Pharmacophore

ISSN-2229-5402

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

A DRUG CANDIDATE HAS AN ENVIRONMENTAL PROFILE BEFORE IT BECOMES A MEDICINE: SUSTAINABILITY-AWARE LEAD OPTIMIZATION

Nokuthula Dlamini[1]*, Sipho Zulu[1], Thabo Nkosi[2], Lerato Molefe[3]

1. ***Department of Environmental Profiling and Sustainability-Aware Lead Optimization, Faculty of Pharmacy, University of KwaZulu-Natal, Durban, South Africa.***
2. ***Department of Green Chemistry and Drug Design, Faculty of Pharmacy, University of Cape Town, Cape Town, South Africa.***
3. ***Department of Sustainable Drug Discovery, Faculty of Pharmaceutical Sciences, University of the Witwatersrand, Johannesburg, South Africa.***

ARTICLE INFO

Received:
08 March 2026
Received in revised form:
10 July 2026
Accepted:
12 July 2026
Available online:
28 August 2026

Keywords: Sustainable molecular design, Lead optimization, Environmental fate, Pharmaceutical pollution, Biodegradability, Ecotoxicology

ABSTRACT

Pharmaceutical lead optimization traditionally integrates potency, selectivity, exposure, safety, physicochemical behavior, and synthetic feasibility, yet the environmental implications of a candidate are often considered only after substantial structural and process commitments have been made. This separation is scientifically problematic because molecular features selected during discovery can influence environmental transformation, persistence, mobility, bioaccumulation, ecotoxicological activity, manufacturing burden, and the properties of transformation products. This article develops Sustainability-Aware Lead Optimization as an original conceptual contribution for treating these consequences as a structured part of candidate design rather than as an end-stage annotation. The approach centers on a proposed Candidate Environmental Profile that preserves distinct evidence streams for environmental fate, ecological exposure and effects, therapeutic quality, synthesis, lifecycle burden, uncertainty, and decision provenance. The profile is intended to support multiobjective comparison, targeted evidence generation, and explicit governance gates without collapsing heterogeneous dimensions into a universal sustainability score. The synthesis shows why environmental performance cannot be inferred from a single descriptor, biodegradation result, hazard endpoint, or computational prediction, and why apparent environmental improvement may be offset by reduced therapeutic performance, increased dose, difficult synthesis, hazardous processing, or burden shifting across the lifecycle. The contribution remains conceptual: it does not establish predictive validity, regulatory acceptability, or operational readiness. Its value lies in defining the components, relationships, uncertainty labels, and decision boundaries that future prospective medicinal-chemistry programs can evaluate. Embedding such evidence earlier may help pharmaceutical discovery identify avoidable environmental liabilities while sufficient molecular and process flexibility remains to address them.

To Cite This Article: Dlamini N, Zulu S, Nkosi T, Molefe L. A Drug Candidate Has an Environmental Profile before It Becomes a Medicine: Sustainability-Aware Lead Optimization. Pharmacophore. 2026;17(4):13-25. https://doi.org/10.51847/16ujeGQHG1

Introduction

Pharmaceutical discovery treats a candidate molecule as a multidimensional object long before it becomes a medicine. Its structure is interrogated for target activity, selectivity, physicochemical behavior, permeability, metabolic stability, pharmacokinetic exposure, toxicity, formulation potential, and synthetic tractability. Environmental properties, however, are commonly represented less systematically or deferred until development has narrowed the available structural alternatives. This separation is increasingly difficult to justify because environmental criteria can be incorporated during discovery, structural design can influence the prospect of environmental mineralization, and existing pharmaceutical research parameters remain only partly aligned with the drivers of environmental impact [1-3]. The resulting problem is not merely insufficient environmental reporting. It is a design-architecture gap: the molecular decisions that create therapeutic performance may simultaneously establish persistence, mobility, transformation, ecological exposure, and manufacturing consequences that become more difficult to alter after candidate selection.

Corresponding Author: Nokuthula Dlamini; Department of Environmental Profiling and Sustainability-Aware Lead Optimization, Faculty of Pharmacy, University of KwaZulu-Natal, Durban, South Africa. E-mail: dlaminin@ukzn.ac.za

The objective is not to replace therapeutic optimization with environmental optimization. A medicine must retain adequate efficacy, safety, quality, and clinical usefulness; otherwise, a reduction in one environmental property cannot make it a viable candidate. The relevant scientific challenge is therefore to identify structures and development strategies that preserve therapeutic function while avoiding preventable environmental burdens. Calls for greener drug design have explicitly framed these goals as concurrent rather than substitutable objectives [4]. Yet their coexistence creates difficult questions for computational and AI-based pharmaceutical science. A model may predict high potency while remaining silent about environmental persistence. A candidate may appear readily degradable while requiring a larger dose, generating persistent transformation products, or depending on a resource-intensive synthetic route. Conversely, a structure with an unfavorable screening prediction may remain acceptable after measured fate, exposure, and benefit evidence are considered. Environmental performance must consequently be treated as a conditional, evidence-dependent profile rather than an isolated molecular label.
This article proposes Sustainability-Aware Lead Optimization as a conceptual and methodological architecture for incorporating environmental evidence into candidate design. Its central object is the Candidate Environmental Profile, defined as a structured, versioned, and uncertainty-qualified representation of the environmental properties and lifecycle consequences that may be associated with a pharmaceutical candidate. The profile does not merge heterogeneous evidence into a universal sustainability score. It keeps environmental fate, persistence, mobility, bioaccumulation, ecotoxicological pharmacology, anticipated release, transformation products, synthetic burden, lifecycle context, therapeutic quality, and uncertainty distinguishable. These evidence streams can then be considered within a multiobjective design process in which unacceptable deficits, remediable uncertainties, and genuine trade-offs remain visible. The proposed contribution is therefore organizational rather than predictive: it specifies what evidence should be connected, how claims should be bounded, and where human judgment remains indispensable.
The article develops this contribution in four steps. First, it explains why environmental consequences begin during molecular design, even though their realized magnitude depends on later dose, use, manufacturing, treatment, and receiving-environment conditions. Second, it decomposes environmental performance into fate, persistence, mobility, bioaccumulation, and ecotoxicity so that one property cannot serve as an unsupported proxy for the others. Third, it connects candidate-level environmental evidence with lifecycle, therapeutic, safety, exposure, and synthesis objectives. Fourth, it defines uncertainty-aware profiling and evidence-sensitive governance gates for continuation, redesign, targeted testing, escalation, or stopping. The proposed architecture does not establish that environmental outcomes can already be predicted reliably for arbitrary candidates. It provides a testable intellectual structure through which pharmaceutical scientists can determine what must be measured, modeled, documented, challenged, and prospectively validated.

