



Review Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Evaluation of Turmeric Rhizome Extract's Antiestrogenic Potential in Reducing Weight and Uterine Diameters in Rats

Yuki Tanaka¹, Haruto Suzuki^{2*}, Kenji Nakamura³, Rina Sato⁴, Takeshi Mori⁵

¹Department of Endocrine Pharmacology and Natural Product Research, Graduate School of Pharmaceutical Sciences, University of Tokyo, Tokyo, Japan.

²Department of Pharmacognosy and Medicinal Plant Chemistry, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan.

³Department of Experimental Physiology and Reproductive Biology, Graduate School of Medicine, Osaka University, Osaka, Japan.

⁴Department of Biochemistry and Nutritional Sciences, Graduate School of Life Sciences, Tohoku University, Sendai, Japan.

⁵Department of Toxicology and Safety Assessment, Graduate School of Pharmaceutical Sciences, Nagoya University, Nagoya, Japan.

*Email: haruto.suzuki@pharm.kyoto-u.ac.jp

ABSTRACT

Molecular dynamics simulation has become capable of resolving conformational states, ligand-binding pathways, kinetic events, allosteric coupling and binding thermodynamics at resolutions that increasingly intersect with medicinal-chemistry decisions. Yet greater trajectory length, algorithmic sophistication or apparent molecular detail does not by itself establish that a simulated observation is sufficiently reliable, mechanistically interpretable or transferable to support compound selection. This Critical Review examines the inferential steps separating molecular trajectories from therapeutic decisions. Particular attention is given to conformational-state discovery, binding and unbinding mechanisms, allostery, free-energy estimation, membrane and solvent representation, and integration with structural and biophysical experiments. Across these domains, three recurring limitations emerge: simulated observables depend on representation and sampling choices; different computational outputs answer non-equivalent pharmacological questions; and agreement within a computational workflow is not equivalent to external validation or downstream therapeutic relevance. The review therefore develops a proposed evidence architecture in which trajectory-derived states, kinetic quantities and thermodynamic estimates advance toward compound decisions only through question-specific validation gates. Rather than treating molecular dynamics as a single predictive technology, this framework positions it as a family of conditional inference tools whose usefulness depends on matching the simulated observable, uncertainty structure and validation design to the medicinal-chemistry decision being made.

Keywords: Molecular dynamics, Enhanced sampling, Binding kinetics, Binding free energy, Conformational ensembles, Drug discovery

Received: 17 March 2026; **Revised:** 07 June 2026; **Accepted:** 10 June 2026

INTRODUCTION

Molecular dynamics (MD) now occupies a different position in pharmaceutical research from the one it held when trajectories were used mainly to inspect structural flexibility. Increasing computational scale, enhanced-sampling algorithms and machine-learning-assisted methods are making more demanding molecular questions accessible, including rare conformational transitions and drug-binding processes [1]. At the same time, modern molecular

simulation can distinguish quantities that static structural models cannot: equilibrium populations, free-energy differences, association and dissociation processes, transition pathways and residence-time-related behavior [2]. Dynamic information is particularly consequential for targets such as G protein-coupled receptors (GPCRs), where ligand action can depend on ensembles of conformations rather than a single canonical structure [3].

The resulting opportunity creates an equally important interpretive problem. A simulation can generate a long trajectory without establishing that the relevant biological states were sampled. A rare event can be observed without establishing its population or rate. A numerically converged free-energy estimate can remain wrong because of the force field, protonation state, binding pose or environmental representation. Conversely, simulation-derived hypotheses can become unusually informative when they are connected to orthogonal structural, biophysical or pharmacological evidence.

The central argument of this review is therefore that the pharmaceutical value of MD should not be judged by trajectory production alone. The more useful question is what decision-relevant inference is justified at each stage between molecular sampling and compound prioritization. The framework developed here is explicitly proposed: it separates trajectory generation, molecular-state inference, quantitative observable estimation, mechanistic interpretation, external validation and compound-level action. These levels can reinforce one another, but they are not interchangeable. The distinction becomes particularly important as automated workflows and artificial intelligence make it easier to generate convincing molecular outputs faster than their assumptions can be interrogated.

What molecular-dynamics trajectories actually establish

A trajectory is a finite realization of motion under a defined molecular representation, Hamiltonian, force field, solvent treatment, boundary condition and sampling protocol. It is therefore neither a direct recording of biological reality nor an intrinsic mechanism. Analysis transforms atomic coordinates into higher-level objects such as conformational clusters, collective variables, transition networks, kinetic models or free-energy surfaces. Markov state models illustrate this distinction clearly: they can organize trajectories into metastable states and estimate slow transitions, but their interpretation depends on state decomposition, lag-time selection, sampling adequacy and model validation [4]. The trajectory supplies observations; the state model supplies an inference about those observations.

Machine learning expands this inferential layer rather than eliminating it. Neural potentials, dimensionality-reduction methods, generative models and learned collective variables can accelerate molecular simulation, extract slow coordinates and approximate thermodynamic or kinetic quantities [5]. More recent approaches explicitly target Boltzmann-weighted structural ensembles rather than single predicted conformations [6]. Such methods are important because medicinal chemistry frequently concerns populations and transitions, not simply whether one plausible structure can be generated. Yet an ensemble produced by a learned model still requires evidence that its distribution, domain of applicability and physical constraints are appropriate for the question being asked.

Sampling uncertainty is equally fundamental. Independent MD replicas can diverge sufficiently for single-trajectory analysis to support false-positive conclusions, making replication a basic safeguard against overinterpreting stochastic trajectory behavior [7]. Replication, however, should not be converted into a universal numerical rule. The number and length of trajectories required depend on the landscape, the observable, the expected transition timescale and the intended inference. Repeated short trajectories may be useful for some ensemble quantities but insufficient for a rare activated transition. Conversely, extending one trajectory does not necessarily repair poor initialization or inadequate exploration of disconnected basins.

A useful interpretation hierarchy therefore begins with four questions. Was the molecular representation appropriate? Was the state space relevant to the decision sampled? Is the requested observable identifiable from that sampling? Is uncertainty reported at the same level as the claimed inference? Only after these conditions are addressed should molecular behavior be translated into mechanism or medicinal-chemistry action. This hierarchy is proposed here as an inference discipline rather than as an assertion that one simulation design is universally preferable.

