Pharmacophore

ISSN-2229-5402

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

A FOUNDATION MODEL SHOULD NOT FLATTEN PHARMACEUTICAL EVIDENCE ACROSS MOLECULES, SEQUENCES, OMICS, IMAGES, AND SCIENTIFIC TEXT

Omar Khoury[1]*, Elizabeth Ross[2], Jean-Marc Lefevre[3], Aisha Al-Hammadi[4]

1. ***Department of Multimodal Pharmaceutical Foundation Models, Faculty of Pharmacy, American University of Beirut, Beirut, Lebanon.***
2. ***Department of Evidence Preservation and Modality Integrity, College of Pharmacy, University of Arizona, Tucson, United States.***
3. ***Department of Non-Flattened Molecular and Omics Representation, Faculty of Sciences, University of Montpellier, Montpellier, France.***
4. ***Department of Scientific Text and Image Integration, College of Medicine and Health Sciences, UAE University, Al Ain, United Arab Emirates.***

ARTICLE INFO

Received:
15 October 2024
Received in revised form:
05 December 2024
Accepted:
09 December 2024
Available online:
28 December 2024

Keywords: Multimodal foundation models, Pharmaceutical data science, Evidence preservation, Molecular representation, Multi-omics integration, Conditional fusion

ABSTRACT

Pharmaceutical foundation models are increasingly expected to learn across chemical structures, biological sequences, molecular profiles, experimental images, and scientific language. Although joint representation learning may increase task coverage, naive multimodal fusion can erase the differences that determine what each source actually measures, predicts, or claims. Molecular structures encode chemical identity through representation-dependent abstractions; sequences reflect evolutionary and functional regularities; omics measurements describe context-sensitive biological states; images capture assay-conditioned phenotypes; and scientific text contains reported interpretations shaped by publication, terminology, and evidentiary quality. This article develops a proposed evidence-preserving multimodal pharmaceutical foundation-model architecture for organizing these non-equivalent evidence classes without reducing them to anonymous latent vectors. The architecture combines modality-specific evidence contracts, specialized encoders, typed alignment interfaces, conditional fusion, missing-modality management, disagreement preservation, uncertainty calibration, abstention, and bounded scientific-use controls. Its central principle is that fusion should be authorized by the scientific task, relation type, provenance, biological context, modality availability, uncertainty state, and intended decision rather than applied automatically whenever multiple inputs are available. Evaluation is correspondingly separated into within-modality representation validity, cross-modal alignment, transfer, grounding, calibration, disagreement handling, missingness robustness, and prospective scientific usefulness. The proposed architecture remains conceptual: it does not establish empirical superiority, mechanistic validity, clinical utility, regulatory acceptability, or deployment readiness. Its contribution is a methodological basis for designing pharmaceutical foundation models that preserve evidentiary distinctions, expose unsupported inferential transitions, and support more disciplined evaluation of when multimodal learning may contribute to drug discovery and development.

To Cite This Article: Khoury O, Ross E, Lefevre JM, Al-Hammadi A. A Foundation Model Should Not Flatten Pharmaceutical Evidence across Molecules, Sequences, Omics, Images, and Scientific Text. Pharmacophore. 2024;15(6):108-17. https://doi.org/10.51847/rbFe001wZP

Introduction

Foundation models have extended machine learning from narrowly specified tasks toward reusable representations that can support multiple biomedical objectives. Generalist medical models illustrate how shared pretraining may be adapted across heterogeneous tasks, while multimodal biomedical systems combine information obtained through different acquisition processes, and pharmaceutical machine learning spans target discovery, molecular design, preclinical research, and development decisions with distinct evidentiary demands [1-3]. These developments create a strong incentive to train pharmaceutical models across molecules, sequences, structures, omics profiles, images, experimental records, and scientific text. Yet greater modality coverage does not by itself produce greater scientific coherence. A model may process several

Corresponding Author: Omar Khoury; Department of Multimodal Pharmaceutical Foundation Models, Faculty of Pharmacy, American University of Beirut, Beirut, Lebanon. E-mail: omar.khoury@aub.edu.lb.

evidence types while obscuring the differences that determine whether an output represents chemical similarity, biological association, phenotypic correspondence, textual co-occurrence, predictive inference, or experimentally supported knowledge. Therapeutic foundation-model research has therefore begun to emphasize integration across molecular, cellular, biological, and disease scales rather than model size alone [4]. The unresolved problem is how such integration should occur without treating all inputs as interchangeable observations of one underlying truth. A molecular graph, a protein sequence, a transcriptomic profile, a microscopy image, and a sentence in a journal article differ not only in format but also in scientific status. Some encode defined entities, some record measurements, some summarize biological states, some depict phenotypes, and some communicate claims made by investigators. Their reliability, uncertainty, provenance, dimensionality, temporal scale, and relationship to mechanism are correspondingly different. When these distinctions disappear inside a common embedding, the resulting representation may become computationally convenient while becoming scientifically difficult to interpret.

