Multitech Journal of Science and Technology (MJST)
Vol.3, No.8, 2026:917-938

Transforming Computational Pharmacy in Drug Discovery and Development: Integrating CADD, Bioinformatics, Artificial Intelligence, and Pharmacometric Modeling

Loso Judijanto
IPOSS Jakarta
Corresponding Author: Loso Judijanto losojudijantobumn@gmail.com

A R T I C L E I N F O

Keywords: Psychometrics; Psychological Measurement; Validity; Reliability; Item Response Theory; Computerized Adaptive Testing; Process Data; Artificial Intelligence; Algorithmic Fairness; Digital Ethics

Received : 2 June
Revised : 20 July
Accepted: 20 August

A B S T R A C T

Computational pharmacy has evolved from using molecular modeling and quantitative structure–activity relationship (QSAR) to a multidisciplinary ecosystem that integrates computer-aided drug design (CADD), bioinformatics, artificial intelligence (AI), molecular simulation, as well as pharmacokinetic and systems pharmacology modeling. This article aims to analyze how the convergence of these technologies shapes the workflow of modern drug discovery and development, and to identify opportunities, limitations, and research priorities that determine the translation of computational results into credible pharmaceutical decisions. The writing uses a qualitative literature review with an interpretative synthesis of reputable journal articles mainly published between 2020 and 2026. The review shows that the main value of computational pharmacy doesn't come from a single algorithm, but from the integration of the data–model–experiment cycle: databases and structure predictions expand target space; virtual screening, molecular dynamics, and free-energy calculations prioritize candidates; AI speeds up prediction and generative design; while PBPK and quantitative systems pharmacology bridge molecular findings to predictions of exposure, response, and populations. However, data quality, domain applicability, bias, interpretability, reproducibility, experimental validation, and regulatory acceptance still remain major challenges. The article concludes that the future of computational pharmacy will be shaped by hybrid physics- and data-based models, closed-loop experimentation, validated digital twins, and transparent, auditable model governance

DOI: https://doi.org/10.59890/mjst.v3i8.313
E-ISSN: 3089-9966
https://slamultitechpublisher.my.id/index.php/mjst/index

INTRODUCTION

The discovery and development of drugs is a high-risk process because decisions need to be made step by step under biological, chemical, pharmacokinetic, toxicological, and clinical uncertainty. The high failure rate in clinical development shows that just having more candidates in the early stages doesn't automatically lead to a successful drug; what's more important is improving decision quality, target selection, and the ability to predict a candidate's limitations before development costs rise. In this context, computational pharmacy has grown into a strategic tool to narrow the search space, integrate multidimensional evidence, and prioritize the most informative experiments, rather than replace experimental or clinical research (D. Sun et al., 2022; Sadybekov and Katritch, 2023; Vincent et al., 2022).

Advances in computing have changed the scale of problems we can tackle. Structure-based virtual screening can now explore libraries of compounds on a huge scale, while hierarchical synthon-based approaches cut down the number of molecules that need to be fully evaluated. Prospective studies show that ultra-large screening can lead to new chemotypes that are later confirmed experimentally, making computers function as prioritization machines that connect the digital chemical space with synthesis and biological testing (Gorgulla et al., 2020; Stein et al., 2020; Sadybekov et al., 2022).

The next transformation comes from AI and machine learning. Deep learning models are expanding QSAR, property prediction, molecular representation, generative design, and unstructured data analysis. However, recent literature also emphasizes that high performance on benchmarks doesn't necessarily mean prospective success: models can pick up dataset biases, lose accuracy when applied outside the training domain, or produce predictions that are hard for scientists and regulators to explain. So, the scientific urgency isn't just about developing more complex algorithms, but also about building validation processes, uncertainty estimation, interpretability, and comparisons with strong baselines (Schneider et al., 2020; Bender and Cortés-Ciriano, 2021a, 2021b; Jiménez-Luna et al., 2020; Pandey et al., 2022).

On the biological side, better and more connected data is changing how drug targets are understood. AlphaFold makes protein structure models more accessible, while PubChem, UniProt, the Protein Data Bank, and pharmacology databases offer complementary information about structure, function annotations, interactions, and chemical properties. The result is a shift from one-molecule–one-target analysis to integrating structure, genomics, proteomics, biological networks, and phenotypic evidence. Still, predictive structure models need to be assessed for biological context, conformational states, ligands, cofactors, and model quality before being used for medicinal chemistry decisions (Jumper et al., 2021; Tunyasuvunakool et al., 2021; Kim et al., 2021; UniProt Consortium, 2021; Burley et al., 2021; Akdel et al., 2022).

The expansion of computational pharmacy also happens after the hit and lead stages. Physiologically based pharmacokinetic (PBPK), pharmacometrics, and quantitative systems pharmacology (QSP) connect drug properties with physiological systems, biomarkers, disease mechanisms, dosage, and population

variations. This integration is important because translating molecular affinity into therapeutic success requires understanding exposure, tissue distribution, target engagement, dose–response relationships, and patient heterogeneity. The use of QSP in clinical development and regulatory documents shows that computational pharmacy is increasingly moving from a preclinical exploration tool toward model-informed drug development (Jamei, 2020; Nigam et al., 2020; Bai et al., 2020, 2021).

Even so, the literature still often talks about CADD, bioinformatics, AI, and pharmacometrics as separate methodological streams. This fragmentation risks creating unrealistic expectations for certain technologies and blurring the more important questions: at what decision point does a model add value, what evidence is needed to trust its predictions, and how are those predictions combined with experiments? Challenges related to data quality, reproducibility, model applicability, synthesis, protein flexibility, prospective validation, and cross-disciplinary communication are still identified as major hurdles for CADD and AI in drug discovery (Medina-Franco, 2021; Tropsha et al., 2024; Gao and Coley, 2020).

Based on that background, this article positions computational pharmacy as a decision-making system that connects chemical-biological data, molecular modeling, AI, and translation models. The analysis isn't just a technology inventory, but rather looks at the conceptual relationships between approaches and the conditions that make their integration scientific, reproducible, and valuable for drug development (Sadybekov and Katritch, 2023; Tropsha et al., 2024).

The research questions (RQ) being answered are:

RQ1. How have the evolution and integration of CADD, bioinformatics, AI/deep learning, molecular simulations, and pharmacometric modeling shaped modern computational pharmaceutics across the spectrum from drug discovery to development?

RQ2. What are the opportunities, limitations, and research priorities that determine how computational pharmaceutics methods can produce more predictive, explainable, reproducible, experimentally validated, and translationally and regulatorily acceptable decisions?

In line with these two questions, the goal of this article is: (1) to put together a conceptual synthesis of the evolution and convergence of main computational pharmacy methods; (2) to assess the contributions and limitations of each approach in the stages of target identification, hit discovery, lead optimization, ADME/Tox, and model-informed development; and (3) to outline a research agenda that places validation, uncertainty, data interoperability, and the integration of data–model–experiment as the foundation for the next steps in computational pharmacy development.

