Pharmacophore

ISSN-2229-5402

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

FROM MOLECULE TO PATIENT: LEARNING FORMULATION BEHAVIOR ACROSS MATERIAL, DOSAGE-FORM, PHYSIOLOGICAL, AND EXPOSURE SCALES

Adamu Sani[1]*, Grace Nwachukwu[2], Kehinde Ogunleye[3], Bola Ajayi[4]

1. ***Department of Multi-Scale Formulation Modeling, Faculty of Pharmacy, Ahmadu Bello University, Zaria, Nigeria.***
2. ***Department of Material and Dosage-Form Behavior, Faculty of Pharmaceutical Sciences, University of Ibadan, Ibadan, Nigeria.***
3. ***Department of Physiological and Exposure Scale Learning, Faculty of Pharmacy, Federal University of Agriculture Abeokuta, Abeokuta, Nigeria.***
4. ***Department of Formulation-to-Patient Translation, Faculty of Pharmacy, Nnamdi Azikiwe University, Awka, Nigeria.***

ARTICLE INFO

Received:
16 April 2025
Received in revised form:
14 August 2025
Accepted:
14 August 2025
Available online:
28 August 2025

Keywords: Multiscale pharmaceutics, Formulation learning, Excipient interactions, Material organization, Dosage-form transformation, Physiological transport

ABSTRACT

Pharmaceutical formulation behavior is commonly evaluated through isolated measures such as molecular solubility, physical stability, dissolution, release, permeability, or systemic exposure. These measures are scientifically useful, but none independently represents the sequence of transformations through which a formulated drug becomes an absorbed and systemically distributed entity. The unresolved problem is therefore not simply how to predict individual formulation attributes, but how to learn the conditional relationships connecting molecular properties, excipient interactions, material organization, dosage-form transformation, physiological transport, absorption, and exposure. This article proposes a Molecule-to-Patient Formulation Learning Model as an original multiscale pharmaceutics construct. The model represents formulation performance as a series of connected, time-dependent states rather than as a direct mapping from composition to a single endpoint. Its architecture separates molecular, material, dosage-form, physiological, absorption, and exposure layers while linking them through explicitly defined state, flux, boundary-condition, and uncertainty interfaces. Mechanistic constraints and pharmaceutical knowledge may be incorporated without assuming that physical admissibility establishes causal correctness or prospective validity. The central contribution is an evidence-organizing architecture that distinguishes prediction from mechanism, concentration from mass flux, model confidence from calibrated uncertainty, and benchmark performance from pharmaceutical usefulness. The proposed construct may support more interpretable formulation hypotheses, cross-scale experimental planning, and identification of the scale at which a formulation-development assumption fails. However, it is not an empirically validated predictive system, universal formulation ontology, clinical decision tool, regulatory framework, or autonomous optimization platform. Its value depends on route-specific implementation, temporally aligned evidence, transparent provenance, applicability assessment, and prospective experimental evaluation.

To Cite This Article: Sani A, Nwachukwu G, Ogunleye K, Ajayi B. From Molecule to Patient: Learning Formulation Behavior across Material, Dosage-Form, Physiological, and Exposure Scales. Pharmacophore. 2025;16(4):81-90. https://doi.org/10.51847/hh5s4yLm1O

Introduction

Machine learning and mathematical modeling are increasingly used to organize formulation variables, predict product attributes, guide optimization, and connect experimental information with pharmaceutical-development decisions. Current applications extend beyond a single modeling task and include property prediction, composition selection, process understanding, experimental prioritization, and lifecycle-oriented product knowledge. Nevertheless, the resulting evidence remains fragmented across endpoints, data structures, development stages, and intended uses, while formulation-specific data quality and validation continue to determine whether a computational result is scientifically meaningful [1-3]. A model that

Corresponding Author: Adamu Sani; Department of Multi-Scale Formulation Modeling, Faculty of Pharmacy, Ahmadu Bello University, Zaria, Nigeria. E-mail: adamu.sani@abu.edu.ng.

accurately predicts one measured attribute may therefore remain uninformative about the transformations that connect that attribute to drug release, absorption, or exposure.

The central scientific difficulty arises because a formulation does not retain one stable behavior as it moves from manufacture to administration and then through a biological environment. Molecular properties influence interactions with excipients, but those interactions are expressed through composition-dependent material organization. Material organization then conditions wetting, hydration, disintegration, erosion, phase separation, dissolution, precipitation, particle formation, or depot transformation. These processes determine which molecular, colloidal, particulate, or associated species become locally available for transport. Physiological conditions further alter these species before absorption and systemic disposition generate an exposure profile. Consequently, solubility, stability, release, permeability, and exposure are not interchangeable measures of one underlying property; they are observations at different positions within a conditional sequence.

This distinction is particularly important for computational and artificial-intelligence-based pharmaceutical science. Applications described under the same technological label differ substantially in data provenance, prediction target, representation, mechanistic content, validation design, and proximity to pharmaceutical use. They should not therefore be interpreted as interchangeable evidence of translational readiness [4]. Many learning systems map formulation inputs directly to an endpoint while leaving intervening states implicit. Conversely, mechanistic models may describe a selected process in considerable detail while remaining weakly connected to upstream material organization or downstream exposure. The unresolved gap is an architecture capable of preserving scale-specific scientific meaning while enabling information to pass between these otherwise separated representations.