The risk is particularly consequential in pharmaceutical science because multimodal outputs can influence decisions that lie at different distances from experimental confirmation. A model may be asked to retrieve related compounds, predict binding, infer a target, prioritize a pathway, interpret a cellular phenotype, propose a biomarker, summarize literature, or recommend a next experiment. The validity conditions for these tasks are not equivalent. Biomedical multimodality can improve information coverage, but its outputs remain conditional on how the contributing modalities were generated, aligned, selected, and validated [2]. Similarly, performance in a drug-discovery benchmark cannot be assumed to establish synthesizability, selectivity, developability, safety, therapeutic relevance, or prospective utility [3]. The central problem is therefore not whether modalities can be mathematically combined, but whether the combination preserves the evidence needed to interpret the resulting claim.

This article proposes an Evidence-Preserving Multimodal Pharmaceutical Foundation Model, abbreviated EP-MPFM, as a conceptual architecture for addressing this problem. The contribution is not a new trained model or an empirically validated fusion algorithm. It is an architectural and interpretive specification that requires each modality to retain an evidence contract describing what the input represents, how it was generated, which context applies, what uncertainty accompanies it, and which claims it may legitimately support. Modality-specific encoders produce evidence-bearing representations; typed interfaces distinguish identity, derivation, association, prediction, description, and contextual co-occurrence; and a conditional fusion mechanism determines whether evidence should be combined, compared, kept separate, or rejected as insufficient. Disagreement, missing modalities, calibration, abstention, scientific usefulness, and deployment boundaries are treated as core architectural properties rather than downstream reporting details.

Why Naive Multimodal Fusion Erases Evidentiary Differences

Multimodal fusion is often described through the computational stage at which information is combined: raw inputs may be joined before representation learning, modality-specific features may be merged in an intermediate latent space, or separate predictions may be aggregated at the decision level. These strategies impose different assumptions about scale, correspondence, and information sharing. Biomedical fusion research shows that integration design can privilege dominant modalities or suppress information that is specific to one view; multi-omics factor modeling demonstrates the importance of separating shared variation from modality-specific structure; and single-cell multi-omics research shows that assays differ in sparsity, resolution, noise, pairing, and biological interpretation [5-7]. Naive fusion erases evidentiary differences when it treats these distinctions as nuisance variation rather than as information required to understand what the model has learned.

The first flattening mechanism is distributional equivalence. Concatenating or projecting all inputs into a common space can imply that their dimensions have comparable observational meanings even when one modality contains counts, another contains categorical tokens, another contains continuous image intensities, and another contains extracted textual claims. Joint probabilistic modeling of single-cell RNA and protein measurements illustrates a more defensible alternative: shared latent structure can be learned while each measurement type retains a likelihood suited to its own data-generating process [8]. The architectural implication is not that every pharmaceutical modality requires a probabilistic generative model of the same form. Rather, a foundation model should preserve sufficient information about measurement scale, preprocessing, censoring, technical noise, and acquisition so that cross-modal learning does not silently convert distinct observation processes into apparently homogeneous features.

A second mechanism is context stripping. Relationships between modalities are often conditional on specimen, organism, tissue, cell state, time point, dose, perturbation, assay platform, image channel, publication context, or experimental objective. Single-cell multi-omics technologies demonstrate that even ostensibly matched biological measurements may differ in whether they are paired within the same cell, measured in separate samples, inferred computationally, or connected through a reference atlas [9]. The same concern applies across pharmaceutical data. A molecule may be linked to a protein through an assay-specific activity value, to an image through a treatment-induced phenotype, to an omics profile through a particular dose and time point, and to text through an author-reported interpretation. Removing these relation conditions can transform a qualified association into an apparently direct and universal connection.

A third mechanism is evidence-status collapse. Measured observations, model-derived predictions, database annotations, literature extractions, inferred missing values, and expert interpretations may enter the same architecture without durable labels identifying their origin. When this occurs, the model can propagate a prediction as though it were an observation, use an extracted textual assertion as though it were experimental confirmation, or average contradictory sources into a confident latent

representation. EP-MPFM therefore proposes that evidence identity must remain attached through encoding and fusion. The architecture should record whether an input was measured, predicted, imputed, extracted, curated, or generated; preserve provenance and uncertainty; and expose unresolved disagreement rather than automatically compressing it. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why naive multimodal fusion erases evidentiary differences within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why naive multimodal fusion erases evidentiary differences

