



Predicting Polymorph Risk Before Formulation Locks in a Natural Product Candidate Through Crystal Landscape Reasoning and Solid-State Uncertainty Assessment

David Thompson^{1*}, Sarah Mitchell¹, Rachel Adams², James Brown³, Michael Lee⁴

¹Department of Polymorph Risk Prediction, Faculty of Pharmacy, University of Toronto, Toronto, Canada.

²Department of Crystal Landscape Reasoning, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, Canada.

³Department of Solid-State Uncertainty Assessment, Faculty of Pharmacy, McGill University, Montreal, Canada.

⁴Department of Formulation Lock-In Prevention, Faculty of Pharmacy, University of Guelph, Guelph, Canada.

ABSTRACT

Polymorphism creates a development problem before it becomes a formulation problem. A natural product candidate may appear to have an acceptable solid form while its accessible crystal landscape remains incompletely sampled, its energetic ordering uncertain, or its crystallization behavior sensitive to solvent, temperature, and molecular conformation. This article develops a proposed decision framework for assessing polymorph risk before formulation choices become difficult to reverse. The analysis separates four questions that are often conflated: which crystal structures are plausible, which are thermodynamically competitive under relevant conditions, which are kinetically accessible, and which alternative forms would materially affect development. Recent advances in crystal structure prediction, machine-learned intermolecular potentials, finite-temperature free-energy treatment, and orthogonal solid-state characterization strengthen early risk assessment, but none eliminates uncertainty in landscape completeness or experimental accessibility. The framework therefore treats polymorph risk as an evidence-integration problem rather than a request for a single predicted structure or energy ranking. Molecular flexibility, packing compensation, solvent-dependent stability, finite-temperature landscape reduction, and late or elusive polymorph discovery are interpreted as distinct uncertainty sources that should alter the intensity of experimental escalation. Natural-product examples further show why complex intermolecular recognition cannot be reduced to simple flexibility or hydrogen-bond counts. The proposed framework is nonempirical and requires prospective validation across chemically diverse candidates. It is intended to support earlier, explicitly qualified solid-state decisions, not to replace experimental screening, establish universal risk thresholds, or predict formulation performance with validated accuracy.

Key Words: Polymorphism, Crystal structure prediction, Natural products, Solid-state uncertainty, Crystal energy landscape, Preformulation risk

eJPPR 2026; 16(4):30-39

HOW TO CITE THIS ARTICLE: Thompson D, Mitchell S, Adams R, Brown J, Lee M. Predicting Polymorph Risk Before Formulation Locks in a Natural Product Candidate Through Crystal Landscape Reasoning and Solid-State Uncertainty Assessment. Int J Pharm Phytopharmacol Res. 2026;16(4):30-9. <https://doi.org/10.51847/PhJ4S82s8>

INTRODUCTION

Solid-form selection is often treated as a downstream characterization task, yet the practical consequence of polymorphism can be determined much earlier. Once a

candidate has accumulated process knowledge, formulation assumptions, analytical methods, and stability expectations around one crystal form, discovery of a competing form can force expensive reinterpretation. Pharmaceutical solid-form selection is therefore not

Corresponding author: David Thompson

Address: Department of Polymorph Risk Prediction, Faculty of Pharmacy, University of Toronto, Toronto, Canada.

E-mail: ✉ david.thompson@utoronto.ca

Received: 03 April 2026; **Revised:** 16 July 2026; **Accepted:** 17 July 2026

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



merely the enumeration of forms but a conditional choice among structures that can differ in stability, accessibility, and development relevance [1]. For natural products, conformational complexity, dense hydrogen-bonding possibilities, stereochemical richness, or limited material availability may make this question especially consequential when such features are present; none, however, should be presumed to confer high polymorph risk in every molecule. The relevant question is whether the available solid-state evidence is strong enough to justify committing development to the currently observed form.

