Pharmacophore

ISSN-2229-5402

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

TRACING PROTEIN–LIGAND PREDICTIONS BACK TO ATOMIC CONTACTS, BINDING CONTEXT, STRUCTURAL ALTERNATIVES, AND MODEL UNCERTAINTY

Giuseppe Bianchi[1]*, Marie Dubois[2], Pedro Reyes[3], Chiara Lombardi[4]

1. ***Department of Protein–Ligand Prediction Traceability, Faculty of Pharmacy, University of Naples Federico II, Naples, Italy.***
2. ***Department of Atomic Contacts and Binding Context, Faculty of Pharmacy, University of Avignon, Avignon, France.***
3. ***Department of Structural Alternatives and Conformational Ensembles, Faculty of Pharmacy, University of Chile, Santiago, Chile.***
4. ***Department of Model Uncertainty in Molecular Recognition, Faculty of Pharmaceutical Sciences, University of Pisa, Pisa, Italy.***

ARTICLE INFO

Received:
03 November 2024
Received in revised form:
09 February 2025
Accepted:
12 February 2025
Available online:
28 February 2025

Keywords: Protein–ligand binding, Binding-affinity prediction, Atomic contacts, Binding-site water, Protonation states, Structural uncertainty

ABSTRACT

Computational models can assign credible binding-affinity or interaction predictions to protein–ligand complexes while providing little defensible information about the atomic relationships, hydration patterns, chemical microstates, or conformational states on which those predictions depend. This distinction limits the scientific interpretation of model outputs in pharmaceutical research because a numerically accurate prediction does not necessarily identify a transferable binding hypothesis or explain how a molecular modification may alter affinity or selectivity. This Original Molecular Evidence-Tracing Article proposes the Protein–Ligand Molecular Evidence Trace, a conceptual record that connects a versioned model output to atomic-contact hypotheses, binding-context assumptions, structural alternatives, evidence provenance, and multidimensional uncertainty. The approach treats direct contacts, water-mediated interactions, protonation and tautomer states, ligand poses, receptor conformations, and model-derived contributions as related but non-interchangeable evidence objects. It further distinguishes model faithfulness from mechanistic confirmation and separates numerical confidence from uncertainty about structural state, applicability, and external evidence quality. The proposed construct is intended to support later perturbation testing, counterfactual analysis, structural corroboration, and bounded interpretation in hit assessment, selectivity reasoning, and lead optimization. Its principal contribution is not a new affinity predictor, but an explicit evidentiary architecture for showing what a prediction refers to, which structural explanation it assumes, what alternatives remain plausible, and which claims are scientifically admissible. The construct remains conceptual and requires architecture-specific, system-specific, and prospective validation. It cannot establish binding mechanism, causal interaction importance, experimental confirmation, pharmaceutical developability, clinical utility, or operational readiness on its own.

To Cite This Article: Bianchi G, Dubois M, Reyes P, Lombardi C. Tracing Protein–Ligand Predictions Back to Atomic Contacts, Binding Context, Structural Alternatives, and Model Uncertainty. Pharmacophore. 2025;16(1):71-80. https://doi.org/10.51847/eQYRUJt7YT

Introduction

Computational models increasingly infer protein–ligand affinity, interaction likelihood, pose quality, or rank ordering from molecular structures. Three-dimensional convolutional architectures demonstrate that spatial information surrounding a bound ligand can be mapped to an affinity estimate, but the resulting scalar does not independently reveal which atom–atom relations generated the prediction or whether those relations correspond to a defensible binding explanation [1]. This distinction matters because pharmaceutical interpretation generally requires more than a predicted value. A medicinal chemist may need to know whether an apparent affinity advantage is attributed to a direct hydrogen bond, hydrophobic packing, displacement of an unfavorable water, stabilization of a receptor conformation, or a ligand microstate that may not predominate under the relevant

Corresponding Author: Giuseppe Bianchi; Department of Protein–Ligand Prediction Traceability, Faculty of Pharmacy, University of Naples Federico II, Naples, Italy. E-mail: giuseppe.bianchi@unina.it.

conditions. The prediction and the molecular hypothesis are therefore related scientific objects, but they are not the same object.

Some modeling strategies attempt to expose internal associations alongside affinity prediction. Interpretable compound–protein architectures can identify learned sequence regions or relational features, yet such outputs remain different from experimentally supported three-dimensional atomic evidence [2]. Similarly, graph neural representations can embed spatial relations within a protein–ligand complex, providing a natural basis for indexing atoms, residues, and intermolecular edges [3]. Nevertheless, representational explicitness alone does not establish that a highlighted edge faithfully explains the model's calculation, that the interaction is physically important, or that the observed contact persists across alternative poses, receptor states, protonation assignments, or water configurations. An explanation must therefore be evaluated both as a statement about the model and as a conditional statement about molecular structure.

The evidentiary problem is amplified by the construction of structure-based datasets and benchmarks. Cross-docked evaluation has shown that model behavior depends on pose generation, structural pairing, dataset design, and the degree of novelty represented in the test set [4]. A model may consequently appear accurate under one benchmark arrangement while relying on regularities that do not support interpretation in an unfamiliar target, scaffold, receptor conformation, or binding environment. Benchmark performance remains necessary for characterizing predictive behavior, but it cannot substitute for a trace showing which structural information was available, which information the model used, and whether the resulting interaction hypothesis survives chemically meaningful challenge.

