Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2026, 15(1):110-122
https://doi.org/10.51847/akeqidLQGG

Review Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Enhanced Sampling for Pharmaceutical Molecular Dynamics: A Systematic Review of Accuracy, Efficiency, Reproducibility, and Mechanistic Validity

Zsolt Kovács[1]*, Katalin Szabo[1], Gabor Nagy[2], Eszter Toth[3]

[1]Department of Enhanced Sampling and MD Accuracy, Faculty of Pharmacy, Semmelweis University, Budapest, Hungary
[2]Department of Efficiency and Reproducibility, Faculty of Pharmacy, University of Debrecen, Debrecen, Hungary
[3]Department of Mechanistic Validity and Systematic Review, Faculty of Pharmacy, University of Pécs, Pécs, Hungary.

***Email:** zsolt.kovacs@semmelweis.hu*

ABSTRACT

Molecular dynamics can characterize conformational ensembles, energy landscapes, ligand-recognition pathways, allosteric transitions, and molecular mechanisms relevant to pharmaceutical research, but conventional trajectories frequently undersample rare events and slow collective motions. Enhanced-sampling methods address this limitation through biased potentials, generalized ensembles, replica exchange, pathway sampling, adaptive trajectory generation, or learned collective variables. This systematic methodological review evaluates how these approaches support four distinct claims: accurate recovery of equilibrium or kinetic properties, computational efficiency, reproducibility across implementations, and mechanistic validity. A protocol-driven search and selection process was organized around explicit review questions, reproducible database syntax, prespecified eligibility criteria, duplicate handling, structured extraction, and qualitative appraisal. The evidence was synthesized by method family, target observable, estimator requirements, implementation context, convergence assessment, and validation stage. The reviewed literature indicates that acceleration is not equivalent to correctness: apparent exploration, convergence, or computational speed may coexist with collective-variable omission, force-field error, reweighting instability, implementation sensitivity, or distorted kinetics. Collective-variable and machine-learning approaches can improve access to slow modes, whereas replica, tempering, pathway, and adaptive methods offer complementary strategies for overcoming barriers or constructing transition ensembles. However, comparative performance remains conditional on the molecular system, observable, initial information, algorithmic settings, and validation reference. Reproducibility further depends on transparent software versions, parameters, randomization, input files, analysis definitions, and uncertainty assessment. Mechanistic interpretation is strongest when sampled ensembles and pathways are consistent with independent experimental observables and when alternative explanations are examined. Enhanced sampling may therefore support pharmaceutical molecular dynamics when method choice, estimand, convergence, uncertainty, and validation are aligned, but no method family can be considered universally accurate, efficient, reproducible, or mechanistically valid.

Key words: *Molecular dynamics, Conformational ensemble, Energy landscape, Sampling, Drug–target recognition, Simulation validation*

Received: 17 January 2026; **Revised:** 19 March 2026; **Accepted:** 21 March 2026

INTRODUCTION

Molecular dynamics provides atomistic, time-resolved representations of molecular motion and can connect static structures with conformational exchange, ligand recognition, hydration, membrane interactions, and allosteric communication. In pharmaceutical applications, these trajectories may help interpret how a drug encounters a binding site, how a protein redistributes among functional states, or how mutations and chemical modifications alter an interaction network. Such interpretation nevertheless remains constrained by the accessible trajectory duration, the accuracy of the molecular model, and the rarity of many biologically consequential transitions [1].

Enhanced sampling modifies trajectory generation or statistical weighting to increase access to configurations that conventional simulation visits infrequently. Metadynamics, for example, introduces a history-dependent bias along selected collective variables to discourage repeated sampling of previously explored regions and facilitate reconstruction of an underlying free-energy landscape. Its usefulness therefore depends not only on acceleration, but also on whether the collective variables resolve the slow process, whether the bias is deposited appropriately, and whether reweighting or free-energy reconstruction remains reliable [2].

The term *enhanced sampling* also covers methods pursuing different estimands. Some approaches target equilibrium conformational probabilities or free-energy differences, whereas others estimate pathways, transition mechanisms, binding kinetics, or residence times. These outputs are not interchangeable. A trajectory may identify a plausible dissociation route without recovering an unbiased rate, and a converged free-energy estimate may remain inaccurate if relevant states are absent or the force field is deficient. Method selection must therefore follow the thermodynamic, kinetic, or mechanistic question being asked rather than the general desire to generate more configurations [3].

The resulting literature is methodologically heterogeneous: studies differ in molecular systems, collective variables, bias forms, simulation engines, convergence definitions, reference data, and intended pharmaceutical use. This review consequently examines evidence at the level of the claim being supported rather than treating all enhanced-sampling applications as directly comparable. It uses a PRISMA-compatible systematic-review structure to connect explicit questions, eligibility criteria, structured extraction, critical appraisal, and qualitative framework synthesis [4].

Systematic-review questions and protocol

The review addressed six questions. First, what precisely defined evidence question should enhanced-sampling studies answer when accuracy, efficiency, reproducibility, and mechanistic validity are considered jointly? Second, which publications provide methodologically relevant and extractable evidence? Third, what limitations or risks of bias affect their conclusions? Fourth, how consistent are findings across molecular systems, algorithms, and implementation contexts? Fifth, which determinants explain successful recovery, failure, or variation? Sixth, which conclusions are supported and which remain uncertain? These questions were prespecified to maintain a transparent link between the review objective, study selection, extraction, and synthesis [4].