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Modality identity	What scientific object does each input represent?	Molecules, sequences, omics measurements, images, and text encode different entities, observations, or claims.	Establishes why multimodality is not equivalent to multiple interchangeable views of one object.	A shared embedding may be interpreted as proof that modalities have the same meaning.	Common representation does not establish semantic equivalence.
Data-generating process	How was the evidence produced?	Assays, imaging systems, sequencing technologies, curation practices, and language corpora have distinct noise and bias structures.	Requires modality-specific likelihoods, preprocessing records, and quality descriptors.	Technical artifacts may become fused with biological or chemical signals.	Learned associations remain conditional on acquisition and processing.
Provenance and evidence status	Is the input measured, predicted, imputed, extracted, curated, or generated?	Evidence sources differ in their distance from direct observation and experimental confirmation.	Supports persistent evidence contracts and traceable outputs.	Predicted or extracted information may be presented as observed fact.	Transformations must not erase the status of the originating evidence.
Cross-modal relation type	What relation connects two modalities?	Identity, derivation, association, description, prediction, and contextual co-occurrence support different interpretations.	Motivates typed rather than anonymous alignment edges.	High alignment may be misread as causality, mechanism, or interchangeability.	Alignment strength does not determine the scientific meaning of the relation.
Context dependence	Under which biological, chemical, experimental, or textual conditions does the relation hold?	Dose, time, tissue, organism, assay, cell state, instrument, and study design alter interpretation.	Makes context part of the encoded evidence rather than external metadata.	Context-specific findings may be generalized beyond their valid domain.	Cross-modal relations are valid only within supported contexts.
Modality dominance	Does one modality determine the fused output because of scale, dimensionality, abundance, or label density?	Fusion architectures can privilege high-dimensional or frequently observed inputs.	Requires modality-level ablation, contribution analysis, and routing controls.	Apparent multimodal performance may reflect one dominant source and hidden degradation elsewhere.	Fusion benefit must be compared with valid unimodal and restricted-fusion baselines.
Missing modalities	Why is an expected modality absent?	Absence may be structural, selective, censored, unavailable, or outside reference coverage.	Supports separate treatment of missingness mechanisms and abstention.	Imputed evidence may be treated as measured, or missingness may leak outcome information.	Missing evidence cannot be silently replaced by a neutral value or unlabelled estimate.
Cross-modal disagreement	What should happen when modalities support incompatible interpretations?	Conflict may reflect biological heterogeneity, technical error, contextual differences, or model failure.	Introduces a disagreement ledger rather than forced averaging.	Contradictions may be hidden inside a confident fused prediction.	Preserved disagreement does not establish which source is correct.
Uncertainty	Is uncertainty associated with the input, alignment, representation, fusion decision, or output?	Different uncertainty sources require different assessments and cannot be summarized by confidence alone.	Connects evidence preservation to calibration and abstention.	A confident model output may be interpreted as low scientific uncertainty.	Predictive confidence does not establish calibrated or decision-relevant uncertainty.
Intended pharmaceutical use	What decision or scientific action is the fused representation expected to support?	Retrieval, prioritization, mechanistic inference, experimental planning, and deployment have different evidence thresholds.	Makes fusion conditional on a declared task and consequence.	Benchmark improvement may be presented as pharmaceutical usefulness or readiness.	Fusion is permissible only for the defined task, evidence state, and validation scope.

Distinct Semantics of Molecules, Sequences, Omics, Images, and Text

Molecular evidence begins with a chemically defined object but is never independent of representation. A compound may be encoded as a graph, line notation, descriptor vector, fingerprint, conformer ensemble, reaction participant, or learned token sequence. These encodings preserve different aspects of connectivity, stereochemistry, geometry, electronic environment, substructure, or synthetic context. Comparative analyses of molecular representations show that predictive behavior depends on both the chosen encoding and the downstream property task [10]. An EP-MPFM should therefore not treat a molecular embedding as a representation of complete chemical truth. The embedding must remain linked to the original structure,

representation type, stereochemical completeness, conformational assumptions, protonation or tautomer state where relevant, and the domain over which the encoder was trained. Even a highly transferable molecular representation cannot alone establish synthesizability, assay activity, selectivity, developability, safety, or therapeutic value.

Sequence evidence has a different semantic basis. Protein or nucleic-acid sequences are ordered biological polymers shaped by evolution, genomic context, organism, isoform, processing, and functional constraint. Large-scale sequence pretraining can learn representations associated with structural and functional regularities, demonstrating that biological information may emerge from self-supervised analysis of extensive sequence corpora [11]. Nevertheless, a sequence-derived representation remains distinct from an experimentally measured structure, biochemical activity, cellular function, or disease mechanism. The same amino-acid sequence can participate in different conformational, interaction, localization, or post-translational states, while closely related sequences may have context-dependent functional differences. An evidence-preserving architecture should retain organism, isoform, sequence provenance, annotation status, and any distinction between observed and predicted structural information. It should also prevent sequence similarity from being reported as proof of functional equivalence or therapeutic relevance.

Omics evidence represents context-sensitive measurement rather than fixed biological identity. Genomic, epigenomic, transcriptomic, proteomic, metabolomic, and related assays capture different layers of biological organization through technology-specific sampling, normalization, and inference processes. Deep-learning methods in genomics have shown that model interpretation depends on sequence context, regulatory architecture, assay design, cell type, and the relationship between measured labels and the biological process of interest [12]. Omics profiles are further shaped by tissue composition, cell state, disease stage, exposure, dose, time, technical batch, and analytical pipeline. Their high dimensionality does not make them comprehensive descriptions of a biological system. In EP-MPFM, an omics encoder should therefore retain assay type, biological scale, sample context, preprocessing history, measurement uncertainty, and reference coverage. Associations between omics features, compounds, targets, or phenotypes may support hypothesis generation, but they do not alone demonstrate mechanism, causal mediation, treatment response, or clinical utility.

Image and text evidence require additional separation because both can appear semantically rich while remaining indirect. Image-based profiling in drug discovery captures morphology, localization, spatial organization, or tissue architecture under acquisition conditions that determine which phenotypes are visible [13]. A learned image representation may connect perturbations with cellular states, yet morphological similarity is not molecular identity and does not establish a shared mechanism. Scientific text is further removed from direct measurement: biomedical language models encode terminology, reported relations, discourse patterns, and claims extracted from selected corpora [14]. Text may summarize experimental evidence, repeat established knowledge, communicate uncertainty, or propagate unsupported interpretations. EP-MPFM should consequently label textual elements by document provenance, section, claim type, evidentiary support, temporal context, and extraction confidence. It should permit text to contextualize, retrieve, or qualify molecular and biological evidence without allowing linguistic fluency, citation frequency, or textual co-occurrence to substitute for experimental confirmation.