The central problem addressed in this article is therefore not simply whether a protein–ligand prediction is accurate, but whether it can be traced back to a scientifically bounded account of atomic contacts, binding context, structural alternatives, and uncertainty. Controlled analyses of deep affinity models have demonstrated that apparently strong performance can coexist with limited acquisition of transferable protein–ligand interaction information [5]. In response, this article proposes the Protein–Ligand Molecular Evidence Trace as an original conceptual and methodological construct. The trace is designed to connect a versioned prediction to contact-level hypotheses, local chemical context, alternative structural explanations, evidence provenance, and differentiated uncertainty. It is not presented as an empirically validated standard or a replacement for structural biology, biophysics, simulation, medicinal chemistry, or experimental pharmacology. Its purpose is to define what must be represented and tested before an affinity prediction can responsibly influence a molecular interpretation.

Why Accurate Affinity Prediction Can Remain Structurally Uninformative

Affinity prediction is structurally informative only when the model output can be connected to interaction evidence that is relevant to the stated molecular question. A model may predict an affinity value accurately because the ligand encodes chemical-series membership, size, lipophilicity, functional-group patterns, or other correlates of measured activity. Ligand-derived information can materially improve affinity prediction even when detailed receptor–ligand geometry is not the dominant source of predictive signal [6]. Such behavior is not necessarily a modeling error: ligand chemistry genuinely constrains binding. The interpretive error occurs when predictive success is treated as evidence that the model has learned the specific atomic recognition process represented by the supplied complex. The proposed evidence trace therefore requires the model output to be linked to the structural features actually available to the model and to evidence that those features contribute to the output.

Dataset composition can further weaken the relationship between performance and structural understanding. Molecular benchmarks containing closely related compounds across training and test partitions may reward memorization of series-specific patterns rather than generalization to new chemistry [7]. Debiasing or repartitioning protein–ligand data can substantially alter apparent generalization, demonstrating that the evaluation split is part of the scientific evidence rather than a neutral administrative choice [8]. For an evidence-tracing article, the relevant question is consequently not only whether a model achieved a favorable metric, but also whether the evaluation required transfer across ligand scaffolds, target families, binding-site environments, or structural states. A prediction produced within a familiar chemical series may be useful, yet it should not be accompanied by the same interpretive claim as a prediction made for a structurally novel target–ligand relationship.

Bias-control experiments are needed because a model can exploit properties that correlate with benchmark labels without representing the intended recognition problem. Structure-based virtual-screening studies have shown that chemical datasets can contain systematic biases that permit models to distinguish classes through unintended signals [9]. A structurally plausible visualization produced after prediction does not resolve this issue. The visualization may simply project a molecular story onto a decision that was driven by ligand identity, dataset frequency, target similarity, or another shortcut. The Protein–Ligand Molecular Evidence Trace therefore separates four questions: whether the output is predictive, whether the reported contact explanation is faithful to the model, whether the proposed contacts are physically plausible in the specified binding context, and whether independent evidence supports their pharmaceutical interpretation. None of these questions can be answered by the others.

Similarity between training and test complexes provides a related source of interpretive inflation. Protein sequence and structural similarity can materially affect the apparent performance of machine-learning scoring functions [10]. Accordingly, a prediction trace should state whether the target, pocket, ligand scaffold, interaction motif, and structural template fall within the model's represented domain. It should also distinguish a model-confidence estimate from evidence that the binding-site hypothesis is transferable. In the proposed construct, "structurally uninformative" does not mean that the prediction is useless

or necessarily inaccurate. It means that the output cannot yet support a specific statement about why the ligand binds, which contacts matter, which structural state is relevant, or how a molecular edit will alter the system. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why accurate affinity prediction can remain structurally uninformative within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why accurate affinity prediction can remain structurally uninformative

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Predictive performance	Does the model estimate the stated affinity, ranking, or interaction outcome within a defined domain?	Task-appropriate retrospective or prospective evaluation with clearly specified data partitions	Establishes whether the underlying prediction warrants further interpretation	Treating one aggregate metric as evidence of molecular understanding	Predictive adequacy does not establish contact-level explanation or binding mechanism
Training–test relationship	How similar are the test ligand, target, pocket, and interaction environment to the training domain?	Scaffold, sequence, structural, pocket, and interaction-pattern comparisons	Defines the applicability context of the prediction	Presenting interpolation as broad structural generalization	Applicability must be stated separately for chemistry, target, structure, and task
Ligand-derived signal	Could the output be produced substantially from ligand identity or chemical-series information?	Ligand-only controls, feature ablation, and matched data partitions	Tests whether receptor-context information is necessary for the prediction	Attributing a ligand-driven output to specific receptor contacts	Ligand signal may be useful, but it does not demonstrate use of binding-site geometry
Structural representation	Which atoms, residues, coordinates, distances, and interaction edges were available to the model?	Versioned input structures and architecture-specific representation records	Defines the structural evidence that the model could potentially use	Discussing waters, contacts, or conformations absent from the actual input	A model cannot provide evidence about a structural variable that it did not represent
Explanation faithfulness	Do the reported contacts or residues participate in the model's calculation?	Model-internal perturbation, mediation, deletion, insertion, or architecture-aware attribution tests	Connects the explanation to the predictive computation	Treating a chemically plausible illustration as a faithful explanation	Faithfulness to the model does not establish physical or biological truth
Molecular plausibility	Are the proposed interactions compatible with geometry, chemistry, solvation, and steric constraints?	Structural validation, interaction geometry, microstate analysis, and physically informed assessment	Filters impossible or internally inconsistent hypotheses	Equating geometric proximity with energetic importance	Plausibility is necessary but not sufficient for mechanistic interpretation
External evidentiary support	Do independent structures, perturbations, biophysical measurements, or structure–activity relationships support the hypothesis?	Orthogonal structural, experimental, and computational evidence with documented provenance	Determines whether a model-derived hypothesis can influence pharmaceutical reasoning	Counting correlated evidence as independent confirmation	External support remains system-specific and does not establish universal transferability
Multidimensional uncertainty	What is uncertain about the model, structural state, chemical microstate, applicability domain, and supporting evidence?	Calibration, alternative-state enumeration, domain analysis, and evidence-quality assessment	Prevents one confidence score from concealing distinct uncertainties	Interpreting model confidence as probability that a structural explanation is true	Uncertainty components must retain separate meanings and validation requirements