Four evaluation domains were distinguished. Accuracy referred to agreement of a stated equilibrium, free-energy, kinetic, or structural estimand with an appropriate reference and was not inferred from trajectory length alone. Efficiency referred to useful independent information obtained relative to computational cost, rather than raw simulation speed alone. Reproducibility referred to the capacity to recover compatible findings under sufficiently specified computational conditions, distinct from exact trajectory replication. Mechanistic validity required more than molecular plausibility; it concerned whether the proposed states, pathways, or interactions were supported by appropriate physical analysis and, where available, independent observations. Thermodynamic and kinetic estimators were appraised separately because their assumptions and validation requirements differ [3].

The protocol specified structured narrative and framework synthesis rather than quantitative pooling. Meta-analysis was not planned because the evidence spans noncommensurate systems, observables, algorithms, estimators, validation references, and performance definitions. Appraisal focused on the molecular model, sampling adequacy, collective-variable or progress-coordinate justification, estimator validity, convergence diagnostics, sensitivity analysis, implementation transparency, external validation, and correspondence between the reported conclusion and the study design. No registration identifier, numerical quality score, certainty grade, or pooled effect estimate was assigned.

Search strategy, eligibility criteria, and study selection

Information sources comprised PubMed-indexed records, DOI and publisher metadata, targeted searches of relevant molecular-simulation and chemical-science journals, and citation-based verification of method and application papers. The core search logic combined molecular dynamics with enhanced sampling, metadynamics,

replica exchange, weighted ensemble, adaptive sampling, collective variables, free energy, accuracy, convergence, reproducibility, validation, kinetics, residence time, and allostery. A reproducible PubMed-compatible string was: ("molecular dynamics"[Title/Abstract]) AND ("enhanced sampling"[Title/Abstract] OR metadynamics OR "replica exchange" OR "weighted ensemble" OR "adaptive sampling" OR "collective variable" OR "free energy") AND (accuracy OR convergence OR reproducib* OR validation OR kinetic* OR "residence time" OR alloster*). Equivalent syntax was defined for broader bibliographic databases but was not represented as executed. Reporting followed PRISMA-compatible transparency principles [4].

Eligible publications were peer-reviewed journal articles with verifiable DOIs and direct relevance to at least one prespecified evidence domain. Method-development studies, comparative benchmarks, software and implementation papers, force-field validation studies, integrated computational–experimental investigations, and high-relevance methodological reviews were eligible when they supplied extractable evidence for a specific claim, table entry, figure component, or synthesis category. Publications were excluded when they addressed only adjacent simulation topics, lacked usable methodological detail, duplicated another journal record, offered unsupported promotional claims, or could not be mapped to the taxonomy and evaluation framework. Journal standing was checked in a relevant subject category, with the latest available ranking used where publication-year rankings were not yet available.

Duplicates were identified using DOI, normalized title, publication status, and correction or preprint relationships, with the definitive journal article retained. Screening was conducted against the same prespecified criteria at title-and-abstract and full-record levels, followed by a verification pass that rechecked bibliographic identity, eligibility, and exact claim fit. Extraction captured system type, enhanced-sampling family, collective-variable dependence, bias or target distribution, estimator and reweighting requirements, comparator, outcome, validation stage, software and hardware context, convergence evidence, sensitivity analysis, principal finding, and limitation. The review does not report database-hit totals, exclusion totals, or flow counts because these were not derived from database-native exports.

Method taxonomy and data-extraction framework

A useful taxonomy must describe what a method changes, what probability distribution or trajectory ensemble it targets, which coordinates or states must be supplied, and what estimator is required afterward. Biasing approaches may be organized by the distribution they seek to construct rather than solely by inherited algorithm names, thereby distinguishing the sampling objective from the particular form of the evolving bias [5]. **Figure 1** presents the original conceptual synthesis for taxonomy of enhanced-sampling methods, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