Earlier computational involvement can widen that evidence base. Digital pharmaceutical design increasingly links molecular structure, predicted crystal forms, and crystallization-process reasoning rather than treating them as isolated stages [2]. The value of such integration is illustrated by ritonavir: retrospective crystal structure prediction exposed a landscape containing risk-relevant alternatives, while experimental work showed that access to the historically problematic Form I remained strongly dependent on crystallization solvent and conditions [3, 4]. These observations support a distinction between latent polymorph possibility and practical polymorph accessibility. Neither a predicted low-energy structure nor a negative experimental screen, considered alone, resolves both questions.

This article therefore proposes an uncertainty-aware decision framework for polymorph risk before formulation commitment. The framework is not a classifier with validated probabilities, and it does not assume that all predicted structures are realizable. Instead, it asks whether computational landscape multiplicity, energetic uncertainty, molecular packing behavior, experimental coverage, and kinetic accessibility converge sufficiently to justify either provisional commitment or further escalation. The central premise is that early solid-state decisions should be based on the strength and agreement of heterogeneous evidence, with unresolved disagreement treated as information rather than averaged away. Risk is not synonymous with the predicted existence of another form. It becomes consequential when credible alternatives combine with uncertain stability or accessibility and could disrupt manufacturability, dissolution, or formulation assumptions. The proposed framework therefore separates evidence for solid-form multiplicity from evidence for development consequence, preventing a computationally crowded landscape from being interpreted automatically as high risk.

Crystal energy landscapes and solid-form multiplicity

A crystal energy landscape is most useful in development when interpreted as a structured set of plausible packings

with uncertain energetic and physical relationships, not as a definitive catalogue of polymorphs. Stability under realistic conditions cannot always be inferred from static lattice energies alone; finite-temperature and environmental contributions may alter the relative free energies of competing crystal forms [5]. Consequently, the lowest calculated minimum at 0 K is evidence about one thermodynamic approximation, not automatic evidence of the form that will dominate storage, processing, or crystallization.

The same landscape can also encode kinetically disconnected outcomes. Molecular simulations increasingly describe crystalline states through free-energy basins, transitions, disorder, and pathway-dependent behavior, emphasizing that thermodynamic preference and kinetic access are different questions [6]. For a preformulation decision, this means that a dense cluster of low-energy structures can represent genuine unresolved alternatives, artifacts of the modeling level, thermally equivalent states, or forms separated by barriers that are irrelevant under available processing conditions.

Benchmark construction makes this ambiguity explicit. The CPOSS209 set combines experimental and hypothetical polymorphic structures to test how computational models distinguish physically realized forms from plausible but unobserved alternatives [7]. Such work cautions against using landscape population itself as a risk score. In the proposed framework, landscape multiplicity is therefore an uncertainty signal whose weight depends on structural distinctness, energy-model confidence, finite-temperature persistence, and experimental correspondence. A sparse landscape can still conceal an important missed form, while a crowded landscape need not imply many developmentally accessible polymorphs. Before lock-in, experimental observability remains conditional on solvent, temperature, seeding history, and analytical sensitivity. The landscape should therefore be read as a hypothesis space whose members gain development significance when thermodynamic credibility, kinetic plausibility, and experimental evidence align. Disagreement among these layers should preserve uncertainty and redirect experiments rather than be collapsed into one rank.

Molecular flexibility, packing, and interaction motifs

Molecular flexibility influences polymorphism because crystal formation couples intramolecular conformation to intermolecular packing. Simulations of closely related hepatitis C virus drug analogues show that differences in accessible three-dimensional conformations and crystal interactions can track markedly different solid-state development behavior [8]. The implication is not that flexibility alone predicts polymorphism, but that

conformational populations define which recognition patterns can participate in nucleation, packing, and lattice stabilization.

Natural-product-related solids provide useful caution against oversimplifying this relation. Xanthone, despite a comparatively rigid aromatic framework, exhibits experimentally distinguishable polymorphism and condition-dependent crystallization behavior [9]. Conversely, flexible molecules may accommodate higher-energy conformations in the crystal when favorable intermolecular contacts compensate the intramolecular distortion. Large-scale lattice-energy partitioning demonstrates this energetic trade-off across flexible molecular crystals [10]. Thus, counting rotatable bonds, hydrogen-bond donors, or acceptors cannot by itself establish polymorph risk.