This article addresses that gap by proposing the Molecule-to-Patient Formulation Learning Model, an original conceptual and methodological construct for learning formulation behavior across connected scales. The contribution has four elements: scale-specific state representations; typed interfaces that transmit states, fluxes, boundary conditions, and uncertainties; mechanistic constraints that restrict inadmissible behavior without being mistaken for validation; and an interpretation layer that returns bounded formulation hypotheses for human and experimental review. The objective is not to replace established molecular modeling, formulation testing, biopharmaceutics, or pharmacokinetic modeling. Rather, it is to define how their outputs may be organized as interoperable but non-equivalent forms of evidence. The resulting model treats exposure as a downstream consequence of interacting transitions and treats formulation design as a problem of identifying which upstream state or interface must change to produce a defensible downstream effect.

Why Formulation Behavior Emerges Across Multiple Interacting Scales

Formulation behavior is described here as emergent when its relevant downstream properties cannot be assigned to an isolated molecular, material, dosage-form, or physiological variable without considering interactions and state transitions at other scales. Evidence from multiscale drug-delivery modeling indicates that formulation inputs propagate through coupled transport barriers, local tissue or fluid conditions, absorption processes, and systemic disposition rather than acting directly on exposure. Mechanistic local-to-systemic models for injected products further illustrate that local transport and absorption can be connected to systemic pharmacokinetics, while particle behavior remains conditional on tissue structure and multiphysics transport [5-7]. These examples do not establish a universal formulation model, but they support the general proposition that exposure-relevant behavior is generated across levels.

The proposed model represents this proposition as a sequence of conditional states. A molecular state contains physicochemical and structural information relevant to interactions. An interaction and material state represents composition, phase behavior, morphology, mobility, heterogeneity, and processing history. A dosage-form state represents time-dependent transformation after contact with the administration environment. A physiological state describes route-specific fluids, tissues, transit, transport barriers, perfusion, and biological variability. An absorption state represents the amount, form, location, and timing of material entering systemic circulation, after which disposition generates exposure. These states are not presumed to follow a simple linear chain. Feedback, competing pathways, irreversible transformations, delayed effects, and context-dependent transitions can make identical upstream measurements compatible with different downstream outcomes.

Cross-scale learning therefore requires more than placing heterogeneous variables in one feature matrix. Each interface must specify what is being transferred and what scientific assumptions permit that transfer. The molecular-to-material interface translates descriptors and interaction evidence into a representation of organization; the material-to-dosage-form interface converts an initial solid, semisolid, liquid, particulate, or depot state into transformation rules; the release-to-physiology interface transmits species-specific concentrations or fluxes under local boundary conditions; and the absorption-to-exposure interface transmits a time- and route-dependent systemic input. Information can be lost or misassigned at each handoff. A latent representation may correlate across adjacent scales without identifying the physical quantity responsible for the relationship, and an apparently accurate terminal prediction may conceal compensating errors between modules.

Physiologically based biopharmaceutics modeling demonstrates the value of connecting drug-substance properties, formulation characteristics, dissolution, gastrointestinal physiology, absorption, and pharmacokinetics while also showing that such integration remains limited by assumptions, parameter identifiability, verification, and context of use [8]. The proposed construct generalizes this integrative principle beyond one administration route or one modeling formalism. It does not assume that every dosage form requires the same variables or that all cross-scale relationships can be learned from existing data. Instead, it defines a framework for asking whether the evidence at each scale is suitable for the downstream claim, whether

uncertainty has been propagated rather than hidden, and whether a failure can be localized to a state representation, transition, or interface.

Table 1 organizes the evidence, constructs, and interpretation boundaries needed to develop why formulation behavior emerges across multiple interacting scales within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for why formulation behavior emerges across multiple interacting scales

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Molecular state	Which drug properties are relevant to interaction, organization, and subsequent transformation?	Physicochemical and structural properties provide upstream constraints but are measurement- and context-dependent.	Defines the initial pharmaceutical information entering the model.	Treating molecular plausibility as sufficient evidence of formulatability or exposure.	Molecular properties alone do not establish material behavior, product performance, safety, or therapeutic value.
Excipient interactions and material organization	How do composition and interactions generate a mesoscopic material state?	Drug–excipient relationships, mobility, phase behavior, morphology, heterogeneity, and processing history jointly condition bulk behavior.	Bridges molecular information to dosage-form transformation.	Inferring product behavior from one interaction measurement or bulk average.	An observed interaction is not automatically causal, and an average material property may conceal locally important structures.
Dosage-form transformation and release	Which time-dependent processes generate locally available drug species and mass flux?	Wetting, hydration, swelling, disintegration, erosion, dissolution, precipitation, particle formation, and depot change alter the administered state.	Converts the manufactured product into transport-relevant species and fluxes.	Treating a static formulation label or one release measurement as a complete process description.	In vitro transformation does not alone identify the dominant in vivo pathway.
Physiological transport and absorption	How do local fluids, tissues, barriers, transit, and uptake processes condition availability?	Route-specific physiology modifies dosage-form transformation, local transport, permeability, retention, degradation, and absorption.	Provides the biological conditions connecting local availability to absorbed input.	Assuming population-average or simplified physiology represents every subject and administration context.	Physiological parameters and transport mechanisms require route-, population-, and formulation-specific justification.
Systemic exposure	How is absorbed input translated into plasma or target-site exposure?	Exposure reflects both absorption and downstream distribution, metabolism, and elimination.	Supplies the terminal pharmacokinetic readout and a source of upstream design constraints.	Attributing exposure error entirely to formulation when disposition-model error may also contribute.	Exposure prediction does not establish efficacy, safety, clinical utility, or regulatory acceptability.
Cross-scale interfaces and uncertainty	What quantity passes between scales, and how is uncertainty retained?	State, flux, boundary-condition, provenance, and applicability information are required to interpret transitions.	Prevents heterogeneous measurements from being treated as directly interchangeable features.	Producing a coherent latent representation that lacks physical identifiability or fails under domain shift.	Cross-scale consistency is not equivalent to mechanistic truth, causal validity, or prospective performance.