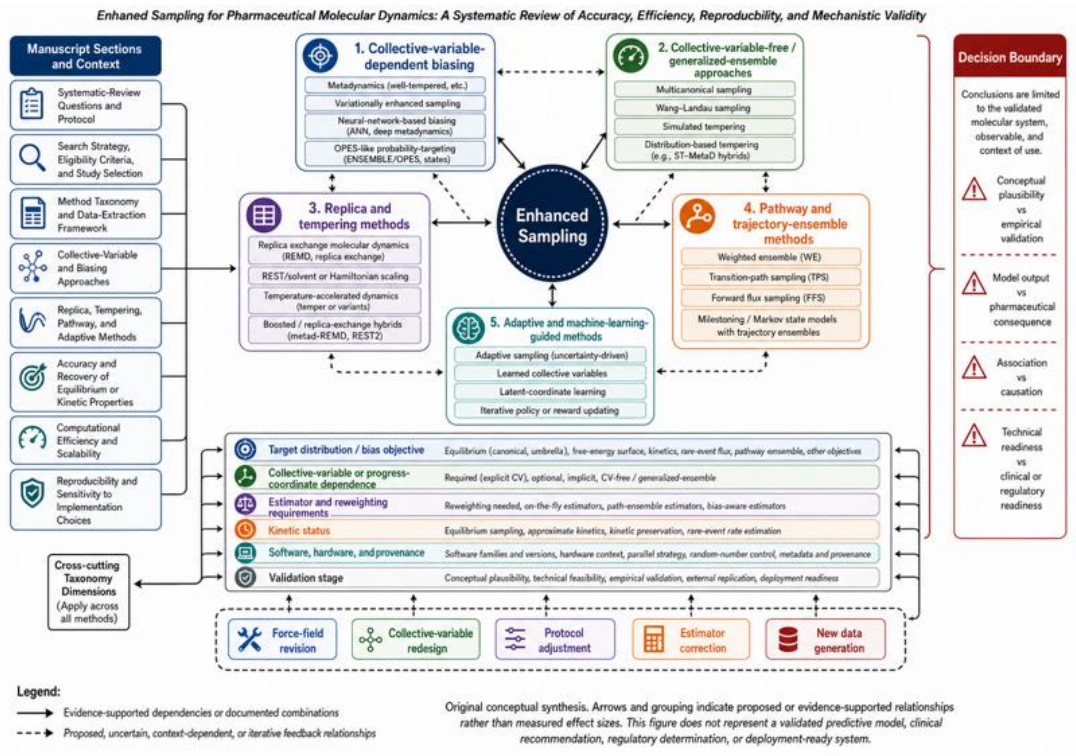

Figure 1. Taxonomy of Enhanced-Sampling Methods.

A second taxonomic dimension concerns whether a strategy can be represented within a common probability-targeting formalism. Such unification can clarify relationships among biasing schemes and expose shared requirements for target-distribution design, normalization, and reweighting, but it does not make their practical assumptions identical [6]. Hybrid methods reinforce this point by combining replicas, complementary target distributions, or exploration and convergence phases within a single workflow. Their reported advantages support classification by functional components rather than by a single mutually exclusive method label [7].

The extraction framework therefore treats algorithm identity as only one variable. It additionally records the biased or tempered degrees of freedom, adaptive update rule, replica or trajectory structure, estimator, kinetic interpretation, initialization, stopping criteria, software version, hardware architecture, and provenance. GPU-accelerated and interoperable platforms can improve practical accessibility and throughput, but implementation capability does not establish sampling correctness or mechanistic validity [8]. The resulting taxonomy is descriptive rather than hierarchical: collective-variable-based, probability-targeting, and hybrid approaches overlap as composable design dimensions, while their relative value remains conditional on the molecular system, intended observable, and validation context [5-7].

Collective-Variable and Biasing Approaches

A collective variable is a reduced descriptor used to represent progress among molecular states. Its usefulness depends on whether it captures the slow degrees of freedom governing the transition of interest. Time-lagged independent component analysis can identify slowly decorrelating combinations of structural descriptors and use them to guide metadynamics, improving exploration when the precursor trajectories already contain informative fluctuations [9].

Variational conformational dynamics provides a related route by optimizing candidate descriptors against slow dynamical modes. This gives collective-variable selection a kinetic basis, but it remains conditional on the descriptor set, lag time, training data, and representativeness of the sampled states. A variationally optimal coordinate within an incomplete descriptor space may still omit the true rate-limiting process [10].

Machine-learning approaches extend collective-variable discovery into nonlinear latent spaces. Reweighted autoencoded variational Bayes iteratively learns a reduced coordinate while correcting for previously introduced bias, offering an adaptive connection between trajectory generation and representation learning [11]. The latent coordinate should nevertheless be interpreted as an operational sampling variable rather than direct proof of a physical mechanism.

Neural-network bias representations can approximate complex free-energy structures more flexibly than fixed analytical functions, but their performance depends on optimization stability, reweighting, training coverage, and diagnostic transparency [12]. Across these approaches, increased state discovery cannot alone establish equilibrium recovery, unbiased kinetics, or causal mechanism. **Table 1** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for eligibility, extraction, and quality-appraisal framework for enhanced sampling for pharmaceutical molecular dynamics.

Table 1. Eligibility, extraction, and quality-appraisal framework for Enhanced Sampling for Pharmaceutical Molecular Dynamics

Evidence domain or review construct	Review question	Eligible evidence type	Key extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Review-method transparency	Were questions, eligibility, selection, extraction, and synthesis reproducibly specified?	Systematic-review methods and reporting guidance	Information sources, search syntax, eligibility, screening, extraction, synthesis	Methods distinguish completed procedures from planned procedures [4]	Establishes review credibility	Reporting completeness does not validate scientific findings	Transparent review conduct is necessary but not evidence of method accuracy
Molecular model	Does the potential-energy model represent the relevant states and interactions?	Force-field development and validation studies	Force field, solvent, protonation, parameters, experimental reference	Agreement across relevant conformational and interaction observables	Separates sampling failure from model error	Validation domains may not cover the pharmaceutical system	Enhanced sampling cannot correct an inaccurate energy function
Collective variable	Does the coordinate resolve the slow process of interest?	CV-development and comparative method studies	Descriptors, lag time, training trajectory, dimensionality, physical interpretation	Slow-mode relevance, stability, state coverage, sensitivity [9]	Explains method success or failure	Important hidden coordinates may be absent	A useful CV is not automatically a causal reaction coordinate
Bias and target distribution	Is the imposed distribution appropriate for the intended estimand?	Metadynamics, variational, and probability-targeting studies	Bias form, deposition rule, target distribution, convergence criterion	Recoverability of unbiased probabilities and estimator assumptions [2]	Links acceleration to thermodynamic inference	Aggressive bias may impair reconstruction or kinetics	Exploration is not equivalent to accurate free energy
Equilibrium or kinetic estimator	Does the analysis estimate the property actually claimed?	Free-energy, pathway, and kinetic studies	Reweighting, rate estimator, calibration, uncertainty	Compatibility between trajectory generation and estimator [3]	Prevents endpoint conflation	Different estimators are not directly commensurate	Free energy, pathway probability, and residence time are distinct claims
Reproducibility	Can compatible findings be recovered under specified conditions?	Cross-engine benchmarks and reproducibility studies	Code, version, parameters, seeds, inputs, analysis workflow	Protocol completeness and cross-implementation agreement [13]	Assesses implementation dependence	Exact trajectory replication is generally inappropriate	Reproducibility concerns distributions and conclusions, not identical coordinates
Convergence	Has the relevant distribution or observable stabilized?	Diagnostic and statistical studies	Replicates, block analysis, state populations,	Multiple diagnostics address the claimed estimand [14]	Identifies unsupported convergence claims	Selected observables may appear stable while hidden modes	Apparent stationarity cannot establish