Atomic Contacts, Water Networks, Protonation, and Conformational Alternatives

An atomic contact is not a self-sufficient explanation of binding. Protein–ligand scoring functions transform structural observations and modeled interaction features into numerical terms through choices about descriptors, reference states, contact definitions, and available structural data [11]. A close donor–acceptor pair may be classified as a hydrogen bond, yet its interpretation depends on orientation, protonation, competing partners, solvent exposure, local dielectric conditions, and the energetic cost of arranging the participating groups. Hydrophobic or aromatic contacts similarly depend on packing, conformational strain, desolvation, and alternative states. The proposed evidence trace therefore records a contact as a typed and conditional hypothesis: it identifies the ligand atom, receptor atom or residue, interaction class, geometry, source structure, model component, and binding-context state in which the contact was inferred. It does not describe the contact as causal or energetically dominant unless separate evidence supports that stronger claim.

Water is a central part of this context because binding-site waters can bridge polar groups, stabilize local protein conformations, complete hydrogen-bond networks, or impose displacement costs that influence ligand recognition [12]. Water networks can also account for affinity differences that are not apparent from a simple comparison of direct protein–ligand contacts [13]. Two ligands may therefore appear to make similar direct interactions while organizing the surrounding solvent differently. Conversely, removal of an apparently unfavorable water may be offset by disruption of a broader network or by an unsatisfied polar group. For evidence tracing, a water should not be represented merely as present or absent. Its record should distinguish experimentally resolved water, predicted hydration site, simulation-derived occupancy, conserved network member, proposed bridging role, and uncertain placement. The trace must also preserve whether the model explicitly represented water or whether the water explanation was introduced afterward by a human or auxiliary calculation.

Binding-site water can become a design hypothesis when a ligand modification is proposed to conserve, reorganize, or displace a localized hydration feature. Structure-guided ligand design has demonstrated that a defined water may be targeted through chemically specific modifications, provided that its location and role are sufficiently supported [14]. The evidentiary emphasis, however, should remain on the conditional chain: a water is observed or predicted; the ligand modification is expected to alter its occupancy or network; the resulting structural change is expected to affect binding; and an external experiment or validated calculation is required to test that expectation. The evidence trace should retain each link rather than collapsing the chain into the statement that "water displacement improves affinity." This formulation permits contradictory observations, alternative hydration patterns, and target-state dependence to remain visible.

Protonation and tautomerism further complicate contact interpretation because changing a ligand microstate can reorganize local electrostatics, hydrogen bonding, and water structure [15]. The same heavy-atom pose may therefore support different interaction explanations depending on which atoms are proton donors, acceptors, charged centers, or tautomer-specific sites. Receptor residues may also change protonation or orientation in response to ligand binding, and conformational alternatives may expose or occlude different contact networks. The proposed evidence trace consequently links each contact map to a defined ligand pose, ligand conformer, protonation or tautomer state, receptor conformation, side-chain arrangement, water configuration, ion state, and structural-preparation procedure. When multiple states remain plausible, they are retained as competing explanations rather than averaged into one apparently definitive interaction diagram. This approach does not claim that all relevant states can be exhaustively enumerated. It instead makes the incompleteness of structural representation explicit and prevents a single prepared complex from being mistaken for the uniquely established binding context.

Evidence Tracing From Model Output to Binding-Site Hypotheses

The proposed Protein–Ligand Molecular Evidence Trace (PL-MET) begins with a versioned prediction object rather than an isolated affinity value. The object records the model and representation used, the prepared protein–ligand complex, the prediction task, the output, and the applicable molecular domain. Interaction-aware graph architectures show that bonded and spatial relations can be represented before their aggregation into a molecular-level prediction [16]. PL-MET extends this representational principle into an evidentiary requirement: every proposed explanation should identify the model component, molecular entities, and structural state from which it was derived. This does not make the explanation correct, but it makes its relationship to the computation inspectable.

The next layer comprises typed atomic-contact hypotheses. Physics-informed pairwise models illustrate how a complex-level affinity estimate can be decomposed into atom-pair contributions associated with physically motivated interaction classes [17]. In PL-MET, each contact object records the ligand atom, protein atom or residue, interaction type, geometry, directionality, source structure, model-derived contribution, and relevant context assumptions. A model-derived contribution is described as evidence about the model's calculation, not as a measured interaction energy. Contact hypotheses therefore remain conditional until they have survived structural-validity checks, perturbation tests, and comparison with independent evidence.

Intermolecular graph edges provide explicit representational units through which ligand atoms, protein atoms, residues, and their spatial relations can be linked to a binding-site interpretation [18]. Interaction-based inductive biases further demonstrate that model architecture can privilege three-dimensional cross-molecular relations, making the origin of a contact hypothesis an essential part of the trace [19]. PL-MET consequently distinguishes contacts obtained from intrinsic pairwise terms, attention or attribution procedures, predicted intermediate contact maps, post hoc structural inspection, and external simulation. These sources cannot be treated as interchangeable because they differ in their connection to the model and in their evidentiary strength. **Figure 1** presents the atomic evidence traceback chamber, showing how the article's key components and boundaries are connected within evidence tracing from model output to binding-site hypotheses.