For decision purposes, the important variable is the degree of packing competition generated by the molecule: whether several conformers and interaction motifs can produce comparably credible crystal environments, and whether small changes in conditions could favor different solutions. This is a proposed interpretive construct rather than an established scalar descriptor. It directs attention to conformational degeneracy, interaction-motif alternatives, intramolecular versus intermolecular energetic compensation, and stereochemical constraints. In natural products, where multiple functional groups may participate in competing hydrogen-bond networks or conformationally coupled contacts, these questions justify targeted analysis; they do not justify presuming that chemically complex structures necessarily possess many polymorphs. Interaction motifs likewise are not immutable molecular labels: a donor–acceptor pair dominant in one packing arrangement may be frustrated, internally satisfied, or replaced in another. The decision-relevant question is whether alternative packings remain competitive after conformational cost and lattice stabilization are considered together. Screening should therefore emphasize *competing structurally coherent solutions* rather than simple motif counts.

Computational crystal structure prediction

Computational crystal structure prediction provides the broadest available means of asking whether an experimentally observed solid form sits within a larger set of plausible alternatives. Recent large-scale validation across drug-like molecules, including blind-study components, demonstrates that contemporary CSP workflows can recover experimentally known structures and meaningfully rank candidate packings across substantial chemical space [11]. This supports the use of CSP as an early evidence source, but not as proof that a

generated landscape is exhaustive or that every low-energy minimum is experimentally accessible.

Machine-learned intermolecular potentials are changing the practical scale of this search. AIMNet2-based workflows show that neural-network potentials can accelerate crystal-structure prediction and stability assessment while retaining useful agreement with higher-level reference calculations within tested domains [12]. For polymorph-risk assessment, the consequence is important: more candidates can be generated, optimized, and compared before formulation decisions are fixed. Yet increased throughput also increases the importance of model applicability, structural redundancy, and uncertainty in energetic ranking. A larger predicted landscape is not automatically a more accurate description of the physically accessible one.

The remaining methodological frontier is concentrated in exactly the molecular features that often complicate natural-product development. Current CSP scholarship identifies conformational flexibility, complex intermolecular interactions, larger search spaces, and demanding energetic discrimination among the continuing challenges for pharmaceutical crystal prediction [13]. Accordingly, the proposed framework treats CSP output as graded evidence. Confidence rises when independent search strategies, accurate energetic treatments, finite-temperature refinement, and experimental structures converge; it falls when results depend strongly on model choice or when many near-degenerate structures remain unresolved. CSP should therefore narrow and organize experimental questions rather than terminate them. Candidate generation, energy ranking, duplicate reduction, finite-temperature refinement, and experimental comparison should remain distinguishable because each contributes a different uncertainty signal. An uncertain search should not be disguised as precise ranking, nor should precise ranking of an incomplete search be mistaken for landscape closure.

Experimental evidence and incomplete landscape sampling

Experimental screening converts the hypothetical landscape into observed solid-state evidence, but observation is not equivalent to landscape closure. Even extensively studied pharmaceuticals can yield previously unrecognized forms when crystallization conditions or analytical capabilities change. A newly reported κ -form of indomethacin, CSP-informed isolation of elusive meloxicam polymorphs, and the computationally anticipated late-appearing form of rotigotine collectively show that apparently mature solid-form knowledge can remain incomplete [14–16]. These cases do not establish how frequently late forms occur; they establish the

narrower point that historical non-observation is insufficient evidence of nonexistence.

This distinction matters most when negative screening is used to justify formulation commitment. A broad screen can substantially reduce uncertainty, yet its evidential meaning depends on the solvents, supersaturation histories, temperatures, seeding conditions, crystallization modes, sample amounts, and analytical methods actually explored. CSP can identify structurally credible regions that deserve targeted experiments, whereas advanced diffraction or spectroscopic methods can resolve forms that routine workflows may not characterize cleanly. The proposed interpretation is therefore asymmetric: positive experimental identification establishes that a form is realizable under at least one set of conditions, while failure

to observe a predicted form only constrains accessibility within the investigated domain.

