



Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Causal QSAR Requires Separation of Molecular Association from Mechanistic Explanation

Fatima El Idrissi^{1*}, Samira Bennani¹, Youssef Benali², Ahmed Ben Ali³

¹Department of Causal QSAR and Mechanistic Explanation, Faculty of Pharmacy, University of Marrakech, Marrakech, Morocco.

²Department of Molecular Association and Drug Activity, Faculty of Pharmacy, University of Fez, Fez, Morocco.

³Department of Mechanistic Pharmacology, Faculty of Pharmacy, University of Casablanca, Casablanca, Morocco.

*Email: fatima.idrissi@uca.ma

ABSTRACT

Quantitative structure–activity relationship (QSAR) models are often interpreted as if accurate prediction were evidence that a molecular feature participates mechanistically in producing activity. That inference is not generally warranted. Predictive models can exploit correlations arising from scaffold structure, assay context, sampling, representation choice, redundancy, or hidden common causes without identifying the variables through which a structural intervention changes a biological outcome. This Conceptual Perspective separates molecular association, model explanation, mechanistic mediation, and causal effect as distinct inferential objects. Molecular descriptors are treated initially as proxies whose causal status must be established rather than assumed. Dataset confounding, activity cliffs, representation-sensitive explanations, counterfactual edits, and designed perturbations are organized into a proposed causal-QSAR reasoning framework. Stronger causal language is reserved for cases in which structural interventions, mechanistic variables, competing explanations, causal structure, and validation evidence align. The framework is nonempirical and does not claim prospective validation; its purpose is to prevent predictive accuracy or explanation from being mistaken for mechanism.

Keywords: Causal inference, QSAR, Molecular descriptors, Counterfactual explanations, Structure–activity relationships, Causal graphs

Received: 23 March 2026; **Revised:** 11 June 2026; **Accepted:** 14 June 2026

INTRODUCTION

QSAR is useful because molecular structure contains information that can be mapped to biological activity and pharmaceutical properties. The inferential problem begins when predictive success is treated as evidence about why an outcome changes. Causal inference instead asks what would happen under a specified intervention relative to a counterfactual alternative, under assumptions connecting observed data to that intervention [1]. A QSAR model may approximate an outcome accurately while remaining silent about whether changing a descriptor would change that outcome, whether the descriptor is a proxy for a causal molecular state, or whether both reflect another factor.

Machine learning is already valuable across drug discovery, but its impact depends on task definition, data quality, interpretability, and the relationship between a benchmark and the intended decision [2]. Chemical-machine-learning evaluation therefore must match the use being claimed rather than rely only on favorable retrospective

performance [3]. Prediction can be useful without mechanistic meaning, and mechanism can be scientifically important without being the statistically dominant predictor in a heterogeneous dataset.

Work spanning chemical AI, causal machine learning, and causal representation learning supports a common boundary: predictive performance, explanation, and causal identification are not interchangeable [4-6]. This Perspective therefore distinguishes four progressively stronger statements: a feature predicts an outcome; a model relies on the feature; a structural intervention changes a mechanistic variable; and the intervention changes the outcome through that variable. The proposed causal-QSAR framework permits movement between these levels only when corresponding evidence is available.

Predictive association is not causation

Ligand-based benchmarks can reward memorization of chemical-series structure or train–test redundancy rather than genuine generalization [7]. In such settings, a model can classify compounds correctly because the test set resembles the training set along dimensions that also generated the labels. The association remains statistically useful, but it does not establish a transferable structure–activity mechanism.

This problem becomes more important when prediction is translated into medicinal-chemistry action. AI critiques in drug discovery emphasize that surrogate tasks and retrospective data can separate benchmark performance from the downstream decision [8]. Chemical and biological datasets also contain measurement error, assay heterogeneity, project-specific sampling, and other provenance effects [9]. These factors are not all formal confounders, but they can create associations incorrectly attributed to molecular structure alone.

Context can itself carry predictive information. Leakage-controlled activity-cliff analyses showed that predictability from bioactivity profiles depended strongly on neighboring assay relationships [10], while data-integration work showed that benefits depend on cross-source consistency and experimental conditions [11]. Adding assay descriptions to molecular structure can improve prediction [12], and large language models can perform strongly on selected predictive chemistry tasks [13]. None of these results establishes that contextual variables are mechanistic causes. Richer inputs can improve prediction without identifying which molecular variables would remain causal under intervention.