LITERATURE REVIEW

1. The Concept and Evolution of Computational Pharmacy

In this article, computational pharmacy is broadly understood as the use of theories, algorithms, data, and computing infrastructure to support decisions in pharmaceutical science, ranging from target identification and molecule design to ADME/Tox prediction, dose modeling, and clinical translation. This broad definition avoids reducing computational pharmacy to just molecular docking. CADD itself is an umbrella that includes ligand-based and structure-based design, QSAR/QSPR, virtual screening, molecular dynamics, free-energy calculations, cheminformatics, and various AI methods aimed at drug discovery questions (Medina-Franco, 2021; Sadybekov and Katritch, 2023).

The evolution of CADD shows a shift from models with limited chemical space to exploring giga-scale virtual libraries, using both experimental and predictive target structures, and combining data-driven models for screening and optimization. This change doesn't get rid of classic medicinal chemistry principles; on the contrary, the quality of structure–activity hypotheses, synthesis, experimental testing, and mechanistic understanding still determines whether computational outputs have real value (Gorgulla et al., 2020; Stein et al., 2020; Sadybekov et al., 2022).

Expanding chemical space also broadens the diversity of candidate sources. Natural products remain important because they occupy structural spaces that are different from many synthetic libraries, but structural complexity, material access, dereplication, and activity data require the integration of metabolomics, cheminformatics, and AI. So, computational innovation doesn't just increase the number of molecules we can explore, it changes how chemical diversity is organized and prioritized (Atanasov et al., 2021; Santana et al., 2021; Mullowney et al., 2023).

2. CADD: QSAR, Virtual Screening, Molecular Docking, and Molecular Simulation

QSAR links structural representations with the activity or properties of molecules, while the development of deep QSAR adds representations that are learned directly from data and deep learning architectures. The value of QSAR lies in its ability to quickly prioritize and map structure–activity relationships, but its validity heavily depends on the quality of the endpoints, representations, data splits, applicability domain, and external evaluation. Therefore, complex models still need to be compared with simpler baselines and tested against truly independent data (Tropsha et al., 2024; Bender and Cortés-Ciriano, 2021b).

Structure-based virtual screening and molecular docking predict the orientation and compatibility of ligands at the binding site and are really helpful for cutting down the number of candidates before testing. With bigger libraries, better docking algorithms, and automation, we can now screen billions of structures, but docking scores shouldn't be treated like actual binding energies. Things like protonation, tautomers, structural water, receptor flexibility, pocket choice, and scoring function limitations can change the rankings, so it's still important to use consensus methods and experimental validation (Gorgulla et al., 2020; Murail et al., 2021; Medina-Franco, 2021).

Molecular dynamics (MD) and free-energy calculations add a time dimension and conformational ensembles that a single docking pose can't capture. MD can check the stability of complexes, protein flexibility, interaction networks, and access to alternative conformational states, while alchemical or endpoint free-energy methods give estimates closer to the thermodynamics of binding during lead optimization. Higher accuracy usually comes with higher computational cost and more sampling, so a tiered strategy—docking for screening, then MD/free-energy for a subset of candidates—makes more sense than using just one method for all decisions (King et al., 2021; Brunt et al., 2022). Prospective evidence from ultra-large docking strengthens the argument that computational value emerges when direct predictions are linked to synthesis and testing. Identifying melatonin receptor ligands from over a hundred million structures and developing V-SYNTHES for a library of more than eleven billion compounds show that chemical space searching can be scaled up, but success still depends on the experimental stage that tests affinity, selectivity, and the biological effects of the candidates (Stein et al., 2020; Sadybekov et al., 2022).

3. Bioinformatics, Databases, and Medicinal Chemistry Integration

Bioinformatics provides the biological context needed so that molecule design doesn't get disconnected from disease mechanisms. Sequence data, functional annotations, protein structures, interactions, and phenotypes can be used to choose targets, identify domains or pockets, compare homologs, and connect targets with biological pathways. PubChem, UniProt, and the Protein Data Bank are examples of knowledge infrastructures that support chemical and biological analysis across studies, but using public data requires curating compound identities, standardizing endpoints, handling duplicates, and tracking provenance (Kim et al., 2021; UniProt Consortium, 2021; Burley et al., 2021).

Databases that link pharmacokinetics, targets, interactions, and chemical space expand cross-scale integration capabilities. PK-DB facilitates structured pharmacokinetic data, PROMISCUOUS 2.0 organizes drug–protein relationships for drug repositioning, while DrugSpaceX expands the searchable structural space for compound design and exploration. Its scientific value isn't just in the size of the database, but in interoperability, annotation quality, and the ability to track data sources so analyses can be reproduced (Grzegorzewski et al., 2021; Gallo et al., 2021; Yang et al., 2021).

Medicinal chemistry acts as a translator layer between computational outputs and molecules that can actually be made, are stable, active, selective, and have a viable ADME profile. Because of this, integrating computational pharmaceutics with medicinal chemistry needs to treat synthetic accessibility, novelty, liabilities, stereochemistry, and multi-parameter optimization as shared goals. Generative models that propose structures without considering ease of synthesis or the context of medicinal chemistry risk producing mathematical solutions that can't be followed up on (Gao and Coley, 2020; Grebner et al., 2020).

4. AI, Deep Learning, and Generative Molecular Design

AI in computational pharmacy covers supervised learning for predicting properties and activities, unsupervised or self-supervised learning for representations, graph neural networks for molecular structures, language models for sequential representations, and generative models for suggesting new molecules. Deep learning is most useful when there's enough data, well-defined endpoints, and evaluation reflects real-world use; with small or heterogeneous datasets, simpler methods can perform just as well or be more stable (Elbadawi et al., 2021; Vijayan et al., 2022; Liu et al., 2021).

Generative molecular design shifts the process from picking candidates from an available library to creating new structures that meet a specific objective function. Studies on low-data generative design, scaffold-constrained generation, and target-oriented deep learning show the flexibility of this approach in optimizing multiple properties at once. However, incomplete objective functions can push models to exploit weaknesses in predictors, so generative design has to integrate synthesizability, uncertainty, meaningful novelty, and experiments as real constraints (Moret et al., 2020; Grebner et al., 2020; Tong et al., 2021; Gao and Coley, 2020).

The impact of AI becomes stronger when models are connected with prospective experiments. The discovery of RIPK1 inhibitors using generative deep learning shows the possibility of combining model-based design, synthesis, and biological evaluation in an iterative cycle. That said, one prospective success can't be generalized into a claim that AI always speeds up drug development; results should be evaluated based on the type of target, data quality, number of experimental cycles, comparators, and downstream developability (Li et al., 2022; Bender and Cortés-Ciriano, 2021a; Moingeon et al., 2022).

Interpretability becomes a central issue because models can be used to select molecules, explain SAR, or support high-risk decisions. Explainable AI tries to identify features or substructures that influence predictions, but post hoc explanations aren't necessarily identical to causal mechanisms. That's why interpretability needs to be paired with uncertainty estimates, confirmatory experiments, and domain knowledge rather than being used as automatic justification for complex models (Jiménez-Luna et al., 2020; Schneider et al., 2020; Leite et al., 2021).