Molecular Properties, Excipient Interactions, and Material Organization

The first substantive layer of the proposed model concerns the transition from molecular information to an organized pharmaceutical material. Polymer choice, drug–polymer interactions, molecular-level characterization, composition, mobility, moisture exposure, and processing history can jointly condition physical stability and dissolution-related behavior. The available evidence also shows that increased stability and improved dissolution need not change in parallel, because an interaction that restricts molecular mobility may affect drug release differently depending on polymer characteristics and material state [9-11]. The model therefore does not encode an excipient as a fixed categorical input or an interaction as a universally beneficial bond. It represents both as context-dependent contributors to a material state.

The molecular state should include properties that influence association, phase behavior, and response to the administration environment, while preserving their measurement context and uncertainty. Relevant information may include ionization behavior, structure, conformational flexibility, hydrogen-bonding capacity, polarity, solubility-related tendencies, and interaction-compatible surface features. These variables should not be interpreted as immutable causal determinants. Apparent interaction strength can differ with analytical method, composition, hydration, temperature, physical form, and timescale. A molecular representation that performs well for one endpoint may therefore omit information needed to predict miscibility,

mobility, phase separation, or release. Within the proposed architecture, molecular descriptors are inputs to an interaction model rather than substitutes for the resulting material state.

The excipient-interaction state represents pairwise and higher-order relationships among drug, excipients, residual solvents, water, and other formulation constituents. Its purpose is to organize evidence about association, preferential contact, miscibility, mobility restriction, plasticization, and competition for intermolecular interactions. These relationships become pharmaceutically consequential through material organization, defined here as the mesoscopic arrangement produced by composition and processing. Material organization includes phase continuity, domain size, surface composition, amorphous or crystalline regions, porosity, local mobility, and spatial heterogeneity. This layer is essential because identical nominal compositions can develop different structures after differences in processing, storage, aging, or environmental exposure. A bulk assay may consequently characterize the average material while failing to detect a minority domain that controls nucleation, hydration, or release.

Drug load and molecular interactions can alter the mechanism of dissolution, meaning that performance cannot be reduced to a single descriptor of interaction strength, physical stability, or nominal composition [12]. The proposed model accordingly represents the material layer as a structured state containing composition, interaction evidence, phase and morphological attributes, mobility, processing history, and uncertainty. Its output is not a predicted clinical outcome but an initial condition for the dosage-form transformation layer. This distinction prevents the architecture from assuming that a stable material must release adequately, that strong molecular association must improve performance, or that one experimentally observed interaction establishes causality. The material representation remains formulation-specific and requires validation against time-resolved structural and performance measurements before it can support downstream prediction.

Dosage-Form Transformation and Release Processes

The dosage form should be represented as a dynamic pharmaceutical state rather than as a static label assigned at manufacture. After administration, the initial material encounters fluids, interfaces, mechanical forces, enzymes, tissues, or extracellular structures that can alter hydration, porosity, phase continuity, mobility, surface composition, and spatial organization. These changes determine whether drug and excipient components remain associated, dissolve together, separate, precipitate, form particles, or follow different release pathways. Evidence from amorphous solid dispersions shows that phase separation and morphology can condition whether drug and polymer release congruently or through distinct mechanisms [13]. In the proposed model, the material state is therefore translated into a time-dependent transformation state before any exposure-related prediction is attempted.

Release must also be distinguished from the concentration measured in a dissolution medium. A concentration profile reflects the net result of dissolution, precipitation, stabilization, partitioning, particle formation, and sampling conditions, whereas absorptive availability depends on mass transfer toward and across a relevant barrier. Coupled dissolution and membrane-transfer measurements demonstrate that a high apparent solution concentration does not independently establish a correspondingly high absorptive flux [14]. The proposed architecture consequently separates local concentration, thermodynamic or kinetic state, colloidal association, and membrane-directed mass flux. This separation prevents a learning module from treating a favorable dissolution endpoint as sufficient evidence of improved absorption.

Particle formation further complicates the transition from dosage-form release to physiological transport. Drug-rich particles, colloidal structures, or polymer-associated species may preserve supersaturation, change local transport, act as temporary reservoirs, or contribute indirectly to absorption through dissolution near a biological interface. Human pharmacokinetic evidence indicates that particle-forming behavior and solution-phase processes may influence absorption, but the dominant interpretation remains dependent on the compound, formulation, analytical method, and physiological setting [15]. The model must therefore retain species identity and uncertainty rather than merging dissolved, associated, and particulate drug into an undifferentiated measure of available material.

The proposed dosage-form module contains transformation operators for wetting, hydration, swelling, disintegration, erosion, dissolution, precipitation, association, particle generation, and release. Only processes supported by the dosage form and intended context should be activated. Its outputs comprise time-resolved species, concentrations, mass fluxes, particle or colloidal states, and uncertainty descriptors for transfer to the physiological layer. Release performance arises from the interaction of drug–polymer relationships, polymer behavior, composition, and environmental conditions rather than from one isolated formulation attribute [16]. Accordingly, this module is intended to generate competing, testable transformation hypotheses; it does not establish the in vivo dominance of a mechanism without route-relevant transport and absorption evidence.

