Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

NEURO-SYMBOLIC CHEMISTRY RECONNECTS LEARNED MOLECULAR REPRESENTATIONS WITH REACTION RULES, BIOLOGICAL MECHANISMS, AND SCIENTIFIC CONSTRAINTS

Juergen Hofmann[1]*, Jan Novotny[2], Eva Horvathova[3], Lukas Novak[4]

1. ***Department of Neuro-Symbolic Chemistry and Learned Representations, Faculty of Pharmacy, University of Veterinary Medicine Vienna, Vienna, Austria.***
2. ***Department of Reaction Rules and Biological Mechanisms, Faculty of Pharmacy, Mendel University in Brno, Brno, Czech Republic.***
3. ***Department of Scientific Constraints in Neuro-Symbolic Models, Faculty of Pharmaceutical Sciences, University of Pécs, Pécs, Hungary.***
4. ***Department of Neuro-Symbolic Reasoning in Drug Discovery, Faculty of Pharmacy, University of South Bohemia, České Budějovice, Czech Republic.***

ARTICLE INFO

Received:
17 May 2025
Received in revised form:
06 October 2025
Accepted:
09 October 2025
Available online:
28 October 2025

Keywords: Neuro-symbolic artificial intelligence, Molecular representation learning, Reaction rules, Mechanistic knowledge graphs, Chemical constraints, Uncertainty quantification

ABSTRACT

Machine-learning systems can encode molecular structures, predict chemical transformations, generate candidate compounds, and infer biomedical relationships, yet these capabilities are commonly distributed across models that do not share a coherent scientific reasoning framework. Learned representations can identify statistical regularities without preserving explicit chemical syntax, reaction applicability, biological context, or the evidentiary status of mechanistic claims. Conversely, symbolic systems can express rules and ontological relationships but may be brittle, incomplete, and poorly adapted to uncertain or previously unseen molecular contexts. This article proposes a neuro-symbolic molecular architecture designed to reconnect these representational modes through bidirectional reasoning. The proposed system contains learned molecular and reaction encoders, explicit chemical and biological knowledge layers, neural-to-symbolic grounding, symbolic-to-neural feedback, constraint adjudication, uncertainty estimation, contradiction preservation, provenance tracking, and task-specific reasoning interfaces. Its central contribution is not a new predictive score but an architectural account of how molecular hypotheses could be generated, checked, revised, qualified, or deferred through interactions among embeddings, reaction rules, mechanistic knowledge, and scientific constraints. The framework distinguishes syntactic validity from reaction feasibility, association from mechanism, model confidence from calibrated uncertainty, and benchmark performance from prospective pharmaceutical usefulness. It also specifies how the architecture should be evaluated separately in molecular design, synthesis planning, and mechanism inference. The proposed system remains conceptual: it has not been empirically validated, does not establish causal or therapeutic claims, and cannot substitute for experimental confirmation, medicinal-chemistry judgement, or regulatory assessment. Its intended value is to provide a disciplined foundation for developing molecular AI systems whose outputs are not merely plausible or fluent, but scientifically inspectable, constraint-aware, and appropriately bounded.

To Cite This Article: Hofmann J, Novotny J, Horvathova E, Novak L. Neuro-Symbolic Chemistry Reconnects Learned Molecular Representations with Reaction Rules, Biological Mechanisms, and Scientific Constraints. Pharmacophore. 2025;16(5):67-77. https://doi.org/10.51847/rv7Ot5GqYS

Introduction

Machine learning now contributes to molecular property prediction, inverse design, virtual screening, reaction modeling, knowledge organization, and other stages of pharmaceutical research. Contemporary pharmaceutical machine learning spans property prediction, inverse molecular design, and explainability, but these capabilities are usually developed as separate representational functions rather than as a unified chemically governed reasoning system [1-3]. A molecular encoder may therefore learn a useful latent organization of chemical space while remaining unable to state which valence rule, reaction condition, ontology relation, or biological mechanism supports a particular output. The problem is not simply that a model is

Corresponding Author: Juergen Hofmann; Department of Neuro-Symbolic Chemistry and Learned Representations, Faculty of Pharmacy, University of Veterinary Medicine Vienna, Vienna, Austria. E-mail: juergen.hofmann@vetmeduni.ac.at.

difficult to interpret. It is that the learned state and the scientific objects used by chemists and pharmacologists—structures, transformations, conditions, mechanisms, exceptions, and evidence traces—often remain computationally disconnected.

Chemistry has already provided important precedents for combining statistical prediction with explicit symbolic operations. Neural-symbolic approaches to retrosynthesis and reaction prediction demonstrated that learned models and structured chemical knowledge can be joined within a common reasoning task [4]. Yet this precedent has not matured into a general molecular architecture spanning candidate design, chemical feasibility, mechanistic inference, evidentiary contradiction, and uncertainty-aware decision support. Reaction prediction and route search remain narrower than pharmaceutical reasoning because a route that appears chemically plausible does not establish developability, biological mechanism, efficacy, or safety. A broader architecture must therefore retain the strengths of learned representations while exposing their outputs to multiple scientific knowledge systems that operate at different scales and with different standards of evidence.

The unresolved gap is architectural rather than merely algorithmic. Current models are commonly optimized for a defined endpoint, such as a molecular property, product structure, reaction template, drug–target link, or generative objective. These task formulations can be valuable, but they do not automatically preserve the distinction between molecular validity and pharmaceutical usefulness. An interpretable feature attribution may reveal which molecular substructures influenced a prediction, yet it does not establish why a transformation is feasible or whether a predicted biological relation is mechanistically causal. Likewise, a knowledge graph may provide a path between a compound and a phenotype, but graph connectivity alone cannot establish that the path represents an experimentally confirmed mechanism. The scientific challenge is consequently to create controlled interfaces among predictions, rules, mechanisms, uncertainties, and evidence rather than to assume that any one representation contains them all.