5. Protein Structure Prediction, Multi-omics, and Precision Drug Discovery

Breakthroughs in protein structure prediction have expanded access to three-dimensional models for targets that previously had no experimental structure. AlphaFold shows high accuracy across many protein classes, and applying it to the human proteome creates an initial resource for forming structural hypotheses. However, predictive models mostly depict specific structural states and don't always capture flexibility, ligand-induced conformations, complexes, post-translational modifications, or the relevant environment for pharmacology (Jumper et al., 2021; Tunyasuvunakool et al., 2021; Cramer, 2021).

Community evaluations of AlphaFold2 emphasize that confidence scores, local quality, domain boundaries, and biological context need to be considered before using the model for docking or mechanistic inference. AlphaFill shows

one way to improve it, by enriching structural models with ligands and cofactors based on homologous structures. Because of that, integrating structure prediction into CADD is best done as a workflow with quality control, ensemble modeling, and verification against experimental data when available (Akdel et al., 2022; Hekkelman et al., 2023; Thornton et al., 2021).

At the system level, multi-omics and AI expand target discovery as well as natural product discovery by linking biosynthetic gene clusters, metabolomics, biological activity, and network context. This approach helps find patterns that are hard to see from a single type of data, but the results still depend on harmonizing molecular identities, batch effects, annotation completeness, and functional validation. So, precision drug discovery requires data integration that is not just big, but also comparable and biologically meaningful (Atanasov et al., 2021; Mullowney et al., 2023).

6. PBPK, QSP, Virtual Patients, and Digital Twins

PBPK uses a physiological representation of organs and blood flow to connect drug properties with distribution and elimination, while QSP combines biological mechanisms, pharmacology, and disease dynamics to simulate system responses. Both expand computational pharmacology from the question 'does a molecule bind?' to 'how do exposure, target engagement, and response change with dose, time, therapy combinations, and patient characteristics?' (Jamei, 2020; Nigam et al., 2020).

QSP has particular value in multiscale and combinatorial problems, like immuno-oncology and neuroscience, because molecular-level mechanisms need to be mapped to observable biomarkers and outcomes. The modular QSP-IO platform and QSP reviews in neuroscience show the role of virtual populations, sensitivity analysis, and scenario simulations in evaluating hypotheses that are difficult to test one by one in clinical settings. Increased regulatory application also makes model credibility, context of use, calibration, validation, and documentation just as important methodologically as the model structure itself (Sové et al., 2020; Bloomingdale et al., 2021; Bai et al., 2020, 2021).

The digital twin concept takes this integration further by proposing a dynamic digital representation of patients or pharmaceutical processes that is updated with data. In precision medicine, health digital twins could potentially combine mechanistic models, longitudinal data, and sensors; in pharmaceutical manufacturing, digital twins can be used to understand and control processes. However, a useful digital twin needs to have a clear definition of state, update mechanisms, uncertainty, verification/validation, data governance, and limits of use; without that, the term digital twin risks just becoming a new label for ordinary simulation models (T. Sun et al., 2022; Venkatesh et al., 2022; Moreno-Benito et al., 2022).

7. Computing Infrastructure and Quantum Computing Prospects

GPUs and parallel computing have become key enablers for deep learning and large-scale simulations. These infrastructure changes boost the throughput of computational experiments, but they also bring trade-offs in energy, cost, portability, and reproducibility. Since hardware and software evolve rapidly, research papers need to document model versions, parameters, seeds, data, and

computing environments so that performance can be replicated and compared fairly (Pandey et al., 2022).

Quantum computing is often seen as a future technology for quantum chemistry, optimization, and graph problems related to drug discovery. Experimental evidence like programmable Gaussian boson sampling and hybrid quantum–classical studies show possible proofs of concept, while advances in neural-network wavefunctions also strengthen the connection between machine learning and high-precision electronic calculations (Yu et al., 2023; Gircha et al., 2023; Hermann et al., 2023).

However, the claim that quantum computing will soon replace classical simulations isn't widely supported. Analyses of ground-state quantum chemistry show that the evidence for a generic exponential quantum advantage is still lacking. So, quantum computing should be seen as an emerging research path that needs benchmarking against the best classical methods, rather than assuming automatic speed or accuracy improvements (Lee et al., 2023).

METHODOLOGY

This article uses a qualitative literature review with a narrative/state-of-the-art synthesis orientation, not a systematic literature review. This approach was chosen because the goal of the article is to explain conceptual evolution, connect heterogeneous methodological streams, compare assumptions and limitations, and formulate a research agenda, rather than calculate pooled effects or claim exhaustive literature coverage. A qualitative literature review allows researchers to read the literature interpretatively and develop an argumentative synthesis, while a state-of-the-art review emphasizes the current state of knowledge, the developmental paths that shaped it, and possible next directions (Barry et al., 2022; Sukhera, 2022; Furlong and Lester, 2023).

The search was conducted iteratively with a primary focus on journal articles from 2020–2026 relevant to computational drug discovery, CADD, QSAR, molecular docking, molecular dynamics, free-energy methods, cheminformatics, bioinformatics, AI/machine learning, generative molecular design, PBPK/QSP, digital twins, and quantum computing. Sources were prioritized if they were published in reputable peer-reviewed journals, had a DOI or verifiable publisher page, presented influential conceptual reviews, relevant methods, or prospective/empirical evidence clarifying the benefits and limitations of the approaches. Title searches, keywords, related citations, and backward/forward citation checking were carried out step by step until the main themes and differences in perspective could be adequately explained; PRISMA diagrams or claims that all potentially relevant articles had been identified were not used.

The analysis was done through reading and thematic coding of the role of technology in the drug development decision chain. Evidence was grouped into six themes: (1) CADD and molecular simulation; (2) databases and bioinformatics; (3) AI, explainability, and generative design; (4) protein structure and multi-omics data; (5) PBPK/QSP, virtual patients, and digital twins; and (6) computing infrastructure and emerging technologies. The synthesis then compared the contributions, uncertainties, validation needs, and translation

readiness of each theme to answer RQ1 and RQ2. Bibliographic consistency was checked by matching article metadata, year, journal, DOI, and access address; references that could not be verified were removed or replaced with traceable sources.

RESULTS AND DISCUSSION

1. Technological Convergence is Turning Computational Pharmacy into a Decision-Making System

Literature synthesis shows that the evolution of computational pharmacy is best understood as convergence, not the replacement of one technology by the next. Docking didn't become obsolete when deep learning came along; QSAR didn't disappear because of graph neural networks; and experiments weren't replaced by virtual screening. Instead, these methods form a hierarchy of evidence and a hierarchy of cost. Fast methods are used to narrow down the search space, more expensive physical methods are used for selected candidates, and experiments are used to test the most valuable hypotheses (Sadybekov and Katritch, 2023; Tropsha et al., 2024).