Molecular descriptors as proxies

Descriptors are representations chosen for prediction, even when their names are chemically intuitive. Comparative studies show that fixed and learned representations differ across tasks, transformer behavior depends on tokenization and pretraining, and similar downstream scores can conceal substantially different latent organizations of chemical space [14-16]. A descriptor, embedding coordinate, or attributed substructure should therefore be treated initially as a proxy rather than a mechanistic variable.

ADMET benchmarking further shows that feature representation materially affects model performance across datasets [17]. A physically richer representation does not remove this problem. Geometry-aware learning can improve molecular-property prediction [18], and physics-related denoising objectives can improve learned representations and downstream performance [19], but neither result establishes that a learned latent feature is a mediator between structural change and pharmacological outcome.

Explanation stability provides a practical boundary. Molecular explainability methods can change materially when the same molecule is represented differently [20], and chemical-space diagnostics can reveal local model cliffs or weak applicability that global performance masks [21]. For causal-QSAR, a representation-sensitive explanation is evidence that the model has not yet isolated an intervention-stable molecular variable. Mechanistic interpretation should therefore require support beyond descriptor importance or embedding geometry.

Confounding in chemical datasets

Activity cliffs expose the limits of smooth association learning. Across multiple algorithms and targets, many molecular-learning approaches struggled with cliffs, and added model complexity did not eliminate the problem [22]. A cliff may reflect a real mechanistic switch, altered binding mode, stereochemistry, solvation, assay noise, or another context-specific factor. It is therefore a high-information discrepancy, not proof of one causal explanation.

Prediction uncertainty must also be distinguished from causal uncertainty. QSAR methods can identify weakly supported chemical regions and qualify confidence for data-poor compounds [23]. Yet a precise prediction can still reflect scaffold membership, assay conditions, or an unobserved mechanistic variable. Conversely, a causally relevant feature may remain difficult to predict in sparse chemical space.

Formal causal tools can be applied to chemical variables but remain assumption-sensitive. A bounded environmental-chemical study combined machine learning, causal inference, and refutation procedures to prioritize chemical properties [24]. Separately, causal-discovery simulations for industrial chemical processes showed that temporal aggregation, subsampling, and unobserved confounding can produce incorrect recovered topology [25]. These studies do not validate causal QSAR for drug molecules; they show why causal conclusions depend on what is measured, what remains hidden, and how observations are generated. Chemical series, assay context, target state, synthesis strategy, and selection processes should therefore be represented according to the specific causal question rather than automatically “controlled for” as a generic confounder list.

Structural interventions and counterfactuals

Molecular counterfactuals form a bridge between prediction and causal reasoning because they ask how a model responds when structure is changed. Model-agnostic methods can generate chemically constrained edits that change a prediction and expose reliance on unexpected features [26]. Such a counterfactual is initially a model intervention: it establishes how the model responds to a represented structural change, not what a biological system would do if that molecule were physically changed.

Contrastive explanations similarly identify features that drive differences between model decisions [27]. They are useful for selecting discriminating structural hypotheses, but the perturbed feature may be causal, a proxy, or one of several correlated encodings. Causal force increases when an intervention is physically executed and the relevant process is measured. Experimental protein-corona work provides a bounded example in which controlled perturbations were used to move beyond correlation toward causal attribution [28]; its value here is methodological rather than as direct evidence for ligand QSAR.

Activity-cliff-aware molecular design supplies another useful stress case because small structural edits can correspond to large activity differences [29]. Such cases force a mechanistic hypothesis to explain local discontinuity rather than only reproduce a smooth trend. Still, the cliff does not identify its mediator. Structural sensitivity answers whether something changed; mechanism requires evidence about what changed in the causal pathway and why.

The distinctions are summarized in **Table 1** as a boundary-condition register rather than a numerical causal hierarchy. The table is needed because the same structural modification can carry different evidentiary meaning depending on whether it exists only in model space, is experimentally realized, or discriminates among competing mechanistic explanations.