This article develops an original conceptual architecture termed the neuro-symbolic chemistry reasoning bridge. The proposed contribution comprises two coupled representational spaces: a learned space that captures molecular, reaction, and biological regularities, and an explicit scientific space containing chemical syntax, reaction constraints, ontological types, mechanistic relations, provenance, and contextual qualifiers. Between them, typed translation and adjudication components convert latent proposals into inspectable scientific objects, test those objects against available knowledge, preserve unresolved contradictions, and return qualified feedback to the learned system. This is not presented as an empirically validated model or universal solution. Rather, it is a methodological construct for specifying what must be connected, what must remain distinct, how claims should be evaluated, and where the system must abstain or defer to expert and experimental review.

Why Learned Representations and Scientific Rules Remain Disconnected

Learned molecular representations are typically optimized to preserve information useful for a selected predictive or generative task. Their scientific meaning is therefore conditional on the training objective, available data, molecular representation, and evaluation design. Analyses of learned molecular representations and critical appraisals of pharmaceutical AI show that benchmark performance, representation quality, and practical impact are not interchangeable [5-7]. A vector that supports accurate prediction on one dataset may not encode reaction applicability, stereochemical consequences, target context, or causal biological structure. Even when chemically similar molecules are positioned near one another, the geometry of the latent space need not correspond to explicit scientific categories. Similarity learned from assay labels, for example, can differ from similarity defined by mechanism, scaffold transformation, synthetic accessibility, or physicochemical behavior.

The disconnection is reinforced by the way pharmaceutical problems are divided into optimization tasks. Molecular generators are rewarded for satisfying calculated properties, predictive models are evaluated against held-out labels, and explanation systems are assessed by whether they identify influential features. These formulations rarely require the model to maintain an explicit account of why a proposed molecule is chemically permissible, what evidence supports a predicted mechanism, or which assumptions fail outside the training domain. AI-enabled drug design therefore remains dependent on iterative medicinal-chemistry reasoning, experimental interpretation, and project context rather than functioning as autonomous candidate generation [8]. The architectural implication is that expert knowledge should not be treated as a final visual explanation appended to a completed prediction. It must participate in hypothesis formation, constraint checking, revision, and deferral.

Prospective prediction studies demonstrate that learned representations can have substantive scientific value when they are embedded in carefully defined workflows and followed by experimental evaluation. Deep learning has, for example, been used to prioritize compounds for antibacterial testing and thereby support application-specific discovery [9]. Such results establish that learned models can guide experimental selection under particular task conditions. They do not establish that the model possesses a transferable theory of chemical reactivity, biological mechanism, developability, or therapeutic effectiveness. Predictive usefulness is therefore narrower than general scientific understanding. The proposed architecture preserves this distinction by allowing learned models to contribute candidate hypotheses while requiring separate evidence for syntactic validity, reaction feasibility, mechanistic coherence, uncertainty, and prospective value.

Scientific rules are also difficult to integrate because they are heterogeneous. Some rules are effectively hard constraints, such as basic graph well-formedness or impossible valence assignments. Others are defeasible patterns derived from reaction precedent, empirical medicinal chemistry, biological association, or incomplete ontology structure. Still others depend on solvent, catalyst, tissue, assay, disease state, dose, or temporal context. Encoding all such knowledge as fixed prohibitions would make the system brittle, whereas treating every constraint as a soft penalty would allow serious scientific violations to be traded against optimization rewards. The proposed contribution therefore separates hard constraints, soft preferences,

contextual conditions, and unresolved claims. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why learned representations and scientific rules remain disconnected within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why learned representations and scientific rules remain disconnected

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Learned molecular representations	What information is retained in an embedding, and for which task?	Representation quality depends on molecular encoding, training objective, dataset composition, and evaluation split.	Defines the flexible statistical component of the proposed architecture.	Treating a task-performing embedding as a complete representation of chemistry.	Predictive utility does not establish reaction knowledge, mechanism, causality, or pharmaceutical value.
Chemical and biological data suitability	Are available labels sufficiently reliable, contextualized, and relevant to the intended claim?	Assay heterogeneity, missing context, measurement error, selection bias, and inconsistent identifiers shape model behavior.	Motivates provenance, context qualification, and evidence-aware uncertainty.	Equating data abundance with scientific suitability.	Available data may remain inadequate for causal, mechanistic, or translational inference.
Explicit scientific rules	Which constraints must be represented as hard, soft, conditional, or revisable?	Chemical syntax, valence, stereochemistry, transformation precedent, ontologies, and mechanistic relations differ in certainty and scope.	Defines the symbolic knowledge space and its internal stratification.	Encoding uncertain empirical regularities as universal laws.	Rule status must remain visible, versioned, and conditional on scope.
Interpretability and explanation	Does an explanation reveal model dependence, scientific rationale, or both?	Feature attribution and concept activation can describe model behavior without establishing mechanistic truth.	Separates explanation interfaces from mechanism and causality claims.	Presenting a plausible explanation as experimental evidence.	Explanation fidelity is distinct from mechanistic validity.
Experimental and workflow context	How does a prediction enter a design–make–test–analyze cycle?	Pharmaceutical usefulness depends on expert review, experimental selection, assay design, and iterative correction.	Positions the architecture as decision support rather than autonomous discovery.	Converting retrospective or benchmark performance into prospective readiness.	Candidate prioritization requires independent experimental and project-specific validation.
Uncertainty and domain shift	When should the system qualify, revise, defer, or abstain?	Error likelihood may increase under sparse data, unfamiliar chemistry, missing rules, or distribution shift.	Supports selective reasoning and controlled escalation.	Treating confidence as proof of correctness or mechanistic support.	Model confidence, calibrated uncertainty, evidence uncertainty, and rule coverage must remain distinct.
Human scientific oversight	Which judgements cannot be delegated to model optimization?	Chemists, biologists, pharmacologists, and translational scientists interpret exceptions, trade-offs, and decision consequences.	Provides the final review and correction boundary in the architecture.	Using human presence as a substitute for validation or as justification for opaque automation.	Expert oversight is necessary but does not establish that the architecture is empirically validated.