In the early stages, expanding the digital chemical space increases the chances of finding chemotypes different from traditional collections. VirtualFlow and V-SYNTHES show that screening scale can jump significantly if algorithms, library preparation, and infrastructure are designed for high throughput. However, increasing the library size only matters if chemical sampling, scoring, filters, and synthetic accessibility are well managed; a bigger search space can also lead to more false positives if quality control doesn't keep up with the scale (Gorgulla et al., 2020; Sadybekov et al., 2022; Gao and Coley, 2020).

AI speeds up several steps that used to be bottlenecks, like property prediction, prioritization, representation learning, and hypothesis generation. GPU computing makes training models and analyzing large datasets more practical. That said, computational gains shouldn't be used as a stand-in for pharmacological benefits. A model that speeds up predictions by a thousand times but doesn't improve decision quality or reduce failed experiments might have limited translational value. So, evaluation metrics need to shift from just benchmark accuracy to things like decision utility, prospective enrichment, calibration, and impact on the design–make–test–analyze cycle (Pandey et al., 2022; Bender and Cortés-Ciriano, 2021a).

2. Integration of Computational Pharmacy, Bioinformatics, and Medicinal Chemistry

The integration of bioinformatics has shifted CADD from optimizing known target structures to a more upstream process: target identification, variant mapping, functional annotation, isoform selection, and evaluation of disease association evidence. Experimental structures from the PDB and predictive models from AlphaFold can provide initial geometry, while UniProt and chemical data from PubChem help link targets to protein identity, ligands, and activity. That way, the quality of the target model becomes part of drug design quality; annotation errors or wrong conformational state choices can carry over

into docking and downstream AI (Burley et al., 2021; Jumper et al., 2021; Kim et al., 2021; UniProt Consortium, 2021).

AlphaFold expands the scope of structural hypothesis generation, but its integration with medicinal chemistry needs to be selective. High confidence in the backbone doesn't always mean the pocket is ready for docking, especially if the binding site is affected by ligand-induced fit, ions, metals, the membrane environment, or oligomerization. Community evaluations and efforts like AlphaFill show the importance of adding ligand/cofactor context and comparing models with homolog structures and biochemical data (Akdel et al., 2022; Hekkelman et al., 2023; Thornton et al., 2021).

Medicinal chemistry then functions as a reality check mechanism that filters computational output through questions about synthesizability, chemical stability, selectivity, permeability, metabolic liabilities, and multi-parameter trade-offs. This explains why the most valuable generative chemistry isn't the model that churns out the largest number of molecules, but rather the model that proposes a series of compounds that can be made, tested, their SAR studied, and optimized over a few cycles. Constraints on scaffold, synthesis, and developability properties therefore need to be part of the objective function from the start (Grebner et al., 2020; Tong et al., 2021; Gao and Coley, 2020).

This integration also expands opportunities for natural products. AI and omics can help link biosynthetic potential with structure and activity, but natural-product drug discovery still requires dereplication, isolation/production, structure determination, and biological validation. So, computational natural-product discovery is a clear example that data science and experiments need to be designed as one workflow, not as two separate stages (Atanasov et al., 2021; Mullowney et al., 2023).

3. AI's Contributions and Limits in Drug Discovery

AI literature shows two types of contributions that need to be distinguished. First, AI can improve operational efficiency by automating representation, classification, ranking, and pattern searching. Second, AI can expand the hypothesis space through de novo design or language-model-based generation. The first type of contribution is relatively easier to validate by comparing workflows, while the second type requires stronger proof because structural novelty does not necessarily come with pharmacological or developability novelty (Liu et al., 2021; Leite et al., 2021; Moingeon et al., 2022).

Generative deep learning gives an important example of both the potential and the risks. Models can optimize molecular representations toward target properties, but the designs will follow whatever is defined by the objective function and the predictive model. If the predictor is wrong or easy to exploit, the generator can produce structures with high computational scores but low actual quality. Low-data approaches, scaffold constraints, and synthesizability assessments have been developed to tackle these issues, while prospective studies like RIPK1 inhibitors show the need to close the loop through synthesis and bioassays (Moret et al., 2020; Gao and Coley, 2020; Li et al., 2022).

The second problem is generalization. Pharmaceutical data often isn't identically distributed: one project might have a different scaffold, target, assay format, lab, or endpoint than the training data. Random splits can give overly optimistic estimates if very similar analogs are spread across train and test. More relevant evaluation requires scaffold splits, temporal splits, external test sets, and prospective testing according to the context of use. This argument lines up with the critique that data quality and structure often matter more than AI architecture choice (Bender and Cortés-Ciriano, 2021b; Tropsha et al., 2024).

The third issue is explainability and uncertainty. Easily visualizable feature explanations are useful for communicating with medicinal chemists, but interpretations shouldn't go beyond what the model has actually learned. Confidence also needs to be calibrated, especially for out-of-domain compounds. Strong model governance should track data lineage, model version, training procedure, applicability domain, uncertainty, and human reasoning when the final decision is made. This kind of framework is more realistic for translation and regulation than treating the model as an autonomous black box (Jiménez-Luna et al., 2020; Schneider et al., 2020; Vijayan et al., 2022).

4. From Molecule to Patient: PBPK, QSP, and Model-Informed Development

One downside of computational pharmacy discussions that focus too much on discovery is the tendency to equate higher affinity with higher chances of clinical success. In reality, a candidate needs to hit the target at the right exposure, have a safety margin, maintain effects in a complex biological system, and produce relevant outcomes. PBPK and QSP fill this gap by incorporating physiology, mechanisms, and system dynamics into decision modeling (Jamei, 2020; Nigam et al., 2020).

PBPK is particularly useful for testing the consequences of changes in drug properties or physiology on tissue exposure and drug interactions, while QSP can integrate pathways, biomarkers, disease progression, and therapy combinations. QSP-IO shows how modular models can be used for immuno-oncology scenarios, and developments in neuroscience highlight the need to connect cross-scale data with network models, imaging, and digital biomarkers (Sové et al., 2020; Bloomingdale et al., 2021).

From a regulatory perspective, the increased use of mechanistic models demands adjustments to evidence standards. The context of use needs to be stated explicitly: models for hypothesis exploration require a different level of credibility compared to models that support dose decisions or replace certain data. Calibration and validation should use data that isn't identical to the data used to build the model, and uncertainty needs to be translated into decision risk. Submission analyses to the FDA show that QSP has entered various development contexts, but the details and depth of the models are still quite mixed (Bai et al., 2020, 2021).

Discovery–development integration is also a response to attrition. High clinical failure rates can't be fixed just by finding hits faster; the quality of target validation, translational models, biomarker strategy, patient selection, and dose rationale all need to advance together. Because of that, mature computational pharmaceutics should connect molecular decisions with exposure and

mechanism models as early as possible, so that developability or translatability issues can be spotted before investments ramp up (D. Sun et al., 2022).

5. Future Innovation: Closed-loop Modeling, Digital Twins, and Quantum Computing

The most promising direction isn't AI working on its own, but rather closed-loop modeling: the model chooses the next experiment, the experiment updates the data, uncertainty is evaluated, and the model is recalibrated. This principle can be applied to active learning for screening, lead optimization, formulation, or even clinical models. The advantage of a closed loop is that each iteration generates information to reduce uncertainty, so the goal of computation shifts from predicting as much as possible to picking experiments that best increase the project's knowledge (Sadybekov and Katritch, 2023; Tropsha et al., 2024).