Why Environmental Consequences Begin During Molecular Design

Environmental consequences begin during molecular design because structure influences what happens after a pharmaceutical has served, or failed to serve, its intended clinical purpose. Functional groups, ionization state, molecular size, stereochemistry, conformational behavior, lipophilicity, polarity, and metabolic susceptibility can affect aqueous mobility, sorption, transformation, biodegradation, uptake, and interaction with biological targets. These properties do not independently determine environmental harm, but they contribute to the pathways through which a candidate may enter and persist in environmental systems. Source-oriented molecular design is important because downstream treatment cannot be assumed to remove every pharmaceutical or transformation product, pharmaceutical residues can enter aquatic and riparian food webs, and pharmacological activity may extend to non-target organisms possessing susceptible biological targets [5-7]. The discovery-stage question is therefore not whether molecular structure alone predicts ecosystem outcomes. It is whether structural decisions create environmental tendencies that should be investigated before structural flexibility is lost.
This distinction between an environmental tendency and an environmental outcome is central to the proposed Candidate Environmental Profile. A candidate's structure may make hydrolysis, photolysis, microbial transformation, or sorption more or less plausible, but realized behavior depends on pH, temperature, redox conditions, microbial communities, exposure duration, wastewater treatment, co-contaminants, and receiving-environment characteristics. Similarly, a pharmacologically potent compound may create biologically plausible concern in a non-target taxon, yet an adverse outcome also requires environmentally relevant exposure and organism-level sensitivity. The profile should therefore connect structure with testable environmental hypotheses rather than declare a molecule environmentally safe or unsafe. Practical development of environmentally biodegradable drugs remains constrained by limited design rules, assay availability, interdisciplinary expertise, organizational incentives, and the requirement to preserve pharmaceutical performance [8]. These constraints make early profiling difficult, but they do not make late consideration scientifically preferable.
The environmental profile must also distinguish properties that medicinal chemistry can influence directly from consequences that emerge only through a larger system. Molecular modification may change ionization, metabolic transformation, or biodegradability, whereas prescribed dose, treatment duration, patient adherence, excretion, manufacturing location, formulation, disposal, and wastewater infrastructure affect the magnitude and geography of release. A compound cannot therefore be described as environmentally preferable solely because it performs well in one degradation assay. Sustainable chemical design requires attention to systems-level consequences and to the possibility that improvement in one location, stage, or endpoint shifts burden elsewhere [9]. In the proposed architecture, molecular design is the beginning of environmental responsibility, not its entire scope. The Candidate Environmental Profile records structure-dependent evidence while explicitly reserving space for use-dependent, process-dependent, and location-dependent variables.

Early consideration matters because lead optimization is a period of unusually high decision leverage. Chemists may still alter substituents, scaffold features, ionization, stereochemistry, route choice, formulation assumptions, or expected exposure while comparing multiple viable series. Once a candidate has accumulated extensive efficacy, safety, manufacturing, and development investment, environmental redesign may carry greater scientific, financial, and therapeutic costs. Sustainability-aware optimization should consequently ask whether a candidate possesses a plausible environmental liability, whether that liability is supported by measured or predicted evidence, whether the uncertainty is decision relevant, and whether a structurally or procedurally feasible alternative exists. It should not reject candidates merely because environmental evidence is incomplete, nor should it allow absence of data to be interpreted as absence of concern. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why environmental consequences begin during molecular design within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why environmental consequences begin during molecular design

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Structure-dependent environmental tendencies	Which molecular features may influence transformation, mobility, sorption, uptake, or biological activity?	Physicochemical properties, local structure–property relationships, mechanistic hypotheses, and matched molecular comparisons	Establishes why environmental inquiry can begin before a medicine is developed	Correlative molecular descriptors may be treated as proven environmental mechanisms	Structural features generate hypotheses and prioritization signals; they do not independently establish environmental outcomes
Anticipated release and exposure	How might dose, use pattern, excretion, disposal, and treatment affect environmental entry?	Pharmacokinetic disposition, anticipated clinical use, removal evidence, and release scenarios	Connects candidate properties with plausible environmental exposure	Early clinical assumptions may be represented as fixed or universally applicable	Discovery-stage exposure estimates are provisional scenarios rather than measured population-level release
Environmental transformation	Does the parent compound disappear, transform, or mineralize under relevant conditions?	Biodegradation, hydrolysis, photolysis, transformation-product identification, and mass-balance evidence	Prevents parent-compound disappearance from being equated with benign environmental resolution	Partial transformation may be mislabeled as complete biodegradation	Loss of the parent structure does not alone establish mineralization or absence of persistent products
Pharmacological ecotoxicity	Could intended or off-target pharmacology affect environmentally exposed species?	Target conservation, comparative pharmacology, internal exposure, organism-level endpoints, and adverse-outcome evidence	Links therapeutic mechanism to environmentally plausible biological effects	Target similarity may be presented as proof of ecological harm	Mechanistic plausibility requires exposure and organism-level confirmation before an adverse environmental claim is made
Design flexibility	Can an identified liability be altered without unacceptable loss of therapeutic value?	Medicinal-chemistry series data, potency, selectivity, exposure, safety, and synthetic evidence	Defines the practical value of identifying environmental properties during lead optimization	Environmental improvement may be assumed achievable for every scaffold	Redesign feasibility is conditional on chemical series, biological mechanism, and therapeutic constraints
Evidence provenance and maturity	Is each environmental claim measured, predicted, inferred, transferred, or missing?	Assay metadata, model lineage, applicability information, versioning, and expert review	Prevents heterogeneous evidence from appearing equally certain	Model outputs, expert judgments, and experiments may be collapsed into one status label	Every profile element requires an explicit evidence type and maturity label
System and lifecycle context	Could molecular improvement shift burden to synthesis, formulation, manufacturing, or	Process chemistry, material use, energy, waste, supply-chain, and receiving-	Extends molecular design without claiming that the molecule	A favorable molecular property may be interpreted as	Candidate-level profiling complements but cannot replace product-level

	another environmental compartment?	environment evidence	determines the complete lifecycle	overall product sustainability	lifecycle assessment
Decision relevance	Would additional environmental information change candidate choice, experiment selection, redesign, or escalation?	Comparative alternatives, uncertainty, reversibility, consequence, and information-value reasoning	Links environmental evidence to a defined discovery action	Profiling may become a data-collection exercise without decision impact	Evidence should be collected proportionately to its capacity to alter a consequential decision

Environmental Fate, Persistence, Bioaccumulation, and Ecotoxicity

Environmental fate describes the distribution and transformation of a compound after release, whereas persistence concerns the duration over which the parent compound or relevant transformation products resist removal or degradation. Mobility concerns transport through water, soil, sediment, and related environmental compartments. These concepts overlap but are not interchangeable. A substance may be persistent and mobile without meeting conventional bioaccumulation criteria, creating prolonged and spatially distributed exposure that would be underestimated by a framework centered only on hydrophobic accumulation. Research on persistent, mobile, and toxic substances therefore supports joint prioritization of persistence and mobility while retaining the underlying dimensions as independently inspectable evidence [10, 11]. For a pharmaceutical candidate, the profile should record the environmental conditions, test methods, thresholds, and data provenance behind each classification rather than transferring an undifferentiated concern label across contexts.