Conformational-state discovery

Drug targets often occupy multiple conformational basins whose populations and accessibility are not recoverable from one experimental structure. Computational studies increasingly exploit this property to identify alternative conformations, cryptic pockets and transiently accessible binding geometries [8–10]. These developments

broaden the structural search space available to medicinal chemistry. They also create a distinction that is easily lost: detecting a geometrically plausible state does not establish that the state is appreciably populated under physiological conditions, and detecting a pocket does not establish that it is chemically tractable or pharmacologically consequential.

Alternative-state prediction using modern machine-learning approaches provides a useful example. Models can recover ligand-specific protein conformations from unbound inputs and thereby generate structures that conventional rigid-receptor docking would miss [11]. Such predictions may be excellent hypotheses for docking, design or experimental interrogation, but their evidentiary status differs from that of a transition repeatedly sampled under a defined dynamical model. The distinction matters because structural plausibility, equilibrium population and transition kinetics are different quantities.

Trajectory-based state discovery has its own representation problem. Collective variables and metastable-state definitions can reorganize the same underlying trajectories in different ways, especially when the coordinates chosen for dimensional reduction do not align with the slow process relevant to the scientific question [12]. A chemically intuitive coordinate is not necessarily the kinetically optimal coordinate, and a statistically separable cluster is not necessarily a pharmacologically distinct state. State discovery should therefore be treated as an iterative model-building problem in which alternative decompositions, kinetic separation, structural interpretation and external observables are tested against one another.

For medicinal chemistry, the actionable unit should consequently be neither “a new conformation was found” nor “a pocket opened.” A more defensible unit is a decision state that records what was observed, how strongly sampling supports it, what mechanistic interpretation remains unresolved and what new observation would change the proposed action. This preserves potentially useful structural discoveries without silently converting them into claims of occupancy, mechanism or druggability.

The practical distinctions are compacted in **Table 1** because they represent alternative evidentiary states rather than a simple list of conformational-analysis methods.

Table 1. Decision states for converting conformational discoveries into medicinal-chemistry hypotheses

Decision context	Evidence and limits	Permitted action
Mechanistic interpretation permitted	Sampling or structural support available	decision state / Unresolved question before compound action
A reproducible state exists within the simulated model	State appears across replicate sampling and remains structurally coherent	Manuscript-derived decision state: Recurrent state within independently sampled trajectories Unresolved question before compound action: Is its population and interconversion behavior consistent with experiment?
The state is accessible under the applied biasing and reconstruction procedure	Transition is observed after deliberate acceleration of exploration	Manuscript-derived decision state: Rare state reached only after enhanced sampling Unresolved question before compound action: Is the recovered population or pathway robust to alternative sampling choices?
The model nominates a candidate receptor configuration	Ligand-conditioned structure is geometrically plausible and benchmark-compatible	Manuscript-derived decision state: Alternative conformation generated by a learned structural model Unresolved question before compound action: Does an independent dynamical or experimental observable support that state?
A conditional binding opportunity may exist	Pocket formation is coupled to selected conformational fluctuations	Manuscript-derived decision state: Transient cavity or cryptic pocket detected during conformational exploration Unresolved question before compound action: Is the pocket sufficiently occupied, chemically accessible and functionally relevant?
The molecular landscape is incompletely resolved at the chosen representation level	Multiple representations fit substantial portions of the trajectory	Manuscript-derived decision state: State assignment changes with collective-variable or clustering choice

		Unresolved question before compound action: Which decomposition best predicts an independent kinetic, structural or pharmacological observable?
The state remains a simulation-supported hypothesis	Internal simulation reproducibility is adequate	Manuscript-derived decision state: State persists computationally but lacks orthogonal support
		Unresolved question before compound action: What experiment can discriminate the proposed state from competing interpretations?

Binding and unbinding mechanisms

Ligand binding is a rare-event problem in many pharmaceutically relevant systems. Conventional MD may describe local fluctuations well while failing to observe association or dissociation on accessible timescales. Enhanced-sampling methods, machine learning and carefully designed collective variables have therefore been combined to increase the throughput of binding and unbinding simulations [13]. The principal advance is not simply more events. These methods can generate path-resolved hypotheses linking encounter complexes, intermediate states, pocket rearrangements and final binding configurations.

Residence time illustrates both the potential and the difficulty. A state-predictive information-bottleneck approach coupled with enhanced sampling has been used to estimate protein-ligand residence times spanning a broad range of timescales in benchmark complexes [14]. Ligand Gaussian accelerated MD likewise seeks repeated binding and unbinding while recovering both thermodynamic and kinetic quantities [15]. These studies demonstrate that kinetics need not remain outside atomistic simulation, but they do not imply that one protocol has uniform accuracy across targets and ligands.

The importance of preserving failure cases is shown by quantitative work on PYK2 ligands. Umbrella sampling combined with milestoning yielded dissociation-rate estimates close to experimental values for two related ligands but produced a much larger error for a third [16]. Such heterogeneity is analytically valuable. It shows why success on a subset of compounds should not be converted into a target-independent performance claim and why reaction-coordinate quality, pathway completeness and state definition must be interrogated ligand by ligand.

Association introduces another level of scale. Long-range diffusion and electrostatic steering can determine access to the protein before atomistic interactions near the pocket become dominant. Coupling Brownian dynamics to MD provides one way of separating these regimes when estimating association rates [17]. At the opposite end of the workflow spectrum, automated residence-time methods now combine pathway discovery with enhanced-sampling estimation across congeneric compound series [18]. Their value for discovery depends not only on accuracy but also on whether the calculation is reproducible, sufficiently automated and fast enough to be performed before the medicinal-chemistry decision has already been made.

Mechanistic reconstruction can also benefit from combining state-based and path-based methods. Integration of a Markov state representation with path metadynamics has been used to characterize ligand binding to β 2-adrenergic receptor [19]. The deeper lesson is that pathways, rates and state populations are related but distinct outputs. An observed egress pathway does not by itself determine the dissociation rate; a calculated residence time does not establish that a single molecular pathway dominates experimentally; and kinetic discrimination between compounds need not imply an efficacy difference in vivo. Binding kinetics are therefore most useful when treated as one component of a broader compound profile rather than as a replacement for affinity, pharmacodynamics or exposure.