Modality-Specific Encoders and Evidence-Preserving Interfaces

An evidence-preserving architecture begins with encoders designed for the scientific properties of their source modalities rather than with a universal tokenizer that assumes every object can be interpreted through the same representational grammar. Chemical reactions, for example, require representations that retain reactant identity, reagent and condition information, product relationships, stereochemical details, and uncertainty about alternative outcomes. Transformer-based reaction modeling demonstrates that domain-specific chemical tokenization can support reaction prediction while making uncertainty an explicit evaluation concern [15]. In the proposed EP-MPFM, a molecular or reaction encoder should therefore emit not only a learned vector but also an evidence-bearing representation linked to the original chemical object, its encoding, its training-domain coverage, and the conditions under which the representation may be interpreted.

Large-scale molecular language pretraining further shows that reusable chemical representations can transfer across several downstream property tasks [16]. Such transfer is valuable, but it does not remove the need to distinguish the representation from the pharmaceutical evidence it supports. A molecular embedding may encode substructural, physicochemical, or corpus-derived regularities without revealing which properties are reliable in a new chemical neighborhood. EP-MPFM consequently proposes an interface that preserves structure provenance, representation type, stereochemical completeness, conformational assumptions, and domain-distance information. Downstream components should be able to determine whether a representation arose from a molecular graph, a line notation, a predicted conformer, a reaction context, or a transformed derivative rather than receiving an anonymous vector whose scientific origin is no longer recoverable.

The proposed interface layer is an evidence contract. Each encoded object would carry a compact, machine-readable declaration of what the object represents, how it was obtained, which transformations have been applied, what context remains applicable, what uncertainty is known, and which claims the representation may support. Evidence contracts would also distinguish measured, predicted, extracted, curated, generated, and imputed inputs. This design does not require every modality to use an identical metadata schema. Instead, a shared upper-level structure would permit modality-specific fields: molecular contracts could preserve stereochemistry and chemical state; omics contracts could preserve assay, biological scale, and preprocessing; image contracts could preserve acquisition conditions; and text contracts could preserve document provenance, claim type, and extraction confidence.

Specialization is equally important for biological encoders. Protein-structure prediction depends on architectures and priors suited to sequence, evolutionary, and geometric information; transcriptomic transfer learning depends on cell-state and network contexts; and single-cell foundation modeling uses gene- and cell-specific objectives to support multi-omic, perturbational, and annotation tasks [17-19]. These achievements do not imply that protein structures, transcriptomic states, and cellular profiles should share an undifferentiated representation. EP-MPFM instead proposes evidence-bearing latent tokens whose modality identity persists after encoding. Cross-modal interfaces may transform these tokens, but every transformation should remain traceable to the originating evidence object, its context, and the uncertainty introduced by the encoder or transfer operation.

Conditional Fusion, Alignment, Disagreement, and Missing Modalities

Conditional fusion begins by asking whether the available evidence is suitable for the declared task. Reference-mapping methods for single-cell data illustrate that transfer can be useful when a query is compatible with a reference while remaining vulnerable to incomplete reference coverage, biological novelty, and study effects [20]. EP-MPFM generalizes this principle by requiring the fusion controller to inspect modality availability, domain coverage, provenance, context, relation type, and uncertainty before combination. The controller may authorize joint processing, retain modalities as parallel evidence, request additional information, or abstain. Fusion is therefore treated as a scientific decision governed by evidence conditions rather than as an unavoidable stage of multimodal modeling.

Alignment must likewise remain typed. Molecule–text models show that contrastive learning can support retrieval and text-guided editing by connecting chemical structures with natural-language descriptions [21]. However, such alignment does not make a molecular graph and a textual statement semantically equivalent. A description may refer to a property, use, assay result, hypothesis, or literature claim, and each relation supports a different interpretation. EP-MPFM proposes directional alignment labels such as identity, derivation, association, prediction, description, and contextual co-occurrence. The model should preserve which relation was trained, which evidence created the pair, and whether the connection was directly observed, curated, inferred, or generated.

Bidirectional molecular modeling further demonstrates that structures and properties can be mapped in both directions within a shared architecture [22]. Bidirectionality can improve retrieval or generation, but it also creates a risk that a predicted property becomes embedded as though it were an intrinsic and experimentally established attribute of the molecule. Evidence-preserving fusion should therefore retain directionality and transformation history. A structure-to-property prediction and a property-conditioned structure generation are different scientific operations, even when they use the same model. Their errors, uncertainties, validation requirements, and permissible downstream claims must remain separable.

Natural-language interaction can make molecular discovery tools more accessible by translating textual instructions into retrieval or editing operations, but generated molecular outputs still require chemical and experimental review [23]. Similarly, multimodal prediction of molecule–disease relations can combine heterogeneous relational evidence without establishing that a predicted link is causal or therapeutically actionable [24]. EP-MPFM therefore adds two control structures. A disagreement ledger records incompatible modality-specific predictions or claims instead of averaging them away. A missing-modality manager distinguishes evidence that is unmeasured, unavailable, censored, structurally absent, or outside reference coverage. Imputed information must remain visibly imputed, and unresolved conflict should be able to trigger abstention, repeat measurement, expert review, or a proposed next experiment.