For natural-product candidates, this asymmetry is especially useful when material is scarce or chemically complex. Rather than demanding exhaustive experimentation, the early decision can ask whether independent evidence has converged sufficiently to make an unobserved competitor unlikely to alter the development choice. Where low-energy alternatives remain structurally distinct, experimentally unexplored, or energetically uncertain, incompleteness should remain visible.

The mechanism–evidence alignment emerging from the manuscript to this point is summarized in **Table 1**.

Table 1. Mechanism–evidence alignment for interpreting an incompletely sampled solid-form landscape before formulation commitment

Manuscript-derived issue	What is directly evidenced	What remains unresolved	Decision interpretation
Landscape multiplicity	Multiple plausible or hypothetical packings can be generated and compared	Which minima correspond to experimentally distinct forms	Treat multiplicity as uncertainty, not a form count
Conformational competition	Molecular conformation and lattice interactions can compensate one another	Whether competing conformers are populated during nucleation	Escalate when several coherent packing solutions remain credible
Experimental non-observation	Screening constrains forms accessible under tested conditions	Landscape completeness outside the sampled condition space	Do not interpret absence as proof of nonexistence
Late or elusive forms	Previously unknown forms can emerge after targeted computation or advanced characterization	Frequency and timing across new candidates	Increase caution when consequences of a missed form are large
Cross-method convergence	Computation and experiment can mutually constrain structural hypotheses	Residual model and sampling error	Confidence increases through agreement of independent evidence

Note: Entries marked by their wording as decision interpretations are proposed syntheses of this article; the empirical statements are derived from the evidence discussed in the preceding sections.

Figure 1 translates the same evidence into a visual distinction between what the landscape suggests, what experiments have sampled, and what remains uncertain.

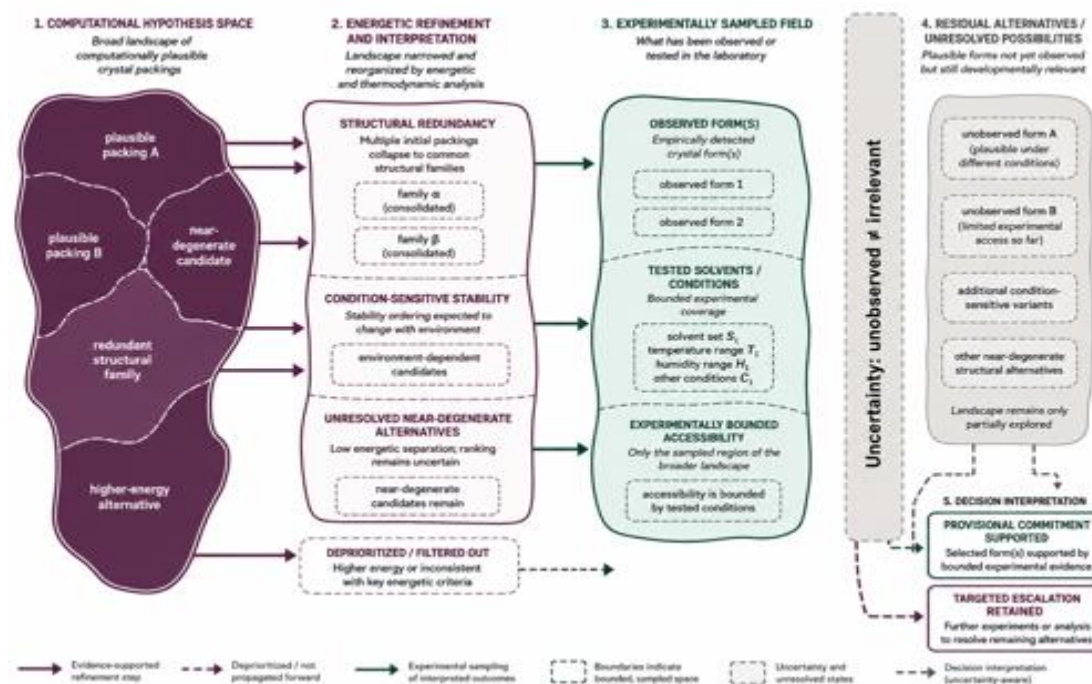


Figure 1. From a predicted crystal landscape to experimentally bounded polymorph evidence