Allostery

Allostery poses an even greater interpretive challenge because correlated molecular motion can be mistaken for causal communication. Simulations can reveal residues, pathways and transient pockets whose behavior covaries with activation, but mechanistic confidence rises substantially when such predictions are experimentally perturbed. An integrative machine-learning and MD analysis of β 2-adrenergic receptor identified a candidate allosteric region and coupled computational predictions to experimental interrogation [20]. Work on the vasopressin V2 receptor similarly identified a dynamic allosteric site along the receptor activation landscape and supported its functional relevance with experimental analyses [21].

These examples support a bounded conclusion: simulation-derived allosteric hypotheses can become experimentally testable and, in specific systems, can survive such testing. They do not establish that every correlated residue network or transient pocket constitutes a causal allosteric mechanism. Several competing

explanations may produce similar trajectory correlations, including global conformational rearrangement, common coupling to a third structural coordinate or artifacts of the analysis representation.

The proposed evidentiary distinction is therefore between dynamic association, structural pathway and functional allostery. Dynamic association requires reproducible covariance or state dependence within the modeled ensemble. A structural pathway requires a more specific, physically interpretable sequence or network connecting sites. Functional allostery requires evidence that perturbing the proposed pathway or site alters a relevant functional output. This separation is particularly important for drug discovery because the attractive label “allosteric site” can otherwise move a computationally detected region directly into compound-design logic before ligandability and functional leverage have been established.

Free-energy estimation

Binding free energy occupies a central position between molecular description and medicinal-chemistry ranking because it converts a modeled ensemble into a scalar thermodynamic quantity. Yet the scalar can conceal substantial modeling structure. Contemporary free-energy practice spans alchemical transformations, pathway-based methods, nonequilibrium protocols and hybrid strategies, each with assumptions about state definition, sampling and convergence [22]. The critical question is therefore not whether a workflow returns a value for ΔG , but what molecular comparison that value represents and how convincingly uncertainty in the underlying states has been controlled.

Scale is increasing rapidly. Automated workflows have been developed to generate molecular-dynamics ensembles and absolute binding-free-energy estimates across large collections of protein-ligand complexes and binding poses [23]. Such scale is important for constructing physics-informed training data and for extending free-energy calculations toward screening-like settings. It does not remove the distinction between a computational label and an experimental affinity, particularly when pose preparation, solvent treatment or convergence is automated.

At the campaign level, free-energy calculations can contribute directly to medicinal chemistry. Kinome-wide calculations in a Wee1 inhibitor program demonstrate how energetic predictions can be incorporated into selectivity-oriented compound design rather than evaluated only as retrospective benchmark exercises [24]. This represents a meaningful decision-level use case, but its transferability remains conditional on target class, structural information, chemical series and the prospective workflow in which calculations are embedded.

Path-based nonequilibrium approaches broaden the types of conformational change that can be represented. Applications to Abl-Gleevec binding and RNA-ligand systems show that pathway construction can accommodate systems with substantial structural rearrangement, while also revealing sensitivity to environmental choices such as the water model [25]. Related work on nonequilibrium binding calculations shows that dissipation and convergence depend on how the path and simulation conditions are constructed [26]. Consequently, a free-energy method cannot be characterized adequately by estimator name alone; the path, solvent, endpoint states and sampling schedule belong to the evidentiary description.

Foundational relative and absolute binding-free-energy work reached the same practical conclusion from different directions. Relative binding free-energy methods can provide useful predictions within appropriately constructed ligand series, rigorous absolute protocols can estimate standard binding thermodynamics when the bound state and force-field representation are suitable, and alchemical-transfer approaches can extend calculations to more structurally diverse ligand transformations [27–29]. None of these methods eliminates the requirement to define the chemical comparison correctly.

Automation and machine learning may reduce some of the intervention required to construct potential-of-mean-force or binding-free-energy workflows [30]. However, benchmark design remains essential because methodological improvements can otherwise be judged on systems selected after the method is known to work. Community benchmarking has therefore remained a central requirement for evaluating advances in binding-free-energy prediction [31].

This review consequently proposes that free-energy estimates should be treated as gated evidence. The output may advance toward compound prioritization only when the modeled chemical state, convergence behavior, environmental representation, uncertainty and decision context are simultaneously adequate. **Figure 1** places free-energy estimation at the center of this inference problem rather than at the end of a generic simulation pipeline.

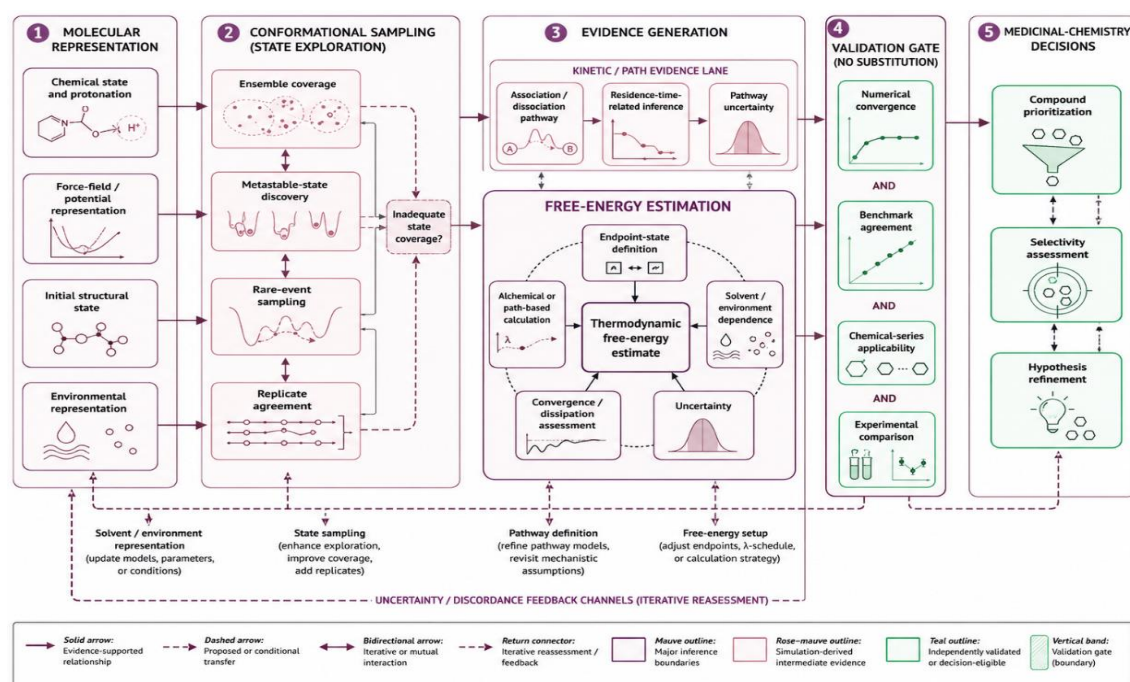


Figure 1. Free-energy evidence gates separating molecular sampling from compound-level action