Bioconcentration, bioaccumulation, and trophic exposure add another layer. Bioconcentration generally concerns uptake from the surrounding medium, whereas bioaccumulation may include dietary and other exposure routes. Neither concept is equivalent to toxicity. A compound can accumulate without producing an observed adverse effect under the studied conditions, while a potent bioactive substance may affect organisms at low internal concentrations without exhibiting extreme accumulation. Experimental work with carbamazepine in *Daphnia magna* demonstrates why internal accumulation and physiological or biochemical responses must be represented as connected but distinct dimensions [12]. In the proposed Candidate Environmental Profile, evidence of uptake should therefore be linked to exposure route, organism, life stage, metabolism, depuration, duration, and measured response. A single partition coefficient or predicted accumulation category cannot substitute for this context.

Ecotoxicity similarly requires more than a generic toxicity flag. Pharmaceutical compounds are designed to interact with biological systems, and their environmental effects may involve conserved therapeutic targets, off-target interactions, endocrine signaling, neurobehavioral pathways, reproduction, growth, immune function, or indirect ecological relationships. The relevant question is not simply whether an assay produces an effect, but whether the tested endpoint, concentration, exposure duration, species, and mechanism are informative for plausible environmental conditions. Aquatic pharmaceutical evidence shows that chemistry, occurrence, transformation, environmental effects, and removal processes jointly determine concern [13]. Accordingly, the profile should connect intrinsic activity with realistic exposure and should distinguish molecular association, target-based plausibility, organism-level response, population consequence, and ecosystem implication. Evidence at one level should not be presented as validation of the next.

Field evidence imposes an additional boundary on computational simplification. Bioaccumulation observed in wild fish can reflect continuous effluent exposure, chemical mixtures, species-specific physiology, feeding behavior, metabolism, seasonal conditions, and local hydrology [14]. Such findings demonstrate that candidate-level descriptors and standardized assays are necessary but incomplete representations of environmental behavior. The proposed architecture therefore treats laboratory tests, mechanistic models, in-silico predictions, mesocosm evidence, monitoring data, and field observations as different evidence classes rather than interchangeable measurements. Their disagreements should trigger investigation rather than automatic averaging. A sustainability-aware candidate profile must remain updateable as evidence moves from structural hypotheses to measured fate, transformation-product characterization, internal-exposure studies, organism-level effects, and context-specific environmental observations. This layered interpretation preserves the usefulness of early profiling without claiming that discovery-stage evidence can already determine real-world ecological safety.

Integrating Lifecycle Evidence into Lead-Optimization Objectives

A Candidate Environmental Profile cannot be limited to the fate of the active pharmaceutical ingredient after patient use. Environmental burdens arise across raw-material acquisition, reagent and solvent production, synthesis, purification, formulation, packaging, distribution, administration, excretion, disposal, wastewater treatment, and environmental transformation. Evidence and responsibility are consequently distributed among medicinal chemists, process chemists, formulators, pharmacologists, environmental scientists, manufacturers, suppliers, healthcare systems, and waste-management actors. Pharmaceutical-sector perspectives show that environmental challenges are recognized across the lifecycle but are approached through different information systems, priorities, and organizational responsibilities [15]. Sustainability-Aware Lead Optimization is proposed to create an early evidence interface among these functions, not to imply that discovery teams can complete a full lifecycle assessment for every analogue.

The first integration requirement is to keep molecular environmental performance separate from production performance. A candidate may have favorable biodegradation properties yet depend on a route with high material consumption, hazardous reagents, difficult separations, or substantial solvent use. Conversely, an efficient synthesis does not resolve persistence or ecotoxicological activity after release. Process mass intensity, waste generation, energy demand, yield, solvent selection, and material efficiency therefore represent distinct evidence dimensions. The E-factor and related green-chemistry measures demonstrate why reaction yield alone is an incomplete representation of process sustainability [16]. Within the proposed profile, such measures should be attached to a specific route, scale, and process version because an early medicinal-chemistry synthesis may differ substantially from a later manufacturing process.

The second requirement is to treat process evidence as design-relevant rather than administratively downstream. Pharmaceutical green-chemistry priorities identify route design, catalysis, solvent use, separations, hazardous transformations, and scalable operations as areas where environmental burden can be altered [17]. These variables should not be converted into an immutable penalty attached to a candidate structure, because routes can change and process innovation may remove an apparent disadvantage. Instead, the profile should distinguish intrinsic molecular constraints from route-contingent burdens. A structural feature that requires a persistently hazardous transformation across credible routes may constitute a candidate-level concern, whereas a high-burden discovery synthesis with plausible alternatives may justify process research rather than molecular rejection.

The third requirement is to preserve the local and conditional nature of environmental structure–property relationships. Experimental work can reveal structural features associated with biodegradation within a defined chemical series, but such rules may not transfer across scaffolds, organisms, test systems, or environmental conditions [18]. Lifecycle analysis likewise shows that burden can shift among manufacturing operations and geographic supply-chain configurations [19]. The proposed integration layer therefore records the source, chemical domain, process scenario, and uncertainty of each relationship. It supports comparison of candidate–route combinations rather than assuming that a molecule possesses one fixed lifecycle score. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop integrating lifecycle evidence into lead-optimization objectives within the article's central argument.