Physiological Transport, Absorption, and Systemic Exposure

The physiological layer defines the environment in which released species are transported, transformed, retained, degraded, or absorbed. Its representation must be specific to the route of administration and the model's intended use. Relevant conditions may include fluid composition, volume, pH, transit, mixing, tissue architecture, extracellular transport, vascular or lymphatic access, permeability, perfusion, metabolism, and population variability. Best-practice recommendations for physiologically based biopharmaceutics modeling emphasize that model purpose, parameter provenance, verification, validation, variability, and uncertainty must be specified explicitly [17]. The proposed architecture adopts these requirements as interface conditions

rather than assuming that physiological complexity can be represented adequately by adding population-average covariates to an endpoint model.

For orally administered products, the location and timing of release can be as important as the total amount released. Regional absorption is conditional on dosage-form state, permeability, transit, fluid availability, dissolution or precipitation behavior, and the adequacy of the gastrointestinal assumptions used in the model [18]. Analogous route dependence applies to injected, implanted, inhaled, ocular, or transdermal products, although the controlling variables and validation evidence differ. The architecture therefore prohibits direct transfer of a physiological module across routes merely because its input and output dimensions appear compatible. Each route requires a justified ontology of local states, boundary conditions, and transport pathways.

Physiological perturbations may also change multiple formulation-relevant processes simultaneously. Food-effect prediction, for example, can depend on changes in fluid composition, gastric emptying, intestinal transit, solubilization, dissolution, precipitation, and model calibration, and reported predictive performance is not uniform across compounds or scenarios [19]. A learning system trained only on terminal exposure may fit these combined effects without identifying which physiological or formulation process produced them. The proposed model instead represents physiological scenarios as structured boundary conditions and tests whether the inferred upstream response remains consistent with known changes in transformation and transport.

The absorption output is defined as a time-, route-, species-, and location-dependent systemic input rather than as an unqualified prediction of bioavailability. A downstream pharmacokinetic module may then translate that input into plasma or target-site exposure using an explicitly defined disposition structure. Physiologically based biopharmaceutics models can connect in vitro product attributes with in vivo performance for defined product-development questions, but such use remains bounded by model credibility and context of use [20]. The proposed architecture therefore permits exposure-informed formulation comparison only when formulation, physiological, absorption, and disposition uncertainties remain distinguishable. An accurate exposure trajectory does not independently validate the local mechanism, establish therapeutic benefit, or demonstrate readiness for regulatory or clinical use.

Architecture of the Multiscale Learning Model

The Molecule-to-Patient Formulation Learning Model is organized as a modular architecture in which each scale has a defined scientific task, input domain, state representation, transformation rule, output, and validation requirement. The molecular module does not attempt to predict the entire formulation-to-exposure pathway directly. It generates representations appropriate to specific interaction or material questions. Benchmarking principles from molecular machine learning show why tasks, representations, data partitions, and endpoint-appropriate metrics must be specified rather than compressed into one aggregate performance score [21]. The proposed architecture extends this principle across formulation scales by requiring every module to declare what it predicts and what its output means pharmaceutically.

The interaction and material modules may use descriptor-based, graph-based, geometric, experimental, or hybrid representations, but the resulting latent state must be evaluated against chemically and pharmaceutically meaningful alternatives. Learned molecular representations can perform differently under changes in chemical space and should be compared with strong descriptor-based baselines rather than assumed to be intrinsically superior [22]. Within the proposed model, representation transfer across scales is permitted only through typed interfaces. These interfaces identify whether the transferred quantity is a physical state, probability distribution, mass flux, boundary condition, learned embedding, or absorbed input, together with its units, provenance, uncertainty, and applicability domain.

Mechanistic knowledge may enter the architecture through data generation, model structure, transition equations, constraints, loss functions, priors, or hybrid coupling between learning and simulation. Physics-informed machine learning provides several strategies for incorporating scientific knowledge, but imposing a physical constraint does not automatically establish that the selected mechanism is complete or causally correct [23]. The architecture therefore distinguishes admissibility constraints, such as positivity and conservation, from mechanistic hypotheses, such as the assumed rate-limiting transformation. Constraint satisfaction is evaluated separately from predictive validity, and alternative mechanisms must remain testable where the evidence does not identify one unique explanation.

Graph representations can support relational learning among molecular entities, excipient interactions, material domains, biological barriers, and other connected pharmaceutical components, although their scientific meaning remains dependent on data quality and task definition [24]. The complete architecture contains six linked layers: molecular representation; interaction and material organization; dosage-form transformation and release; physiological transport and absorption; systemic disposition and exposure; and interpretation and design review. Each layer can combine learned and mechanistic components, while provenance, uncertainty, and applicability checks operate across the full model. The architecture returns bounded formulation hypotheses and sensitivity relationships to a human review gate rather than producing autonomous formulation decisions.

Figure 1 presents the molecule-to-patient formulation scale bridge, showing how the article's key components and boundaries are connected within architecture of the multiscale learning model.