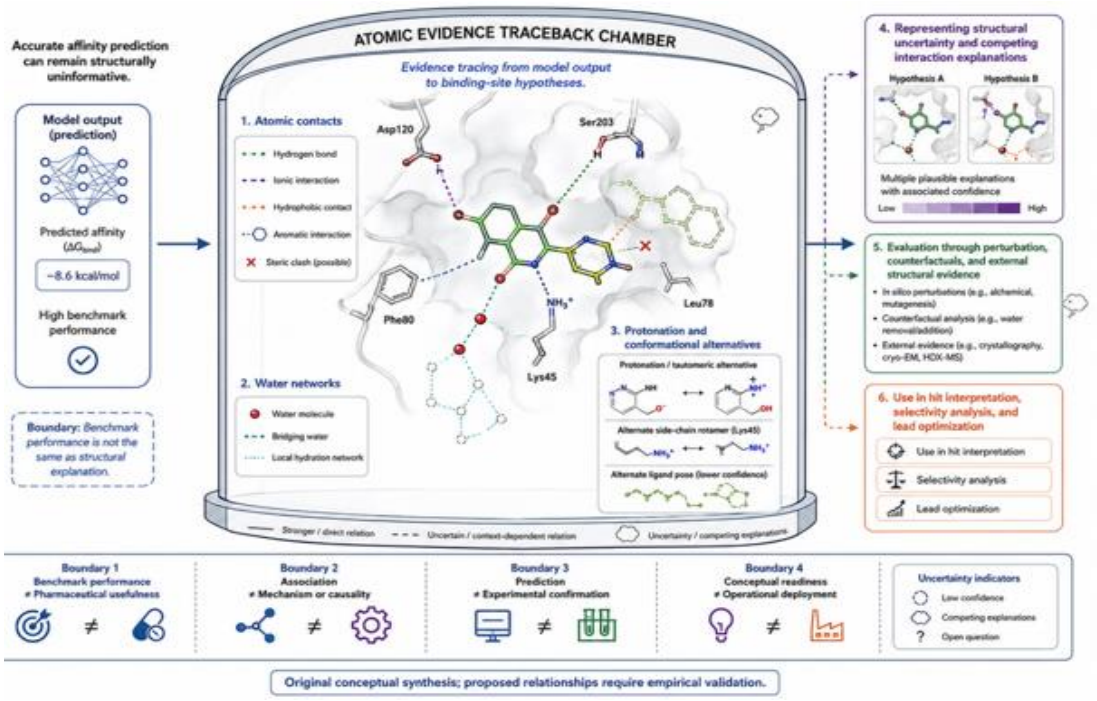

Figure 1. Atomic Evidence Traceback Chamber.

The figure is an original conceptual synthesis that organizes the article's central contribution across accurate affinity prediction can remain structurally uninformative, atomic contacts, water networks, protonation, and conformational alternatives, atomic contacts, water networks, conformational alternatives, evidence tracing from model output to binding-site hypotheses. 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.

A complete trace must also preserve how local hypotheses relate to the global output. Joint affinity-and-contact prediction demonstrates that atomic contacts can be modeled as explicit intermediate targets rather than added only as post hoc illustrations [20]. PL-MET generalizes this idea into a proposed evidence graph connecting the prediction object, contact hypotheses, binding-context states, structural alternatives, uncertainty fields, and external evidence. A trace may contain supportive, contradictory, unresolved, and unavailable links. It should not force these into a single explanation when the available evidence is insufficient. The scientific output is therefore a testable binding-site hypothesis with documented provenance and alternatives, not a declaration that the model has discovered the binding mechanism.

Representing Structural Uncertainty and Competing Interaction Explanations

Uncertainty in a protein–ligand prediction is multidimensional. Neural uncertainty-estimation methods differ in calibration, error ranking, and behavior under distribution shift, meaning that an uncertainty value is interpretable only when its estimation method and validation context are known [21]. PL-MET therefore separates uncertainty about model parameters and training data from uncertainty about the molecular structure being interpreted. A model may be confident within its learned domain while the ligand pose, receptor state, water network, or protonation assignment remains unresolved. Conversely, a structurally well-characterized complex may still receive an unreliable prediction because it lies outside the model's represented chemical or target domain.

Aleatoric, epistemic, and out-of-domain uncertainties require different evaluation strategies and cannot reliably be compressed into one generic confidence score [22]. PL-MET adds structural-alternative uncertainty, microstate uncertainty, and external-evidence uncertainty to this distinction. Structural-alternative uncertainty concerns which pose, receptor conformation, side-chain arrangement, or water network is relevant. Microstate uncertainty concerns protonation and tautomer assignments. Evidence uncertainty concerns the independence, resolution, and applicability of supporting structures, assays, or simulations. These fields may interact, but they retain separate meanings because they imply different corrective actions.

Evidential neural models can produce uncertainty parameters together with molecular predictions, but such outputs do not enumerate the structural explanations compatible with the prediction [23]. Bayesian graph models similarly provide predictive distributions that can help characterize uncertainty in learned molecular relationships [24]. PL-MET treats these methods as possible sources for the model-uncertainty field, while requiring a separate alternative-state record. Each alternative state is linked to its own contact pattern, binding context, physical-validity assessment, model support, and external evidence. Alternatives should not be ranked as though their numerical scores were calibrated probabilities unless an appropriate validation procedure has established that interpretation.