Table 2. Components, relationships, uncertainties, and validation needs within Integrating lifecycle evidence into lead-optimization objectives

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Candidate molecular layer	Structure, ionization, stereochemistry, physicochemical properties, anticipated metabolism	Represents molecular features plausibly affecting environmental behavior and therapeutic performance	Versioned candidate description linked to testable hypotheses	Structure–property relationships may be scaffold specific	Prospective evaluation within medicinal-chemistry series
Environmental fate layer	Biodegradation, hydrolysis, photolysis, sorption, mobility, transformation-product data	Characterizes environmental persistence, movement, and transformation	Endpoint-specific fate profile with conditions and provenance	Laboratory conditions may not reproduce receiving environments	External testing across relevant environmental conditions
Ecological exposure and effect layer	Release scenarios, uptake, depuration, target conservation, organism-level assays	Connects environmental presence with biologically plausible effects	Exposure–effect evidence map	Exposure and effect evidence may derive from different species or contexts	Tiered confirmation from mechanism to organism and population levels
Therapeutic quality layer	Potency, selectivity, pharmacokinetics, safety, anticipated dose, formulation	Protects therapeutic value during environmental redesign	Candidate-specific therapeutic constraints and objectives	Early exposure and dose estimates may change during development	Reassessment as pharmacology and clinical-use assumptions mature
Route and process layer	Reagents, solvents, yield, separations, energy, waste, scale assumptions	Represents burdens associated with producing the candidate	Route-specific process profile	Discovery routes may not represent manufacturing routes	Comparative route evaluation and process-development confirmation
Supply-chain and geography layer	Supplier locations, energy systems, transportation, manufacturing distribution	Identifies burden shifting across organizations and locations	Scenario-based lifecycle context	Data may be confidential, incomplete, or rapidly changing	Supplier-supported inventories and scenario sensitivity analysis
Transformation-product layer	Product identification, formation kinetics,	Prevents parent disappearance from	Parent–product environmental profile	Unknown products and incomplete mass	Analytical confirmation and

	persistence, mobility, hazard evidence	being interpreted as benign resolution		balance may obscure burden	adequate mass-balance studies
Evidence-provenance layer	Source, method, date, version, assay conditions, model applicability	Distinguishes measured, predicted, inferred, and missing evidence	Traceable evidence record	Metadata completeness does not establish scientific validity	Independent provenance and reproducibility audit
Multiobjective integration layer	Environmental, therapeutic, process, lifecycle, and uncertainty evidence	Preserves separate objectives while revealing conflicts and alternatives	Nondominated candidate–route options and unresolved tensions	Results depend on objective definitions and weighting choices	Sensitivity analysis and prospective decision studies
Lifecycle update mechanism	New process, formulation, use, monitoring, and fate evidence	Revises the profile as the development context changes	Version-controlled environmental profile	Later information may invalidate early assumptions	Longitudinal case evaluation across development stages

Trade-Offs with Potency, Exposure, Safety, and Synthesis

The central design problem is not the independent optimization of environmental properties but the management of interactions among environmental and pharmaceutical objectives. Green medicinal chemistry has emphasized that compound design and synthetic-process design should be considered together [20]. Sustainability-Aware Lead Optimization extends this principle by treating potency, selectivity, pharmacokinetic exposure, safety, anticipated dose, environmental fate, ecotoxicity, synthesis, and lifecycle burden as distinct but connected dimensions. Some may function as objectives to improve, others as minimum constraints, and others as evidence gaps requiring testing. The architecture rejects unrestricted compensation: a severe safety liability, loss of therapeutic efficacy, or unresolved high-consequence environmental concern should not disappear because another property improves.

Potency illustrates why simple environmental preferences can be misleading. A highly potent and selective candidate may achieve therapeutic exposure at a lower administered dose, potentially reducing the quantity entering production, use, and waste streams. Yet potency alone does not determine clinical dose, because absorption, distribution, clearance, target engagement, tissue penetration, safety margins, formulation, and treatment duration also matter. Holistic compound-quality approaches show that potency-centered indices become more informative when pharmacokinetic characteristics are explicitly represented [21]. The Candidate Environmental Profile therefore should not reward lower potency or higher potency in isolation. It should examine whether a structural modification changes the amount used, the fraction excreted, the identity of metabolites, therapeutic selectivity, and the environmental activity of the parent and transformation products.

Synthesis creates a second major trade-off. Computational optimization can generate molecules that satisfy a formal objective yet are difficult to synthesize or require implausible routes, demonstrating that molecular desirability and synthetic feasibility are separate properties [22]. Environmental optimization can create an analogous failure if a candidate is predicted to be less persistent but requires hazardous reagents, excessive protecting-group operations, difficult purification, or supply-constrained starting materials. Synthesizability should consequently be represented through more than a scalar accessibility estimate. Route availability, precedent, selectivity, expected waste, hazardous transformations, scalability, and the maturity of process evidence should remain inspectable. A candidate with uncertain synthesis may warrant route exploration; a candidate whose environmental improvement depends on a consistently high-burden route may warrant redesign.

Computational design can help explore these interactions, but it should not silently determine which trade-off is acceptable. Drug-design scholarship emphasizes that model outputs must remain connected to medicinal-chemistry judgment, experimental evidence, and the practical constraints of discovery [23]. Likewise, success on a standardized molecular-design objective demonstrates performance on the specified task rather than prospective pharmaceutical usefulness [24]. A sustainability-aware system must therefore report which objectives were optimized, which constraints were imposed, what evidence was omitted, how candidate rankings change under alternative assumptions, and which decisions remain human responsibilities. **Table 3** organizes the evidence, constructs, and interpretation boundaries needed to develop trade-offs with potency, exposure, safety, and synthesis within the article's central argument.

Table 3. Components, relationships, uncertainties, and validation needs within Trade-offs with potency, exposure, safety, and synthesis

Evaluation dimension	What must be tested	Suitable evidence or assessment	Failure signal	Interpretive limitation	Decision relevance
Potency–dose relationship	Whether increased potency plausibly reduces administered and environmentally released mass	Exposure–response evidence, anticipated dose, target engagement, pharmacokinetics	Potency improves while dose or duration does not decrease	Early dose estimates may change substantially	Determines whether potency improvement carries environmental relevance
Selectivity–ecotoxicity relationship	Whether therapeutic selectivity reduces	Comparative pharmacology, target conservation, off-	Human selectivity is assumed to	Species biology and internal	Guides mechanism-informed ecotoxicity testing