Chemical Syntax, Reaction Constraints, and Mechanistic Knowledge

Chemical syntax is the first explicit layer required to convert a learned molecular proposal into an inspectable scientific object. At this level, the system must determine whether atoms, bonds, charges, aromaticity, stereochemistry, and molecular connectivity form a coherent representation. Sequence-based reaction models demonstrate that reaction mappings can be learned from molecular strings and that uncertainty can be attached to predicted products [10]. However, a high-probability sequence does not guarantee that the underlying structure is chemically valid, that the reaction is feasible under available conditions, or that the predicted product reflects the correct mechanistic pathway. The proposed architecture therefore treats syntax validation as a necessary gate rather than as a sufficient certificate of chemical plausibility. Invalid structures may be rejected or repaired, whereas unusual but permissible structures should be retained with appropriate uncertainty rather than excluded merely because they are rare in the training data.

Reaction constraints operate at a higher level than syntax because they concern transformations rather than isolated structures. Graph-based reactivity models can identify likely reaction centers and score candidate transformations from local molecular environments [11]. These learned proposals are useful because explicit reaction libraries alone may not cover new substrates or transformation variants. Yet reactivity is also conditional on catalysts, reagents, solvents, temperature, protection state, competing functional groups, and laboratory context. Models that predict reaction conditions show that these variables constitute a distinct information problem rather than incidental metadata [12]. In the proposed system, reaction rules are therefore represented as typed and scoped objects. Each rule should record the transformation pattern, applicability conditions, known exceptions, evidence provenance, and confidence in transfer beyond the documented domain.

Multi-step synthesis planning further demonstrates why a hybrid architecture is necessary. Learned reaction models can prioritize likely disconnections, while symbolic search can assemble those disconnections into candidate routes [13]. Neither component is sufficient alone. A purely symbolic planner can be brittle when its rule library is incomplete, whereas an unconstrained neural model may propose locally plausible steps that do not compose into a coherent route. The proposed reasoning bridge separates step proposal, rule matching, condition qualification, route search, and route-level adjudication. It also requires the planner to distinguish a known precursor from an accessible precursor, a precedent transformation from a robust transformation, and a computational route from an experimentally executable synthesis. Route scores should consequently be interpreted as search and prioritization signals rather than as expected yields, scale-up guarantees, or evidence of laboratory success.

Mechanistic knowledge extends the architecture beyond reaction feasibility into biological interpretation. Heterogeneous biomedical networks can connect drugs, targets, genes, pathways, phenotypes, and diseases and can support the prioritization of repurposing hypotheses [14]. Such networks provide explicit relational structure that latent molecular representations usually lack, but their edges differ in evidence type, directionality, context, and causal meaning. A drug–target relation may be experimentally measured, computationally inferred, literature extracted, or transferred from a related system; a target–disease relation may indicate association rather than causation. The proposed mechanistic layer therefore stores typed relations with provenance, biological context, and evidentiary status. It permits the system to generate candidate mechanism paths while preventing path existence from being reported as mechanistic confirmation. Chemical syntax, reaction rules, and biological mechanisms are thus connected but not collapsed: a molecule may be syntactically valid yet unsynthesizable, synthesizable yet biologically inactive, associated with a pathway yet unsupported as causal, or mechanistically plausible yet unsuitable as a pharmaceutical intervention.

Architecture of the Neuro-Symbolic Molecular System

The proposed neuro-symbolic molecular system begins from the premise that neural and symbolic components should exchange typed scientific objects rather than unstructured scores. Its learned layer may contain molecular-graph encoders, sequence models, reaction encoders, three-dimensional representations, or biological-state embeddings, but these representations remain task-conditioned statistical states. A neural-to-symbolic grounding interface converts selected latent outputs into candidate structures, reaction centers, transformation rules, ontology classes, drug–target relations, or mechanism paths. Probabilistic logic programming provides one formal precedent for treating neural outputs as uncertain facts that can participate in explicit inference [15]. In the proposed architecture, this interface would also preserve the source model, input domain, confidence estimate, representational assumptions, and transformation history associated with every grounded object. Grounding therefore does not certify an output as scientifically correct; it makes the output available for inspection, constraint testing, and evidence qualification.

The central adjudication layer distinguishes hard constraints, soft constraints, contextual conditions, and unresolved claims. Logic Tensor Networks illustrate how first-order logical expressions can be converted into differentiable satisfaction objectives that interact with learned representations [16]. Such a formalism may enable symbolic knowledge to influence training or inference, but chemical use requires additional semantic discipline because an ontology restriction, a valence condition, a reaction precedent, and a mechanistic hypothesis should not be assigned the same logical status. T-norm-derived loss functions provide mechanisms for graded satisfaction and differentiable knowledge integration [17]. Their use nevertheless requires explicit reporting of connective choice, aggregation behavior, constraint weighting, and exception handling. Otherwise, an apparently interpretable constraint score could conceal trade-offs in which a serious chemical violation is offset by multiple weakly satisfied preferences.

The explicit knowledge layer contains chemical syntax rules, reaction transformations, condition qualifiers, chemical and biological ontologies, mechanistic relations, and provenance records. Biological neuro-symbolic representation learning has shown that ontology axioms and graph relations can influence learned representations rather than functioning solely as post hoc labels [18]. This supports a symbolic-to-neural feedback interface through which accepted relations, violated rules, contextual qualifiers, and unresolved contradictions can alter candidate rankings or subsequent representation learning. Chemical ontology classification likewise demonstrates that structural information can be mapped to explicit chemical categories [19]. The proposed system extends this idea by keeping ontology class, structural similarity, and functional prediction separate. A molecule can resemble members of a class without satisfying every defining condition, and an assigned chemical class does not establish reaction behavior or biological function.