Membrane and solvent effects

The environment surrounding a target is part of the molecular model, not passive visual context. This point is particularly important for membrane proteins, whose conformational energetics and ligand interactions can depend on lipid composition, hydration, ions and receptor orientation. Computational drug development for membrane targets therefore requires representations that preserve relevant environmental constraints rather than assuming that conclusions from soluble proteins transfer unchanged [32].

Solvent can also function as a probe of binding opportunity. Mixed-solvent MD applied to CYP3A4 identified multiple candidate ligand-binding regions beyond the canonical active site, illustrating how receptor flexibility and probe distribution can reveal sites that may be poorly represented by static-pocket analysis [33–35]. Yet three different observations must remain separate: preferential probe occupancy, displacement or stabilization of structured water, and a ligand pose supported by experimental density. Each can constrain molecular interpretation, but none is interchangeable with functional pharmacology.

Water-network effects are particularly instructive because they can alter both pose interpretation and predicted energetics. A ligand may appear geometrically compatible with a pocket while disrupting a favorable network, or an apparently unfavorable direct contact may be compensated by water-mediated interactions. Differences between computational predictions and experiment therefore provide information about the adequacy of the solvent model rather than simply constituting model “noise.” Likewise, cryo-EM-guided ligand modeling can use density and explicit water placement to constrain poses, but limited ligand density and conformational heterogeneity can leave more than one molecular explanation consistent with the map.

For decision making, environmental uncertainty should be represented upstream of compound ranking. If changing water model, membrane representation or probe conditions reverses the molecular interpretation, the appropriate response is not to average the conclusions into a false consensus. The decision should remain conditional until the environmental variable responsible for the instability is identified or bounded.

Table 2 therefore registers environmental decision states rather than presenting membrane and solvent effects as generic simulation “considerations.”

Table 2. Environmental decision states that determine whether a simulated binding interpretation is transportable

Evidence and limits	Permitted action
Observable supporting the current interpretation / Competing explanation / unresolved uncertainty	Environmental decision state / Decision consequence

Observable supporting the current interpretation: Core ligand contacts and pocket geometry remain coherent Competing explanation: Stability may reflect restraints or an overly rigid receptor Main unresolved uncertainty: Whether hydration free energies and local water structure are realistic	Environmental decision state: Binding geometry stable across justified solvent representations Decision consequence: Molecular interpretation can advance to energetic or experimental testing
Observable supporting the current interpretation: Explicit hydration pattern differentiates candidate poses or ligands Competing explanation: Force-field or water-model artifact Main unresolved uncertainty: Whether experimentally observable hydration-sensitive behavior supports the model	Environmental decision state: Water-network rearrangement changes predicted ligand preference Decision consequence: Retain ligand ranking as conditional
Observable supporting the current interpretation: Probe occupancy recurs with receptor flexibility Competing explanation: Probe chemistry may overfavor a region not used by drug-like ligands Main unresolved uncertainty: Whether conventional ligands occupy or perturb the site	Environmental decision state: Mixed-solvent probe enrichment reveals an additional cavity Decision consequence: Treat site as a hypothesis, not a validated pocket
Observable supporting the current interpretation: State occupancy responds to membrane environment Competing explanation: Difference may arise from insufficient equilibration or composition choice Main unresolved uncertainty: Which membrane variables reflect the relevant biological setting	Environmental decision state: Membrane composition alters receptor-state populations Decision consequence: Do not transfer soluble or single-bilayer conclusions automatically
Observable supporting the current interpretation: Cryo-EM map constrains placement Competing explanation: Several poses may fit limited ligand density Main unresolved uncertainty: Whether simulation, water placement or complementary data resolve orientation	Environmental decision state: Experimental density supports a ligand region but not a unique pose Decision consequence: Use an ensemble or competing-pose interpretation
Observable supporting the current interpretation: Direction of a key interaction or state preference is model-dependent Competing explanation: Multiple physically plausible environmental regimes Main unresolved uncertainty: Which representation is externally supported	Environmental decision state: Environmental model changes reverse the proposed mechanism Decision consequence: Return the inference upstream rather than rank compounds

Integrating simulation with structural and biophysical experiments

The strongest use of simulation is not necessarily independent prediction. Often it is the construction of a molecular explanation that can be constrained by measurements with different sensitivities. NMR spectroscopy, cryo-electron microscopy and MD each observe different projections of protein dynamics and operate across different effective spatial and temporal resolutions [36]. Their integration can reveal atomic-level pictures not accessible to any one method, but integration should not be confused with forcing all modalities to support one preferred mechanism.

Structural data can constrain starting states and ligand poses; NMR can report populations, exchange or local environmental changes; cryo-EM can expose conformational heterogeneity; mutational or functional experiments can test consequences of perturbing residues identified by simulation. Conversely, MD can connect structural snapshots, suggest transitions that motivate experimental state classification and identify molecular variables whose perturbation should discriminate competing explanations. The evidentiary value comes from this complementarity.

Discordance is therefore informative. A conformational basin that is stable in MD but unsupported by structural or spectroscopic evidence may reflect an inaccessible experimental timescale, a low-population state, an inaccurate molecular representation or insufficient experimental sensitivity. A cryo-EM class absent from simulation may reflect incomplete sampling, incorrect starting conditions or a compositional state not represented computationally. The appropriate analytical response is to locate the source of disagreement, not to designate one modality automatically correct.

This review proposes a cross-scale validation rule: confidence should increase most when an independently measured observable was not used merely to construct the same computational model and nevertheless discriminates among competing molecular interpretations. Validation becomes weaker when the experimental quantity is effectively recycled as a restraint and then presented as independent confirmation. Likewise, agreement

on one structural observable should not be generalized automatically to binding kinetics, free energy or functional consequence.