Evaluation of Transfer, Grounding, Calibration, and Scientific Usefulness

Evaluation should begin with task-specific validity rather than a single aggregate score. Molecular benchmarking frameworks demonstrate that datasets, split strategies, endpoints, and metrics materially affect the interpretation of model performance [25]. An EP-MPFM evaluation should therefore separate within-modality representation quality from cross-modal alignment and fusion benefit. Molecular, sequence, omics, image, and text encoders should first be evaluated against valid modality-specific baselines. Fusion should then be compared with the strongest applicable unimodal and restricted-fusion alternatives through ablations that reveal whether every included modality contributes useful evidence or whether one dominant modality accounts for the apparent gain.

Transfer must be assessed across the dimensions that matter to the intended pharmaceutical use. Ligand-based benchmarks can reward memorization when closely related compounds appear across training and evaluation sets, creating performance estimates that do not reflect chemical generalization [26]. For a multimodal foundation model, leakage can additionally occur through duplicated structures, homologous sequences, repeated samples, near-identical images, shared patients, overlapping articles, database-derived labels, or relation graphs that indirectly expose the target. Evaluation should therefore include scaffold, temporal, assay, laboratory, instrument, species, tissue, population, and corpus splits where appropriate. Success on one transfer axis must not be reported as proof of transportability across the others.

Uncertainty evaluation must determine whether a model knows when its evidence is inadequate. Comparative studies of uncertainty estimation in molecular prediction show that methods differ in calibration, ranking performance, and scalability across datasets and tasks [27]. EP-MPFM requires uncertainty to be localized rather than collapsed into one confidence value. Relevant sources include measurement uncertainty, encoder uncertainty, domain distance, ambiguous alignment, missing modalities, cross-modal disagreement, and uncertainty in the final task output. Evaluation should test whether these signals

deteriorate under distribution shift and whether abstention occurs for scientifically appropriate reasons rather than merely for difficult but familiar cases.

Model confidence must also be distinguished from calibrated uncertainty, particularly outside the training domain [28]. Grounded explanations can contribute to scientific usefulness only when they remain faithful to the model, traceable to the underlying evidence, chemically or biologically coherent, and explicit about interpretive limits [29]. Scientific usefulness should ultimately be evaluated prospectively through a defined decision, such as evidence retrieval, compound prioritization, hypothesis comparison, or next-experiment selection, rather than inferred from retrospective benchmark performance. **Figure 1** presents the multimodal evidence-preserving foundation core, showing how the article's key components and boundaries are connected within evaluation of transfer, grounding, calibration, and scientific usefulness.

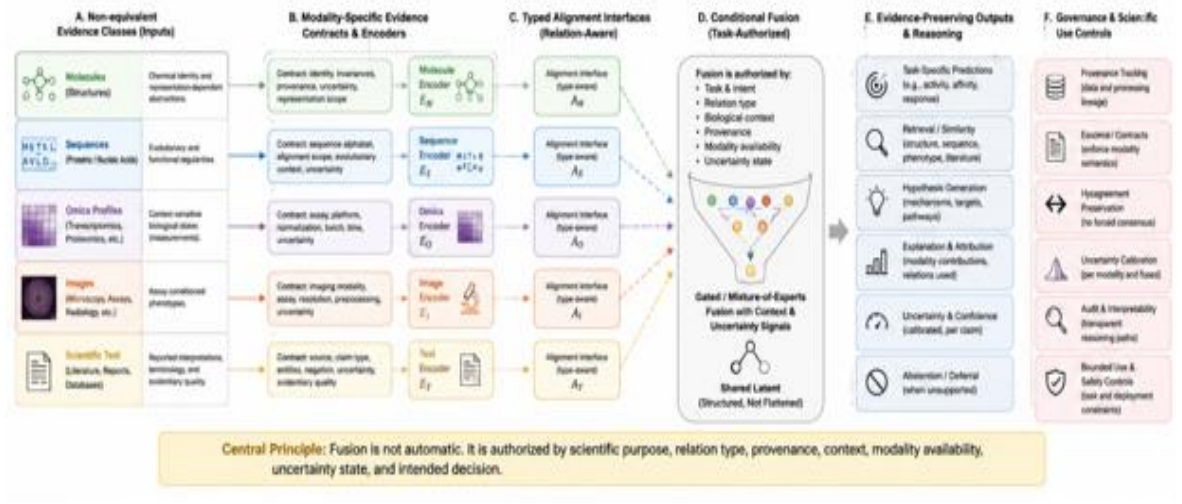

Figure 1. Multimodal Evidence-Preserving Foundation Core

The figure is an original conceptual synthesis that organizes the article's central contribution across naive multimodal fusion erases evidentiary differences, distinct semantics of molecules, sequences, omics, images, and text, distinct semantics of molecules, modality-specific encoders and evidence-preserving interfaces, conditional fusion, alignment, disagreement, and missing modalities, conditional fusion. 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.