	effects in non-target species	target panels, organism assays	imply ecological selectivity	exposure remain uncertain	
Exposure–persistence relationship	Whether lower release compensates for prolonged environmental residence	Release scenarios, degradation kinetics, mobility, monitoring analogues	Small release is treated as negligible despite cumulative persistence	Release and degradation estimates may come from incompatible contexts	Identifies candidates requiring sustained-exposure evaluation
Biodegradation–transformation-product relationship	Whether parent removal produces benign products	Product identification, mineralization, mass balance, product hazard testing	Parent disappears while persistent or active products remain	Unknown products may escape analytical detection	Prevents false environmental improvement claims
Potency–safety relationship	Whether potency gains narrow or improve the therapeutic margin	Safety pharmacology, exposure margins, selectivity, toxicology evidence	High potency masks increased off-target or systemic toxicity	Preclinical margins do not establish clinical utility	Defines noncompensable therapeutic constraints
Molecular design–synthesizability relationship	Whether proposed structures can be prepared through credible routes	Retrosynthesis, reaction precedent, route scouting, experimental synthesis	Optimization generates inaccessible or unstable structures	Computational accessibility does not prove laboratory feasibility	Determines whether environmental redesign is actionable
Route feasibility–process burden relationship	Whether feasible routes remain acceptable in waste, solvent, energy, and hazard terms	Process mass measures, solvent and reagent assessment, scale scenarios	Synthesis is possible only through consistently high-burden operations	Early route data may not represent optimized manufacture	Distinguishes route-research needs from candidate-level liabilities
Environmental profile–therapeutic value relationship	Whether environmental improvement preserves adequate therapeutic function	Integrated potency, PK, safety, formulation, dose, and environmental evidence	Candidate appears greener but becomes clinically nonviable	Therapeutic value cannot be reduced to one model score	Prevents environmental criteria from displacing patient benefit
Multiobjective stability	Whether conclusions persist across reasonable weights and constraints	Pareto analysis, sensitivity analysis, alternative decision rules	Candidate ranking reverses under minor modeling changes	Mathematical stability does not establish ethical acceptability	Identifies fragile rankings that require deliberation
Prospective usefulness	Whether profiling improves experiment selection or candidate decisions	Prospective design–make–test–analyze comparison	Better benchmark scores do not alter or improve real decisions	Controlled evaluation may not generalize across programs	Establishes whether the architecture adds practical scientific value

Figure 1 presents the environmental profile lead-optimization compass, showing how the article's key components and boundaries are connected within trade-offs with potency, exposure, safety, and synthesis.

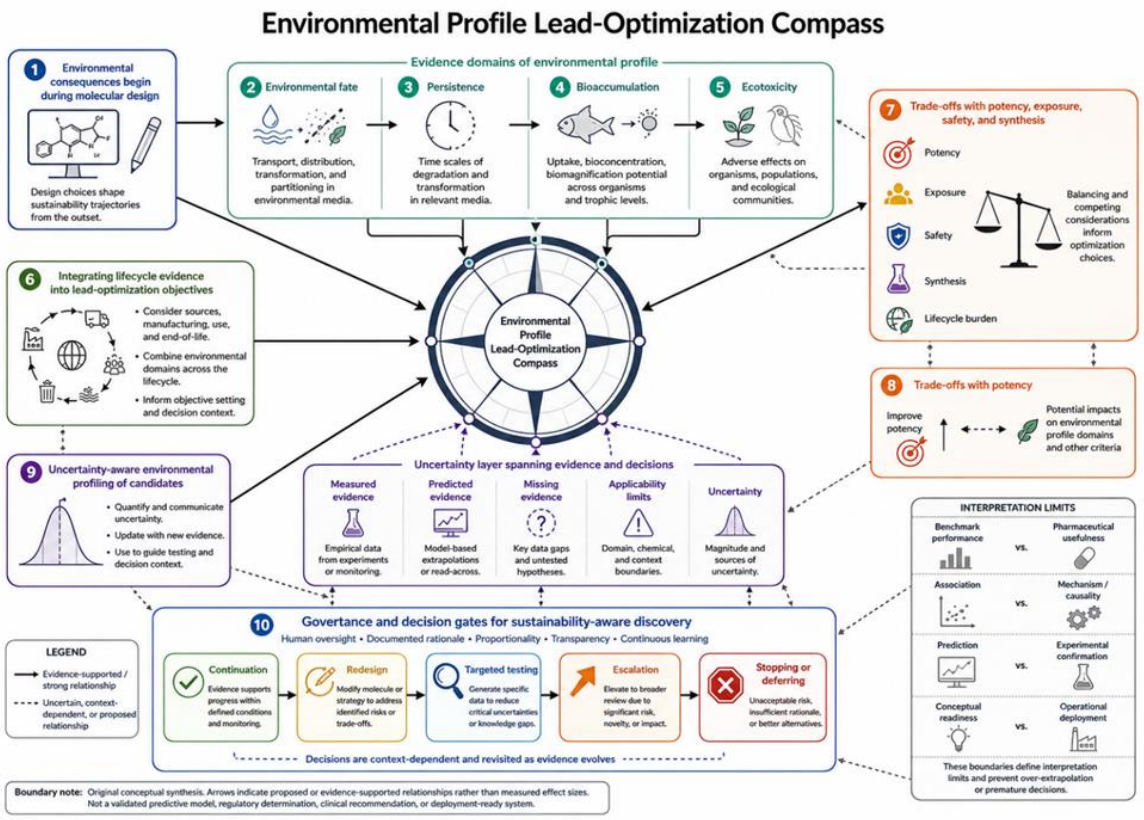

Figure 1. Environmental Profile Lead-Optimization Compass

The figure is an original conceptual synthesis that organizes the article's central contribution across environmental consequences begin during molecular design, environmental fate, persistence, bioaccumulation, and ecotoxicity, environmental fate, integrating lifecycle evidence into lead-optimization objectives, trade-offs with potency, exposure, safety, and synthesis, trade-offs with potency. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Uncertainty-Aware Environmental Profiling of Candidates

Environmental predictions are especially vulnerable to false precision because evidence is sparse, endpoint definitions vary, and relevant chemical or ecological conditions may fall outside model training domains. A Candidate Environmental Profile should therefore report not only predicted values but also the uncertainty associated with each prediction, the chemical and experimental domain supporting it, and the consequence of being wrong. Comparative work on neural-network molecular-property prediction shows that uncertainty quantification can provide useful reliability information but must be evaluated rather than assumed [25]. The environmental profile should label whether uncertainty concerns measurement variability, irreducible heterogeneity, missing mechanisms, model parameters, sparse chemical-space coverage, or uncertain use and release scenarios.

No single uncertainty estimator is likely to perform adequately across persistence, mobility, bioaccumulation, ecotoxicity, synthesis, and lifecycle endpoints. Evaluations of scalable uncertainty methods show that performance depends on the estimator, model architecture, dataset, and prediction task [26]. A profile should consequently avoid a common uncertainty scale unless calibration has been demonstrated for each endpoint. It should retain endpoint-specific intervals, applicability labels, evidence quality, and model disagreement. A narrow interval from an out-of-domain model should not be interpreted as strong evidence, and disagreement among credible methods should remain visible rather than being removed through unexamined averaging.