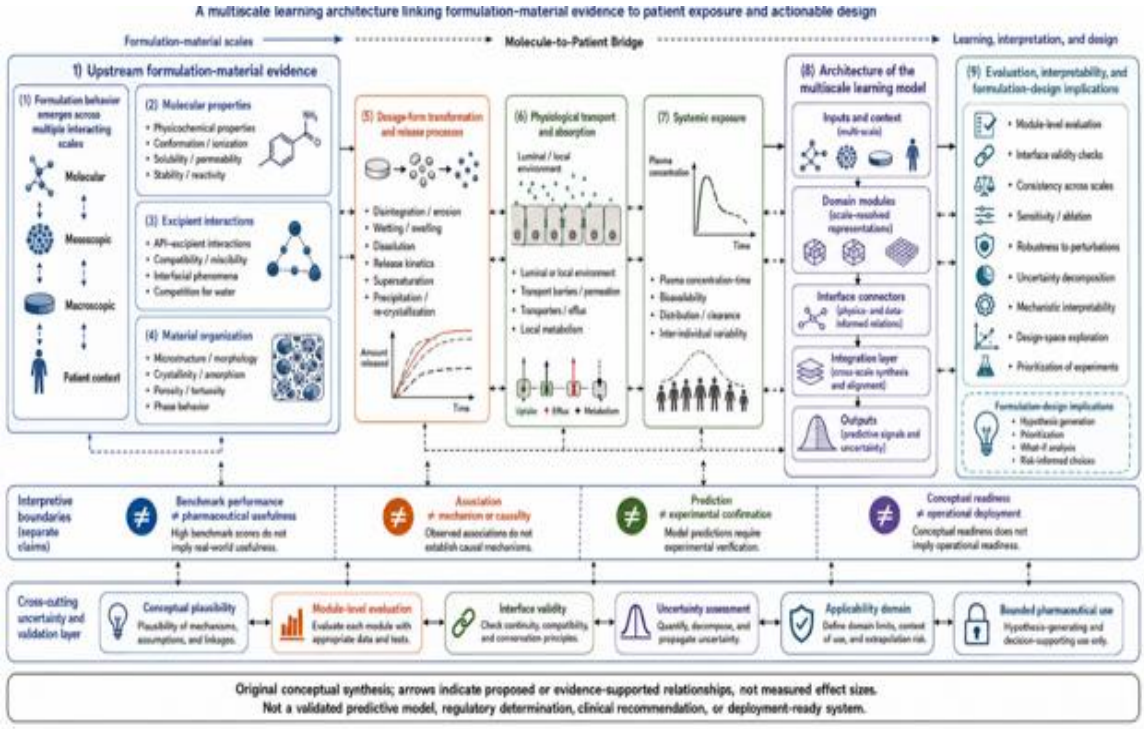

Figure 1. Molecule-to-Patient Formulation Scale Bridge

The figure is an original conceptual synthesis that organizes the article's central contribution across formulation behavior emerges across multiple interacting scales, molecular properties, excipient interactions, and material organization, molecular properties, excipient interactions, material organization, dosage-form transformation and release processes. 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.

Evaluation, Interpretability, and Formulation-Design Implications

Evaluation must be conducted at the level of modules, interfaces, integrated predictions, and intended formulation decisions. Interpretability is likewise question-dependent: an explanation suitable for identifying influential molecular descriptors may not explain a material transition, release mechanism, physiological sensitivity, or exposure difference. Explainable artificial intelligence in drug discovery requires explanations to be matched to the scientific question and the needs of the intended user, while post hoc attribution should not be treated as equivalent to mechanistic or causal explanation [25]. The proposed architecture therefore separates feature attribution, sensitivity, counterfactual analysis, mechanism testing, and experimental confirmation.

Uncertainty is represented as a collection of distinguishable sources rather than as one confidence score. Measurement uncertainty concerns the precision and reproducibility of observed inputs and outputs. Aleatory variability describes heterogeneity that remains under the defined model. Epistemic uncertainty reflects limited evidence or insufficient representation, while structural uncertainty concerns missing or misspecified mechanisms. Applicability uncertainty arises when a formulation, process, route, or population differs from the validated domain. Scalable uncertainty-estimation methods differ in calibration and out-of-domain behavior and must be evaluated for the particular task in which they are used [26]. A narrow interval from one method is therefore not accepted as proof of reliable extrapolation.

The architecture evaluates uncertainty against observed errors, domain shifts, repeated measurements, and formulation perturbations. No uncertainty-quantification approach is universally superior, and confidence estimates or error-ranking behavior must be validated for their intended use [27]. An uncertainty score may help prioritize experiments, identify abstention conditions, or communicate evidence gaps, but it cannot guarantee safety or scientific correctness. The model should report which uncertainty source dominates, which assumptions were held fixed, and what new evidence could reduce uncertainty. Human reviewers must remain able to reject an output that is numerically confident but pharmaceutically implausible.

The principal formulation-design implication is a shift from optimizing a terminal measure to interrogating the complete chain of states and interfaces. A proposed composition change should be assessed for its effects on material organization, transformation dynamics, species-specific release, physiological transport, absorption input, and exposure—not merely for its effect on one benchmark endpoint. Benchmark performance, data availability, prospective usefulness, and pharmaceutical impact constitute different evidentiary levels [28]. The model may therefore support formulation prioritization, mechanism-discriminating experiments, sensitivity analysis, and scenario comparison, but its recommendations remain hypotheses until confirmed in the relevant experimental and development context.

Table 2 organizes the evidence, constructs, and interpretation boundaries needed to develop evaluation, interpretability, and formulation-design implications within the article's central argument.