The complete architecture contains five interacting zones: learned molecular states; neural-to-symbolic grounding; explicit chemical and biological knowledge; constraint, uncertainty, and contradiction adjudication; and task-specific reasoning

outputs. Molecular design, synthesis planning, and mechanism inference share the representational bridge but retain separate objectives, rules, evidence requirements, and stopping conditions. A provenance layer records which model, rule version, ontology term, database relation, or literature-derived claim contributed to an output. An abstention and expert-review gate receives cases involving out-of-domain chemistry, unresolved contradiction, insufficient rule coverage, or poorly calibrated uncertainty. The resulting design is deliberately modular: it does not assume that a single knowledge representation, reasoning formalism, or molecular encoder will be optimal for every task. **Figure 1** presents the neuro-symbolic chemistry reasoning bridge, showing how the article's key components and boundaries are connected within architecture of the neuro-symbolic molecular system.

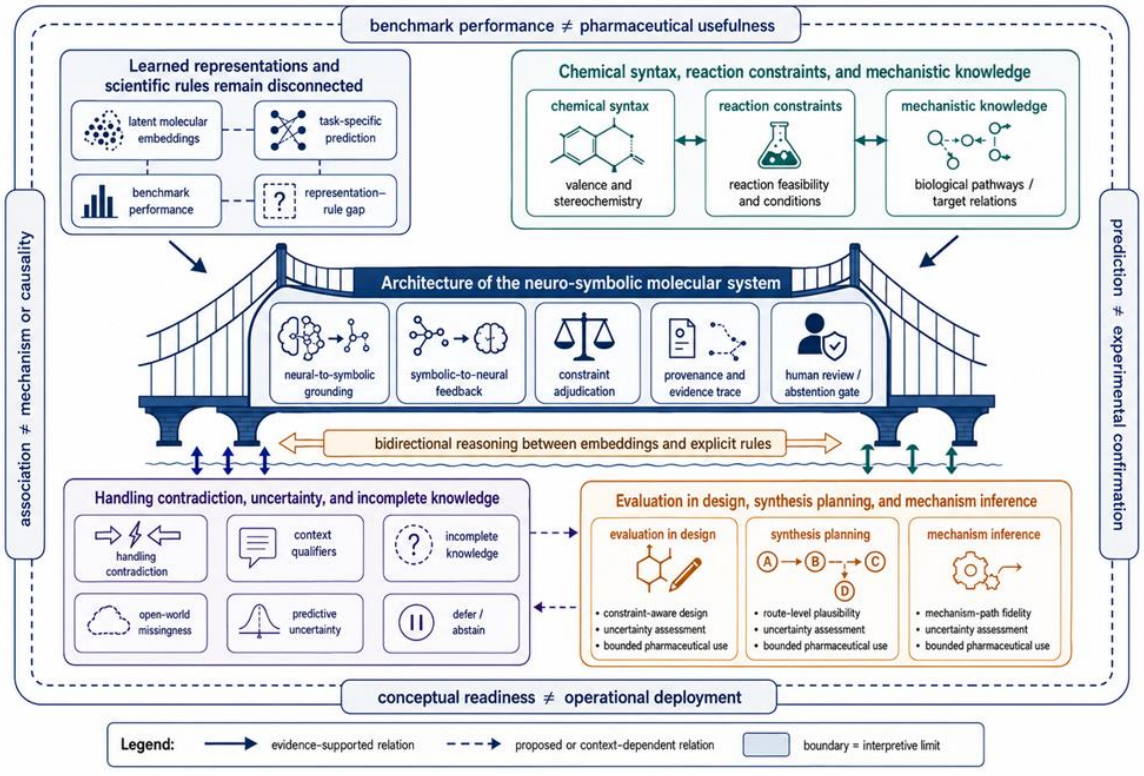

Figure 1. Neuro-Symbolic Chemistry Reasoning Bridge

The figure is an original conceptual synthesis that organizes the article's central contribution across learned representations and scientific rules remain disconnected, chemical syntax, reaction constraints, and mechanistic knowledge, chemical syntax, reaction constraints, mechanistic knowledge, architecture of the neuro-symbolic molecular system. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Bidirectional Reasoning between Embeddings and Explicit Rules

Bidirectional reasoning requires more than applying symbolic filters after a neural model has completed its prediction. The forward direction begins when an embedding is decoded into an explicit candidate hypothesis, such as a molecular structure, reaction transformation, ontology category, or mechanistic relation. Learned reaction embeddings can organize transformation space into chemically meaningful neighborhoods and reveal regularities that may support rule induction [20]. The architecture would use such patterns to propose candidate symbolic relations while retaining their statistical origin. A cluster of transformations may suggest a common reaction family, but the corresponding rule should not enter the governed rule library until its scope, exceptions, reaction-center definition, condition dependence, and evidence provenance have been examined. Neural-to-symbolic translation is therefore hypothesis generation, not automatic knowledge acquisition.

The reverse direction begins when an explicit rule or relation changes the interpretation of a latent candidate. Atomic-environment representations can connect local molecular structure with predicted retrosynthetic transformations [21]. Within the proposed bridge, a predicted disconnection would be tested against atom mapping, reaction-center consistency, substrate scope, protection state, precursor availability, and competing transformations. The resulting symbolic assessment could then revise the neural ranking, constrain a subsequent decoding step, or identify which local representation requires reconsideration. This feedback differs from a static rejection filter because the model receives structured information about why a candidate failed, which conditions were missing, and whether the failure reflected a hard violation, a weak precedent, or an unresolved context.

