning in Drug Discovery. *Journal of chemical information and modeling*, *66*(6), 3288–3301. https://doi.org/10.1021/acs.jcim.5c02504
120. Baskin, I. I., Winkler, D., & Tetko, I. V. (2016). A renaissance of neural networks in drug discovery. *Expert Opinion on Drug Discovery*, *11*(8), 785–795. https://doi.org/10.1080/17460441.2016.1201262
121. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell* **2**, 573–584 (2020). https://doi.org/10.1038/s42256-020-00236-4
122. Fuoco, D. (2026). First Case of Viral Transmission Computer-To-Human: How Should It Be Classified? The Emerging of "Technonotic Viruses". Available at SSRN. http://dx.doi.org/10.2139/ssrn.6880403
123. Cinquin O. (2024). Steering veridical large language model analyses by correcting and enriching generated database queries: first steps toward ChatGPT bioinformatics. *Briefings in bioinformatics*, *26*(1), bbaf045. https://doi.org/10.1093/bib/bbaf045
124. Sharma, P., Chung, W. T., Akoush, B., & Ihme, M. (2023). A Review of Physics-Informed Machine Learning in Fluid Mechanics. *Energies*, *16*(5), 2343. https://doi.org/10.3390/en16052343
125. Moya, B., Badías, A., González, D. *et al.* Computational Sensing, Understanding, and Reasoning: An Artificial Intelligence Approach to Physics-Informed World Modeling. *Arch Computat Methods Eng* **31**, 1897–1914 (2024). https://doi.org/10.1007/s11831-023-10033-y
126. Tiwari, A., Mishra, S., & Kuo, T. R. (2025). Current AI technologies in cancer diagnostics and treatment. *Molecular cancer*, *24*(1), 159. https://doi.org/10.1186/s12943-025-02369-9
127. Rahdar, A., Mohammadzadeh, V., Razzaq, S., Shirzad, M., Fathi-Karkan, S., Bakhshi, A., Behzadmehr, R., Kharaba, Z., & Ferreira, L. F. R. (2026). Physics-informed deep learning sharpens nano diagnostics for elusive pancreatic cancer. *Seminars in oncology*, *53*(1), 152427. https://doi.org/10.1016/j.seminoncol.2025.152427
128. Higgins, D. C., & Johner, C. (2023). Validation of Artificial Intelligence Containing Products Across the Regulated Healthcare Industries. *Therapeutic innovation & regulatory science*, *57*(4), 797–809. https://doi.org/10.1007/s43441-023-00530-4

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.