Candidate-level error estimates are particularly important because discovery decisions concern individual structures rather than average benchmark behavior. Frameworks for estimating prediction-specific errors demonstrate how confidence intervals can accompany molecular-property predictions [27]. In sustainability-aware optimization, such intervals may determine whether a predicted environmental difference between candidates is meaningful, whether two options are effectively indistinguishable, or whether an experiment is required. However, prediction intervals are conditional on assumptions about data distribution, model behavior, and domain similarity. Their presence does not establish that an environmental endpoint has been represented adequately.

Uncertainty becomes most useful when it changes the next action. It can identify unreliable regions of chemical space and prioritize additional evidence collection [28]. Evidential molecular modeling further illustrates how different uncertainty

sources may inform guided discovery and active learning [29]. The proposed profile therefore links uncertainty to decision consequence: high uncertainty about a low-consequence, reversible choice may be tolerable, whereas moderate uncertainty about a persistent, difficult-to-reverse environmental liability may justify targeted testing or redesign. This is a proposed decision principle rather than a validated threshold system. Its evaluation requires prospective comparison of uncertainty-guided choices with conventional lead-optimization practice.

Governance and Decision Gates for Sustainability-Aware Discovery

Environmental evidence will not influence discovery merely because it is generated. Implementation depends on ownership, incentives, shared terminology, evidence access, cross-functional expertise, and clear points at which environmental information can alter a decision. Research on greener-pharmaceutical implementation identifies coordination among actors and institutional conditions as central challenges [30]. Sustainability-Aware Lead Optimization therefore proposes stage-specific gates at hit triage, series selection, lead optimization, experiment selection, candidate nomination, and major route commitment. Each gate should state which environmental questions are relevant, what minimum evidence is required, who reviews it, what uncertainties remain, and which actions are available.

The governance model should support iteration rather than binary pass–fail classification. Computer-aided safe-and-sustainable redesign demonstrates the conceptual feasibility of introducing environmental hazard information into repeated chemical redesign [31]. In pharmaceutical discovery, a concerning result might lead to structural modification, transformation-product testing, alternative route exploration, improved exposure estimation, or explicit acceptance of a residual concern because no therapeutically adequate alternative exists. The gate should preserve the reasoning behind such decisions, including rejected alternatives and unresolved evidence. It should not convert screening descriptors into automatic candidate rejection.

Decision authority must also remain explicit when computational systems generate structures, predictions, explanations, or rankings. Levels of automated chemical design distinguish automated ideation from the delegation of consequential scientific decisions [32]. Under the proposed architecture, models may suggest candidates, estimate properties, identify trade-offs, or recommend experiments, but accountable scientists authorize synthesis, escalation, continuation, and nomination. The level of permissible automation should decrease as environmental consequence, uncertainty, novelty, irreversibility, or therapeutic importance increases. This allocation is a governance proposal and requires empirical study of decision quality, workload, error detection, and accountability.

Interpretability can support review but cannot replace validation. Explainable AI methods may reveal features associated with a model output without proving that those features are causal environmental determinants [33]. In compound optimization, explanations may help experts inspect property predictions, identify implausible associations, and formulate structural hypotheses, but their usefulness remains dependent on stability, model quality, and domain interpretation [34]. Governance gates should therefore ask whether an explanation is reproducible, chemically plausible, supported by independent evidence, and capable of changing a decision. **Table 4** organizes the evidence, constructs, and interpretation boundaries needed to develop governance and decision gates for sustainability-aware discovery within the article's central argument.

Table 4. Evaluation, implementation, and research priorities arising from A Drug Candidate Has an Environmental Profile before It Becomes a Medicine

Research or implementation priority	Unresolved gap	Required methodological work	Evidence needed	Responsible actors	Expected contribution
Minimum Candidate Environmental Profile	No agreed discovery-stage environmental evidence structure	Define mandatory, conditional, and optional profile fields with provenance and uncertainty labels	Cross-program case analysis, expert consensus, retrospective candidate evidence	Medicinal chemists, environmental scientists, data stewards	Creates a consistent evidence object without implying a universal score
Prospective environmental lead optimization	Limited evidence that environmental objectives can be improved within viable pharmaceutical series	Conduct prospective design–make–test–analyze studies	Matched molecular series with therapeutic, fate, ecotoxicity, and synthesis measurements	Discovery teams and external environmental laboratories	Tests whether environmental design hypotheses are actionable
Environmental assay strategy	Existing assays may not match discovery cadence or material availability	Develop tiered, miniaturized, and decision-focused testing strategies	Assay concordance, reproducibility, throughput, and domain coverage	Ecotoxicologists, analytical chemists, assay developers	Enables earlier evidence generation with explicit limits
Transformation-product integration	Parent disappearance is often easier to measure than	Link transformation prediction, analytical identification,	Time-resolved product profiles and adequate mass balance	Environmental chemists and metabolite scientists	Prevents false claims of benign degradation

	complete environmental resolution	persistence, and hazard testing			
Endpoint-specific uncertainty validation	UQ methods are rarely validated for candidate environmental endpoints	Compare calibration, sharpness, applicability, and prospective error	External chemical series and prospective assay outcomes	Cheminformaticians and environmental modelers	Determines when uncertainty can support decisions
Multiobjective decision evaluation	Weighting and compensation rules may remain hidden	Compare Pareto, constraint-based, deliberative, and information-value approaches	Decision consistency, sensitivity, expert reasoning, and error analysis	Multidisciplinary discovery committees	Makes trade-offs and noncompensable failures explicit
Molecule–route integration	Candidate and process assessments are often separated	Evaluate candidate–route combinations under alternative scale scenarios	Route scouting, process mass, solvent, hazard, energy, and lifecycle evidence	Medicinal and process chemists	Identifies burden shifting and route-contingent liabilities
Governance-gate testing	Proposed gates have not been shown to improve decisions	Pilot staged review with documented overrides and escalation	Decision logs, delays, evidence use, error detection, and user experience	Portfolio leaders, project teams, governance specialists	Tests proportionality and organizational feasibility
Explanation-quality assessment	Plausible explanations may generate unjustified trust	Assess stability, faithfulness, chemical plausibility, and decision impact	Model perturbation, independent validation, and expert studies	Model developers and domain reviewers	Defines when explanations add scrutiny rather than persuasion
Interoperable provenance infrastructure	Environmental evidence may be fragmented across systems and vendors	Develop versioned schemas, ontologies, and exchange standards	Assay metadata, model lineage, route versions, and lifecycle scenarios	Data scientists, informaticians, laboratories, suppliers	Supports cumulative learning and auditable updates
Post-selection learning	Later development and monitoring evidence rarely feeds back into discovery design rules	Establish longitudinal profile updating and retrospective reconciliation	Manufacturing, use, environmental monitoring, and fate evidence	Development, manufacturing, environmental, and discovery functions	Improves future hypotheses while exposing failed early assumptions
Human capability development	Environmental interpretation is not consistently embedded in medicinal-chemistry training	Develop cross-disciplinary education and review practices	Competency evaluation and decision-case studies	Universities, companies, professional societies	Reduces dependence on opaque scores and isolated specialists