Deployment Boundaries for Pharmaceutical Foundation Models

Deployment boundaries begin with the distinction between technical performance and real-world impact. Clinical AI research has shown that effective translation depends on workflow integration, target-population definition, usability, prospective assessment, data quality, and changes in the operational environment rather than on model accuracy alone [30]. Pharmaceutical foundation models face analogous constraints even when they are used before clinical deployment. A model that performs well in retrospective molecular, omics, imaging, or text benchmarks may still fail to support a valid discovery decision. Its outputs must be evaluated in relation to the intended user, scientific stage, decision consequence, available alternatives, and follow-up evidence.

Responsible deployment also requires lifecycle controls extending beyond initial validation. A health-care machine-learning roadmap emphasizes problem formulation, stakeholder involvement, validation, monitoring, and harm mitigation throughout the model lifecycle [31]. Applied to EP-MPFM, these principles require versioned evidence contracts, controlled model and data updates, traceable output generation, role-specific interfaces, documented human review, and mechanisms for withdrawing or revising unsupported claims. The architecture should identify which outputs may support exploratory research, which require expert confirmation, and which uses remain prohibited. Human oversight must involve authority to challenge, defer, or reject a model output rather than functioning as ceremonial approval after an automated decision.

Generalisability must be framed as a relationship among a model, a task, a setting, and a data-generating process rather than as a permanent model property [32]. A pharmaceutical model may transfer across compounds but not assays, across cell lines but not tissues, across imaging instruments but not staining protocols, or across literature corpora but not newly emerging terminology. Deployment documentation should therefore specify the validated chemical, biological, technical, population, and temporal domains separately. Claims should narrow when the model encounters unfamiliar evidence contracts, unsupported relation types, unresolved modality disagreement, substantial domain distance, or missing data that differ from the conditions used during evaluation.

External validation studies demonstrate that models can perform differently across institutions because they exploit site-specific features and prevalence patterns [33]. Robustness is further constrained by deliberate or incidental perturbations that can create confident but unsafe outputs [34]. Pharmaceutical multimodality expands this attack and failure surface through

corrupted molecular records, altered images, batch-dependent omics signals, poisoned relation graphs, misleading text, and manipulated prompts. Deployment therefore requires monitoring, security testing, abstention, change control, and incident-response procedures proportionate to the intended use. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop deployment boundaries for pharmaceutical foundation models within the article's central argument. **Figure 2** presents the evidence, validation, and translation boundary observatory for a foundation model should not flatten pharmaceutical, showing how the article's key components and boundaries are connected within deployment boundaries for pharmaceutical foundation models.

Table 2. Evaluation, implementation, and research priorities arising from A Foundation Model Should Not Flatten Pharmaceutical Evidence across Molecules, Sequences, Omics

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Modality evidence contract	Source object, provenance, acquisition method, context, transformation history, evidence status	Declare what the input represents and which claims it may support	Traceable, modality-specific evidence record	Metadata may be incomplete, inconsistent, or itself erroneous	Schema completeness, provenance auditing, and expert review across modalities
Modality-specific encoder	Molecular, sequence, omics, image, or text evidence and its contract	Learn reusable representations without discarding modality identity	Evidence-bearing latent tokens	Representations may encode shortcuts, corpus bias, artifacts, or limited domain coverage	Within-modality probing, external datasets, perturbation analysis, and domain-specific baselines
Typed alignment interface	Paired or relational evidence with direction and relation type	Connect modalities while preserving whether the relation is identity, derivation, association, prediction, or description	Calibrated, interpretable cross-modal relation	Pairing may be incomplete, noisy, asymmetric, or based on co-occurrence	Hard-negative testing, relation-specific retrieval, provenance audits, and expert adjudication
Conditional fusion router	Encoded evidence, task, context, uncertainty, domain coverage, modality availability	Select which modalities may be combined, compared, retained separately, or rejected	Task-bounded fused representation or abstention	Routing may privilege dominant modalities or fail under unfamiliar missingness	Unimodal comparisons, controlled ablations, modality-drop tests, and cross-context evaluation
Disagreement ledger	Conflicting predictions, measurements, annotations, or textual claims	Preserve and localize cross-modal conflict	Explicit contradiction record and review trigger	Conflict may reflect true heterogeneity, measurement error, or model failure	Adjudicated discordant cases, repeat measurements, and resolution-trace auditing
Missing-modality manager	Availability map and reason for absence	Distinguish unavailable, censored, structurally absent, and out-of-reference evidence	Conditional inference, warning, evidence request, or abstention	Missingness mechanism may be unknown or informative	Real and simulated missingness studies stratified by mechanism
Calibration and abstention layer	Predictive distributions, ensembles, domain distance, disagreement, quality signals	Estimate task-relevant uncertainty and stop unsupported outputs	Calibrated prediction, risk statement, or non-answer	Calibration can deteriorate under distribution shift	Reliability analysis, risk–coverage testing, shift evaluation, and continuous recalibration
Grounding and provenance layer	Input evidence, alignment paths, model traces, supporting records	Link each output to the evidence and transformations that produced it	Auditable evidence trace and bounded explanation	Plausible explanations may be unfaithful or selectively grounded	Counterfactual evidence removal, trace reproducibility, and claim-support audits
Scientific-usefulness gate	Grounded output, uncertainty, intended decision, alternatives, human interpretation	Determine whether the output can improve a defined scientific action	Bounded recommendation, hypothesis, retrieval result, or next-experiment proposal	Usefulness is role-, stage-, and consequence-dependent	Prospective, blinded workflow studies with predefined decisions and error consequences
Deployment-boundary controller	Validation scope, user role, workflow, consequence, security, monitoring, governance	Restrict use to settings and claims supported by the evidence	Allowed-use statement, human-review requirement, or prohibition	Contexts and evidence can change after deployment	External and prospective validation, lifecycle monitoring, security testing, and change control
Versioning and audit infrastructure	Model, data, ontology, interface, and calibration versions	Preserve reproducibility and identify the impact of updates	Versioned records, regression results, rollback pathway	Silent upstream changes may invalidate prior outputs	Longitudinal regression testing, drift monitoring, and controlled release governance
Staged implementation programme	Research tasks, evidence maturity, user needs, risk classification	Move from conceptual testing to bounded research use without implying readiness	Sequenced development and validation plan	No stage guarantees success at the next stage	Independent replication and explicit advancement criteria at every stage