Route reasoning illustrates the importance of maintaining an explicit symbolic state. Retrosynthetic planning systems such as AiZynthFinder combine reaction proposals with tree search, stock constraints, stopping criteria, and route-scoring policies [22]. These elements cannot be reduced to a molecular embedding because they represent the evolving state of a multi-step search. The proposed architecture would additionally track whether route alternatives depend on the same uncertain transformation, whether a precursor is commercially listed or practically accessible, and whether several paths inherit a common data-source bias. Symbolic search would therefore provide compositional structure, while neural components would prioritize expansions and generalize beyond exact rule matches. Failure in either direction should remain inspectable: a neural proposal may violate a route constraint, and a symbolic search may fail because its rule inventory is incomplete.

Bidirectionality also applies to biological reasoning. Knowledge-graph embeddings can prioritize candidate drug–target relations by learning latent patterns over explicit entities and relation types [23]. Such scores can guide hypothesis formation, but the proposed system would return every candidate relation to a graph containing provenance, biological context, and competing paths. Conversely, biologically structured neural models demonstrate that pathway knowledge can constrain internal connectivity and support pathway-level interpretation of drug-response predictions [24]. The architecture therefore allows explicit mechanistic organization to shape learned representations while preventing a pathway-informed prediction from being reported as a confirmed mechanism. Latent inference and explicit reasoning are complementary only when their evidentiary statuses remain distinguishable and when information can move in both directions without erasing uncertainty.

Handling Contradiction, Uncertainty, and Incomplete Knowledge

Contradiction and uncertainty are related but non-equivalent. Predictive uncertainty and mechanistic contradiction are distinct evidence problems: an uncertainty estimate qualifies a model output but does not by itself establish the contextual or causal account needed to reconcile competing pharmacological claims [25-27]. A model may be uncertain because a molecule lies outside its training distribution, because the available labels are noisy, or because multiple model parameterizations produce different outputs. A knowledge system may contain contradiction because two studies report opposing effects under different tissues, doses, assay designs, disease states, or temporal conditions. The proposed architecture therefore maintains separate records for predictive uncertainty, rule-coverage uncertainty, evidence uncertainty, and explicit conflict among scientific claims.

Incomplete knowledge must be represented as an open-world condition rather than automatically converted into negative evidence. Biomedical datasets and the knowledge graphs assembled from them contain coverage gaps, inconsistent identifiers, ambiguous relations, and uneven provenance [28]. The absence of an edge between a molecule and a target may therefore mean that the relation is false, untested, unpublished, inaccessible, represented under another identifier, or omitted during integration. The architecture assigns distinct states to supported, contradicted, unsupported, unknown, and context-dependent claims. It also records whether a relation is directly measured, computationally inferred, extracted from text, or transferred from a related biological system. This prevents missing knowledge from being silently interpreted as evidence of absence.

Uncertainty handling should influence system behavior rather than appear only as an auxiliary numerical output. Chemistry-specific analyses emphasize that uncertainty can arise from data limitations, model assumptions, representation choice, domain shift, and task formulation [29]. The proposed system uses these sources to determine whether it should continue reasoning, request additional context, return multiple qualified hypotheses, or abstain. A structurally valid molecule with high property-prediction uncertainty might remain eligible for exploration, whereas a low-uncertainty prediction that violates a hard chemical constraint should not. Similarly, a mechanistic path with strong graph connectivity but weak provenance should be reported differently from a path supported by direct perturbation evidence. Uncertainty is thus linked to the type of claim being made and the consequence of acting on it.

The contradiction ledger is the architecture's explicit mechanism for preserving disagreement. Each conflicting claim is stored with direction, source, evidence type, biological context, model or assay system, and unresolved status. The system may identify that two relations become compatible after adding tissue or dose context, but it must also allow contradictions to remain unresolved. Automatic reconciliation is inappropriate when the available evidence does not justify choosing one account. In such cases, the output should state what is known, what conflicts, which context is missing, and what additional evidence would discriminate among alternatives. This design resists the tendency of optimization systems to collapse multiple possibilities into a single ranked answer and instead treats scientific uncertainty as information that must remain available to subsequent experimental and expert reasoning.

Evaluation in Design, Synthesis Planning, and Mechanism Inference

Evaluation must be stratified by task and by claim type. Molecular benchmarks provide standardized datasets, splits, and metrics for comparing predictive models, but their interpretation depends on whether the split reflects scaffold novelty, temporal separation, source variation, or random sampling [30]. A neuro-symbolic system should therefore report conventional predictive performance separately from structural validity, constraint satisfaction, uncertainty calibration, explanation fidelity, and out-of-domain behavior. An improvement in one dimension should not compensate silently for failure in another. For example, a property model that performs well while producing poorly calibrated confidence estimates may still be unsuitable for selective prediction, and a constraint-aware model that rejects chemically unusual but valid structures may reduce scientific exploration despite improving an aggregate validity rate.

Evaluation in molecular design must determine whether optimization respects scientific constraints without exploiting model artifacts. Generative benchmarks distinguish distribution-learning tasks from goal-directed molecular optimization [31]. The proposed architecture adds assessments of rule compliance, structural diversity, novelty relative to training data, predictor disagreement, uncertainty, reason-code fidelity, and the frequency with which candidates are deferred for expert review. These measures remain proxies until compounds are synthesized and tested. High predicted activity, drug-likeness, or constraint satisfaction cannot establish developability, safety, exposure, or therapeutic value. Design evaluation should therefore include adversarial objectives and counterfactual challenges that reveal whether the system is satisfying the intended scientific objective or merely exploiting weaknesses in its predictors.

Synthesis-planning evaluation must operate at both reaction-step and route levels. Transformer-based retrosynthesis combined with hyper-graph exploration demonstrates how single-step predictions can be composed into candidate pathways [32]. A neuro-symbolic planner should be assessed for reaction-center accuracy, applicability of transformation rules, condition compatibility, search efficiency, route diversity, precursor assumptions, and the propagation of uncertainty across steps. Retrospective recovery of documented routes can test whether a planner locates known solutions, but it cannot establish laboratory executability. Prospective route evaluation should document attempted steps, failed transformations, condition modifications, expert interventions, and alternative routes rather than reporting only successful endpoints.