Uncertainty in relative stability and kinetic accessibility

Energetic uncertainty remains after structures have been generated and observed. Relative stability can depend on solvent environment rather than being an invariant property of two isolated crystal lattices [17]. Finite-temperature calculations likewise show why a 0 K ranking should not be treated as a development verdict: vibrational and thermal contributions can alter free-energy differences among polymorphs [18]. In addition, finite-temperature molecular dynamics can collapse several static CSP minima into fewer dynamically equivalent states, reducing apparent landscape complexity [19].

These observations imply two separate uncertainties. The first is thermodynamic ranking uncertainty: whether the available model resolves the relevant free-energy ordering under the intended temperature, pressure, and environment. The second is kinetic accessibility uncertainty: whether a competitive form can nucleate, grow, transform, or persist on development timescales. The proposed framework keeps them orthogonal because a highly stable form may be difficult to access, while a metastable form may be readily produced and kinetically persistent. Their interaction also matters: a form whose calculated stability changes with environment but that is repeatedly obtained experimentally represents a different uncertainty profile from a consistently low-energy form that has never been isolated. The first calls for condition-specific thermodynamic clarification; the second calls for testing accessibility and the adequacy of the search space.

Accordingly, a narrow lattice-energy separation should trigger interpretation rather than automatic classification. Confidence should be lower when ranking changes with computational level, solvent treatment, or temperature, and higher when independent calculations and experiments converge. No universal energy window is proposed. The decision variable is the robustness of the conclusion to reasonable changes in physical treatment, not a fixed numerical cutoff.

Risk classification before formulation commitment

Early classification should therefore represent decision uncertainty, not merely the number of predicted polymorphs. Artificial-intelligence approaches are increasingly used to assist pharmaceutical crystallization, but their utility remains dependent on data quality, process representation, and the physical meaning of the target being predicted [20]. A difficult AZD1305 polymorph-screening case further shows that near-degenerate energetic conclusions can depend on the level of thermodynamic treatment [21]. Rapid descriptor-based crystal prediction provides another computational route, illustrating that different representations can change how efficiently candidate structures are explored [22].

The framework consequently proposes three qualitative decision states: provisional convergence, material unresolved competition, and insufficiently bounded evidence. These are not validated classes. Provisional convergence means that experimentally observed behavior and sufficiently robust computational evidence point

toward the same development-relevant form without a credible unresolved competitor. Material unresolved competition means that at least one structurally distinct alternative remains plausible and could matter if realized. Insufficiently bounded evidence means that sampling, model applicability, or characterization is too incomplete to distinguish the first two states. Classification should also be consequence aware. A credible alternative that would not change handling, stability, dissolution assumptions, or manufacturability need not trigger the same response as an alternative capable

of invalidating the selected form's development rationale. This does not require predicting a formulation outcome before formulation studies exist. It requires identifying which solid-state differences would be decision-relevant if the alternative appeared, and then matching experimental intensity to that uncertainty.

The practical boundary conditions for those states are summarized in **Table 2**. They deliberately avoid numerical thresholds because the approved evidence does not establish universal cutoffs.