			transition counts			remain unsampled	correctness alone
Mechanistic validation	Are proposed states or pathways supported independently?	Integrated computational –experimental studies	Experimental observable, comparator, predicted interaction or pathway	Concordance across orthogonal evidence and rival explanations [15]	Bounds mechanistic interpretation	Agreement may be nonunique or system-specific	Consistency supports a bounded mechanism, not universal causality
Uncertainty and data provenance	Are stochastic, parametric, and model uncertainties traceable?	Uncertainty and FAIR-data studies	Replicates, parameter variation, metadata, trajectories, scripts	Explicit uncertainty decomposition and reusable records	Supports independent verification	Repository deposition does not guarantee scientific correctness	Data availability is necessary but insufficient for validity

Replica, tempering, pathway, and adaptive methods

Replica-based methods distribute sampling across modified thermodynamic conditions or bias states and exchange configurations between them. Replica-exchange Gaussian accelerated molecular dynamics combines boosted potentials with exchange to improve barrier crossing and free-energy exploration in selected biomolecular systems [16]. Its cost and accuracy remain dependent on replica design, exchange behavior, boost parameters, and reweighting quality.

Replica exchange with solute tempering limits temperature-like scaling to a selected region, reducing the number of replicas relative to full-system tempering. The definition of that region is consequential: an overly narrow selection may leave coupled barriers unresolved, whereas an overly broad region increases computational cost and may weaken exchange efficiency [17].

Pathway-oriented methods construct ensembles of trajectories rather than only flattening a configurational landscape. Weighted-ensemble simulation can parallelize rare-event generation and estimate pathway or kinetic observables when resampling preserves statistical weights, although progress-coordinate and binning choices affect efficiency [18]. Adaptive resetting offers another acceleration mechanism by redirecting unproductive trajectories, but reduced first-passage time does not by itself establish recovery of natural kinetics [19].

Accuracy and recovery of equilibrium or kinetic properties

Accuracy must be defined against a stated estimand and comparator. Cross-engine free-energy benchmarks show that nominally equivalent molecular models can produce discrepant results unless package-specific settings, long-range interactions, constraints, soft-core potentials, and analysis conventions are carefully harmonized [13]. Numerical agreement therefore depends on implementation as well as theory.

Pharmaceutical relevance requires evaluation beyond simplified benchmarks. Large-scale retrospective assessment across active drug-discovery projects found that binding free-energy performance varied among targets, chemical series, structural inputs, and project settings [20]. Aggregate accuracy can consequently conceal failures important to a particular compound series or decision.

Kinetic recovery imposes additional requirements. Accelerated unbinding simulations using tau-random acceleration molecular dynamics can estimate relative residence-time trends and reveal dissociation pathways, but the applied force and calibration procedure constrain interpretation [21]. Relative ranking within a defined series is a narrower claim than transferable prediction of absolute residence time.

Metadynamics-based investigation of a p38 kinase inhibitor demonstrated that enhanced sampling can connect barriers, intermediate states, and experimentally consistent unbinding kinetics in a specified system [22]. Such examples support bounded mechanistic inference rather than universal kinetic validity. More generally, reported accuracy depends jointly on implementation harmonization, pharmaceutical context, and kinetic calibration [13, 20, 21].

Computational efficiency and scalability

Computational efficiency has at least three meanings: wall-clock throughput, statistical efficiency per unit cost, and time to a decision-relevant result. High-performance molecular-dynamics engines can scale large systems across CPU and GPU architectures, substantially increasing trajectory throughput [23]. This does not establish that the additional trajectory time contains proportionally more independent rare-event information.

Modern GPU frameworks also enable integration of machine-learning potentials and custom forces. OpenMM demonstrates that such models can be incorporated with practical computational overhead in selected applications [24]. Their scientific value still depends on training-domain validity, energy and force accuracy, portability, and compatibility with the enhanced-sampling estimator.

Platforms designed specifically for high-performance biomolecular simulation can support replica exchange, free-energy calculations, coarse-grained models, and hybrid quantum-mechanical workflows within a shared infrastructure [25]. This breadth may improve workflow consistency and reduce implementation barriers, but feature availability is not comparative evidence that one platform or method is more accurate.

Adaptive workflows can allocate new simulations toward undersampled states, retrain slow-mode models, and iteratively update the sampling policy. Combining adaptive sampling, kinetic dimensionality reduction, and probability-enhanced sampling improved exploration in studied biomolecular transitions [26]. The gain must be weighed against training, orchestration, communication, and restart costs. **Table 2** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for characteristics and article-specific classification of the included evidence for enhanced sampling for pharmaceutical molecular dynamics.