Resampling-based uncertainty estimates can be efficient and informative, but their relationship to prediction error must be assessed and calibrated rather than assumed [25]. The same principle applies to the entire uncertainty ledger. A high-

uncertainty trace may indicate sparse training support, disagreement among model instances, unstable contact attributions, competing poses, uncertain protonation, or contradictory external evidence. These possibilities should remain distinguishable. The proposed representation therefore favors a structured uncertainty profile over a single certainty label. Its purpose is not to eliminate ambiguity, but to expose where ambiguity enters the interpretation and which evidence would be needed to reduce it.

Evaluation Through Perturbation, Counterfactuals, and External Structural Evidence

Evaluation of molecular evidence tracing must begin by separating predictive tasks that are often conflated. Comparative scoring-function assessments distinguish scoring power, ranking power, docking power, and screening power because success in one does not guarantee success in another [26]. PL-MET adds further non-substitutable dimensions: prediction fidelity, contact localization, explanation faithfulness, context sensitivity, alternative-state coverage, uncertainty calibration, external corroboration, and bounded decision usefulness. A model may rank ligands adequately while producing unstable contact traces, or may reproduce a pose while assigning the wrong chemical interpretation to it.

Structural plausibility requires independent examination. Docking evaluations have shown that a predicted pose may resemble a reference while violating stereochemical, geometric, steric, or energetic constraints, and that performance may deteriorate on novel protein sequences [27]. Every structural alternative in PL-MET should therefore undergo physical-validity checks before it is used as explanatory evidence. These checks should examine clashes, bond geometry, ligand strain, interaction geometry, receptor preparation, microstate consistency, and unresolved atoms or waters. Passing such checks establishes admissibility for further testing, not proof that the pose is biologically dominant.

Perturbation tests assess whether an explanation responds to molecular changes in a directionally coherent manner. Activity-cliff pairs provide one useful design because small chemical changes associated with large activity differences can test whether feature attributions identify locally consequential molecular regions [28]. Quantitative explainability benchmarks further show that explanation methods require explicit assessment against known or constructed targets rather than visual inspection alone [29]. PL-MET extends this logic to protein–ligand systems through ligand-atom edits, residue substitutions, contact deletion, water removal or conservation, protonation changes, pose replacement, and receptor-conformation changes. **Figure 2** presents the evidence, validation, and translation boundary observatory for tracing protein–ligand predictions back to atomic contacts, showing how the article's key components and boundaries are connected within evaluation through perturbation, counterfactuals, and external structural evidence.

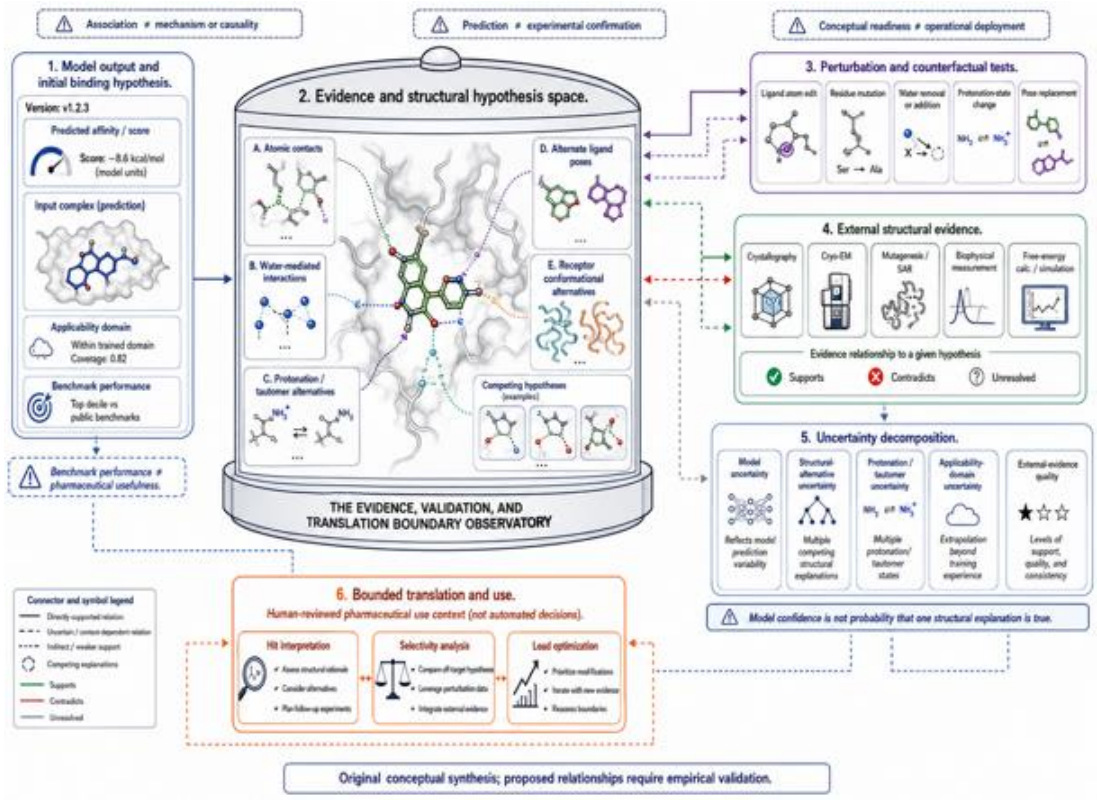

Figure 2. Evidence, Validation, and Translation Boundaries.

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.

Counterfactual evaluation must remain chemically and structurally constrained. Activity-cliff analyses demonstrate that favorable aggregate prediction can conceal serious local failures around small molecular changes [30]. A useful counterfactual should therefore ask whether a feasible ligand modification, residue change, water-network alteration, or microstate transition changes the prediction and trace in a manner consistent with the proposed hypothesis. External evidence should then determine whether that response corresponds to the molecular system rather than merely to model sensitivity. Appropriate evidence may include an independent structure, mutagenesis, matched-series activity data, binding thermodynamics, or a validated simulation. Agreement strengthens a bounded hypothesis; disagreement should remain visible as contradiction rather than being removed from the trace.