Table 1. Boundary conditions separating model-level structural perturbation from mechanistically interpretable intervention

Source state and scale	Transfer conditions	Qualification endpoint	
Structural perturbation basis	Competing explanation / uncertainty	Decision case / Mechanistic interpretation permitted	Validation discriminator / Consequence for causal-QSAR reasoning
Chemically constrained change evaluated only by the trained model	Competing explanation: Representation artifact, scaffold association, extrapolation Principal uncertainty: Model specification and local domain support	Decision case: Prediction-changing virtual edit Mechanistic interpretation permitted: The model is sensitive to the edited feature or correlated representation	Validation discriminator: Test the edited structure and assess representation invariance Consequence for causal-QSAR reasoning: Treat as a proposed model counterfactual, not a causal molecular effect
Feature or substructure alteration chosen to reverse a model decision	Competing explanation: Correlated descriptor, redundant feature, nonphysical perturbation Principal uncertainty: Explanation method and feature encoding	Decision case: Contrastive feature perturbation Mechanistic interpretation permitted: The feature contributes to local model discrimination	Validation discriminator: Reproduce across valid representations and test a realizable structural change Consequence for causal-QSAR reasoning: Use to prioritize hypotheses, not to name a mediator
Small structural difference associated with a large measured activity difference	Competing explanation: Assay noise, binding-mode change, stereochemistry, solvation, contextual effect	Decision case: Activity-cliff pair Mechanistic interpretation permitted: A local structural	Validation discriminator: Replicate the cliff and test candidate pathway variables under matched conditions

	Principal uncertainty: Measurement and mechanistic ambiguity	change is associated with discontinuous response	Consequence for causal-QSAR reasoning: Treat as a high-information causal test case
Deliberate physical intervention followed by mediator or outcome measurement	Competing explanation: Uncontrolled co-intervention or omitted mediator Principal uncertainty: Intervention specificity and measurement completeness	Decision case: Executed molecular perturbation with process readout Mechanistic interpretation permitted: Stronger causal interpretation within the manipulated system	Validation discriminator: Orthogonal perturbation, rescue, mediation, or competing-hypothesis test Consequence for causal-QSAR reasoning: Permit bounded causal language when alternatives are sufficiently constrained
Model and experiment disagree on pathway response	Competing explanation: Proxy feature, compensatory pathway, incorrect mediator Principal uncertainty: Model–biology mismatch	Decision case: Prediction changes but expected mediator does not Mechanistic interpretation permitted: Predictive association is insufficient for the proposed route	Validation discriminator: Test alternative mediators and revise the causal graph Consequence for causal-QSAR reasoning: Downgrade mechanism while retaining validated predictive utility
Same structural modification produces context-dependent outcomes	Competing explanation: Effect modification, assay context, state-dependent mechanism Principal uncertainty: Transportability	Decision case: Effect reverses across assay or molecular context Mechanistic interpretation permitted: Causal effect may be conditional rather than universal	Validation discriminator: Repeat under prespecified contexts and model interactions Consequence for causal-QSAR reasoning: Restrict the causal claim to supported contexts

These states are represented in **Figure 1** as a lattice because stronger claims require explicit interpretation and intervention boundaries rather than a linear progression.