Digital twins can be understood as an advanced form of closed-loop modeling if the digital representation is actually updated by data from the physical system. For patients, this has the potential to support personal simulations and adaptive decision support; in pharmaceutical processes, a digital twin can help with process understanding and control. However, clinical or industrial benefits require interoperability, real-time data quality, identifiability, uncertainty quantification, cybersecurity, and a clear definition of responsibility when model predictions don't match actual conditions (T. Sun et al., 2022; Venkatesh et al., 2022; Moreno-Benito et al., 2022).

Quantum computing needs to be approached with more caution. Proofs of concept in Gaussian boson sampling and hybrid quantum–classical generative chemistry show that some problem formulations in drug discovery can be explored with quantum devices. On the other hand, theoretical studies and benchmarks warn that exponential advantages in quantum chemistry aren't automatically guaranteed, and classical methods keep improving. A rational research strategy is to identify clear problem classes, compare them with the best classical baselines, and measure the total resource cost, not just the number of qubits or kernel runtime (Yu et al., 2023; Gircha et al., 2023; Lee et al., 2023).

Progress in neural-network wavefunctions also shows that the line between 'AI' and 'physics-based computation' is getting blurrier. Models can be used to represent wavefunctions or potential energy surfaces with high accuracy, while physics constraints can make machine learning more data-efficient and stable. The convergence of physics-informed or physics-constrained learning with data-driven models is likely to be an important path because it provides a balance between predictive flexibility and scientific consistency (Hermann et al., 2023).

6. Prospects and Challenges of Computational Pharmacy

The first challenge is data. Many model failures come from inconsistent labels, assay heterogeneity, missing data, batch effects, chemical identity errors, or class imbalance. Big datasets aren't automatically good; it's the curation, provenance, ontology, unit standardization, and experimental metadata that determine whether data can actually be combined. Public infrastructures like PubChem, UniProt, PDB, PK-DB, PROMISCUOUS, and DrugSpaceX are important, but users are still responsible for assessing the context and quality of

the data before developing a model (Kim et al., 2021; UniProt Consortium, 2021; Burley et al., 2021; Grzegorzewski et al., 2021; Gallo et al., 2021; Yang et al., 2021).

The second challenge is validation. CADD and AI can easily produce rankings or scores, but pharmaceutical decisions require prospective evidence. Re-docking or internal cross-validation only test certain parts of the problem and don't prove that candidates will be active in biological systems. Stronger study designs need to separate model development from external validation, define endpoints before looking at results, compare with a baseline, and report negative results. This is important so that literature bias doesn't give the impression that every computational method gives high enrichment (Medina-Franco, 2021; Bender and Cortés-Ciriano, 2021a).

The third challenge is reproducibility and transferability. Reproducing results requires data or detailed dataset descriptions, code or algorithms, environment, hyperparameters, random seeds, and preprocessing procedures. Transferability requires defining the applicability domain and monitoring when new data differs from development data. In industrial settings, this issue is made tougher by proprietary data and models, so internal documentation standards, audit trails, and external benchmarking become important (Jiménez-Luna et al., 2020; Pandey et al., 2022).

The fourth challenge is integration into organizations and regulations. Model outputs need to be understandable by medicinal chemists, pharmacologists, pharmacometricians, toxicologists, clinicians, and decision-makers. A model that’s super accurate but not linked to decision points, has no owner, or can’t be audited will be hard to use. QSP experience shows the importance of context of use and model credibility, while health digital twins add issues of data governance, implementation, and oversight (Bai et al., 2020, 2021; Venkatesh et al., 2022).

So the biggest prospect lies in building an integrative platform that brings together curated data, chemical and biological models, physical simulations, AI, and translational models in a workflow that has experimental feedback. Platforms like this let each model be used according to its strengths and limits, reducing the risk of overconfidence, and speeding up project learning. Computational pharma will become even more valuable if success is measured not by the complexity of algorithms, but by how much critical uncertainty can be reduced before expensive decisions are made (Sadybekov and Katritch, 2023; Tropsha et al., 2024).

7. Analysis and Synthesis Answering Research Questions

The answer to RQ1 is that modern computational pharmacy is built through layered integration. On the target and data layer, bioinformatics, databases, omics, and structure prediction create representations of biological problems. On the molecular layer, QSAR, docking, virtual screening, MD, and free-energy methods map the chemical space and prioritize candidates. On the data-driven layer, AI speeds up predictions, representations, and hypothesis generation. On the translation layer, PBPK, pharmacometrics, and QSP connect molecular properties with system exposure and response. The power of this approach shows up when these flows form a cycle that links predictions with

synthesis, experiments, and model updates, rather than when each is used separately (Gorgulla et al., 2020; Sadybekov et al., 2022; Jamei, 2020; Bai et al., 2021).

This convergence also explains why the traditional 'ligand-based versus structure-based' categorization is no longer enough to describe current practices. Deep QSAR can combine structural representations and target data; generative models can be given structure, property, and synthesizability constraints; AlphaFold expands the source of structures; and QSP can take input from biomarkers, omics, as well as pharmacology models. Therefore, a more useful unit of analysis is the decision problem: what decision needs to be made, what data is available, which model fits, how much uncertainty there is, and what experiment would be most informative next (Tropsha et al., 2024; Akdel et al., 2022; Moret et al., 2020).

The answer to RQ2 is that the biggest opportunities lie in scale, integration, and iteration, while the biggest limitation is the quality of evidence. Models that can explore billions of compounds, generate new structures, or simulate virtual patients can significantly expand the decision space, but they're only useful if predictions are calibrated, tested outside of development domains, and linked with experiments or clinical data. Research priorities therefore include data curation and provenance, prospective benchmarking, uncertainty quantification, scientifically relevant explainability, model governance, hybrid physics–AI methods, and reporting standardization (Jiménez-Luna et al., 2020; Gao and Coley, 2020; Bai et al., 2020; Medina-Franco, 2021).

Strategically, the next developments will be determined by the ability to connect discovery and development. Faster hit discoveries need to be followed by early predictions of developability, exposure, safety, and patient variability. Digital twins and quantum computing could be part of this ecosystem, but both need to be evaluated with strict benchmarks and context of use. So, the most credible computational pharma agenda isn’t about 'replacing experiments with computation,' but about optimizing the relationship between computation and experiments so that every cycle produces more informative and accountable decisions (Venkatesh et al., 2022; Lee et al., 2023; Sadybekov and Katritch, 2023).

CONCLUSIONS AND RECOMMENDATIONS

Computational pharmacy has evolved from a set of molecular modeling techniques into a decision-making ecosystem connecting chemistry, biology, data science, pharmacology, and clinical development. This literature review shows that the most important change is not the emergence of a single dominant technology, but the convergence of CADD, databases and bioinformatics, AI/deep learning, structure prediction, molecular simulation, and PBPK/QSP. Each method addresses different types of uncertainty and gets the most value when used sequentially and in a way that corrects each other (Sadybekov and Katritch, 2023; Tropsha et al., 2024).