The resulting architecture requires parallel evidence lanes until their measurement meanings have been preserved. **Figure 2** therefore makes experimental convergence a validation field rather than merging simulation and experiment into a single evidence stream.

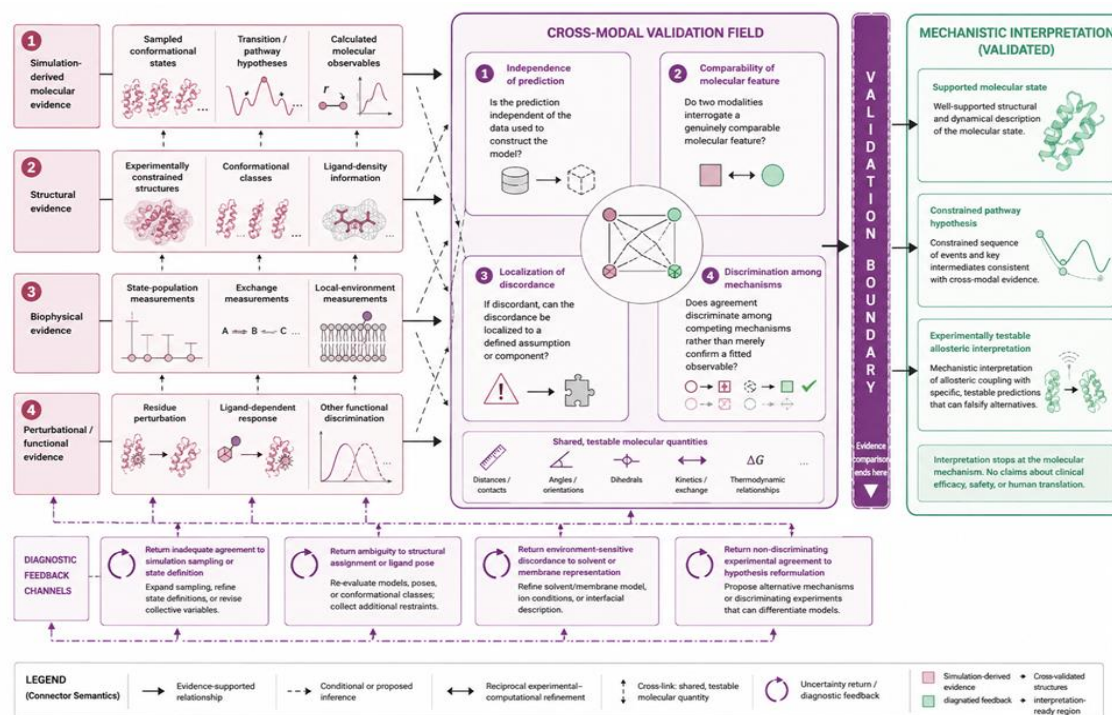


Figure 2. Orthogonal evidence lanes for assigning confidence to simulation-derived molecular states

Uncertainty and reproducibility

Uncertainty in molecular simulation is not a single error term. It arises from molecular representation, initial conditions, finite sampling, stochastic divergence, force-field assumptions, state decomposition, collective-variable choice, environmental modeling and numerical implementation. These sources act at different levels and should not be collapsed into one confidence value. A workflow may be internally reproducible while systematically misrepresenting the molecular system, whereas a physically reasonable model may still produce unstable estimates because a rare transition has been sampled inadequately.

Reproducibility therefore requires more than retaining trajectory coordinates. Workflow provenance, software versions, parameter files, initialization procedures and analysis transformations determine whether another group can reconstruct the computational inference. Martignac illustrates how explicit workflow graphs, traceable intermediate states and reusable simulation components can make coarse-grained molecular-simulation workflows more reproducible and auditable [37]. The broader principle applies beyond Martini simulations: pharmaceutical decisions require reconstruction of how an observable was produced, not merely access to its final numerical value.

Cross-platform comparison provides another level of assurance. Controlled evaluation of alchemical free-energy calculations using different molecular-dynamics engines and estimators has shown that strong inter-engine agreement can be achieved when protocols, ensembles and analysis are carefully harmonized [38]. Such agreement is important because it helps distinguish software-specific implementation effects from more general model behavior. Yet reproducibility and correctness remain different properties. Two engines can reproduce the same result because they share the same physical assumptions, including an inaccurate force field or inappropriate protonation state.

The proposed uncertainty architecture of this review therefore separates sampling uncertainty, model uncertainty, implementation uncertainty and decision uncertainty. Sampling uncertainty asks whether relevant states and transitions have been explored adequately. Model uncertainty concerns the representation of interactions, chemistry and environment. Implementation uncertainty concerns reproducibility across software, workflows and

analysis choices. Decision uncertainty concerns whether the remaining molecular uncertainty is small enough for the particular medicinal-chemistry action being considered.

These categories should propagate forward rather than disappear when a trajectory is converted into a scalar observable. A residence-time estimate that changes substantially with reaction-coordinate construction should remain conditionally interpretable even if one run converges numerically. Likewise, a binding free energy that reproduces across engines but changes when the assumed binding pose or water network is altered should not be treated as fully resolved. The practical objective is not to eliminate uncertainty before every decision, which is unrealistic, but to identify which unresolved uncertainty could reverse the action being considered.

From molecular observation to compound decision

A molecular observation becomes decision-relevant only when its evidentiary meaning matches the medicinal-chemistry question. A newly sampled pocket may justify synthesis of a probe compound, whereas it may be insufficient to justify deprioritizing a chemically validated orthosteric series. A calculated residence time may help discriminate compounds when exposure and target biology make sustained occupancy pharmacologically relevant, but the same quantity may have little value when downstream response is dominated by rapid signaling adaptation. A binding free-energy difference may guide congeneric optimization within a validated domain but become unreliable after a major scaffold or binding-mode change.

This review therefore proposes a decision-readiness hierarchy with six separable stages. The first is molecular representation: the relevant chemical and environmental state must be represented plausibly. The second is sampling adequacy: the states required for the intended observable must be accessible to the simulation. The third is observable estimation: the trajectory must support the particular structural, kinetic or thermodynamic quantity being claimed. The fourth is mechanistic interpretation: the observable must be distinguished from competing molecular explanations. The fifth is external validation: an independent measurement should test a discriminating implication where feasible. The sixth is compound action: only then should the molecular result influence ranking, redesign, experimental escalation or termination.