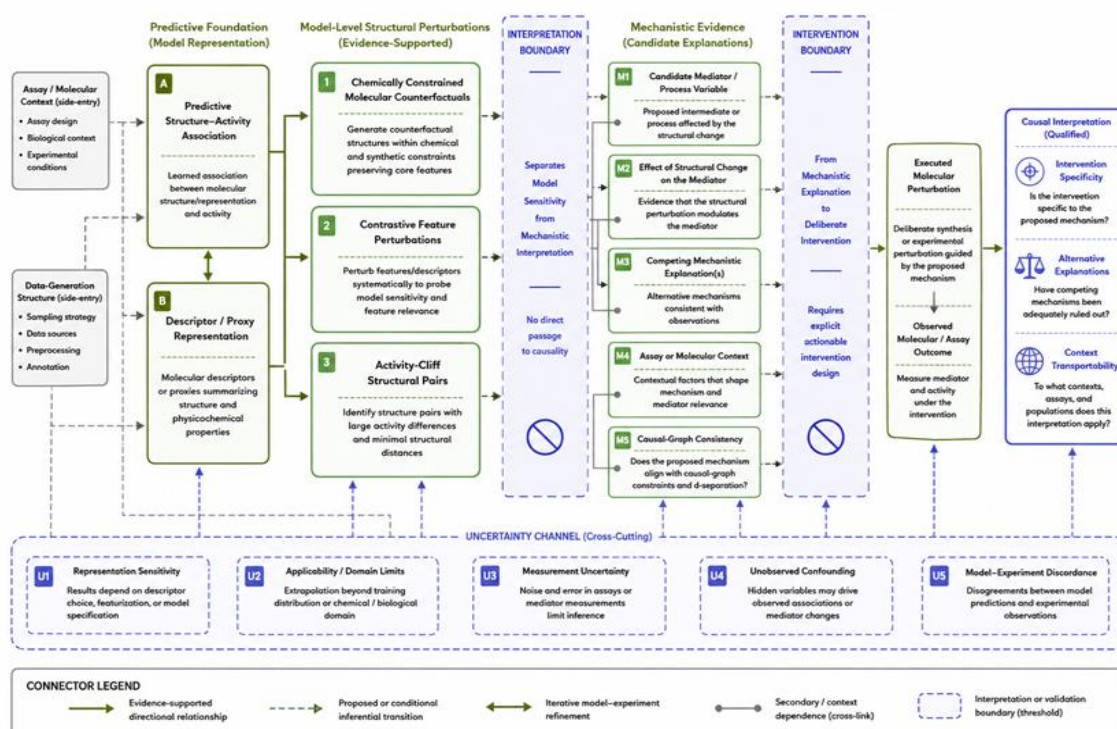


Figure 1. Structural perturbations acquire causal meaning only after crossing model, mechanism, and intervention boundaries

Mediators and mechanistic variables

A mechanistic explanation requires more than identifying a feature that influences model output. Molecular explainability research distinguishes explanations of prediction models from scientific explanations of the

underlying chemical system [30]. A highly attributed substructure may genuinely influence the model yet still fail to qualify as a mediator unless it occupies an appropriate causal position between structural intervention and outcome.

Uncertainty strengthens this requirement. Combining attribution with uncertainty quantification can make molecular explanations more informative by separating well-supported from unstable model behavior [31]. The proposed causal-QSAR rule is therefore two-dimensional: a candidate mediator should be intervention-relevant and uncertainty-qualified. A feature that is highly attributed but unstable, poorly supported, or replaceable by a correlated representation should remain a candidate rather than a mechanistic conclusion.

Prediction uncertainty must also be distinguished from uncertainty in causal structure. Deep-learning approaches can quantify uncertainty for elementary-reaction property prediction [32], but these methods address confidence in a prediction rather than whether the graph connecting molecular state, mediator, context, and outcome is correct. Causal-QSAR should therefore carry separate uncertainty channels for what the model predicts and why the measured outcome changes. When no measurable mediator is available, the scientifically defensible endpoint may remain predictive association rather than a forced mechanism.

Causal graphs for structure–activity questions

Causal graphs provide a formal language for keeping molecular roles separate. A recent molecular-dynamics study used hierarchical causal inference to distinguish system-dependent lipid–protein effects and found that some relations were reproducible whereas others were weak or inconsistent [33]. This does not validate one graph for ligand QSAR; it demonstrates that molecular causal analysis can explicitly represent directional hypotheses, system variation, and uncertainty.

For a structure–activity question, the proposed graph should begin with a manipulable structural intervention. Candidate mediators can occupy downstream causal positions, while descriptors remain measurements or proxies unless stronger evidence supports another role. Assay conditions, target state, chemical series, and data-generation choices may act as causes, effect modifiers, or selection variables depending on the question. The measured activity should be represented in its assay context rather than as an abstract label detached from how it was generated.

Arrows must carry evidentiary meaning. A descriptor–activity edge inferred from prediction is not equivalent to a structure–mediator edge supported by perturbation. Hidden common causes should remain visible when plausible, and competing graphs should be retained when the data cannot discriminate among them. The proposed causal-QSAR graph is therefore not a picture of mechanistic truth; it is an explicit statement of causal commitments that must survive designed perturbations.

Conditions for causal claims

Causal language should become stronger only when the evidence changes the inferential problem rather than merely increasing predictive confidence. Uncertainty-guided active learning illustrates the distinction. Bayesian acquisition can identify compounds whose labels are especially informative for improving a molecular model [34]. That is an intervention on information acquisition, not necessarily an intervention on the biological system. A newly measured compound can refine a decision boundary while leaving the causal structure responsible for activity unresolved.

The first condition for a causal-QSAR claim is therefore intervention relevance. The proposed structural change must correspond to a physically meaningful manipulation rather than only an alteration of a descriptor vector. Second, the intervention must occur inside an adequately supported chemical and model domain. Applicability-domain uncertainty, attribution uncertainty, and predictive uncertainty address different questions and should not be collapsed into one confidence measure [23, 31, 32]. A prediction can be locally well supported while its proposed mechanism remains uncertain.

Third, the causal interpretation must survive plausible alternative explanations. Benchmark redundancy can generate apparently strong structure–activity regularity [7]. Activity-cliff predictability can depend on surrounding assay relationships [10], and feature explanations can change when molecular representation changes [20]. These observations imply an exception-first approach: before asserting mechanism, ask which data-generation, representation, assay, or chemical-series conditions could reproduce the same apparent relationship without the proposed causal pathway.