Limitations and Boundary Conditions

The proposed contribution is limited first by the evidence base available for connecting pharmaceutical structure to real environmental outcomes. Many environmental datasets are sparse, heterogeneous, biased toward known contaminants, or produced under conditions that differ from discovery-stage chemical space. Relationships between structure and biodegradation, mobility, uptake, or ecotoxicity may be local to a scaffold or test system. Field outcomes additionally depend on use patterns, mixtures, wastewater infrastructure, hydrology, species, climate, and ecological interactions. The Candidate Environmental Profile can organize these uncertainties, but it cannot remove them or establish environmental safety from incomplete evidence.

A second limitation concerns the scope of candidate-level intervention. Molecular design can influence environmental behavior, but it cannot independently determine manufacturing burden, healthcare use, disposal, treatment efficiency, or regional exposure. Lifecycle evidence introduced during lead optimization will often rely on provisional routes, anticipated doses, and scenario assumptions. Early comparisons may therefore change as clinical, process, and supply-chain information matures. The proposed architecture should not be used to claim that a molecule is sustainable in an absolute sense. It supports conditional comparison among specified candidates, routes, evidence states, and use assumptions.

A third boundary concerns governance and computational authority. Multiobjective optimization, uncertainty estimation, and explanation can improve visibility without determining what trade-off should be accepted. Objective weights may encode organizational values, and apparently neutral rankings may hide compensation between incommensurable outcomes. The framework cannot establish clinical benefit–risk, regulatory acceptability, causal environmental mechanisms, or population-level ecological impact. It is also not an operational candidate-selection instrument. Before implementation, its components

require endpoint-specific validation, prospective evaluation, human-factors assessment, data-governance safeguards, and demonstration that additional profiling improves decisions proportionately to its burden.

Research Agenda and Implementation Priorities

The first research priority is prospective scientific validation within active medicinal-chemistry programs. Studies should follow viable candidate series through repeated structural modification while measuring therapeutic activity, pharmacokinetics, safety-relevant properties, synthesis, biodegradation, persistence, mobility, transformation products, and selected ecotoxicological endpoints. The purpose should not be to prove that every property can be optimized simultaneously. It should determine which environmental relationships are locally learnable, which conflicts are recurrent, when early predictions fail, and whether environmental evidence changes experiment selection or candidate choice. Negative and inconclusive outcomes are essential because they reveal where environmental design rules do not transfer.

The second priority is infrastructure for traceable, uncertainty-aware evidence. Environmental assays and models require standardized metadata for structure, stereochemistry, speciation, test conditions, organism, exposure, analytical method, route version, model version, applicability domain, and missingness. Reporting should distinguish direct measurement, transferred evidence, computational prediction, expert inference, and scenario assumption. Shared data structures should enable the profile to evolve without erasing previous versions or the reasoning behind decisions. Infrastructure development must also address commercial confidentiality, supplier data, quality control, and the risk that standardized fields create superficial comparability between scientifically incompatible measurements.

The third priority is staged implementation with deliberately narrow decision roles. Initial use could focus on evidence visibility, transformation-product questions, or identification of high-uncertainty environmental endpoints rather than automated candidate ranking. Pilot programs should document how often the profile changes a decision, what additional experiments are requested, whether project timelines are affected, how disagreements are resolved, and whether later evidence confirms or contradicts early judgments. Expansion should occur only when a use case demonstrates scientific value, interpretable uncertainty, and accountable oversight. Such staging would preserve the ambition of sustainability-aware discovery while resisting premature claims of operational maturity.

Conclusion

A drug candidate acquires environmentally relevant tendencies through the same molecular and process decisions that create its therapeutic identity. Sustainability-Aware Lead Optimization is proposed to make those tendencies visible before candidate selection by organizing them within a Candidate Environmental Profile that preserves environmental fate, persistence, mobility, bioaccumulation, ecotoxicological pharmacology, therapeutic quality, synthesis, lifecycle context, provenance, and uncertainty as distinct but connected evidence domains. Its central contribution is not a new universal score but a disciplined way to expose trade-offs, identify consequential evidence gaps, and connect candidate design with accountable decision gates. The architecture remains conditional on data quality, endpoint relevance, chemical domain, use scenario, lifecycle assumptions, and prospective validation. It cannot determine environmental safety, clinical value, regulatory acceptability, or the correct balance between therapeutic and environmental priorities. Its practical promise lies in enabling pharmaceutical scientists to ask environmental questions while meaningful design alternatives remain available and to treat uncertainty as a reason for targeted learning rather than as justification for either unsupported reassurance or automatic rejection.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Moermond CTA, Puhlmann N, Brown AR, Owen SF, Ryan J, Snape J, et al. GREENER pharmaceuticals for more sustainable healthcare. Environ Sci Technol Lett. 2022;9(9):699-705. doi:10.1021/acs.estlett.2c00446
2. Puhlmann N, Mols R, Olsson O, Slootweg JC, Kümmerer K. Towards the design of active pharmaceutical ingredients mineralizing readily in the environment. Green Chem. 2021;23(14):5006-23. doi:10.1039/D1GC01048D
3. Vidaurre R, Bramke I, Puhlmann N, Owen SF, Angst D, Moermond CTA, et al. Design of greener drugs: Aligning parameters in pharmaceutical R&D and drivers for environmental impact. Drug Discov Today. 2024;29(7):104022. doi:10.1016/j.drudis.2024.104022
4. Brodin T, Bertram MG, Arnold KE, Boxall ABA, Brooks BW, Cerveny D, et al. The urgent need for designing greener drugs. Nat Sustain. 2024;7(8):949-51. doi:10.1038/s41893-024-01374-y