Mechanism inference requires an evaluation standard distinct from endpoint prediction. Knowledge networks can prioritize drug mechanisms through explicit paths connecting compounds, targets, processes, and disease-related entities [33]. Curated mechanism databases can provide structured reference paths and provenance for evaluating whether a system recovers known mechanistic chains [34]. These resources are valuable but incomplete, and recovery of a curated path is not causal confirmation. Evaluation should therefore distinguish relation-type accuracy, path coherence, evidence coverage, contradiction sensitivity, contextual specificity, and experimental support. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop evaluation in design, synthesis planning, and mechanism inference within the article's central argument.

Table 2. Evaluation, implementation, and research priorities arising from Neuro-Symbolic Chemistry Reconnects Learned Molecular Representations with Reaction Rules, Biological Mechanisms, and Scientific Constraints

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Molecular representation layer	Molecular graphs, strings, three-dimensional descriptors, assay-linked features	Learn task-relevant molecular states and candidate hypotheses	Embeddings, predictions, generated candidates, latent similarity relations	Task dependence, domain shift, representation bias, limited interpretability	Scaffold-, temporal-, and source-separated evaluation with external-domain testing
Chemical syntax layer	Candidate structures, atom types, bonds, charges, aromaticity, stereochemistry	Check structural well-formedness and identify repairable violations	Validated structure, explicit violation, repair proposal, or deferral	Rule incompleteness and rejection of unusual but valid chemistry	Curated valid and invalid structures, adversarial cases, and expert review
Reaction-rule and condition layer	Reactants, products, reaction centers, reagents, catalysts, solvents, temperature, substrate context	Assess transformation applicability and contextual feasibility	Qualified reaction proposal with scope, conditions, exceptions, and provenance	Dataset bias, incomplete condition reporting, laboratory dependence	Blind reaction challenges and prospective experimental testing
Route-search layer	Reaction proposals, stock lists, search state, route constraints, protection-state information	Compose and rank multi-step retrosynthetic alternatives	Candidate routes, unresolved dependencies, route-level uncertainty, reason codes	Retrospective route recovery may not predict executability	Prospective synthesis attempts with documented failures and expert interventions
Mechanistic knowledge layer	Drugs, targets, pathways, phenotypes, diseases, ontologies, curated relations	Generate and compare explicit biological mechanism paths	Ranked and evidence-qualified mechanism hypotheses	Graph incompleteness, association–causality confusion, context loss	Curated-path evaluation followed by orthogonal perturbation and causal studies
Contradiction and context ledger	Conflicting claims, assay context, tissue, dose, temporal state, evidence type, provenance	Preserve disagreement and identify potentially reconciling context	Resolved contextual distinction, unresolved contradiction, or request for evidence	Context may be absent, incompatible, or insufficient for adjudication	Purpose-built contradiction benchmarks and expert adjudication studies
Uncertainty and abstention layer	Predictive distributions, model ensembles, rule	Qualify outputs and decide whether to	Calibrated uncertainty profile	Calibration is task- and distribution-dependent	External calibration, risk–coverage analysis,

	coverage, domain-shift indicators, evidence quality	continue, revise, defer, or abstain	and action-specific deferral decision		and longitudinal monitoring
Provenance and explanation layer	Model traces, rules, ontology versions, database records, source articles, transformations	Link each claim to its computational and evidentiary basis	Auditable reason codes, evidence traces, and versioned inference record	Traceability does not establish correctness or causal validity	Intervention-based explanation testing and independent audit
Human-review interface	Candidate outputs, uncertainty, contradictions, violated constraints, evidence traces	Support expert correction, prioritization, or rejection	Reviewed decision, requested experiment, revised constraint, or documented abstention	Automation bias, cognitive burden, and reviewer disagreement	Prospective human-in-the-loop studies using realistic decisions and error cases
Integrated task evaluation	Outputs from design, synthesis, and mechanism components	Determine whether the architecture supports bounded pharmaceutical reasoning	Task-specific performance profile rather than a single score	Cross-task metrics may conceal negative transfer or incompatible objectives	Modular ablation, cross-task comparison, and staged prospective validation

Figure 2 presents the evidence, validation, and translation boundary observatory for neuro-symbolic chemistry reconnects learned molecular representations with, showing how the article's key components and boundaries are connected within evaluation in design, synthesis planning, and mechanism inference.