For RQ1, this integration forms a layered workflow from target identification to model-informed development. Biological and structural data build the target context; QSAR, docking, and virtual screening narrow down the candidate space; MD and free-energy methods improve the understanding of

interactions; AI speeds up predictions and hypothesis generation; while PBPK and QSP translate candidate properties into exposure, biomarkers, dosage, and population responses. This convergence makes computational pharma relevant not just for hit discovery, but also for reducing translational uncertainty in later stages (Jamei, 2020; Bai et al., 2021; King et al., 2021).

For RQ2, the main limit is no longer just computational capacity. Data quality, domain applicability, bias, uncertainty, interpretability, reproducibility, synthetic feasibility, and prospective validation determine whether a model can be trusted. So, improvements in benchmark accuracy should be seen as early evidence, not conclusive proof. This manuscript emphasizes that strong computational claims should have a verification path to experiments, external data, or clinical evidence according to the context of use (Bender and Cortés-Ciriano, 2021a, 2021b; Jiménez-Luna et al., 2020; Medina-Franco, 2021).

The practical implication is the need for a closed-loop research architecture that connects design–make–test–analyze with model updates. Generative models need to be limited by synthesizability and medicinal chemistry; predictive structures need quality control; AI needs uncertainty and governance; and translational models need calibration and validation. Digital twins can extend this cycle to processes and patients if data updates, identifiability, and regulations are clearly defined, while quantum computing is still better treated as an emerging technology that needs to be proven against the classical baseline (Venkatesh et al., 2022; Hermann et al., 2023; Lee et al., 2023).

Future research should prioritize five agendas: building prospective benchmarks across projects and labs; developing hybrid models that combine physical constraints with data learning; standardizing uncertainty, data provenance, and model reporting; early integration of discovery with PBPK/QSP to assess developability and translation; and evaluating digital twins and quantum methods based on a clear context of use. These agendas will help shift computational pharma from algorithm competition to decision science that is more transparent, reproducible, and focused on reducing drug development risks.

FURTHER STUDY

This research still has limitations, so further studies on the topic of Transforming Computational Pharmacy in Drug Discovery and Development: Integrating CADD, Bioinformatics, Artificial Intelligence, and Pharmacometric Modeling are needed to improve the research and add insights for both the writers and the readers.

REFERENCES

Akdel, M. et al. (2022) 'A structural biology community assessment of AlphaFold2 applications', Nature Structural & Molecular Biology, 29, pp. 1056–1067. doi: 10.1038/s41594-022-00849-w. Available at: https://doi.org/10.1038/s41594-022-00849-w (Accessed: 14 August 2026).

Atanasov, A.G. et al. (2021) 'Natural products in drug discovery: advances and opportunities', Nature Reviews Drug Discovery, 20, pp. 200–216. doi: 10.1038/s41573-020-00114-z. Available at: https://doi.org/10.1038/s41573-020-00114-z (Accessed: 14 August 2026).

Bai, J.P.F. et al. (2021) 'Quantitative systems pharmacology: landscape analysis of regulatory submissions to the US Food and Drug Administration', CPT: Pharmacometrics & Systems Pharmacology, 10, pp. 1479–1484. doi: 10.1002/psp4.12709. Available at: https://doi.org/10.1002/psp4.12709 (Accessed: 14 August 2026).

Bai, J.P.F., Earp, J.C., Strauss, D.G. and Zhu, H. (2020) 'A perspective on quantitative systems pharmacology applications to clinical drug development', CPT: Pharmacometrics & Systems Pharmacology, 9, pp. 675–677. doi: 10.1002/psp4.12567. Available at: https://doi.org/10.1002/psp4.12567 (Accessed: 14 August 2026).

Barry, E.S., Merkebu, J. and Varpio, L. (2022) 'State-of-the-art literature review methodology: a six-step approach for knowledge synthesis', Perspectives on Medical Education, 11, pp. 281–288. doi: 10.1007/s40037-022-00725-9. Available at: https://doi.org/10.1007/s40037-022-00725-9 (Accessed: 14 August 2026).

Bender, A. and Cortés-Ciriano, I. (2021a) 'Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet', Drug Discovery Today, 26(2), pp. 511–524. doi: 10.1016/j.drudis.2020.12.009. Available at: https://doi.org/10.1016/j.drudis.2020.12.009 (Accessed: 14 August 2026).

Bender, A. and Cortés-Ciriano, I. (2021b) 'Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data', Drug Discovery Today, 26(4), pp. 1040–1052. doi: 10.1016/j.drudis.2020.11.037. Available at: https://doi.org/10.1016/j.drudis.2020.11.037 (Accessed: 14 August 2026).

Bloomingdale, P. et al. (2021) 'Quantitative systems pharmacology in neuroscience: novel methodologies and technologies', CPT: Pharmacometrics & Systems Pharmacology, 10(5), pp. 412–419. doi: 10.1002/psp4.12607. Available at: https://doi.org/10.1002/psp4.12607 (Accessed: 14 August 2026).

Brunt, D., Lakernick, P.M. and Wu, C. (2022) 'Discovering new potential inhibitors to SARS-CoV-2 RNA dependent RNA polymerase via high throughput virtual screening and molecular dynamics simulations', Scientific Reports, 12, 19986. doi: 10.1038/s41598-022-24695-4. Available at: https://doi.org/10.1038/s41598-022-24695-4 (Accessed: 14 August 2026).

Burley, S.K. et al. (2021) 'RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences', Nucleic Acids Research, 49(D1), pp. D437–D451. doi: 10.1093/nar/gkaa1038. Available at: https://doi.org/10.1093/nar/gkaa1038 (Accessed: 14 August 2026).

Cramer, P. (2021) 'AlphaFold2 and the future of structural biology', Nature Structural & Molecular Biology, 28, pp. 704–705. doi: 10.1038/s41594-021-00650-1. Available at: https://doi.org/10.1038/s41594-021-00650-1 (Accessed: 14 August 2026).

Elbadawi, M., Gaisford, S. and Basit, A.W. (2021) 'Advanced machine-learning techniques in drug discovery', Drug Discovery Today, 26(3), pp. 769–777. doi: 10.1016/j.drudis.2020.12.003. Available at: https://doi.org/10.1016/j.drudis.2020.12.003 (Accessed: 14 August 2026).

Furlong, D.E. and Lester, J.N. (2023) 'Toward a practice of qualitative methodological literature reviewing', Qualitative Inquiry, 29(6), pp. 669–677. doi: 10.1177/10778004221131028. Available at: https://doi.org/10.1177/10778004221131028 (Accessed: 14 August 2026).

Gallo, K. et al. (2021) 'PROMISCUOUS 2.0: a resource for drug-repositioning', Nucleic Acids Research, 49(D1), pp. D1373–D1380. doi: 10.1093/nar/gkaa1061. Available at: https://doi.org/10.1093/nar/gkaa1061 (Accessed: 14 August 2026).