Table 2. Characteristics and article-specific classification of the included evidence for Enhanced Sampling for Pharmaceutical Molecular Dynamics

Method, platform, or evidence category	Representative application	Reported strength	Validation context	Generalization concern	Contradictory evidence	Evidence gap	Review interpretation
Metadynamics and probability-targeting bias	Free-energy landscape exploration	Flexible barrier reduction and reconstructable equilibrium distributions [2]	Model and biomolecular applications	CV omission and reweighting instability	Rapid exploration may coexist with inaccurate reconstructed probabilities	Few standardized cross-system comparisons	Appropriate when CV adequacy and estimator diagnostics are demonstrated
Learned collective variables	Slow protein conformational change	Nonlinear or kinetically informed dimensionality reduction [11]	Internal benchmarks	Training data may omit unknown states	Different latent models may yield different coordinates	Limited external mechanistic validation	Operationally useful but not automatically physically unique
Replica exchange and solute tempering	Protein folding and conformational exchange	Exchanges configurations across modified ensembles [17]	Method-specific simulation benchmarks	Replica count and tempered-region choice are system-dependent	Poor overlap can offset theoretical advantages	Limited shared reporting of exchange diagnostics	Efficiency should be assessed from mixing and independent information, not replica number alone
Weighted ensemble	Biomolecular rare-event pathways	Parallel pathway generation with weighted kinetic estimators [18]	Long-timescale applications	Progress-coordinate and resampling design affect cost	Inefficient partitioning can generate many low-value trajectories	Limited head-to-head pharmaceutical benchmarks	Strong for pathway ensembles when statistical-weight preservation is explicit
Accelerated unbinding methods	Drug-target residence-time estimation	Access to dissociation events and relative kinetic trends [21]	Selected protein–ligand systems	Applied acceleration and calibration limit transfer	Faster dissociation may distort natural dynamics	Sparse cross-target prospective validation	Useful for bounded comparative hypotheses rather than universal absolute rates
GPU-accelerated	Large biomolecular	Increased throughput and	Hardware benchmarks	Hardware performance varies by	Faster kernels do not	Few common statistical-	Report wall-clock and scientific

simulation platforms	systems and ML potentials	extensibility [23]		system and algorithm	guarantee better sampling	efficiency benchmarks	efficiency separately
Adaptive ML workflows	Slow conformational transitions	Focuses new computation on uncertain or undersampled states [26]	Internal multi-system demonstrations	Model retraining and orchestration may dominate cost	Learned policies can reinforce initial sampling bias	Limited independent replication	Promising for resource allocation, with explicit uncertainty and stopping rules
Experimental–simulation integration	Membrane-binding and drug-recognition mechanisms	Tests predicted ensembles or interactions against orthogonal observations [15]	System-specific experimental comparison	Agreement may be observable-specific	Alternative mechanisms may explain the same data	Few prospective falsification studies	Strongest basis for bounded mechanistic interpretation

Reproducibility and sensitivity to implementation choices

Reproducibility requires enough information to reconstruct the molecular model, sampling procedure, analysis, and inferential target. Shared input files, versioned software, archived examples, and reusable bias definitions improve transparency in enhanced molecular simulation [27]. These practices reduce ambiguity but cannot ensure that the underlying model or interpretation is correct.

Convergence must also be defined relative to the property being estimated. Stability of energy, root-mean-square deviation, or a selected collective variable may coexist with incomplete sampling of alternative metastable states. Equilibrium claims therefore require distributional and replicate-based evidence appropriate to the stated observable [14].

Implementation sensitivity extends beyond the enhanced-sampling algorithm itself. Protonation schemes, exchange settings, coupling choices, integration details, and parameterization can materially alter accuracy and sampling behavior, as illustrated by methodological comparisons in constant-pH molecular dynamics [28]. Similar sensitivities apply to restraints, solute definitions, bias-update intervals, randomization, and analysis thresholds.

The reproducibility hierarchy proposed here distinguishes computational traceability, numerical repeatability, statistical reproducibility, and mechanistic robustness. A study may satisfy the first two yet fail the latter levels when alternative force fields, initial structures, collective variables, or estimators change the conclusion. Transparent reporting is therefore a prerequisite for evaluation, not a substitute for sensitivity analysis.

Experimental and mechanistic validation

Enhanced sampling operates on a molecular energy model; it cannot independently establish that the model represents the relevant physical system. CHARMM36m improved agreement for folded proteins and intrinsically disordered ensembles, illustrating how force-field validation affects the conformational populations that enhanced sampling may later accelerate [29].

Force-field transferability remains conditional. Development of a99SB-disp sought balanced representation of folded and disordered protein states against diverse experimental observables [30]. Even such broad protein validation does not automatically establish accuracy for every ligand chemistry, protonation state, membrane composition, or binding free energy.

Mechanistic claims are strongest when simulations generate testable predictions and are evaluated against orthogonal observations. Machine-learning-guided enhanced sampling combined with membrane-binding experiments supported a bounded interpretation of aminosterol insertion and affinity in a defined lipid system [15]. Agreement strengthens consistency but cannot exclude every rival mechanism or establish universal transferability.

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for comparative evidence synthesis, methodological limitations, and sources of heterogeneity in enhanced sampling for pharmaceutical molecular dynamics.