Use in Hit Interpretation, Selectivity Analysis, and Lead Optimization

In hit interpretation, PL-MET may help distinguish a high-scoring output accompanied by coherent, testable structural evidence from one dominated by applicability warnings, unstable contacts, or unresolved binding context. Relative binding free-energy methods illustrate how structurally defined hypotheses can support compound prioritization when ligand relationships, structural setup, sampling, and force-field conditions are controlled [31]. PL-MET could precede such calculations by documenting the assumed binding mode, protonation state, water network, receptor conformation, and proposed molecular transformation. It would not demonstrate that the hit is experimentally active or that the simulation setup is correct.

During analog-series interpretation, the trace may connect a chemical modification to changes in direct contacts, desolvation, water organization, conformational strain, or receptor-state preference. Alchemical perturbation methods have a recognized role in medicinal chemistry, but their usefulness remains conditional on transformation scope, sampling adequacy, and the stability of the modeled binding mode [32]. PL-MET provides a proposed way to expose these conditions before a prediction is used to justify synthesis. A molecular edit should therefore be accompanied by the structural assumption it tests, the expected contact or context change, competing explanations, and the experiment or calculation capable of discriminating among them.

Lead-optimization interpretation must also respect the practical ceiling imposed by experimental reproducibility and domain-dependent computational error [33]. A trace should not convert a predicted affinity difference into a precise medicinal-chemistry expectation when the underlying structural alternatives or experimental measurements cannot support that precision.

In selectivity analysis, aligned structural resources can help compare interaction patterns and conformational states across related targets, as demonstrated for curated kinase binding sites [34]. Such comparisons remain bounded to the represented target family, available structures, and assay context. Similar-looking pockets may differ through water networks, local dynamics, protonation, or inaccessible receptor states.

Selectivity is ultimately a comparative property across receptor contexts rather than an attribute of one isolated complex. Alchemical receptor-hopping and receptor-swapping approaches provide one route for estimating relative selectivity while remaining dependent on structural models, sampling, and subsequent validation [35]. PL-MET can organize the evidence entering such comparisons by recording target-specific contacts, alternative conformations, water and microstate assumptions, uncertainty, and contradictory observations. Its practical function is to help experts formulate and prioritize discriminating hypotheses. It cannot establish off-target safety, therapeutic index, pharmacokinetics, synthesizability, developability, or clinical value. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop use in hit interpretation, selectivity analysis, and lead optimization within the article's central argument.

Table 2. Evaluation, implementation, and research priorities arising from Tracing Protein–Ligand Predictions Back to Atomic Contacts, Binding Context, Structural Alternatives, and Model Uncertainty

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Versioned prediction object	Model identity and version; molecular representation; prepared complex; prediction task; output; applicability information	Defines exactly what prediction is being interpreted	Auditable prediction record linked to its computational origin	Model updates, preprocessing changes, and undocumented training-domain boundaries	Reproducibility across software versions and independent reconstruction of the prediction
Atomic-contact hypothesis layer	Protein and ligand atoms; coordinates; interaction types; model-derived contributions	Converts a global output into localized candidate interactions	Typed contact hypotheses with geometry and provenance	Highlighted contacts may be unstable, non-faithful, or energetically unimportant	Architecture-aware perturbation, contact localization, and external structural comparison
Binding-context layer	Waters; ions; cofactors; protonation and tautomer states; local receptor preparation	Conditions the chemical interpretation of every contact	Context-specific contact map and recorded preparation assumptions	Missing or incorrectly assigned waters and microstates may reverse the interpretation	Joint structural, hydration, microstate, and sensitivity evaluation
Structural-alternative layer	Alternative ligand poses; receptor conformers; side-chain states; ligand conformers; water networks	Preserves competing binding explanations	Non-collapsed set of physically admissible alternatives	Enumeration may be incomplete and alternative weights may be uncalibrated	Evaluation on systems with independently supported multiple binding states

Evidence-provenance layer	Experimental structures; mutagenesis; structure–activity relationships; biophysics; simulations; curated resources	Links each hypothesis to supporting, contradictory, or unresolved evidence	Evidence graph with source lineage and independence status	Evidence sources may share templates, assumptions, or data and appear falsely independent	Provenance audits and predefined evidence-independence criteria
Multidimensional uncertainty ledger	Model ensembles or probabilistic outputs; domain distance; pose variation; microstate ambiguity; evidence quality	Separates different reasons for uncertainty	Structured profile of model, applicability, structural, microstate, and evidence uncertainty	No single uncertainty estimator is sufficient across all error regimes	Domain-specific calibration, risk–coverage evaluation, and out-of-domain testing
Perturbation and counterfactual layer	Ligand edits; residue changes; contact masking; water changes; microstate changes; alternative structures	Tests whether the traced explanation responds coherently to controlled changes	Directional and comparative evidence about model faithfulness and hypothesis sensitivity	Perturbations may be chemically impossible or outside the model's domain	Chemically constrained perturbation sets and prospective matched-series studies
Hit-interpretation use	Candidate hit; assay context; structural model; complete evidence trace	Frames a testable explanation for predicted recognition	Prioritized structural or biophysical follow-up hypothesis	A coherent trace may still accompany a false-positive prediction	Prospective experimental confirmation and blinded expert assessment
Selectivity-analysis use	Related targets; aligned binding sites; target-specific structures and assays	Compares contacts, waters, conformations, and uncertainties across receptors	Bounded explanation of potential selectivity determinants	Unrepresented targets, assay differences, or receptor states may invalidate comparison	Target-panel studies with orthogonal structural and pharmacological evidence
Lead-optimization use	Congeneric series; measured activities; structural evidence; feasible chemical modifications	Connects molecular edits to explicit binding hypotheses	Human-reviewed prioritization of testable analog modifications	Binding optimization does not establish synthesis feasibility, exposure, safety, or efficacy	Prospective medicinal-chemistry studies with preregistered predictions and external assays
Lifecycle and governance layer	Model versions; new structures; new assays; contradiction records; user decisions	Maintains trace validity as evidence and models change	Versioned, reviewable, and revalidatable evidence history	An outdated trace may remain persuasive after its assumptions become obsolete	Drift surveillance, periodic revalidation, and controlled retirement of unsupported traces