Gao, W. and Coley, C.W. (2020) 'The synthesizability of molecules proposed by generative models', Journal of Chemical Information and Modeling, 60(12), pp. 5714–5723. doi: 10.1021/acs.jcim.0c00174. Available at: https://doi.org/10.1021/acs.jcim.0c00174 (Accessed: 14 August 2026).

Gircha, A.I. et al. (2023) 'Hybrid quantum-classical machine learning for generative chemistry and drug design', Scientific Reports, 13, 8250. doi: 10.1038/s41598-023-32703-4. Available at: https://doi.org/10.1038/s41598-023-32703-4 (Accessed: 14 August 2026).

Gorgulla, C. et al. (2020) 'An open-source drug discovery platform enables ultra-large virtual screens', Nature, 580, pp. 663–668. doi: 10.1038/s41586-020-2117-z. Available at: https://doi.org/10.1038/s41586-020-2117-z (Accessed: 14 August 2026).

Grebner, C., Matter, H., Plowright, A.T. and Hessler, G. (2020) 'Automated de novo design in medicinal chemistry: which types of chemistry does a generative neural network learn?', Journal of Medicinal Chemistry, 63(16), pp. 8809–8823. doi: 10.1021/acs.jmedchem.9b02044. Available at: https://doi.org/10.1021/acs.jmedchem.9b02044 (Accessed: 14 August 2026).

Grzegorzewski, J. et al. (2021) 'PK-DB: pharmacokinetics database for individualized and stratified computational modeling', Nucleic Acids Research, 49(D1), pp. D1358–D1364. doi: 10.1093/nar/gkaa990. Available at: https://doi.org/10.1093/nar/gkaa990 (Accessed: 14 August 2026).

Hekkelman, M.L., de Vries, I., Joosten, R.P. and Perrakis, A. (2023) 'AlphaFill: enriching AlphaFold models with ligands and cofactors', Nature Methods, 20, pp. 205–213. doi: 10.1038/s41592-022-01685-y. Available at: https://doi.org/10.1038/s41592-022-01685-y (Accessed: 14 August 2026).

Hermann, J. et al. (2023) 'Ab initio quantum chemistry with neural-network wavefunctions', Nature Reviews Chemistry, 7, pp. 692–709. doi: 10.1038/s41570-023-00516-8. Available at: https://doi.org/10.1038/s41570-023-00516-8 (Accessed: 14 August 2026).

Jamei, M. (2020) 'Where do PBPK models stand in pharmacometrics and systems pharmacology?', CPT: Pharmacometrics & Systems Pharmacology, 9(2), pp. 75–76. doi: 10.1002/psp4.12493. Available at: https://doi.org/10.1002/psp4.12493 (Accessed: 14 August 2026).

Jiménez-Luna, J., Grisoni, F. and Schneider, G. (2020) 'Drug discovery with explainable artificial intelligence', Nature Machine Intelligence, 2, pp. 573–584. doi: 10.1038/s42256-020-00236-4. Available at: https://doi.org/10.1038/s42256-020-00236-4 (Accessed: 14 August 2026).

Jumper, J. et al. (2021) 'Highly accurate protein structure prediction with AlphaFold', Nature, 596, pp. 583–589. doi: 10.1038/s41586-021-03819-2. Available at: https://doi.org/10.1038/s41586-021-03819-2 (Accessed: 14 August 2026).

Kim, S. et al. (2021) 'PubChem in 2021: new data content and improved web interfaces', Nucleic Acids Research, 49(D1), pp. D1388–D1395. doi: 10.1093/nar/gkaa971. Available at: https://doi.org/10.1093/nar/gkaa971 (Accessed: 14 August 2026).

King, E., Aitchison, E., Li, H. and Luo, R. (2021) 'Recent developments in free energy calculations for drug discovery', Frontiers in Molecular Biosciences, 8, 712085. doi: 10.3389/fmolb.2021.712085. Available at: https://doi.org/10.3389/fmolb.2021.712085 (Accessed: 14 August 2026).

Lee, S. et al. (2023) 'Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry', Nature Communications, 14, 1952. doi: 10.1038/s41467-023-37587-6. Available at: https://doi.org/10.1038/s41467-023-37587-6 (Accessed: 14 August 2026).

Leite, M.L. et al. (2021) 'Artificial intelligence and the future of life sciences', Drug Discovery Today, 26(11), pp. 2515–2526. doi: 10.1016/j.drudis.2021.07.002. Available at: https://doi.org/10.1016/j.drudis.2021.07.002 (Accessed: 14 August 2026).

Li, Y. et al. (2022) 'Generative deep learning enables the discovery of a potent and selective RIPK1 inhibitor', Nature Communications, 13, 6891. doi: 10.1038/s41467-022-34692-w. Available at: https://doi.org/10.1038/s41467-022-34692-w (Accessed: 14 August 2026).

Liu, Z. et al. (2021) 'AI-based language models powering drug discovery and development', Drug Discovery Today, 26(11), pp. 2593–2607. doi: 10.1016/j.drudis.2021.06.009. Available at: https://doi.org/10.1016/j.drudis.2021.06.009 (Accessed: 14 August 2026).

Medina-Franco, J.L. (2021) 'Grand challenges of computer-aided drug design: the road ahead', Frontiers in Drug Discovery, 1, 728551. doi: 10.3389/fddsv.2021.728551. Available at: https://doi.org/10.3389/fddsv.2021.728551 (Accessed: 14 August 2026).

Moingeon, P., Kuenemann, M. and Guedj, M. (2022) 'Artificial intelligence-enhanced drug design and development: toward a computational precision medicine', Drug Discovery Today, 27(1), pp. 215–222. doi: 10.1016/j.drudis.2021.09.006. Available at: https://doi.org/10.1016/j.drudis.2021.09.006 (Accessed: 14 August 2026).

Moreno-Benito, M. et al. (2022) 'Digital twin of a continuous direct compression line for drug product and process design using a hybrid flowsheet modelling approach', International Journal of Pharmaceutics, 628, 122336. doi: 10.1016/j.ijpharm.2022.122336. Available at: https://doi.org/10.1016/j.ijpharm.2022.122336 (Accessed: 14 August 2026).

Moret, M., Friedrich, L., Grisoni, F., Merk, D. and Schneider, G. (2020) 'Generative molecular design in low data regimes', Nature Machine Intelligence, 2, pp. 171–180. doi: 10.1038/s42256-020-0160-y. Available at: https://doi.org/10.1038/s42256-020-0160-y (Accessed: 14 August 2026).

Mullowney, M.W. et al. (2023) 'Artificial intelligence for natural product drug discovery', Nature Reviews Drug Discovery, 22, pp. 895–916. doi: 10.1038/s41573-023-00774-7. Available at: https://doi.org/10.1038/s41573-023-00774-7 (Accessed: 14 August 2026).

Murail, S., de Vries, S.J., Rey, J., Moroy, G. and Tufféry, P. (2021) 'SeamDock: an interactive and collaborative online docking resource to assist small compound molecular docking', Frontiers in Molecular Biosciences, 8, 716466. doi: 10.3389/fmolb.2021.716466. Available at: https://doi.org/10.3389/fmolb.2021.716466 (Accessed: 14 August 2026).