Fourth, the proposed causal graph must be compatible with intervention and system evidence. Formal causal reasoning requires assumptions about the intervention, measured variables, common causes, and identification

[1]. Controlled experimental perturbations can provide stronger causal leverage within a defined molecular system [28], while molecular-dynamics causal analysis illustrates that even explicitly modeled effects can differ across systems [33]. Causal-QSAR should therefore prefer convergence among structural intervention, mediator behavior, graph consistency, and alternative-explanation testing rather than a single numerical threshold for “causality.”

These requirements are organized in **Table 2** as failure modes because causal claims are most vulnerable when apparently favorable evidence conceals an unresolved exception. The table does not assign causal scores; it identifies what must be discriminated before stronger language is justified.

Table 2. Failure modes that prevent predictive structure–activity evidence from supporting a mechanistic causal claim

Observed configuration	Mechanism and competing diagnosis	Corrective discrimination
Misleading pattern / Observable signature	Likely origin / Competing diagnosis viable	Corrective or discriminating test / Residual consequence after mitigation
Misleading pattern: High predictive accuracy collapses under scaffold or neighborhood shift Observable signature: Strong internal performance but unstable external or local behavior	Likely origin: Memorization, narrow chemical-space support, correlated scaffold features Competing diagnosis that remains viable: Genuine mechanism confined to a narrow chemical series	Corrective or discriminating test: Evaluate structurally distinct compounds and inspect local applicability Residual consequence after mitigation: Restrict inference to the demonstrated chemical domain unless the mechanism transfers
Misleading pattern: Descriptor attribution is strong but changes with representation Observable signature: Different valid encodings produce different explanations for the same molecule	Likely origin: Proxy encoding or model-specific feature allocation Competing diagnosis that remains viable: Multiple representations encode the same mechanistic factor indirectly	Corrective or discriminating test: Test representation invariance and perturb the physical molecular feature Residual consequence after mitigation: Retain descriptor as a proxy until an intervention-stable variable is identified
Misleading pattern: Small structural change causes a large activity change but mediator is unknown Observable signature: Local discontinuity that a smooth model cannot reproduce reliably	Likely origin: Activity cliff, altered binding mode, assay artifact, stereochemical or state change Competing diagnosis that remains viable: Measurement error or context-specific state transition	Corrective or discriminating test: Replicate the activity difference and measure competing mechanistic variables Residual consequence after mitigation: Treat the cliff as an informative causal test case, not a completed mechanism
Misleading pattern: Counterfactual edit reverses model prediction but cannot be physically realized Observable signature: Large predicted effect from an implausible or inaccessible edit	Likely origin: Model-space perturbation outside feasible chemistry or biology Competing diagnosis that remains viable: Representation artifact or extrapolation	Corrective or discriminating test: Constrain edits by synthesis, stereochemistry, assay compatibility, and domain support Residual consequence after mitigation: Counterfactual remains explanatory for the model only
Misleading pattern: Executed structural perturbation changes activity but the proposed mediator does not respond Observable signature: Outcome changes without expected intermediate change	Likely origin: Incorrect mediator, parallel pathway, compensatory process Competing diagnosis that remains viable: Direct effect through another pathway or unresolved confounding	Corrective or discriminating test: Measure alternative mediators, use orthogonal perturbations, revise the graph Residual consequence after mitigation: Downgrade the original mechanistic path while retaining the observed intervention effect
Misleading pattern: Apparent mechanism differs across assays, targets, or molecular states Observable signature: Direction or magnitude changes with experimental context	Likely origin: Effect modification or context-dependent causal structure Competing diagnosis that remains viable: Assay-dependent measurement or target-state dependence	Corrective or discriminating test: Prespecify contextual variables and repeat perturbations under matched conditions Residual consequence after mitigation: Limit the causal statement to the contexts in which the pathway is supported

Misleading pattern: Causal graph is stable only under one sampling or data-processing regime Observable signature: Edge orientation or topology changes after resampling or alternative preprocessing	Likely origin: Hidden confounding, aggregation, selection, or temporal distortion Competing diagnosis that remains viable: Observational artifact	Corrective or discriminating test: Compare plausible graphs, alter observation design, and perform sensitivity analyses Residual consequence after mitigation: Treat edge direction as unresolved unless graph conclusions remain robust
Misleading pattern: Prediction is precise but biological intervention evidence is absent Observable signature: Narrow uncertainty around model output but no intervention-linked mediator evidence	Likely origin: Well-calibrated association without mechanistic identification Competing diagnosis that remains viable: Stable noncausal proxy	Corrective or discriminating test: Conduct discriminating structural or mechanistic perturbations Residual consequence after mitigation: Permit a strong predictive claim but not a mechanistic causal claim

A causal-QSAR reasoning framework

The proposed framework begins by refusing to treat every successful QSAR result as the same kind of evidence. The first state is predictive association: structure or its representation contains information about an outcome under a defined dataset and assay regime. Evaluation should determine whether this association generalizes under conditions relevant to intended use [3]. Benchmark bias and local chemical-space structure must remain visible because apparent generalization may otherwise reflect similarity exploitation [7].