5. Kümmerer K, Dionysiou DD, Olsson O, Fatta-Kassinos D. A path to clean water. Science. 2018;361(6399):222-4. doi:10.1126/science.aau2405
6. Richmond EK, Rosi EJ, Walters DM, Fick J, Hamilton SK, Brodin T, et al. A diverse suite of pharmaceuticals contaminates stream and riparian food webs. Nat Commun. 2018;9(1):4491. doi:10.1038/s41467-018-06822-w
7. Gunnarsson L, Snape JR, Verbruggen B, Owen SF, Kristiansson E, Margiotta-Casaluci L, et al. Pharmacology beyond the patient—the environmental risks of human drugs. Environ Int. 2019;129:320-32. doi:10.1016/j.envint.2019.04.075
8. Bertram MG, Wong BBM, Kümmerer K, Jörg M. Development of environmentally biodegradable drugs: What are the key challenges? Expert Opin Drug Discov. 2025;20(1):1-4. doi:10.1080/17460441.2024.2442746
9. Zimmerman JB, Anastas PT, Erythropel HC, Leitner W. Designing for a green chemistry future. Science. 2020;367(6476):397-400. doi:10.1126/science.aay3060
10. Rüdel H, Körner W, Letzel T, Neumann M, Nödler K, Reemtsma T. Persistent, mobile and toxic substances in the environment: A spotlight on current research and regulatory activities. Environ Sci Eur. 2020;32:5. doi:10.1186/s12302-019-0286-x
11. Hartmann J, Rorije E, Wassenaar PNH, Verbruggen EMJ. Screening and prioritising persistent, mobile and toxic chemicals: Development and application of a robust scoring system. Environ Sci Eur. 2023;35:40. doi:10.1186/s12302-023-00749-w
12. Nkoom M, Lu G, Liu J, Yang H, Dong H. Bioconcentration of the antiepileptic drug carbamazepine and its physiological and biochemical effects on Daphnia magna. Ecotoxicol Environ Saf. 2019;172:11-8. doi:10.1016/j.ecoenv.2019.01.061
13. Patel M, Kumar R, Kishor K, Mlsna T, Pittman CU Jr, Mohan D. Pharmaceuticals of emerging concern in aquatic systems: Chemistry, occurrence, effects, and removal methods. Chem Rev. 2019;119(6):3510-673. doi:10.1021/acs.chemrev.8b00299
14. Grabicova K, Grabic R, Fedorova G, Fick J, Cerveny D, Kolarova J, et al. Bioaccumulation of psychoactive pharmaceuticals in fish in an effluent dominated stream. Water Res. 2017;124:654-62. doi:10.1016/j.watres.2017.08.018
15. Riikonen S, Timonen J, Sikanen T. Environmental considerations along the life cycle of pharmaceuticals: Interview study on views regarding environmental challenges, concerns, strategies, and prospects within the pharmaceutical industry. Eur J Pharm Sci. 2024;196:106743. doi:10.1016/j.ejps.2024.106743
16. Sheldon RA. The E factor 25 years on: The rise of green chemistry and sustainability. Green Chem. 2017;19(1):18-43. doi:10.1039/C6GC02157C
17. Bryan MC, Dunn PJ, Entwistle D, Gallou F, Koenig SG, Hayler JD, et al. Key green chemistry research areas from a pharmaceutical manufacturers' perspective revisited. Green Chem. 2018;20(22):5082-103. doi:10.1039/C8GC01276H
18. Suk M, Haiß A, Westphal J, Jordan A, Kellett A, Kapitanov IV, et al. Design rules for environmental biodegradability of phenylalanine alkyl ester linked ionic liquids. Green Chem. 2020;22(14):4498-508. doi:10.1039/D0GC00918K
19. Tao Y, Zhu S, Smith J, Lakhani N, You F. Environmental sustainability of the globalized pharmaceutical supply chains: The case of tenofovir disoproxil fumarate. ACS Sustain Chem Eng. 2023;11(17):6510-22. doi:10.1021/acssuschemeng.2c06518
20. Castiello C, Junghanns P, Mergel A, Jacob C, Ducho C, Valente S, et al. GreenMedChem: The challenge in the next decade toward eco-friendly compounds and processes in drug design. Green Chem. 2023;25(6):2109-69. doi:10.1039/D2GC03772F
21. Tautermann CS, Borghardt JM, Pfau R, Zentgraf M, Weskamp N, Sauer A. Towards holistic compound quality scores: Extending ligand efficiency indices with compound pharmacokinetic characteristics. Drug Discov Today. 2023;28(11):103758. doi:10.1016/j.drudis.2023.103758
22. Gao W, Coley CW. The synthesizability of molecules proposed by generative models. J Chem Inf Model. 2020;60(12):5714-23. doi:10.1021/acs.jcim.0c00174
23. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64. doi:10.1038/s41573-019-0050-3
24. Brown N, Fiscato M, Segler MHS, Vaucher AC. GuacaMol: Benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59(3):1096-108. doi:10.1021/acs.jcim.8b00839
25. Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. J Chem Inf Model. 2020;60(8):3770-80. doi:10.1021/acs.jcim.0c00502
26. Scalia G, Grambow CA, Pernici B, Li YP, Green WH. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J Chem Inf Model. 2020;60(6):2697-717. doi:10.1021/acs.jcim.9b00975
27. Cortés-Ciriano I, Bender A. Deep Confidence: A computationally efficient framework for calculating reliable prediction errors for deep neural networks. J Chem Inf Model. 2019;59(3):1269-81. doi:10.1021/acs.jcim.8b00542
28. Janet JP, Duan C, Yang T, Nandy A, Kulik HJ. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. Chem Sci. 2019;10(34):7913-22. doi:10.1039/C9SC02298H
29. Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential deep learning for guided molecular property prediction and discovery. ACS Cent Sci. 2021;7(8):1356-67. doi:10.1021/acscentsci.1c00546
30. van Vliet ED, Kannegieter NM, Moermond CTA, Leonardo Alves T. Keystones in the implementation of greener pharmaceuticals: A scoping review. Sustain Chem Pharm. 2025;44:101938. doi:10.1016/j.scp.2025.101938

31. van Dijk J, Flerlage H, Beijer S, Slootweg JC, van Wezel AP. Safe and sustainable by design: A computer-based approach to redesign chemicals for reduced environmental hazards. Chemosphere. 2022;296:134050. doi:10.1016/j.chemosphere.2022.134050
32. Goldman B, Kearnes S, Kramer T, Riley P, Walters WP. Defining levels of automated chemical design. J Med Chem. 2022;65(10):7073-87. doi:10.1021/acs.jmedchem.2c00334
33. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4
34. Rodríguez-Pérez R, Bajorath J. Explainable machine learning for property predictions in compound optimization. J Med Chem. 2021;64(24):17744-52. doi:10.1021/acs.jmedchem.1c01789