Table 3. Comparative evidence synthesis, methodological limitations, and sources of heterogeneity in Enhanced Sampling for Pharmaceutical Molecular Dynamics

Cross-study pattern or configuration	Context or boundary condition	Mechanism or explanatory factor	Observed or reported outcome	Rival explanation	Certainty or rigour concern	Implication	Research priority
Rapid state discovery without verified equilibrium	Strong or imperfectly chosen bias	Barriers are crossed faster than unbiased probabilities are reconstructed	Broader apparent landscape exploration [2]	Bias-induced distortion or omitted coordinate	Convergence tested only in biased space	Separate exploration from equilibrium accuracy	Prespecify reweighting and replicate diagnostics
Learned CV improves transition frequency	Training data contain relevant slow fluctuations	Dimensionality reduction captures a kinetically useful mode [9]	More frequent transitions	Initial trajectories already favor the discovered pathway	Limited external validation of coordinate meaning	Treat learned CVs as operational models	Compare alternative descriptors and falsifiable observables
Replica-based acceleration varies by setup	Exchange overlap and tempered region differ	Mixing depends on energy-distribution overlap	Improved exploration in selected systems [17]	Additional replicas increase cost without independent information	Incomplete reporting of round trips and overlap	Report mixing and effective information gain	Common cross-method efficiency benchmarks
Pathway sampling yields kinetic estimates	Resampling preserves weights and progress variables resolve advancement	Statistical weight is redistributed without discarding pathway probability	Rare-event pathways and rates can be estimated [18]	Poor bins create inefficient branching	Sensitivity to resampling and state definitions	Validate rates and pathway ensembles separately	Robust progress-coordinate diagnostics
Pharmaceutical free-energy performance is heterogeneous	Targets, chemical series, structures, and protocols differ	Model and workflow errors interact	Useful performance in some projects, weaker performance in others [20]	Retrospective selection and project-specific expertise	Aggregate statistics may conceal local failure	Avoid universal accuracy claims	Prospective, target-stratified evaluations
Mechanistic agreement with experiment	Observable directly probes a predicted interaction or state	Simulated ensemble generates experimentally testable consequences	Increased confidence in a bounded mechanism [15]	Multiple ensembles may reproduce the same observable	Nonuniqueness and system specificity	Use orthogonal, perturbational validation	Design experiments that discriminate rival mechanisms
Apparent convergence without correctness	Diagnostic probes only selected observables	Hidden states or force-field error remain undetected	Stable estimates may still be wrong [14]	Insufficient simulation duration or inappropriate model	No universal convergence diagnostic	Require estimator-specific and model-aware validation	Multi-observable, replicate, cross-model tests

Figure 2 presents the original conceptual synthesis for evaluation matrix for accuracy, efficiency, reproducibility, and validity, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

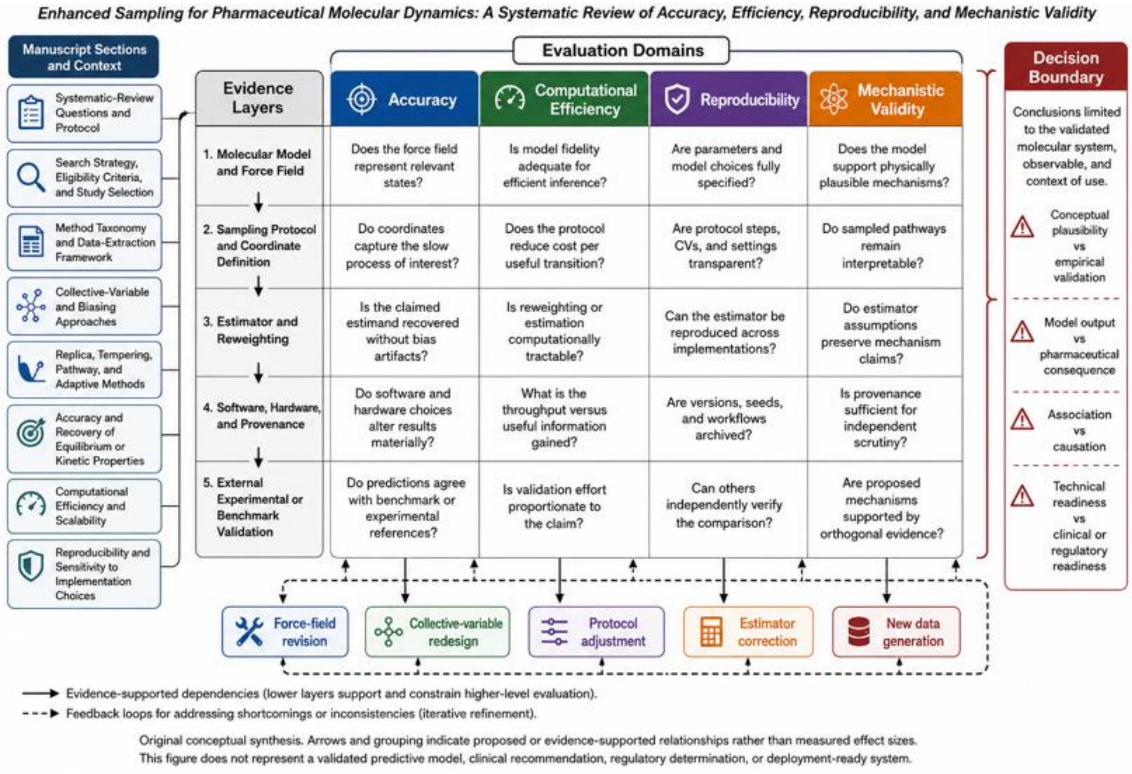

Figure 2. Evaluation Matrix for Accuracy, Efficiency, Reproducibility, and Validity

Comparative evidence synthesis

The evidence does not support a universal ranking of enhanced-sampling methods. Comparative performance changes with biological complexity, landscape topology, available prior information, evaluation target, and definition of correctness. Recent cross-system comparison reinforces that apparent convergence can coexist with context-dependent error [31].