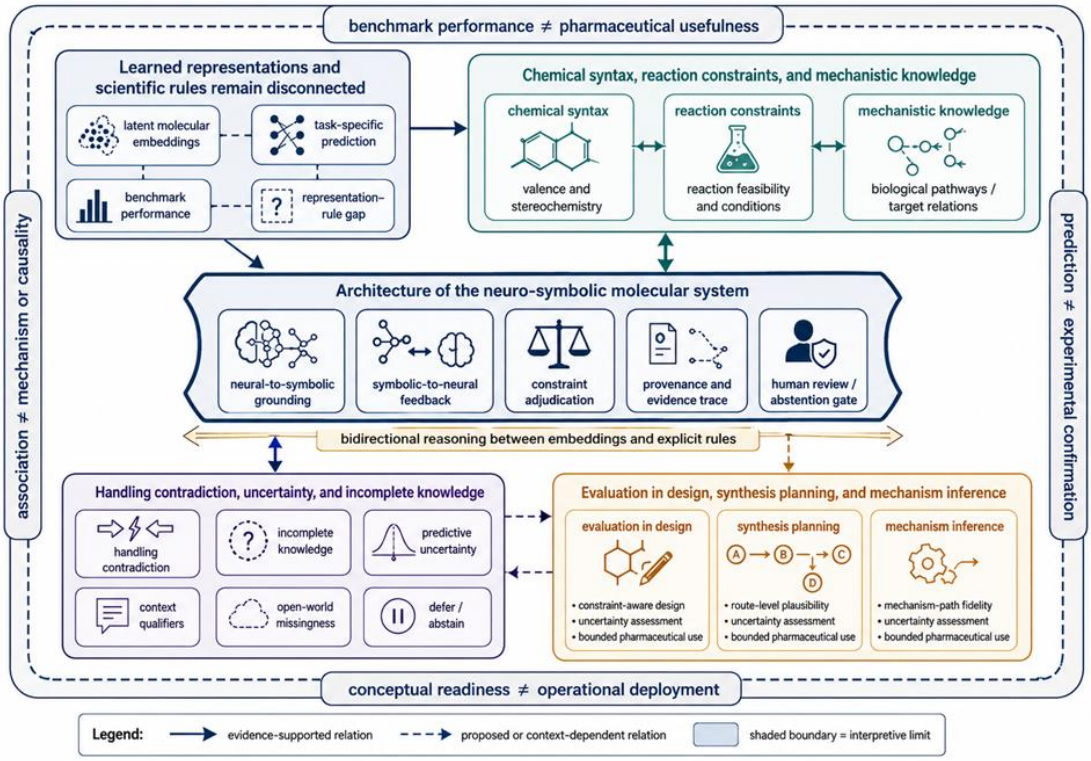

Figure 2. Evidence, Validation, and Translation Boundaries.

The figure is an original conceptual synthesis that shows how claims move from conceptual plausibility through evaluation, uncertainty assessment, and bounded pharmaceutical use. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Limitations and Boundary Conditions

The first limitation is that the proposed architecture is a conceptual specification rather than an implemented and validated molecular system. The selected literature supports individual components—including molecular representation learning, reaction prediction, logical constraints, knowledge graphs, uncertainty estimation, and route planning—but does not establish that their integration will be technically stable or scientifically superior. Representations and constraints may operate at incompatible levels of abstraction, and feedback from symbolic rules may degrade rather than improve a learned model if the rules are incomplete, incorrectly scoped, or over-weighted. The architecture consequently cannot be assumed to outperform

task-specific systems. Its value at this stage lies in exposing design requirements and evaluation boundaries that a future implementation would need to satisfy.

The evidence base is also shaped by the limitations of chemical, biological, and reaction datasets. Data suitability depends on assay context, label quality, molecular coverage, identifier consistency, and provenance [7]. Knowledge graphs inherit missing relations and integration biases from their underlying resources [28]. Uncertainty estimation can identify some forms of model unreliability, but calibration remains conditional on task and domain [25]. A system may therefore produce well-calibrated predictions within a familiar benchmark while remaining unreliable for new reaction classes, biological contexts, or chemical modalities. Explicit rules do not eliminate these limitations because rule libraries and ontologies also reflect incomplete and historically contingent knowledge.

The architecture must not be used to convert molecular plausibility into claims of synthesis success, biological mechanism, therapeutic effectiveness, clinical utility, or regulatory acceptability. Route-planning performance does not establish prospective laboratory execution [32]. A curated mechanism path does not prove causality or clinical relevance [34]. Human oversight can identify some errors and contextual exceptions, but it does not validate the underlying model or remove automation bias. The system should therefore remain bounded as a research-stage reasoning and hypothesis-organization framework until each task-specific component and their interactions have been evaluated prospectively. High-consequence decisions require independent experiments, domain-specific review, and evidence standards appropriate to the claim being made.

Research Agenda and Implementation Priorities

The first research priority is to formalize the interfaces among learned states, explicit rules, and contextual evidence. Future work should establish machine-readable distinctions among hard constraints, soft preferences, defeasible empirical patterns, ontology axioms, and unresolved claims. Studies of differentiable logical losses indicate that aggregation and connective choices can materially influence how knowledge affects learning [17]. Chemistry-specific benchmarks should therefore test whether a system applies each constraint according to its intended semantics, preserves known exceptions, and explains why a candidate was revised or rejected. Bidirectional grounding benchmarks are also needed to determine whether latent molecular patterns can be translated faithfully into atom-level, reaction-level, and mechanism-level symbolic objects and whether symbolic feedback changes the learned state in the expected direction.

A second priority is to create evaluation resources for contradiction, open-world incompleteness, and multi-source uncertainty. Context-sensitive drug-repurposing research shows that apparently conflicting claims may require tissue, disease, dose, or experimental qualifiers rather than automatic resolution [27]. Future benchmarks should contain naturally occurring and adversarial contradictions, known positives, reliable negatives, unresolved relations, missing context, and provenance differences. Chemical uncertainty frameworks should be extended so that predictive uncertainty, rule-coverage uncertainty, evidence uncertainty, and domain shift are reported separately [29]. Evaluation should emphasize selective prediction, risk–coverage behavior, context recovery, inappropriate conflict collapse, and the quality of abstention rather than only average prediction error.

Implementation should proceed in stages. Initial prototypes should focus on narrow, auditable tasks in which rule semantics and validation evidence can be specified clearly, such as syntax-aware molecular generation, reaction-center checking, or curated mechanism-path comparison. Route-planning components can then be evaluated prospectively with explicit records of failed steps and expert intervention [22]. Mechanism components should be tested against curated networks while retaining the distinction between path recovery and causal confirmation [33]. Only after modular validation should shared representations and cross-task feedback be studied. Infrastructure priorities include versioned ontologies, provenance-complete knowledge graphs, reproducible rule libraries, machine-readable evidence types, change-impact analysis, and interfaces that expose uncertainty and contradiction to human reviewers. This staged strategy is intended to test whether the proposed reasoning bridge is useful, not to presume that integration itself guarantees scientific improvement.