Table 2. Boundary conditions for proposed preformulation polymorph-risk states

Proposed state	Evidence pattern	Boundary condition that weakens assignment	Appropriate response
Provisional convergence	Cross-method agreement; no consequential competitor supported	Sparse experimental coverage or model-domain concern	Commit provisionally while preserving surveillance
Material unresolved competition	Distinct plausible forms with uncertain relative stability or accessibility	Alternative collapses under better energetics or is experimentally excluded within relevant conditions	Target the unresolved mechanism or condition
Insufficiently bounded evidence	Search, characterization, or model applicability remains materially incomplete	New orthogonal evidence reduces the dominant uncertainty	Escalate before formulation dependence increases
Any state	Development consequence is considered separately from form existence	Alternative form has negligible practical consequence	Avoid inflating structural multiplicity into development risk

Note: The state names and response logic are proposed in this article. They are qualitative decision constructs, not validated risk categories or regulatory classifications.

The polymorph-risk decision framework

The proposed framework converts these distinctions into an evidence-sensitive decision sequence. High-accuracy electronic-structure calculations can improve relative polymorph energetics [23], while independent first-principles approaches provide another route for testing whether a ranking survives a change in computational treatment [24]. Structural clustering is also necessary because raw CSP landscapes may overpredict distinct structures when closely related minima are counted separately [25]. Machine-learned interatomic potentials can make finite-temperature refinement more tractable, provided their applicability and reference quality are adequate [26].

The framework therefore asks four linked questions. First, is there credible multiplicity? This concerns whether structurally distinct alternatives survive duplicate reduction and reasonable energetic treatment. Second, is the stability conclusion robust? This tests sensitivity to temperature, environment, and model choice. Third, is an alternative accessible? This uses experimental history and kinetic evidence rather than assuming equilibrium selection. Fourth, would the alternative matter? This separates solid-form existence from consequences for dissolution, stability, processing, or formulation assumptions.

No single question dominates by definition. A crowded predicted landscape with strong evidence that alternatives are redundant may support provisional convergence. Conversely, one persistent low-energy alternative with uncertain accessibility and large potential consequence may justify escalation. Evidence is also allowed to change direction: a candidate can move from escalation toward provisional commitment after a predicted alternative collapses under better energetics, or away from commitment after a new form becomes experimentally accessible. The framework is therefore iterative rather than a one-time gate.

Operationally, each escalation should be tied to the uncertainty that generated it. Energetic disagreement calls for improved or orthogonal calculations; unresolved structural multiplicity calls for duplicate reduction or targeted characterization; uncertain accessibility calls for altered crystallization histories; and uncertain consequence calls for comparative solid-state or preformulation testing. This preserves a direct link between evidence deficiency and next action. The framework's distinctive proposal is thus to base action on the *pattern of unresolved uncertainty* rather than on a universal energy threshold or polymorph count.

Figure 2 presents this decision logic as a mirrored comparison between evidence supporting provisional commitment and evidence supporting escalation.