Reaction-coordinate quality emerges as a recurrent determinant. Energy-relaxation-derived coordinates have been reported to accelerate protein transitions and ligand dissociation while preserving more physically natural pathways than selected empirical coordinates [32]. These findings are promising but remain dependent on the tested systems and require broader replication.

Machine learning now contributes to collective-variable discovery, bias construction, generative sampling, adaptive selection, and reinforcement-learning policies. The literature nevertheless spans markedly different levels of methodological maturity, benchmarking, interpretability, and experimental validation [33]. Automation should therefore be evaluated as a collection of task-specific components rather than a unified capability.

Across method families, accuracy is strongest when the estimand is explicit, the molecular model is independently credible, relevant states are sampled, reweighting or kinetic assumptions are tested, and uncertainty is quantified. Efficiency is most informative when reported as independent equilibrium or pathway information per computational cost rather than aggregate trajectory length.

The central synthesis is thus conditional: collective-variable methods may be efficient when slow coordinates are known; replica and tempering methods may be advantageous when exchange overlap is adequate; pathway approaches may support rare-event kinetics when statistical weights are preserved; and adaptive methods may allocate computation effectively when their learned representations remain calibrated. None of these conditions alone establishes pharmaceutical consequence.

Limitations and reporting recommendations

The evidence base is heterogeneous in systems, endpoints, reference standards, computational budgets, and definitions of convergence. Molecular-dynamics uncertainty should be separated into stochastic variability, parameter uncertainty, model-form error, initialization sensitivity, and sampling uncertainty rather than compressed into one trajectory or one error estimate [34].

Independent replicate ensembles are particularly important because chaotic divergence and parameter variation make one prolonged trajectory an inadequate substitute for repeated realizations. Ensemble analysis should report

variation in populations, free energies, pathways, transition counts, or kinetic estimates at the level relevant to the claim [35].

Reproducibility also requires findable, accessible, interoperable, and reusable trajectory data, metadata, parameters, software environments, and analysis workflows [36]. Minimum reporting should include molecular preparation, protonation, force field, solvent conditions, restraints, collective variables, bias parameters, randomization, replica structure, stopping rules, convergence diagnostics, reweighting, uncertainty, and all exclusions from analysis.

These recommendations are proposed methodological safeguards rather than validated regulatory standards. Reliable decision use requires explicit uncertainty decomposition, independent ensemble replication, and FAIR provenance [34-36]. Future studies should prioritize preregistered benchmark questions, shared systems, prospective predictions, orthogonal experimental tests, negative results, and comparisons that separate force-field error from sampling and estimator error.

CONCLUSION

Enhanced sampling expands the molecular events accessible to pharmaceutical molecular dynamics, but acceleration does not itself establish accuracy, equilibrium, unbiased kinetics, reproducibility, or mechanistic truth. The reviewed evidence supports a claim-matched approach in which the method family, collective variables, molecular model, estimator, implementation, uncertainty analysis, and validation reference are evaluated jointly. Collective-variable, replica, tempering, pathway, adaptive, and machine-learning approaches offer complementary strengths under different landscape and decision conditions. Their conclusions should remain bounded to the tested system, observable, and protocol unless broader external validation is available. The principal contribution of this review is therefore not a universal method ranking, but an evidence-based framework for distinguishing increased exploration from reliable inference and molecular plausibility from experimentally supported mechanism.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99(6):1129-43. doi:10.1016/j.neuron.2018.08.011
2. Bussi G, Laio A. Using metadynamics to explore complex free-energy landscapes. Nat Rev Phys. 2020;2(4):200-12. doi:10.1038/s42254-020-0153-0
3. Chipot C. Free energy methods for the description of molecular processes. Annu Rev Biophys. 2023;52:113-38. doi:10.1146/annurev-biophys-062722-093258
4. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372. doi:10.1136/bmj.n71
5. Invernizzi M, Parrinello M. Rethinking metadynamics: From bias potentials to probability distributions. J Phys Chem Lett. 2020;11(7):2731-6. doi:10.1021/acs.jpclett.0c00497
6. Invernizzi M, Piaggi PM, Parrinello M. A unified approach to enhanced sampling. Phys Rev X. 2020;10(4):041034. doi:10.1103/PhysRevX.10.041034
7. Rizzi V, Aureli S, Ansari N, Gervasio FL. OneOPES, a combined enhanced sampling method to rule them all. J Chem Theory Comput. 2023;19(17):5731-42. doi:10.1021/acs.jctc.3c00254
8. Zubieta Rico PF, Schneider L, Pérez-Lemus GR, Alessandri R, Dasetty S, Nguyen TD, et al. PySAGES: Flexible, advanced sampling methods accelerated with GPUs. npj Comput Mater. 2024;10:35. doi:10.1038/s41524-023-01189-z
9. Sultan MM, Pande VS. tICA-metadynamics: Accelerating metadynamics by using kinetically selected collective variables. J Chem Theory Comput. 2017;13(6):2440-7. doi:10.1021/acs.jctc.7b00182