Movement through this hierarchy is not necessarily linear. Failure at a validation stage may require returning to protonation-state assignment, conformational sampling, pathway construction or environmental representation. Nor must every question reach the final stage. A conformational-state simulation used to design a mechanistic experiment may be successful without producing a compound ranking. Conversely, a relative free-energy calculation may support compound prioritization even when the entire microscopic binding pathway remains unresolved, provided the thermodynamic comparison is sufficiently validated for that chemical series.

The central rule is therefore non-interchangeability. Structural plausibility, state population, kinetic accessibility, thermodynamic preference and functional consequence answer different questions. Combining them can strengthen a decision, but one cannot substitute automatically for another.

Translational limits

The transition from molecular simulation to therapeutic decision becomes most fragile when the level of inference changes. A correctly predicted ligand pose does not establish favorable binding kinetics. Accurate affinity does not establish cellular target engagement. Target engagement does not establish pharmacodynamic response. Pharmacodynamic response does not establish clinically useful exposure, safety or therapeutic benefit. Molecular simulation can contribute to several of these stages, but success at one level does not validate the next.

Transferability can fail even before biological scale changes. Learned structural models may perform differently outside their training domain [11]. Ligand-specific kinetic errors can appear within a single target series [16]. Automated residence-time workflows can be encouraging across selected targets while remaining dependent on pathway discovery and chemical context [18]. Nonequilibrium free-energy calculations can change materially with path construction and solvent representation [26]. Benchmarking literature likewise shows that free-energy performance remains dependent on the molecular transformations and systems being tested [31].

Target environment adds further boundaries. Membrane proteins may respond to lipid composition, hydration and receptor-state equilibria in ways not reproduced by a generic soluble-protein model [32]. Experimental integration can constrain such interpretations, but agreement between a simulation and one structural observable does not validate every downstream kinetic or functional claim [36]. Reproducibility across computational engines similarly increases confidence in numerical implementation without establishing that the underlying molecular model is biologically complete [38].

These limits make a single global category of “validated MD prediction” unhelpful. Translational interpretation is better represented by the conditions under which an inference holds, weakens, changes direction or remains indeterminate. **Table 3** formalizes that boundary logic because the decisive information is not simply the limitation itself, but which scientific variable determines whether the molecular conclusion remains actionable.

Table 3. Boundary conditions governing transfer from molecular simulation to therapeutic interpretation

Source state and scale	Transfer conditions	Qualification endpoint
Classification basis / Molecular interpretation that may remain supported	Boundary that weakens or changes that interpretation / Competing explanation	Validation discriminator / Framework endpoint
Classification basis: Chemical series remains within a validated transformation domain Molecular interpretation that may remain supported: Relative energetic ordering may inform compound prioritization	Boundary that weakens or changes that interpretation: Major scaffold, pose or protonation change Competing explanation: Apparent ranking reflects an invalid comparison state	Validation discriminator: Experimental affinity or independently validated calculation under the new chemical regime Framework endpoint: Conditional compound ranking
Classification basis: Kinetic estimate is stable across defensible pathway models Molecular interpretation that may remain supported: Relative residence behavior may inform a target-specific hypothesis	Boundary that weakens or changes that interpretation: Alternative exit pathway materially changes the estimate Competing explanation: Reaction-coordinate construction determines the apparent kinetics	Validation discriminator: Independent kinetic experiment or alternative pathway reconstruction Framework endpoint: Kinetic hypothesis retained or returned for reassessment
Classification basis: Dynamic pocket recurs and is independently constrained Molecular interpretation that may remain supported: The pocket may represent a tractable conformational opportunity	Boundary that weakens or changes that interpretation: Pocket occupancy collapses under alternative environment or sampling Competing explanation: Transient cavity is simulation-specific or too rarely occupied	Validation discriminator: Structural, biophysical or ligand-binding evidence that discriminates pocket occupancy Framework endpoint: Experimental escalation rather than immediate lead prioritization
Classification basis: Membrane or solvent representation preserves the same qualitative state ordering Molecular interpretation that may remain supported: Environmental effects are unlikely to reverse the molecular conclusion	Boundary that weakens or changes that interpretation: Plausible environmental models reverse state or ligand preference Competing explanation: Model-dependent hydration or lipid coupling drives the apparent mechanism	Validation discriminator: Environment-sensitive experimental observable or higher-confidence representation Framework endpoint: Interpretation remains conditional
Classification basis: Multiple computational implementations reproduce the same observable Molecular interpretation that may remain supported: Software-specific numerical error is less likely	Boundary that weakens or changes that interpretation: Shared model assumptions remain untested Competing explanation: Common force-field or state-definition bias	Validation discriminator: Independent experimental observable or orthogonal physical model Framework endpoint: Reproducible but not yet externally validated
Classification basis: Molecular prediction agrees with one experimental modality Molecular interpretation that may remain supported: A specific molecular feature is constrained	Boundary that weakens or changes that interpretation: Agreement does not discriminate among competing mechanisms Competing explanation: Several molecular models explain the same measured signal	Validation discriminator: Orthogonal measurement designed around a different discriminating prediction Framework endpoint: Mechanistic confidence increases only after cross-modal discrimination
Classification basis: Molecular result is robust but downstream pharmacology is untested Molecular interpretation that may remain supported: Molecular mechanism remains a plausible local explanation	Boundary that weakens or changes that interpretation: Cellular, exposure or disease context dominates treatment response Competing explanation: Correct molecular mechanism lacks therapeutic relevance	Validation discriminator: Target engagement, pharmacodynamic and exposure evidence at the appropriate biological level Framework endpoint: Translational claim withheld

The table deliberately terminates several trajectories at experimental escalation or conditional interpretation rather than compound ranking. That outcome is not a failure of simulation. It is the appropriate consequence when the level of molecular evidence does not yet justify a stronger therapeutic claim.

Future standards

Future progress in molecular simulation should be judged less by the maximum trajectory length or nominal complexity of the algorithm than by the quality of the evidence chain linking simulation to decisions. Minimum reporting should include the represented chemical state, environmental model, initialization procedure, replicate design, sampling strategy, analysis transformations, uncertainty definition and exact observable used for decision making. Where simulations influence compound ranking, the domain in which the calculation has been benchmarked should also be explicit.