Table 2. Evaluation, implementation, and research priorities arising from From Molecule to Patient: Learning Formulation Behavior across Material, Dosage-Form, Physiological, and Exposure Scales

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Molecular representation	Molecular structure, physicochemical measurements, assay context, provenance	Encode properties relevant to interaction and material-state questions	Endpoint-specific molecular state with uncertainty	Representation may omit formulation-relevant chemistry or fail under chemical-space shift	Comparison with strong baselines, external compounds, and endpoint-appropriate experimental evidence
Excipient-interaction layer	Drug and excipient identities, composition, interaction measurements, environmental conditions	Represent association, miscibility, mobility restriction, plasticization, and competing interactions	Conditional interaction state rather than a binary compatibility label	Analytical methods may observe different motifs; association does not establish causality	Orthogonal interaction measurements and composition- or condition-perturbation studies
Material-organization layer	Interaction state, processing history, morphology, phase behavior, mobility, storage conditions	Translate molecular and interaction evidence into a mesoscopic pharmaceutical state	Structured material state governing subsequent transformation	Bulk measurements may conceal minority phases, spatial heterogeneity, or aging	Spatial, temporal, and solid-state characterization linked to product-performance measurements
Dosage-form transformation layer	Initial material state, administration environment, time-resolved transformation data	Model wetting, hydration, disintegration, erosion, dissolution, precipitation, and particle formation	Species-specific concentrations, states, and release fluxes	Different mechanisms can produce similar terminal release profiles	Time-resolved orthogonal assays and perturbations that distinguish competing mechanisms
Physiological transport and absorption layer	Released species, route-specific physiology, transport barriers, permeability, transit, perfusion	Condition local transport and generate an absorbed systemic input	Time-, route-, species-, and location-dependent absorption input	Parameter compensation may conceal local-model error; transfer across routes is unsafe without evidence	Independent local transport, regional absorption, and systemic observations in the intended domain
Systemic exposure layer	Absorbed input, disposition model, population variables, pharmacokinetic observations	Translate absorption into plasma or target-site exposure	Exposure trajectories and bounded scenario comparisons	Exposure error may arise from absorption or disposition; exposure does not establish benefit or safety	Prospective pharmacokinetic evaluation across formulations and relevant populations
Uncertainty and applicability layer	Measurement variance, model ensembles, structural alternatives, domain definitions	Decompose uncertainty, detect unsupported use, and trigger abstention or evidence acquisition	Source-specific uncertainty and applicability statement	Calibration may fail in new formulation, route, process, or population domains	External-domain testing, repeated measurements, calibration audits, and predefined stop conditions
Interpretation and design-review layer	Module outputs, sensitivities, counterfactuals, uncertainty, development constraints	Generate bounded hypotheses and identify informative formulation experiments	Human-reviewable formulation rationale and experimental priorities	Explanations may be plausible but unfaithful; optimization may exploit model blind spots	Expert review, intervention tests, prospective comparison with established development workflows, and governed change control

Figure 2 presents the evidence, validation, and translation boundary observatory for from molecule to patient learning formulation behavior, showing how the article's key components and boundaries are connected within evaluation, interpretability, and formulation-design implications.

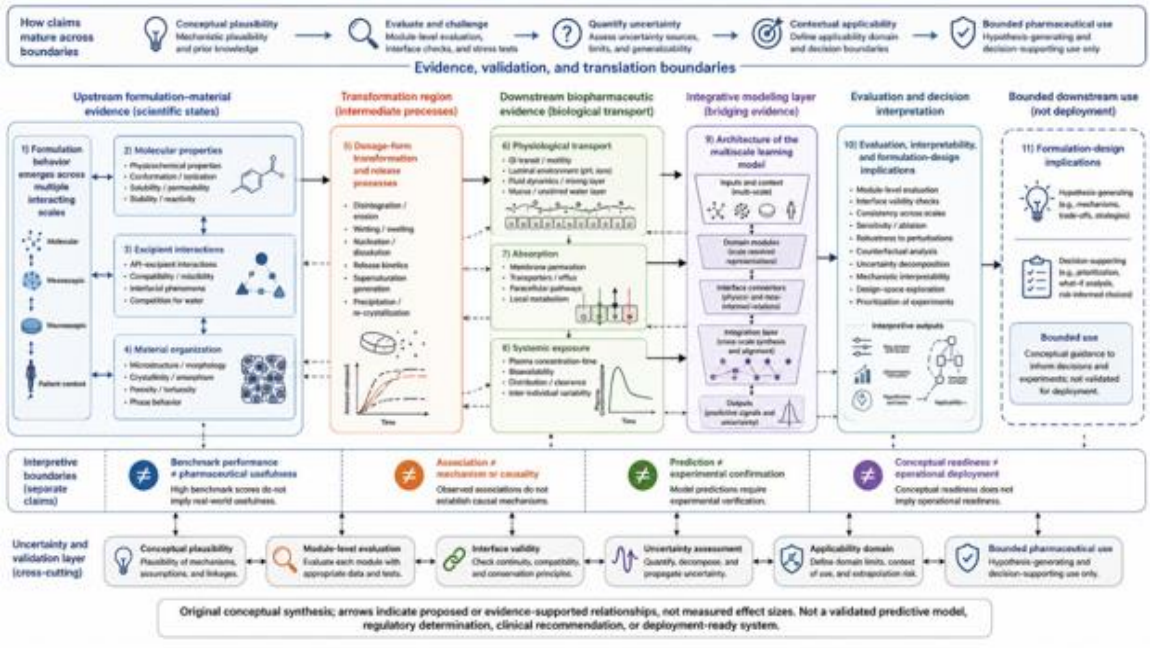

Figure 2. Evidence, Validation, and Translation Boundaries

The figure is an original conceptual synthesis that shows how claims move from conceptual plausibility through evaluation, uncertainty assessment, and bounded pharmaceutical use. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Limitations and Boundary Conditions