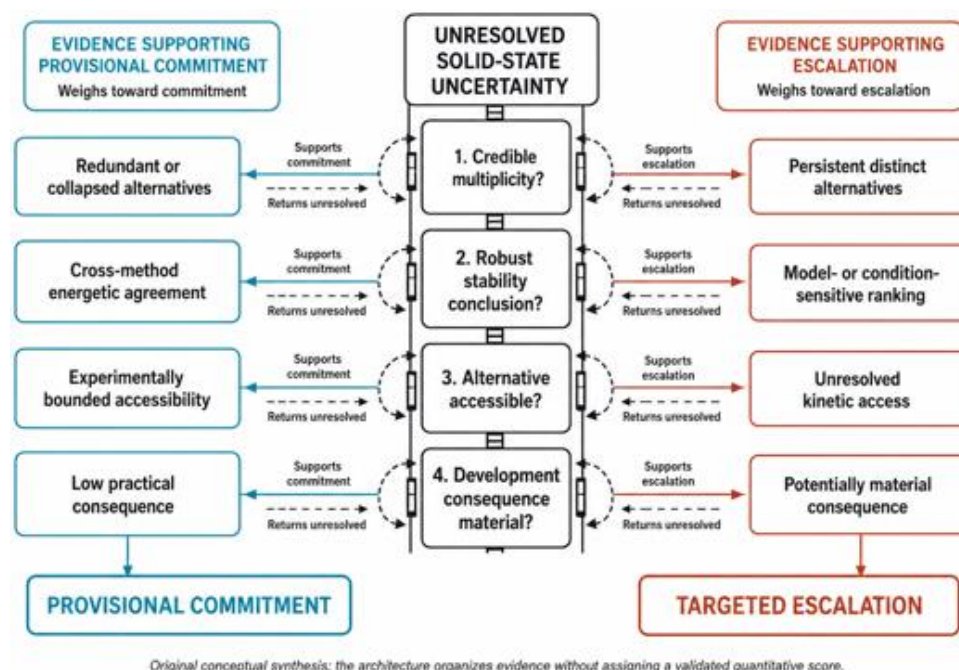


Figure 2. A hinged decision architecture for provisional commitment versus polymorph-risk escalation

Decision scenarios for competing solid forms

The value of the framework is clearest in scenarios where the same number of predicted forms implies different actions. In axitinib, advanced computation has been used to rationalize a challenging family of polymorphs, cocrystals, and salts [27]. In celecoxib, a machine-learned potential enables efficient exploration of a flexible pharmaceutical polymorph landscape [28]. Neither case warrants a universal rule; together they show why computational tractability and structural multiplicity must be interpreted at the compound level. In a candidate with numerous predicted structures but stable ordering across methods, the key question may be experimental accessibility. In a candidate with few structures but unstable ranking, thermodynamic uncertainty may dominate instead.

Natural-product cases sharpen the distinction between structural plausibility and simple molecular descriptors. Pterostilbene demonstrates that an unexpectedly stable high-*Z'* packing arrangement can emerge from a balance of unfavorable packing and favorable hydrogen-bond polarization [29]. Curcumin provides a complementary example in which elusive polymorphs required combined structural characterization and computational analysis [30]. These examples argue against treating natural-product risk as a direct function of flexibility, hydrogen-bond count, or structural complexity.

Across axitinib, celecoxib, and curcumin, computational landscapes become developmentally interpretable only when coupled to compound-specific structural or experimental evidence [27, 28, 30]. A practical scenario distinction follows. If competing predictions collapse

under higher-level energetics or finite-temperature analysis, escalation can focus on confirming the surviving form. If alternatives remain distinct but are not experimentally accessible under relevant conditions, the uncertainty becomes kinetic. If a competitor is both credible and accessible, development consequence becomes decisive. A poorly soluble, unstable, or process-disruptive competitor should command more attention than a structurally distinct alternative with negligible practical effect. This consequence-sensitive step is proposed here; it prevents “polymorphism” from functioning as an undifferentiated hazard label. A fourth scenario is also important: computation and experiment may disagree without either being demonstrably wrong. For example, a persistent predicted form may remain unobserved despite broad screening, or an observed form may rank unexpectedly high in a model. The appropriate response is not forced reconciliation. Such discordance should be retained as a specific uncertainty, prompting examination of model physics, missing environmental variables, kinetic trapping, analytical assignment, or unexplored crystallization pathways.

Boundary conditions and escalation to experiment

Equilibrium reasoning becomes insufficient when crystal growth itself is conformation dependent. Conformational self-poisoning demonstrates that alternative molecular conformers can interfere with productive crystal growth, creating kinetic behavior not captured by stability ranking alone [31]. Such evidence defines an escalation trigger: when a predicted alternative is credible but experimental access remains inconsistent, experiments should probe

kinetic mechanisms rather than merely repeat equilibrium calculations. Escalation can include controlled seeding, solvent and supersaturation variation, temperature cycling, slurry conversion, or growth-focused experiments when those manipulations address the hypothesized barrier; the framework does not assume that any one protocol is universally required.

Integrated computational and experimental screening provides a practical model for such escalation, although available demonstrations remain compound specific [32]. The escalation should target the dominant unresolved uncertainty: broaden crystallization conditions for accessibility, refine free energies for ranking instability, challenge model applicability for computation-sensitive landscapes, or deliberately search for predicted motifs when screening coverage is weak.