Reproducibility standards should increasingly emphasize executable provenance rather than narrative methods descriptions alone. Machine-readable workflows, archived input states, parameter versions and analysis dependencies would make it easier to identify whether disagreement originates in software, molecular representation or interpretation. Comparable standards are needed for enhanced-sampling reconstruction, where reported uncertainty should distinguish statistical convergence from uncertainty in collective-variable or pathway choice.

Validation should also become more question-specific. A conformational-state claim should be tested by an observable sensitive to state occupancy or exchange. A proposed binding pathway should be tested by kinetics or perturbations capable of discriminating alternative pathways. A free-energy workflow should be benchmarked within the chemical and structural domain in which it is expected to support medicinal chemistry. Agreement with a structural endpoint should not be treated as validation of an unmeasured kinetic or thermodynamic quantity.

Artificial intelligence will probably alter this landscape substantially. Boltzmann generators have demonstrated that deep generative models can learn transformations capable of producing equilibrium samples and facilitating free-energy calculations in model molecular systems [39]. Related ensemble-generation approaches may eventually reduce dependence on serial trajectory propagation for some classes of sampling problem. Their value, however, will depend on the same principles that govern conventional simulation: preservation of the target distribution, domain awareness, uncertainty quantification and independent validation.

The most useful future standard may therefore be a decision-centered one. Every simulation used in drug discovery should state not only what was calculated, but which decision the calculation is allowed to influence, which uncertainty could reverse that decision, and which external observation would trigger reassessment. This would move the field away from visually compelling trajectories and isolated benchmark metrics toward auditable computational evidence.

CONCLUSION

Molecular dynamics can expose conformational states, binding pathways, allosteric relationships, kinetic behavior and thermodynamic differences that are inaccessible to static structural reasoning alone. Its pharmaceutical value, however, depends on maintaining the distinction between observing motion in a model and establishing evidence sufficient for a compound decision.

The framework developed in this review treats trajectory generation, state inference, quantitative observable estimation, mechanistic interpretation, external validation and medicinal-chemistry action as separate but connected levels. Uncertainty can enter at each level, and successful performance at one level does not automatically validate the next. Environmental representation, sampling adequacy, model choice, reproducibility and experimental discrimination therefore belong to the decision architecture rather than functioning as secondary methodological details.

The central implication is not that molecular dynamics should be used more cautiously by default, but that it should be used more precisely. A simulation is most valuable when the observable it produces is matched to the decision being made, the relevant uncertainty is visible, and disagreement redirects analysis rather than being concealed by a final scalar estimate. Under those conditions, trajectories can become more than molecular illustrations: they can serve as disciplined, testable components of therapeutic decision making.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Rossetti G, Mandelli D. How exascale computing can shape drug design: A perspective from multiscale QM/MM molecular dynamics simulations and machine learning-aided enhanced sampling algorithms. *Curr Opin Struct Biol.* 2024;86:102814. doi:10.1016/j.sbi.2024.102814
2. Decherchi S, Cavalli A. Thermodynamics and kinetics of drug-target binding by molecular simulation. *Chem Rev.* 2020;120(23):12788-833. doi:10.1021/acs.chemrev.0c00534
3. Conflitti P, Lyman E, Sansom MSP, Hildebrand PW, Gutiérrez-de-Terán H, Carloni P, et al. Functional dynamics of G protein-coupled receptors reveal new routes for drug discovery. *Nat Rev Drug Discov.* 2025;24(4):251-75. doi:10.1038/s41573-024-01083-3
4. Husic BE, Pande VS. Markov state models: From an art to a science. *J Am Chem Soc.* 2018;140(7):2386-96. doi:10.1021/jacs.7b12191
5. Noé F, Tkatchenko A, Müller KR, Clementi C. Machine learning for molecular simulation. *Annu Rev Phys Chem.* 2020;71:361-90. doi:10.1146/annurev-physchem-042018-052331
6. Aranganathan A, Gu X, Wang D, Vani BP, Tiwary P. Modeling Boltzmann-weighted structural ensembles of proteins using artificial intelligence-based methods. *Curr Opin Struct Biol.* 2025;91:103000. doi:10.1016/j.sbi.2025.103000
7. Knapp B, Ospina L, Deane CM. Avoiding false positive conclusions in molecular simulation: The importance of replicas. *J Chem Theory Comput.* 2018;14(12):6127-38. doi:10.1021/acs.jctc.8b00391
8. Bryant P, Noé F. Structure prediction of alternative protein conformations. *Nat Commun.* 2024;15(1):7328. doi:10.1038/s41467-024-51507-2
9. Bemelmans MP, Cournia Z, Damm-Ganamet KL, Gervasio FL, Pande V. Computational advances in discovering cryptic pockets for drug discovery. *Curr Opin Struct Biol.* 2025;90:102975. doi:10.1016/j.sbi.2024.102975
10. Zhang S, Bowman GR. Decrypting cryptic pockets with physics-based simulations and artificial intelligence. *Curr Opin Struct Biol.* 2026;96:103215. doi:10.1016/j.sbi.2025.103215
11. Lu W, Zhang J, Huang W, Zhang Z, Jia X, Wang Z, et al. DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nat Commun.* 2024;15(1):1071. doi:10.1038/s41467-024-45461-2
12. Sittel F, Stock G. Perspective: Identification of collective variables and metastable states of protein dynamics. *J Chem Phys.* 2018;149(15):150901. doi:10.1063/1.5049637
13. Zia SR, Coricello A, Bottegoni G. Increased throughput in methods for simulating protein ligand binding and unbinding. *Curr Opin Struct Biol.* 2024;87:102871. doi:10.1016/j.sbi.2024.102871
14. Lee S, Wang D, Seeliger MA, Tiwary P. Calculating protein-ligand residence times through state predictive information bottleneck based enhanced sampling. *J Chem Theory Comput.* 2024;20(14):6341-9. doi:10.1021/acs.jctc.4c00503
15. Wang J, Miao Y. Ligand Gaussian accelerated molecular dynamics 3 (LiGaMD3): Improved calculations of binding thermodynamics and kinetics of both small molecules and flexible peptides. *J Chem Theory Comput.* 2024;20(14):5829-41. doi:10.1021/acs.jctc.4c00502
16. Spiriti J, Wong CF. Quantitative prediction of dissociation rates of PYK2 ligands using umbrella sampling and milestone. *J Chem Theory Comput.* 2024;20(9):4029-44. doi:10.1021/acs.jctc.4c00192
17. Muñiz-Chicharro A, Ganotra GK, Wade RC. A multiscale simulation approach to compute protein-ligand association rate constants by combining Brownian dynamics and molecular dynamics. *J Chem Inf Model.* 2025;65(20):11215-31. doi:10.1021/acs.jcim.5c01488
18. Smith Z, Branduardi D, Lupyan D, D'Arrigo G, Tiwary P, Wang L, et al. Toward automated physics-based absolute drug residence time predictions. *J Chem Inf Model.* 2025;65(24):13360-73. doi:10.1021/acs.jcim.5c01832
19. Bernetti M, Masetti M, Recanatini M, Amaro RE, Cavalli A. An integrated Markov state model and path metadynamics approach to characterize drug binding processes. *J Chem Theory Comput.* 2019;15(10):5689-702. doi:10.1021/acs.jctc.9b00450