10. McCarty J, Parrinello M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. J Chem Phys. 2017;147(20):204109. doi:10.1063/1.4998598
11. Ribeiro JML, Bravo P, Wang Y, Tiwary P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). J Chem Phys. 2018;149(7):072301. doi:10.1063/1.5025487
12. Bonati L, Zhang YY, Parrinello M. Neural networks-based variationally enhanced sampling. Proc Natl Acad Sci U S A. 2019;116(36):17641-7. doi:10.1073/pnas.1907975116
13. Loeffler HH, Bosisio S, Duarte Ramos Matos G, Suh D, Roux B, Mobley DL, et al. Reproducibility of free energy calculations across different molecular simulation software packages. J Chem Theory Comput. 2018;14(11):5567-82. doi:10.1021/acs.jctc.8b00544
14. Ormeno F, General IJ. Convergence and equilibrium in molecular dynamics simulations. Commun Chem. 2024;7(1):26. doi:10.1038/s42004-024-01114-5
15. Muscat S, Errico S, Danani A, Chiti F, Grasso G. Leveraging machine learning-guided molecular simulations coupled with experimental data to decipher membrane binding mechanisms of aminosterols. J Chem Theory Comput. 2024;20(18):8279-89. doi:10.1021/acs.jctc.4c00127
16. Huang YM, McCammon JA, Miao Y. Replica exchange Gaussian accelerated molecular dynamics: Improved enhanced sampling and free energy calculation. J Chem Theory Comput. 2018;14(4):1853-64. doi:10.1021/acs.jctc.7b01226
17. Kamiya M, Sugita Y. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. J Chem Phys. 2018;149(7):072304. doi:10.1063/1.5016222
18. Russo JD, Zhang S, Leung JMG, Bogetti AT, Thompson JP, DeGrave AJ, et al. WESTPA 2.0: High-performance upgrades for weighted ensemble simulations and analysis of longer-timescale applications. J Chem Theory Comput. 2022;18(2):638-49. doi:10.1021/acs.jctc.1c01154
19. Blumer O, Reuveni S, Hirshberg B. Combining stochastic resetting with metadynamics to speed-up molecular dynamics simulations. Nat Commun. 2024;15(1):240. doi:10.1038/s41467-023-44528-w
20. Schindler CEM, Baumann H, Blum A, Bose D, Buchstaller HP, Burgdorf L, et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. J Chem Inf Model. 2020;60(11):5457-74. doi:10.1021/acs.jcim.0c00900
21. Kokh DB, Amaral M, Bomke J, Gradler U, Musil D, Buchstaller HP, et al. Estimation of drug-target residence times by tau-random acceleration molecular dynamics simulations. J Chem Theory Comput. 2018;14(7):3859-69. doi:10.1021/acs.jctc.8b00230
22. Casasnovas R, Limongelli V, Tiwary P, Carloni P, Parrinello M. Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. J Am Chem Soc. 2017;139(13):4780-8. doi:10.1021/jacs.6b12950
23. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys. 2020;153(4):044130. doi:10.1063/5.0014475
24. Eastman P, Galvelis R, Pelaez RP, Abreu CRA, Farr SE, Gallicchio E, et al. OpenMM 8: Molecular dynamics simulation with machine learning potentials. J Phys Chem B. 2024;128(1):109-16. doi:10.1021/acs.jpcb.3c06662
25. Jung J, Yagi K, Tan C, Oshima H, Mori T, Yu I, et al. GENESIS 2.1: High-performance molecular dynamics software for enhanced sampling and free-energy calculations for atomistic, coarse-grained, and quantum mechanics/molecular mechanics models. J Phys Chem B. 2024;128(25):6028-48. doi:10.1021/acs.jpcb.4c02096
26. Zhang M, Wu H, Wang Y. Enhanced sampling of biomolecular slow conformational transitions using adaptive sampling and machine learning. J Chem Theory Comput. 2024;20(19):8569-82. doi:10.1021/acs.jctc.4c00764
27. The PLUMED Consortium. Promoting transparency and reproducibility in enhanced molecular simulations. Nat Methods. 2019;16(8):670-3. doi:10.1038/s41592-019-0506-8
28. Buslaev P, Aho N, Jansen A, Bauer P, Hess B, Groenhof G. Best practices in constant pH MD simulations: Accuracy and sampling. J Chem Theory Comput. 2022;18(10):6134-47. doi:10.1021/acs.jctc.2c00517
29. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71-3. doi:10.1038/nmeth.4067
30. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci U S A. 2018;115(21). doi:10.1073/pnas.1800690115

31. Kang C, Chen CG, Tang C, Contreras Arredondo S, Zhou M, Fu H, et al. Convergence is not correctness: Context-dependent performance of enhanced-sampling methods across biological complexity. Nat Commun. 2026;17:6245. doi:10.1038/s41467-026-74165-y
32. Li H, Ma A. Enhanced sampling of protein conformational changes via true reaction coordinates from energy relaxation. Nat Commun. 2025;16(1):786. doi:10.1038/s41467-025-55983-y
33. Zhu K, Trizio E, Zhang J, Hu R, Jiang L, Hou T, et al. Enhanced sampling in the age of machine learning: Algorithms and applications. Chem Rev. 2026;126(1):671-713. doi:10.1021/acs.chemrev.5c00700
34. Wan S, Sinclair RC, Coveney PV. Uncertainty quantification in classical molecular dynamics. Philos Trans A Math Phys Eng Sci. 2021;379(2197):20200082. doi:10.1098/rsta.2020.0082
35. Vassaux M, Wan S, Edeling W, Coveney PV. Ensembles are required to handle aleatoric and parametric uncertainty in molecular dynamics simulation. J Chem Theory Comput. 2021;17(8):5187-97. doi:10.1021/acs.jctc.1c00526
36. Amaro RE, Åqvist J, Bahar I, Battistini F, Bellaiche A, Beltran D, et al. The need to implement FAIR principles in biomolecular simulations. Nat Methods. 2025;22(4):641-5. doi:10.1038/s41592-025-02635-0