Nigam, S.K., Bush, K.T., Bhatnagar, V., Poloyac, S.M. and Momper, J.D. (2020) 'The systems biology of drug metabolizing enzymes and transporters: relevance to quantitative systems pharmacology', Clinical Pharmacology & Therapeutics, 108(1), pp. 40–53. doi: 10.1002/cpt.1818. Available at: https://doi.org/10.1002/cpt.1818 (Accessed: 14 August 2026).

Pandey, M. et al. (2022) 'The transformational role of GPU computing and deep learning in drug discovery', Nature Machine Intelligence, 4, pp. 211–221. doi: 10.1038/s42256-022-00463-x. Available at: https://doi.org/10.1038/s42256-022-00463-x (Accessed: 14 August 2026).

Sadybekov, A.A. et al. (2022) 'Synthon-based ligand discovery in virtual libraries of over 11 billion compounds', Nature, 601, pp. 452–459. doi: 10.1038/s41586-021-04220-9. Available at: https://doi.org/10.1038/s41586-021-04220-9 (Accessed: 14 August 2026).

Sadybekov, A.V. and Katritch, V. (2023) 'Computational approaches streamlining drug discovery', Nature, 616, pp. 673–685. doi: 10.1038/s41586-023-05905-z. Available at: https://doi.org/10.1038/s41586-023-05905-z (Accessed: 14 August 2026).

Santana, K. et al. (2021) 'Applications of virtual screening in bioprospecting: facts, shifts, and perspectives to explore the chemo-structural diversity of natural products', Frontiers in Chemistry, 9, 662688. doi: 10.3389/fchem.2021.662688. Available at: https://doi.org/10.3389/fchem.2021.662688 (Accessed: 14 August 2026).

Schneider, P. et al. (2020) 'Rethinking drug design in the artificial intelligence era', Nature Reviews Drug Discovery, 19, pp. 353–364. doi: 10.1038/s41573-019-0050-3. Available at: https://doi.org/10.1038/s41573-019-0050-3 (Accessed: 14 August 2026).

Sové, R.J. et al. (2020) 'QSP-IO: a quantitative systems pharmacology toolbox for mechanistic multiscale modeling for immuno-oncology applications', CPT: Pharmacometrics & Systems Pharmacology, 9(9), pp. 484–497. doi: 10.1002/psp4.12546. Available at: https://doi.org/10.1002/psp4.12546 (Accessed: 14 August 2026).

Stein, R.M. et al. (2020) 'Virtual discovery of melatonin receptor ligands to modulate circadian rhythms', Nature, 579, pp. 609–614. doi: 10.1038/s41586-020-2027-0. Available at: https://doi.org/10.1038/s41586-020-2027-0 (Accessed: 14 August 2026).

Sukhera, J. (2022) 'Narrative reviews: flexible, rigorous, and practical', Journal of Graduate Medical Education, 14(4), pp. 414–417. doi: 10.4300/JGME-D-22-00480.1. Available at: https://doi.org/10.4300/JGME-D-22-00480.1 (Accessed: 14 August 2026).

Sun, D., Gao, W., Hu, H. and Zhou, S. (2022) 'Why 90% of clinical drug development fails and how to improve it?', Acta Pharmaceutica Sinica B, 12(7), pp. 3049–3062. doi: 10.1016/j.apsb.2022.02.002. Available at: https://doi.org/10.1016/j.apsb.2022.02.002 (Accessed: 14 August 2026).

Sun, T., He, X., Song, X., Shu, L. and Li, Z. (2022) 'The digital twin in medicine: a key to the future of healthcare?', Frontiers in Medicine, 9, 907066. doi: 10.3389/fmed.2022.907066. Available at: https://doi.org/10.3389/fmed.2022.907066 (Accessed: 14 August 2026).

Thornton, J.M., Laskowski, R.A. and Borkakoti, N. (2021) 'AlphaFold heralds a data-driven revolution in biology and medicine', Nature Medicine, 27, pp. 1666–1669. doi: 10.1038/s41591-021-01533-0. Available at: https://doi.org/10.1038/s41591-021-01533-0 (Accessed: 14 August 2026).

Tong, X. et al. (2021) 'Generative models for de novo drug design', Journal of Medicinal Chemistry, 64(19), pp. 14011–14027. doi: 10.1021/acs.jmedchem.1c00927. Available at: https://doi.org/10.1021/acs.jmedchem.1c00927 (Accessed: 14 August 2026).

Tropsha, A. et al. (2024) 'Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR', Nature Reviews Drug Discovery, 23, pp. 141–155. doi: 10.1038/s41573-023-00832-0. Available at: https://doi.org/10.1038/s41573-023-00832-0 (Accessed: 14 August 2026).

Tunyasuvunakool, K. et al. (2021) 'Highly accurate protein structure prediction for the human proteome', Nature, 596, pp. 590–596. doi: 10.1038/s41586-021-03828-1. Available at: https://doi.org/10.1038/s41586-021-03828-1 (Accessed: 14 August 2026).

UniProt Consortium (2021) 'UniProt: the universal protein knowledgebase in 2021', Nucleic Acids Research, 49(D1), pp. D480–D489. doi: 10.1093/nar/gkaa1100. Available at: https://doi.org/10.1093/nar/gkaa1100 (Accessed: 14 August 2026).

Venkatesh, K.P., Raza, M.M. and Kvedar, J.C. (2022) 'Health digital twins as tools for precision medicine: considerations for computation, implementation, and regulation', npj Digital Medicine, 5, 150. doi: 10.1038/s41746-022-00694-7. Available at: https://doi.org/10.1038/s41746-022-00694-7 (Accessed: 14 August 2026).

Vijayan, R.S.K., Kihlberg, J., Cross, J.B. and Poongavanam, V. (2022) 'Enhancing preclinical drug discovery with artificial intelligence', Drug Discovery Today, 27(4), pp. 967–984. doi: 10.1016/j.drudis.2021.11.023. Available at: https://doi.org/10.1016/j.drudis.2021.11.023 (Accessed: 14 August 2026).

Vincent, F. et al. (2022) 'Phenotypic drug discovery: recent successes, lessons learned and new directions', Nature Reviews Drug Discovery, 21, pp. 899–914. doi: 10.1038/s41573-022-00472-w. Available at: https://doi.org/10.1038/s41573-022-00472-w (Accessed: 14 August 2026).

Yang, T., Li, Z., Chen, Y., Jiang, H. and Zheng, M. (2021) 'DrugSpaceX: a large screenable and synthetically tractable database extending drug space', Nucleic Acids Research, 49(D1), pp. D1170–D1178. doi: 10.1093/nar/gkaa920. Available at: https://doi.org/10.1093/nar/gkaa920 (Accessed: 14 August 2026).

Yu, S. et al. (2023) 'A universal programmable Gaussian boson sampler for drug discovery', Nature Computational Science, 3, pp. 839–848. doi: 10.1038/s43588-023-00526-y. Available at: https://doi.org/10.1038/s43588-023-00526-y (Accessed: 14 August 2026).