20. Chen X, Wang K, Chen J, Wu C, Mao J, Song Y, et al. Integrative residue-intuitive machine learning and MD approach to unveil allosteric site and mechanism for β 2AR. *Nat Commun.* 2024;15(1):8130. doi:10.1038/s41467-024-52399-y
21. Qiao X, Zhu C, Lan X, Li M, Li N, Huang J, et al. Identification of an unexplored dynamic allosteric site on the activation pathway of the vasopressin V2 receptor. *J Med Chem.* 2025;68(15):16766-80. doi:10.1021/acs.jmedchem.5c01619
22. Ghidini A, Serra E, Cavalli A. On free energy calculations in drug discovery. *Acc Chem Res.* 2025;58(20):3137-45. doi:10.1021/acs.accounts.5c00465
23. Wingbermühle S, Biswas AD, Bonanni D, Shugaeva T, Gadioli D, Beránek J, et al. Molecular dynamics workflows to compute large-scale sets of absolute binding free energies aiding drug candidate and binding pose selection. *J Chem Theory Comput.* 2026;22(10):5122-42. doi:10.1021/acs.jctc.5c02127
24. Knight JL, Clark AJ, Wang J, Placzek A, Bos PH, Bhat S, et al. Harnessing free energy calculations for kinome-wide selectivity in drug discovery campaigns with a Wee1 case study. *Nat Commun.* 2025;16(1):7962. doi:10.1038/s41467-025-62722-w
25. Serra E, Ghidini A, Aguti R, Bernetti M, Decherchi S, Cavalli A. Path-based nonequilibrium binding free energy estimation, from protein-ligand to RNA-ligand binding. *J Chem Inf Model.* 2025;65(12):6057-72. doi:10.1021/acs.jcim.5c00452
26. Serra E, Ghidini A, Decherchi S, Cavalli A. Nonequilibrium binding free energy simulations: Minimizing dissipation. *J Chem Theory Comput.* 2025;21(4):2079-94. doi:10.1021/acs.jctc.4c01453
27. 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
28. Fu H, Chen H, Blazhynska M, Goulard Coderc de Lacam E, Szczepaniak F, Pavlova A, et al. Accurate determination of protein standard binding free energies from molecular dynamics simulations. *Nat Protoc.* 2022;17(4):1114-41. doi:10.1038/s41596-021-00676-1
29. Azimi S, Khuttan S, Wu JZ, Pal RK, Gallicchio E. Relative binding free energy calculations for ligands with diverse scaffolds with the alchemical transfer method. *J Chem Inf Model.* 2022;62(2):309-23. doi:10.1021/acs.jcim.1c01129
30. Bertazzo M, Gobbo D, Decherchi S, Cavalli A. Machine learning and enhanced sampling simulations for computing the potential of mean force and standard binding free energy. *J Chem Theory Comput.* 2021;17(8):5287-300. doi:10.1021/acs.jctc.1c00177
31. Mobley DL, Gilson MK. Predicting binding free energies: Frontiers and benchmarks. *Annu Rev Biophys.* 2017;46:531-58. doi:10.1146/annurev-biophys-070816-033654
32. Li H, Sun X, Cui W, Xu M, Dong J, Ekundayo BE, et al. Computational drug development for membrane protein targets. *Nat Biotechnol.* 2024;42(2):229-42. doi:10.1038/s41587-023-01987-2
33. Feng Y, Gong C, Zhu J, Liu G, Tang Y, Li W. Unraveling the ligand-binding sites of CYP3A4 by molecular dynamics simulations with solvent probes. *J Chem Inf Model.* 2024;64(8):3451-64. doi:10.1021/acs.jcim.4c00089
34. Szalai TV, Bajusz D, Börzsei R, Zsidó BZ, Ilaš J, Ferenczy GG, et al. Effect of water networks on ligand binding: Computational predictions vs experiments. *J Chem Inf Model.* 2024;64(23):8980-98. doi:10.1021/acs.jcim.4c01291
35. Zhao Z, Tajkhorshid E. GOLEM: Automated and robust cryo-EM-guided ligand docking with explicit water molecules. *J Chem Inf Model.* 2024;64(14):5680-90. doi:10.1021/acs.jcim.4c00917
36. Zadorozhnyi R, Gronenborn AM, Polenova T. Integrative approaches for characterizing protein dynamics: NMR, CryoEM, and computer simulations. *Curr Opin Struct Biol.* 2024;84:102736. doi:10.1016/j.sbi.2023.102736
37. Bereau T, Walter LJ, Rudzinski JF. Martignac: Computational workflows for reproducible, traceable, and composable coarse-grained Martini simulations. *J Chem Inf Model.* 2024;64(24):9413-23. doi:10.1021/acs.jcim.4c01754
38. Wade AD, Bhati AP, Wan S, Coveney PV. Alchemical free energy estimators and molecular dynamics engines: Accuracy, precision, and reproducibility. *J Chem Theory Comput.* 2022;18(6):3972-87. doi:10.1021/acs.jctc.2c00114
39. Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science.* 2019;365(6457). doi:10.1126/science.aaw1147