Limitations and Boundary Conditions

PL-MET is a proposed conceptual and methodological architecture rather than an empirically validated model, ontology, software system, or reporting standard. Its components are derived from recurring evidentiary problems in protein–ligand prediction, but their completeness, interoperability, and usefulness require future testing. Different model classes expose different internal objects, and some black-box architectures may not permit a faithful atom-level decomposition. In such cases, the trace should report that limitation rather than create a post hoc molecular explanation with unsupported precision.

The construct is also bounded by the quality and completeness of structural evidence. Experimental structures may contain unresolved atoms, ambiguous waters, engineered constructs, crystal contacts, nonphysiological conditions, or only one member of a conformational ensemble. Docked or simulated structures add further dependence on preparation protocols, force fields, sampling, and microstate assignment. Consequently, a richly documented trace may still be wrong because the represented structural alternatives are incomplete or because all linked evidence shares the same incorrect assumption.

Finally, evidence tracing does not convert association into mechanism, explanation into causality, simulation into experimental confirmation, or molecular plausibility into pharmaceutical value. A contact hypothesis cannot establish synthesizability, formulation feasibility, exposure, safety, efficacy, or regulatory acceptability. The construct may support human reasoning only within an explicitly defined decision-use envelope. Use outside that envelope risks creating false confidence through the visual and organizational coherence of the trace itself.

Research Agenda and Implementation Priorities

The first priority is to define a minimum interoperable representation for prediction objects, contact hypotheses, binding-context states, structural alternatives, provenance links, and uncertainty fields. This work should specify required identifiers, coordinate and microstate conventions, evidence-lineage rules, contradiction handling, and machine-readable boundary statements. Competency tests should determine whether independent groups can reconstruct the same trace from the same model output and whether different modeling architectures can be represented without erasing their methodological differences.

The second priority is empirical evaluation. Contact-level explanations should be tested for model faithfulness, structural plausibility, sensitivity to relevant perturbations, stability to irrelevant perturbations, and agreement with independent evidence. Benchmark systems should include alternative binding modes, conformational changes, water-mediated interactions, protonation ambiguity, activity cliffs, target novelty, and chemically realistic counterfactuals. Prospective studies should predefine the structural hypothesis and discriminating experiment before outcomes are known, reducing the risk that a plausible explanation is fitted retrospectively to observed data.

Implementation should proceed in stages. Early use should focus on research documentation and failure analysis rather than autonomous prioritization. A later stage may evaluate whether access to evidence traces improves expert calibration, decision consistency, experiment selection, and recognition of uncertainty without increasing overtrust. Lifecycle requirements should include model and structure versioning, drift detection, trace revalidation, contradiction logging, and retirement of unsupported hypotheses. Broader pharmaceutical use should be considered only after the relevant trace components and decision contexts have been prospectively validated.

Conclusion

This article proposes the Protein–Ligand Molecular Evidence Trace as a structured way to connect a model output to atomic-contact hypotheses, binding context, structural alternatives, evidence provenance, and multidimensional uncertainty. Its central premise is that accurate affinity prediction may remain structurally uninformative when the output cannot be shown to depend on transferable protein–ligand evidence or when one prepared complex conceals plausible alternatives involving waters, protonation, ligand pose, or receptor conformation. The proposed construct preserves these distinctions and defines how perturbation, counterfactual testing, and external evidence could evaluate them. Its intended contribution is a more inspectable basis for bounded hit interpretation, selectivity analysis, and lead-optimization reasoning, not a claim of mechanism discovery or deployment readiness. The scientific value of the construct will depend on future architecture-specific, molecular-system-specific, human-factors, and prospective pharmaceutical validation.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: Protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model. 2018;58(2):287-96. doi:10.1021/acs.jcim.7b00650
2. Karimi M, Wu D, Wang Z, Shen Y. DeepAffinity: Interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. Bioinformatics. 2019;35(18):3329-38. doi:10.1093/bioinformatics/btz111
3. Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. J Chem Inf Model. 2019;59(9):3981-8. doi:10.1021/acs.jcim.9b00387
4. Francoeur PG, Masuda T, Sunseri J, Jia A, Iovanisci RB, Snyder I, et al. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. J Chem Inf Model. 2020;60(9):4200-15. doi:10.1021/acs.jcim.0c00411
5. Volkov M, Turk J-A, Drizard N, Martin N, Hoffmann B, Gaston-Mathé Y, et al. On the frustration to predict binding affinities from protein–ligand structures with deep neural networks. J Med Chem. 2022;65(11):7946-58. doi:10.1021/acs.jmedchem.2c00487
6. Boyles F, Deane CM, Morris GM. Learning from the ligand: Using ligand-based features to improve binding affinity prediction. Bioinformatics. 2020;36(3):758-64. doi:10.1093/bioinformatics/btz665
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. Sundar V, Colwell LJ. The effect of debiasing protein–ligand binding data on generalization. J Chem Inf Model. 2020;60(1):56-62. doi:10.1021/acs.jcim.9b00415
9. Sieg J, Flachsenberg F, Rarey M. In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening. J Chem Inf Model. 2019;59(3):947-61. doi:10.1021/acs.jcim.8b00712
10. Li Y, Yang J. Structural and sequence similarity makes a significant impact on machine-learning-based scoring functions for protein–ligand interactions. J Chem Inf Model. 2017;57(4):1007-12. doi:10.1021/acs.jcim.7b00049
11. Liu Z, Su M, Han L, Liu J, Yang Q, Li Y, et al. Forging the basis for developing protein–ligand interaction scoring functions. Acc Chem Res. 2017;50(2):302-9. doi:10.1021/acs.accounts.6b00491
12. Spyrakis F, Ahmed MH, Bayden AS, Cozzini P, Mozzarelli A, Kellogg GE. The roles of water in the protein matrix: A largely untapped resource for drug discovery. J Med Chem. 2017;60(16):6781-827. doi:10.1021/acs.jmedchem.7b00057
13. Darby JF, Hopkins AP, Shimizu S, Roberts SM, Brannigan JA, Turkenburg JP, et al. Water networks can determine the affinity of ligand binding to proteins. J Am Chem Soc. 2019;141(40):15818-26. doi:10.1021/jacs.9b06275
14. Matricon P, Suresh RR, Gao ZG, Panel N, Jacobson KA, Carlsson J. Ligand design by targeting a binding site water. Chem Sci. 2021;12(3):960-8. doi:10.1039/D0SC04938G