The principal limitation is that the Molecule-to-Patient Formulation Learning Model is a proposed architecture rather than an implemented or empirically validated system. The selected evidence supports individual relationships, modeling practices, and boundary conditions, but it does not demonstrate that one integrated representation can identify all cross-scale mechanisms. Modular decomposition may improve traceability, yet it can also create artificial boundaries between processes that are experimentally inseparable. The architecture does not establish a universal ontology, required feature set, optimal algorithm, or minimum dataset for every dosage form. Model purpose and context of use must therefore be specified before implementation or validation [17].

The available evidence is uneven across scales. Molecular and formulation datasets are often disconnected from time-resolved material, release, local physiological, absorption, and pharmacokinetic observations. Measurements obtained under different conditions may be semantically similar but scientifically non-equivalent. Sparse local transport data can permit compensation between release, absorption, and disposition parameters, while post hoc explanations can make an incorrect model appear mechanistically coherent. Interpretability outputs should therefore be treated as hypotheses requiring intervention or experimental confirmation rather than as proof of causal structure [25]. Uncertainty estimation may communicate some evidence gaps, but it cannot identify every omitted mechanism or guarantee reliable out-of-domain behavior [26].

Applicability is conditional on dosage-form class, administration route, compound chemistry, excipient system, manufacturing history, physiological population, and intended decision. The model cannot establish synthesizability, manufacturability, safety, efficacy, clinical utility, bioequivalence, or regulatory acceptability from computational coherence alone. It must not be used to eliminate candidates, replace confirmatory experiments, or authorize formulation changes without qualified human review. More broadly, technically impressive benchmark results can remain disconnected from prospective pharmaceutical impact when datasets, validation designs, and operational questions are misaligned [28]. The proposed contribution should therefore be interpreted as an evidence-organizing and hypothesis-generating framework whose value remains contingent on future implementation and validation.

Research Agenda and Implementation Priorities

The first priority is to develop linked, time-resolved datasets in which molecular, interaction, material, transformation, transport, absorption, and exposure observations retain shared formulation identifiers and complete provenance. Prospective perturbation studies should alter one or more interpretable formulation or physiological conditions and measure responses at adjacent scales. These studies are needed to determine whether interfaces transmit meaningful state and flux information or merely learn correlations between disconnected endpoints. Alternative physical constraints should also be compared through ablation, residual analysis, and intervention testing because physically informed architecture does not protect a model from misspecified mechanisms [23].

The second priority is decision-centered validation. Evaluation datasets should contain realistic changes in compounds, drug loads, excipients, processing conditions, dosage forms, administration routes, and physiological populations. Random record-level partitions are insufficient where near-duplicate chemistry or formulation conditions appear across training and evaluation sets. Prospective studies should compare model-assisted and established formulation workflows using predefined decisions, false-elimination risks, experimental costs, and human-override rules. Uncertainty estimates should be evaluated under these shifts, and abstention should be tested as an observable system behavior rather than assumed from an uncertainty score [27].

The third priority is implementation infrastructure that preserves scientific meaning throughout the model lifecycle. A versioned pharmaceutics ontology should define entities, states, processes, units, interfaces, and provenance without forcing unlike dosage forms into one representation. Module replacement should trigger an assessment of which upstream and downstream validation claims are affected. Data lineage, model documentation, applicability definitions, failure logs, and human accountability should accompany every integrated prediction. Product-quality applications of mechanistic biopharmaceutics modeling illustrate the importance of linking a model to a defined development question and evidence standard [20]. Implementation should consequently proceed in stages, beginning with bounded hypothesis generation and interface validation before any prospective decision-support claim is considered.