The second state is proxy qualification. Molecular representations can be predictive while organizing chemical information differently [14], and similar task performance can conceal different latent structures [16]. Explanation instability under alternative representations provides an additional warning [20]. A descriptor should therefore be promoted beyond proxy status only when its interpretation remains chemically meaningful across reasonable representations and its role is consistent with the proposed intervention.

The third state introduces structural counterfactuals. Chemically constrained molecular edits can identify changes that alter model output [26], while contrastive explanations can reveal which features distinguish competing model decisions [27]. These methods are particularly valuable for designing experiments because they convert a diffuse statistical association into a specific structural hypothesis. Their evidentiary status remains model-relative until the proposed modification is experimentally realized or otherwise linked to a physical causal process.

The fourth state is mechanistic intervention. When a structural modification is executed, the causal question changes from “what does the model predict if this representation changes?” to “what happens to the molecular or biological system when this structure is changed?” Controlled perturbational work illustrates how causal attribution can be strengthened when competing explanations are experimentally constrained [28]. The relevant mediator must still be demonstrated rather than inferred from feature importance.

The fifth state is causal-graph qualification. Causal inference requires explicit assumptions about variables and paths [1], and molecular causal modeling shows that effects can be reproducible in one system yet weak or inconsistent in another [33]. Assay conditions, target state, chemical series, measurement process, and unobserved common causes should therefore remain explicit whenever they alter interpretation.

Finally, the framework reaches bounded mechanistic explanation, not unrestricted causal truth. A causal interpretation is strongest when an intervention is physically meaningful, the relevant mediator responds, alternative explanations are challenged, uncertainty is characterized, and the relation transports to the contexts claimed. Failure at any gate does not invalidate the QSAR model. Instead, it determines which statement remains defensible: predictive association, model explanation, intervention effect, or mechanistic causal interpretation.

Validation through designed perturbations

Validation for causal-QSAR should be constructed to discriminate explanations rather than merely to reproduce a prediction metric. Chemical-machine-learning evaluation already emphasizes tests aligned with the intended scientific use [3]. For a causal claim, this means selecting experiments that would produce different observations under competing mechanisms. A counterfactual structure is valuable when it creates precisely such a divergence, but its role is hypothesis generation until the corresponding molecular intervention is feasible and evaluated [26]. Designed perturbations should therefore target candidate mediators. If a model attributes activity to a substituent because the substituent alters a particular interaction or molecular state, an informative experiment should measure that proposed intermediate as well as the endpoint. Experimental intervention evidence is stronger when

orthogonal perturbations or rescue-like designs distinguish the proposed route from alternative pathways [28]. Molecular causal analyses can likewise benefit from comparisons across systems because an effect that disappears when context changes may reveal effect modification rather than a universal mechanism [33].

Activity cliffs are especially useful for falsification-oriented validation. Their local structural discontinuity challenges models that rely on smooth similarity [22], and cliff-aware molecular design demonstrates that such regions can be deliberately incorporated into compound-generation strategies [29]. The proposed causal-QSAR approach uses these regions not because every cliff is mechanistic, but because they concentrate disagreement among competing explanations. A compound pair that separates two mechanistic hypotheses may be more informative than many additional compounds lying along a familiar smooth trend.

Information acquisition can also be designed around uncertainty. Bayesian active learning can reduce the amount of labeled molecular data required to improve selected prediction tasks [34]. Within causal-QSAR, this principle can be repurposed cautiously: prioritize compounds or perturbations expected to distinguish graphs, mediators, or intervention responses rather than simply those expected to improve global accuracy. The combined strategy has not been prospectively validated as a causal-QSAR procedure and should therefore be regarded as proposed.