15. Henderson JA, Harris RC, Tsai CC, Shen J. How ligand protonation state controls water in protein–ligand binding. J Phys Chem Lett. 2018;9(18):5440-4. doi:10.1021/acs.jpclett.8b02440
16. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y, et al. PotentialNet for molecular property prediction. ACS Cent Sci. 2018;4(11):1520-30. doi:10.1021/acscentsci.8b00507
17. Moon S, Zhung W, Yang S, Lim J, Kim WY. PIGNet: A physics-informed deep learning model toward generalized drug–target interaction predictions. Chem Sci. 2022;13(13):3661-73. doi:10.1039/D1SC06946B
18. Jiang D, Hsieh CY, Wu Z, Kang Y, Wang J, Wang E, et al. InteractionGraphNet: A novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. J Med Chem. 2021;64(24):18209-232. doi:10.1021/acs.jmedchem.1c01830
19. Yang Z, Zhong W, Lv Q, Dong T, Chen G, Chen CYC. Interaction-based inductive bias in graph neural networks: Enhancing protein–ligand binding affinity predictions from 3D structures. IEEE Trans Pattern Anal Mach Intell. 2024;46(12):8191-208. doi:10.1109/TPAMI.2024.3400515
20. Karimi M, Wu D, Wang Z, Shen Y. Explainable deep relational networks for predicting compound–protein affinities and contacts. J Chem Inf Model. 2021;61(1):46-66. doi:10.1021/acs.jcim.0c00866
21. 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
22. 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
23. Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential deep learning for guided molecular property prediction and discovery. ACS Cent Sci. 2021;7(8):1356-67. doi:10.1021/acscentsci.1c00546
24. Ryu S, Kwon Y, Kim WY. A Bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification. Chem Sci. 2019;10(36):8438-46. doi:10.1039/C9SC01992H
25. Musil F, Willatt MJ, Langovoy MA, Ceriotti M. Fast and accurate uncertainty estimation in chemical machine learning. J Chem Theory Comput. 2019;15(2):906-15. doi:10.1021/acs.jctc.8b00959
26. Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, et al. Comparative assessment of scoring functions: The CASF-2016 update. J Chem Inf Model. 2019;59(2):895-913. doi:10.1021/acs.jcim.8b00545
27. Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. Chem Sci. 2024;15(9):3130-9. doi:10.1039/D3SC04185A
28. Jiménez-Luna J, Škalič M, Weskamp N. Benchmarking molecular feature attribution methods with activity cliffs. J Chem Inf Model. 2022;62(2):274-83. doi:10.1021/acs.jcim.1c01163
29. Rao J, Zheng S, Lu Y, Yang Y. Quantitative evaluation of explainable graph neural networks for molecular property prediction. Patterns (N Y). 2022;3(12):100628. doi:10.1016/j.patter.2022.100628
30. 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
31. Cournia Z, Allen B, Sherman W. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. J Chem Inf Model. 2017;57(12):2911-37. doi:10.1021/acs.jcim.7b00564
32. Williams-Noonan BJ, Yuriev E, Chalmers DK. Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry. J Med Chem. 2018;61(3):638-49. doi:10.1021/acs.jmedchem.7b00681
33. Ross GA, Lu C, Scarabelli G, Albanese SK, Houang E, Abel R, et al. The maximal and current accuracy of rigorous protein–ligand binding free energy calculations. Commun Chem. 2023;6(1):222. doi:10.1038/s42004-023-01019-9
34. Kanev GK, de Graaf C, Westerman BA, de Esch IJP, Kooistra AJ. KLIFS: An overhaul after the first 5 years of supporting kinase research. Nucleic Acids Res. 2021;49(D1). doi:10.1093/nar/gkaa895
35. Azimi S, Gallicchio E. Binding selectivity analysis from alchemical receptor hopping and swapping free energy calculations. J Phys Chem B. 2024;128(44):10841-52. doi:10.1021/acs.jpcb.4c05732