Conclusion

Neuro-symbolic chemistry offers a conceptual route for reconnecting learned molecular representations with the explicit objects through which chemical and pharmaceutical science reasons: valid structures, reaction transformations, experimental conditions, biological relations, mechanistic paths, uncertainties, contradictions, and evidence provenance. The proposed architecture treats neural models as flexible hypothesis generators and pattern learners while assigning symbolic components the complementary roles of semantic typing, constraint checking, compositional reasoning, and evidence qualification. Its defining feature is bidirectionality: learned states can be grounded into inspectable scientific objects, and explicit rules or unresolved evidence can revise, qualify, or halt latent inference. This organization prevents a single predictive or generative score from standing in for chemical feasibility, mechanism, experimental confirmation, or pharmaceutical usefulness. The contribution remains theoretical and requires modular, task-specific, and prospective validation. It is not a clinical, regulatory, or deployment-ready system and cannot replace experiments or expert scientific judgement. Its practical importance lies in defining a disciplined research program for molecular AI systems that can state not only what they predict, but which rules were applied, what evidence supports the result, where contradictions remain, and when the system should abstain.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5
2. Sánchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: generative models for matter engineering. Science. 2018;361(6400):360-5. doi:10.1126/science.aat2663
3. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4
4. Segler MHS, Waller MP. Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chem Eur J. 2017;23(25):5966-71. doi:10.1002/chem.201605499
5. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370-88. doi:10.1021/acs.jcim.9b00237
6. Bender A, Cortés-Ciriano I. 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 Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009
7. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov Today. 2021;26(4):1040-52. doi:10.1016/j.drudis.2020.11.037
8. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64. doi:10.1038/s41573-019-0050-3
9. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688-702. doi:10.1016/j.cell.2020.01.021
10. Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular Transformer: a model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci. 2019;5(9):1572-83. doi:10.1021/acscentsci.9b00576
11. Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem Sci. 2019;10(2):370-7. doi:10.1039/C8SC04228D
12. Gao H, Struble TJ, Coley CW, Wang Y, Green WH, Jensen KF. Using machine learning to predict suitable conditions for organic reactions. ACS Cent Sci. 2018;4(11):1465-76. doi:10.1021/acscentsci.8b00357
13. Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555(7698):604-10. doi:10.1038/nature25978
14. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife. 2017;6:e26726. doi:10.7554/eLife.26726
15. Manhaeve R, Dumančić S, Kimmig A, Demeester T, De Raedt L. Neural probabilistic logic programming in DeepProbLog. Artif Intell. 2021;298:103504. doi:10.1016/j.artint.2021.103504
16. Badreddine S, d'Avila Garcez AS, Serafini L, Spranger M. Logic Tensor Networks. Artif Intell. 2022;303:103649. doi:10.1016/j.artint.2021.103649
17. Giannini F, Diligenti M, Maggini M, Gori M, Marra G. T-norms driven loss functions for machine learning. Appl Intell. 2023;53:18775-89. doi:10.1007/s10489-022-04383-6
18. Alshahrani M, Khan MA, Maddouri O, Kinjo AR, Queralt-Rosinach N, Hoehndorf R. Neuro-symbolic representation learning on biological knowledge graphs. Bioinformatics. 2017;33(17):2723-30. doi:10.1093/bioinformatics/btx275
19. Hastings J, Glauer M, Memariani A, Neuhaus F, Mossakowski T. Learning chemistry: exploring the suitability of machine learning for the task of structure-based chemical ontology classification. J Cheminform. 2021;13:23. doi:10.1186/s13321-021-00500-8
20. Schwaller P, Probst D, Vaucher AC, Nair VH, Kreutter D, Laino T, et al. Mapping the space of chemical reactions using attention-based neural networks. Nat Mach Intell. 2021;3(2):144-52. doi:10.1038/s42256-020-00284-w
21. Ucak UV, Ashyrmamatov I, Ko J, Lee J. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. Nat Commun. 2022;13:1186. doi:10.1038/s41467-022-28857-w
22. Genheden S, Thakkar A, Chadimová V, Reymond JL, Engkvist O, Bjerrum EJ. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. J Cheminform. 2020;12:70. doi:10.1186/s13321-020-00472-1
23. Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics. 2020;36(2):603-10. doi:10.1093/bioinformatics/btz600
24. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell. 2020;38(5):672-84.e6. doi:10.1016/j.ccell.2020.09.014

25. Scalia G, Grambow CA, Pernici B, Li YP, Green WH. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J Chem Inf Model. 2020;60(6):2697-717. doi:10.1021/acs.jcim.9b00975
26. Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. J Chem Inf Model. 2020;60(8):3770-80. doi:10.1021/acs.jcim.0c00502
27. Sosa DN, Altman RB. Contexts and contradictions: a roadmap for computational drug repurposing with knowledge inference. Brief Bioinform. 2022;23(4):bbac268. doi:10.1093/bib/bbac268
28. Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, Bender A, et al. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. Brief Bioinform. 2022;23(6):bbac404. doi:10.1093/bib/bbac404
29. Heid E, McGill CJ, Vermeire FH, Green WH. Characterizing uncertainty in machine learning for chemistry. J Chem Inf Model. 2023;63(13):4012-29. doi:10.1021/acs.jcim.3c00373
30. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-30. doi:10.1039/C7SC02664A
31. Brown N, Fiscato M, Segler MHS, Vaucher AC. GuacaMol: benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59(3):1096-108. doi:10.1021/acs.jcim.8b00839
32. Schwaller P, Petraglia R, Zullo V, Nair VH, Haeuselmann RA, Pisoni R, et al. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. Chem Sci. 2020;11(12):3316-25. doi:10.1039/C9SC05704H
33. Mayers M, Tu R, Steinecke D, Li TS, Queralt-Rosinach N, Su AI. Design and application of a knowledge network for automatic prioritization of drug mechanisms. Bioinformatics. 2022;38(10):2880-91. doi:10.1093/bioinformatics/btac205
34. Gonzalez-Cavazos AC, Tanska A, Mayers M, Carvalho-Silva D, Sridharan B, Rewers PA, et al. DrugMechDB: a curated database of drug mechanisms. Sci Data. 2023;10:632. doi:10.1038/s41597-023-02534-z