Validation must finally test robustness to context. Assay-neighbor dependence can alter activity-cliff prediction [10], molecular explanations can change under representation perturbation [20], and causal-discovery topology can change with sampling, aggregation, or hidden confounding [25]. A mechanistic interpretation that survives only one representation, one assay regime, or one observation design should be narrowed rather than defended through post hoc explanation. Discordance should trigger graph revision, mediator reassessment, or restriction of the claim.

The intended endpoint of designed perturbation is therefore not an abstract declaration that a model is “causal.” It is a sequence of increasingly discriminating tests that determine which part of the structure–activity statement survives intervention. Prediction may remain useful when mechanism fails. A structural effect may remain real when a proposed mediator fails. A mechanism may remain valid within one target state but not transport to another. Preserving these outcomes is more informative than forcing all evidence into a binary causal/noncausal label.

CONCLUSION

QSAR becomes scientifically overextended when molecular association, model explanation, and mechanistic causation are treated as interchangeable. Predictive accuracy establishes that a representation carries useful information under specified conditions; it does not establish that changing an attributed descriptor will change biology through the mechanism implied by that descriptor.

The proposed causal-QSAR reasoning framework separates predictive association, proxy qualification, structural counterfactuals, mechanistic intervention, causal-graph qualification, and bounded mechanistic explanation. Each transition demands a different form of evidence, and failure to cross a boundary does not erase the value of the preceding level. Counterfactual explanations can reveal model sensitivities, activity cliffs can identify discriminating structural cases, and uncertainty analysis can prioritize where additional evidence is needed, but none alone confers mechanistic status.

The practical consequence is a change in validation philosophy. Instead of asking only whether a model predicts new compounds, causal-QSAR should ask which perturbations discriminate among competing explanations, whether proposed mediators respond as predicted, whether alternative representations and contexts preserve the inference, and where causal claims cease to transport. Such a framework can make QSAR more mechanistically disciplined without pretending that predictive models, by themselves, reveal molecular causation.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Michael T, Zhang JD. Causal inference in drug discovery and development. *Drug Discov Today*. 2023;28(10):103737. doi:10.1016/j.drudis.2023.103737
2. 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
3. Bender A, Schneider N, Segler M, Walters WP, Engkvist O, Rodrigues T. Evaluation guidelines for machine learning tools in the chemical sciences. *Nat Rev Chem*. 2022;6(6):428-42. doi:10.1038/s41570-022-00391-9
4. Bajorath J. From scientific theory to duality of predictive artificial intelligence models. *Cell Rep Phys Sci*. 2025;6(4):102516. doi:10.1016/j.xcrp.2025.102516
5. Feuerriegel S, Frauen D, Melnychuk V, Schweisthal J, Hess K, Curth A, et al. Causal machine learning for predicting treatment outcomes. *Nat Med*. 2024;30(4):958-68. doi:10.1038/s41591-024-02902-1
6. Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Toward causal representation learning. *Proc IEEE*. 2021;109(5):612-34. doi:10.1109/JPROC.2021.3058954
7. Wallach I, Heifets A. Most ligand-based classification benchmarks reward memorization rather than generalization. *J Chem Inf Model*. 2018;58(5):916-32. doi:10.1021/acs.jcim.7b00403
8. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: Ways to make an impact, and why we are not there yet. Part 1. *Drug Discov Today*. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009
9. Bender A, Cortes-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
10. Abe R, Miyao T, Bajorath J. Accurate prediction of activity cliff compounds based on bioactivity profiles depends on assay nearest neighbor relationships. *J Cheminform*. 2026;18(1):72. doi:10.1186/s13321-026-01210-9
11. Parrondo-Pizarro R, Menestrina L, Garcia-Serna R, Fernández-Torras A, Mestres J. Enhancing molecular property prediction through data integration and consistency assessment. *J Cheminform*. 2025;17(1):163. doi:10.1186/s13321-025-01103-3
12. Schuh MG, Boldini D, Sieber SA. Synergizing chemical structures and bioassay descriptions for enhanced molecular property prediction in drug discovery. *J Chem Inf Model*. 2024;64(12):4640-50. doi:10.1021/acs.jcim.4c00765
13. Jablonka KM, Schwaller P, Ortega-Guerrero A, Smit B. Leveraging large language models for predictive chemistry. *Nat Mach Intell*. 2024;6:161-9. doi:10.1038/s42256-023-00788-1
14. 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
15. Sultan A, Sieg J, Mathea M, Volkamer A. Transformers for molecular property prediction: Lessons learned from the past five years. *J Chem Inf Model*. 2024;64(16):6259-80. doi:10.1021/acs.jcim.4c00747
16. Fender I, Gut JA, Lemmin T. Beyond performance: How design choices shape chemical language models. *J Cheminform*. 2025;17(1):173. doi:10.1186/s13321-025-01099-w
17. Kamuntavičius G, Paquet T, Bastas O, Šalkauskas D, Prat A, Abdel Aty H, et al. Benchmarking ML in ADMET predictions: The practical impact of feature representations in ligand-based models. *J Cheminform*. 2025;17(1):108. doi:10.1186/s13321-025-01041-0
18. Dobbelaere MR, Lengyel I, Stevens CV, Van Geem KM. Geometric deep learning for molecular property predictions with chemical accuracy across chemical space. *J Cheminform*. 2024;16(1):99. doi:10.1186/s13321-024-00895-0
19. Ni Y, Feng S, Hong X, Sun Y, Ma WY, Ma ZM, et al. Pre-training with fractional denoising to enhance molecular property prediction. *Nat Mach Intell*. 2024;6:1169-78. doi:10.1038/s42256-024-00900-z
20. Hartog PBR, Krüger F, Genheden S, Tetko IV. Using test-time augmentation to investigate explainable AI: Inconsistencies between method, model and human intuition. *J Cheminform*. 2024;16(1):39. doi:10.1186/s13321-024-00824-1
21. Sosnin S. MolCompass: Multi-tool for the navigation in chemical space and visual validation of QSAR/QSPR models. *J Cheminform*. 2024;16(1):98. doi:10.1186/s13321-024-00888-z
22. van Tilborg D, Alenicheva A, Grisoni F. Exposing the limitations of molecular machine learning with activity cliffs. *J Chem Inf Model*. 2022;62(23):5938-51. doi:10.1021/acs.jcim.2c01073