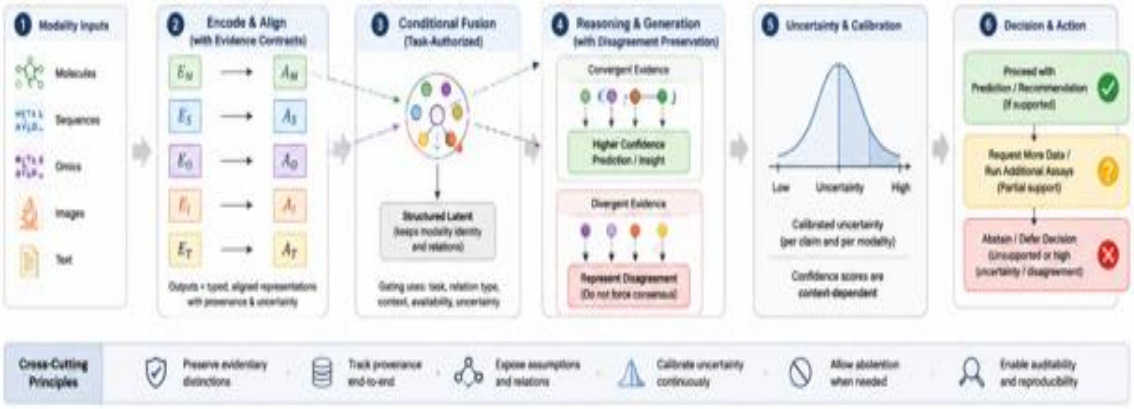

Figure 2. Evidence, Validation, and Translation Boundaries

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 of EP-MPFM is that it is a proposed conceptual architecture rather than an implemented and validated system. The article specifies components, relationships, evaluation dimensions, and claim boundaries but does not establish that evidence contracts, typed alignment, conditional routing, disagreement preservation, or abstention will improve performance or scientific decisions in practice. Different pharmaceutical settings may require different modality definitions, relation ontologies, uncertainty models, and human-review structures. The architecture should therefore be treated as a falsifiable design proposal whose components may need to be revised, separated, or rejected through future empirical work.

The supporting evidence is also heterogeneous. Molecular modeling, sequence learning, omics integration, image-based profiling, biomedical language modeling, health-care translation, and robustness research address different data-generating processes and decision environments. Their combination supports theory building but does not prove that findings transfer unchanged to a unified pharmaceutical foundation model. Several boundary principles are inferred from adjacent biomedical or clinical contexts rather than demonstrated across pharmaceutical workflows. The architecture consequently cannot establish universal transferability, mechanistic correctness, causal validity, clinical utility, economic value, ethical acceptability, regulatory conformity, or safety.

Implementation introduces additional risks. Evidence contracts may be incomplete, expensive to maintain, or inconsistent across organizations. Typed relation systems may oversimplify scientific ambiguity, while complex interfaces may burden users or conceal uncertainty behind technical terminology. Conditional fusion can still learn shortcuts, and disagreement mechanisms may generate excessive deferral or false conflict. Human oversight can fail through automation bias, unclear authority, or inadequate domain expertise. The architecture also cannot prevent misuse when users deliberately ignore its boundaries, remove provenance, override abstention, or present exploratory outputs as validated evidence.

Research Agenda and Implementation Priorities

The first priority is to make evidence preservation testable. Future work should develop interoperable evidence-contract schemas for molecules, reactions, sequences, structures, omics assays, images, and scientific claims. These schemas should define provenance, evidence status, context, relation type, uncertainty, transformation history, and permissible inference. Benchmark resources should then contain typed cross-modal relations, asymmetric mappings, contradictory evidence, hard negatives, and naturally missing modalities. Evaluation should test whether models retain modality identity and relation meaning rather than merely achieving high retrieval or prediction scores.

The second priority is prospective evaluation of conditional fusion. Studies should compare automatic fusion with task-aware routing, unimodal baselines, restricted fusion, explicit disagreement handling, and abstention. Designs should include scaffold, temporal, assay, laboratory, instrument, species, tissue, population, and corpus shifts where relevant. Missingness should be stratified by cause, and imputed information should remain distinguishable from observation. Grounding tests should remove or contradict supporting evidence to determine whether outputs change faithfully. Scientific-usefulness studies should define the intended decision in advance and measure whether the model improves evidence retrieval, hypothesis discrimination, experiment prioritization, or error recognition.