Conclusion

The Molecule-to-Patient Formulation Learning Model reframes formulation development as the study of connected state transitions across molecular properties, excipient interactions, material organization, dosage-form transformation, physiological transport, absorption, and systemic exposure. Its original contribution is a modular architecture that preserves the scientific meaning of each scale while specifying how states, fluxes, boundary conditions, provenance, and uncertainty may be transferred between them. The framework rejects the assumption that one performance measure can represent the entire formulation-to-patient pathway and separates predictive accuracy from mechanism, explanation, experimental confirmation, pharmaceutical usefulness, and operational readiness. It may help formulate testable cross-scale hypotheses, identify where a development assumption fails, and design experiments that reduce uncertainty at the most consequential interface. However, it remains a proposed conceptual and methodological contribution requiring route-specific implementation, linked evidence, prospective evaluation, human oversight, and explicit context-of-use boundaries. It does not establish universal applicability, clinical utility, regulatory acceptance, or deployment readiness.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021;175:113806. doi:10.1016/j.addr.2021.05.016
2. Bao Z, Bufton J, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Revolutionizing drug formulation development: The increasing impact of machine learning. Adv Drug Deliv Rev. 2023;202:115108. doi:10.1016/j.addr.2023.115108
3. Destro F, Barolo M. A review on the modernization of pharmaceutical development and manufacturing—Trends, perspectives, and the role of mathematical modeling. Int J Pharm. 2022;620:121715. doi:10.1016/j.ijpharm.2022.121715
4. Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15(7):1916. doi:10.3390/pharmaceutics15071916
5. Wang N, Zhang Y, Wang W, Ye Z, Chen H, Hu G, et al. How can machine learning and multiscale modeling benefit ocular drug development? Adv Drug Deliv Rev. 2023;196:114772. doi:10.1016/j.addr.2023.114772
6. Zheng F, Hou P, Corpstein CD, Park K, Li T. Multiscale pharmacokinetic modeling of systemic exposure of subcutaneously injected biotherapeutics. J Control Release. 2021;337:407-16. doi:10.1016/j.jconrel.2021.07.043
7. Di J, Hou P, Corpstein CD, Wu K, Xu Y, Li T. Multiphysics modeling and simulation of local transport and absorption kinetics of intramuscularly injected lipid nanoparticles. J Control Release. 2023;359:234-43. doi:10.1016/j.jconrel.2023.05.048
8. Wu D, Li M. Current state and challenges of physiologically based biopharmaceutics modeling (PBBM) in oral drug product development. Pharm Res. 2023;40(2):321-36. doi:10.1007/s11095-022-03373-0
9. Hiew TN, Zemlyanov DY, Taylor LS. Balancing solid-state stability and dissolution performance of lumefantrine amorphous solid dispersions: The role of polymer choice and drug-polymer interactions. Mol Pharm. 2022;19(2):392-413. doi:10.1021/acs.molpharmaceut.1c00481

10. Pugliese A, Tobyn M, Hawarden LE, Abraham A, Blanc F. New development in understanding drug-polymer interactions in pharmaceutical amorphous solid dispersions from solid-state nuclear magnetic resonance. Mol Pharm. 2022;19(11):3685-99. doi:10.1021/acs.molpharmaceut.2c00479
11. Bookwala M, Wildfong PLD. The implications of drug-polymer interactions on the physical stability of amorphous solid dispersions. Pharm Res. 2023;40(12):2963-81. doi:10.1007/s11095-023-03547-4
12. Deac A, Qi Q, Indulkar AS, Purohit HS, Gao Y, Zhang GGZ, et al. Dissolution mechanisms of amorphous solid dispersions: Role of drug load and molecular interactions. Mol Pharm. 2023;20(1):722-37. doi:10.1021/acs.molpharmaceut.2c00892
13. Yang R, Zhang GGZ, Zemlyanov DY, Purohit HS, Taylor LS. Release mechanisms of amorphous solid dispersions: Role of drug-polymer phase separation and morphology. J Pharm Sci. 2023;112(1):304-17. doi:10.1016/j.xphs.2022.10.021
14. Hate SS, Reutzel-Edens SM, Taylor LS. Insight into amorphous solid dispersion performance by coupled dissolution and membrane mass transfer measurements. Mol Pharm. 2019;16(1):448-61. doi:10.1021/acs.molpharmaceut.8b01117
15. Schittny A, Waldner S, Duthaler U, Vorobyev A, Abramovich R, Krähenbühl S, et al. Particle forming amorphous solid dispersions: A mechanistic randomized pharmacokinetic study in humans. Pharmaceutics. 2021;13(3):401. doi:10.3390/pharmaceutics13030401
16. Bapat P, Paul S, Tseng YC, Taylor LS. Interplay of drug-polymer interactions and release performance for HPMCAS-based amorphous solid dispersions. Mol Pharm. 2024;21(3):1466-78. doi:10.1021/acs.molpharmaceut.3c01106
17. Parrott N, Suarez-Sharp S, Kesisoglou F, Pathak SM, Good D, Wagner C, et al. Best practices in the development and validation of physiologically based biopharmaceutics modeling. A workshop summary report. J Pharm Sci. 2021;110(2):584-93. doi:10.1016/j.xphs.2020.09.058
18. Tannergren C, Jadhav H, Eckernäs E, Fagerberg J, Augustijns P, Sjögren E, et al. Physiologically based biopharmaceutics modeling of regional and colon absorption in humans. Eur J Pharm Biopharm. 2023;186:144-59. doi:10.1016/j.ejpb.2023.03.013
19. Li M, Zhao P, Pan Y, Wagner C. Predictive performance of physiologically based pharmacokinetic models for the effect of food on oral drug absorption: Current status. CPT Pharmacometrics Syst Pharmacol. 2018;7(2):82-9. doi:10.1002/psp4.12260
20. Mitra A, Suarez-Sharp S, Pepin XJH, Flanagan T, Zhao Y, Kotzagiorgis E, et al. Applications of physiologically based biopharmaceutics modeling (PBBM) to support drug product quality: A workshop summary report. J Pharm Sci. 2021;110(2):594-609. doi:10.1016/j.xphs.2020.10.059
21. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-30. doi:10.1039/c7sc02664a
22. Yang K, Swanson K, Jin W, Coley CW, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370-88. doi:10.1021/acs.jcim.9b00237
23. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys. 2021;3(6):422-40. doi:10.1038/s42254-021-00314-5
24. Gaudelet T, Day B, Jamasb AR, Soman J, Regep C, Liu G, et al. Utilizing graph machine learning within drug discovery and development. Brief Bioinform. 2021;22(6). doi:10.1093/bib/bbab159
25. 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
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. 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
28. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009