23. Brown TN, Sangion A, Arnot JA. Identifying uncertainty in physical-chemical property estimation with IFSQSAR. *J Cheminform.* 2024;16(1):65. doi:10.1186/s13321-024-00853-w
24. Han M, Liang J, Jin B, Wang Z, Wu W, Arp HPH. Machine learning coupled with causal inference to identify COVID-19 related chemicals that pose a high concern to drinking water. *iScience.* 2024;27(2):109012. doi:10.1016/j.isci.2024.109012
25. Dewantoro H, Smith A, Daoutidis P. Causal discovery for topology reconstruction in industrial chemical processes. *Ind Eng Chem Res.* 2024;63(26):11530-43. doi:10.1021/acs.iecr.4c01155
26. Wellawatte GP, Seshadri A, White AD. Model agnostic generation of counterfactual explanations for molecules. *Chem Sci.* 2022;13(13):3697-705. doi:10.1039/D1SC05259D
27. Lamens A, Bajorath J. Contrastive explanations for machine learning predictions in chemistry. *J Cheminform.* 2025;17(1):143. doi:10.1186/s13321-025-01100-6
28. Rafieioskouei A, Rogale K, Saei AA, Mahmoudi M, Bonakdarpour B. Beyond correlation: Establishing causality in protein corona formation for nanomedicine. *Mol Pharm.* 2025;22(5):2723-30. doi:10.1021/acs.molpharmaceut.5c00262
29. Hu X, Liu G, Zhao Y, Zhang H. Activity cliff-aware reinforcement learning for de novo drug design. *J Cheminform.* 2025;17(1):54. doi:10.1186/s13321-025-01006-3
30. Wellawatte GP, Gandhi HA, Seshadri A, White AD. A perspective on explanations of molecular prediction models. *J Chem Theory Comput.* 2023;19(8):2149-60. doi:10.1021/acs.jctc.2c01235
31. Komissarov L, Manevski N, Groebke Zbinden K, Sach-Peltason L. Explainable graph neural networks in chemistry: Combining attribution and uncertainty quantification. *J Chem Inf Model.* 2025;65(14):7516-28. doi:10.1021/acs.jcim.5c01003
32. Liu Y, Cheng Y. Uncertainty qualification for deep learning-based elementary reaction property prediction. *J Chem Inf Model.* 2024;64(21):8131-41. doi:10.1021/acs.jcim.4c01358
33. Sato T. Computational framework for causal inference in molecular dynamics analysis of lipid-protein interactions. *J Chem Inf Model.* 2026;66(2):1167-78. doi:10.1021/acs.jcim.5c02497
34. Masood MA, Kaski S, Cui T. Molecular property prediction using pretrained-BERT and Bayesian active learning: A data-efficient approach to drug design. *J Cheminform.* 2025;17(1):58. doi:10.1186/s13321-025-00986-6