Implementation should proceed in stages rather than through immediate broad deployment. An initial stage should focus on retrospective, low-consequence research tasks with complete provenance and independent review. A second stage could evaluate prospective scientific workflows under controlled conditions, prespecified abstention rules, versioned models, and documented human authority. Broader use should be considered only after external replication, lifecycle monitoring, security testing, and evidence that the system improves the defined decision without obscuring modality limitations. Advancement

between stages should depend on explicit evidence and failure criteria, not on architecture complexity, model scale, or benchmark leadership.

Conclusion

A pharmaceutical foundation model should not gain apparent generality by discarding the distinctions that make its inputs scientifically interpretable. The proposed EP-MPFM organizes molecules, sequences and structures, omics, images, and scientific text through modality-specific evidence contracts, specialized encoders, typed alignment, conditional fusion, disagreement preservation, missing-modality management, calibration, abstention, and bounded-use controls. Its central contribution is to reposition fusion as an evidence-governed decision rather than an automatic representational operation. The architecture also separates benchmark performance from transfer, grounding, scientific usefulness, experimental confirmation, and deployment. It remains a conceptual proposal requiring empirical, prospective, and context-specific validation, but it provides a structured basis for designing multimodal pharmaceutical models that preserve provenance, expose uncertainty, and resist unsupported inferential flattening.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-65. doi:10.1038/s41586-023-05881-4
2. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773-84. doi:10.1038/s41591-022-01981-2
3. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5
4. Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, et al. Artificial intelligence foundation for therapeutic science. Nat Chem Biol. 2022;18(10):1033-6. doi:10.1038/s41589-022-01131-2
5. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: A review. Brief Bioinform. 2022;23(2). doi:10.1093/bib/bbab569
6. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21(1):111. doi:10.1186/s13059-020-02015-1
7. Lee J, Hyeon DY, Hwang D. Single-cell multiomics: Technologies and data analysis methods. Exp Mol Med. 2020;52(9):1428-42. doi:10.1038/s12276-020-0420-2
8. Gayoso A, Steier Z, Lopez R, Regier J, Nazor KL, Streets A, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nat Methods. 2021;18(3):272-82. doi:10.1038/s41592-020-01050-x
9. Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol. 2023;24(10):695-713. doi:10.1038/s41580-023-00615-w
10. Yang K, Swanson K, Jin W, Coley C, 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
11. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A. 2021;118(15). doi:10.1073/pnas.2016239118
12. Eraslan G, Avsec Z, Gagneur J, Theis FJ. Deep learning: New computational modelling techniques for genomics. Nat Rev Genet. 2019;20(7):389-403. doi:10.1038/s41576-019-0122-6
13. Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE. Image-based profiling for drug discovery: Due for a machine-learning upgrade? Nat Rev Drug Discov. 2021;20(2):145-59. doi:10.1038/s41573-020-00117-w
14. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36(4):1234-40. doi:10.1093/bioinformatics/btz682
15. Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci. 2019;5(9):1572-83. doi:10.1021/acscentsci.9b00576
16. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P. Large-scale chemical language representations capture molecular structure and properties. Nat Mach Intell. 2022;4(12):1256-64. doi:10.1038/s42256-022-00580-7
17. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi:10.1038/s41586-021-03819-2

18. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning enables predictions in network biology. Nature. 2023;618(7965):616-24. doi:10.1038/s41586-023-06139-9
19. Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024;21(8):1470-80. doi:10.1038/s41592-024-02201-0
20. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Buttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022;40(1):121-30. doi:10.1038/s41587-021-01001-7
21. Liu S, Nie W, Wang C, Lu J, Qiao Z, Liu L, et al. Multi-modal molecule structure-text model for text-based retrieval and editing. Nat Mach Intell. 2023;5(12):1447-57. doi:10.1038/s42256-023-00759-6
22. Chang J, Ye JC. Bidirectional generation of structure and properties through a single molecular foundation model. Nat Commun. 2024;15(1):2323. doi:10.1038/s41467-024-46440-3
23. Zeng Z, Yin B, Wang S, Liu J, Yang C, Yao H, et al. ChatMol: Interactive molecular discovery with natural language. Bioinformatics. 2024;40(9). doi:10.1093/bioinformatics/btae534
24. Wen J, Zhang X, Rush E, Panickan VA, Li X, Cai T, et al. Multimodal representation learning for predicting molecule-disease relations. Bioinformatics. 2023;39(2). doi:10.1093/bioinformatics/btad085
25. 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
26. Wallach I, Heifets A. Most ligand-based classification benchmarks reward memorization rather than generalization. J Chem Inf Model. 2018;58(5):916-32. doi:10.1021/acs.jcim.7b00403
27. 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
28. 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
29. Jimenez-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
30. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
31. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
32. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health. 2020;2(9). doi:10.1016/S2589-7500(20)30186-2
33. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. 2018;15(11). doi:10.1371/journal.pmed.1002683
34. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. Science. 2019;363(6433):1287-9. doi:10.1126/science.aaw4399