Structural confirmation is a separate boundary. Pharmaceutical solids that are difficult to grow as conventional single crystals may require combinations of powder diffraction, electron diffraction, spectroscopy, CSP, or other orthogonal methods for defensible structure assignment [33]. Escalation should therefore end not when more data have been collected, but when the evidence relevant to the decision has become sufficiently discriminating. This is deliberately narrower than claiming landscape completeness. A formulation commitment can be rational despite residual unknowns if those unknowns have been bounded and their plausible consequences are acceptable. Conversely, repeated screening without a stated uncertainty target can create more observations without materially improving the decision. The escalation principle is therefore discriminative rather than exhaustive: add the experiment or calculation most capable of separating the competing explanations that currently drive the decision.

Limitations

This framework is conceptual and has not been prospectively calibrated as a predictive model. Its computational evidence layers inherit the limitations of the methods used to estimate polymorph stability. Comparisons of statistical-mechanical approximations show that free-energy calculations involve system-dependent trade-offs between computational cost and accuracy [34]; therefore, no universal tolerance for an acceptable energy difference is justified.

Machine-learned potentials introduce an additional applicability boundary. Transferable organic force fields such as MACE-OFF can achieve broad performance within their intended chemical and interaction domains, but transferability does not remove assumptions concerning reference data, interaction range, charge states, or chemistry outside the validated domain [35]. Natural

products may intensify these concerns when they contain unusual functional groups, high conformational complexity, or interaction motifs underrepresented in training data. The framework also assumes that the chosen analytical methods distinguish forms reliably; misassignment, disorder, solvation, or unresolved mixtures could make apparent convergence misleading.

The evidence base also contains prominent pharmaceutical case studies whose lessons may not transfer uniformly to natural products. Late-form discovery establishes possibility, not frequency. Finally, the proposed states and escalation logic lack prospective outcome data showing that they reduce formulation failure, development cost, or missed polymorphs. Their value should therefore be tested against blinded candidate series before predictive performance is claimed.

CONCLUSION

Polymorph risk before formulation lock-in is best treated as uncertainty about a decision, not as a count of predicted crystal structures. The proposed framework separates credible solid-form multiplicity, robustness of relative stability, kinetic accessibility, and development consequence so that disagreement among these dimensions remains visible. This separation prevents a low calculated energy from being equated with inevitable appearance, a negative screen from being interpreted as landscape closure, or structural multiplicity from being treated automatically as pharmaceutical risk.

For natural-product candidates, the framework provides a disciplined reason to integrate conformational analysis, packing competition, CSP, finite-temperature energetics, crystallization evidence, and orthogonal characterization without assuming that chemical complexity itself predicts polymorphism. Its practical output is deliberately qualitative: provisional commitment when evidence converges, targeted escalation when a material uncertainty remains, and withholding of commitment when evidence is insufficiently bounded. The framework does not replace experimental solid-form screening and does not establish universal thresholds. Its future value depends on prospective testing that records both correct reassurance and missed-risk cases. Its central contribution is an evidence architecture for deciding what remains uncertain, why that uncertainty matters, and which experiment or calculation should address it before downstream formulation choices become difficult to reverse.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Chappa P, Jenjeti RN, Kljun J, Mittapalli S, Pichler A, Desiraju GR. Pharmaceutical solid form screening and selection: Which form emerges? *Cryst Growth Des.* 2025;25(13):4783-94. doi:10.1021/acs.cgd.5c00252
- [2] Burcham CL, Doherty MF, Peters BG, Price SL, Salvalaglio M, Reutzel-Edens SM, et al. Pharmaceutical digital design: From chemical structure through crystal polymorph to conceptual crystallization process. *Cryst Growth Des.* 2024;24(13):5417-38. doi:10.1021/acs.cgd.3c01390
- [3] Iuzzolino L, Kelly AW, Chaudhry MT, Jandl C, Stam D, Lee AY. Predicting the ritonavir crisis by revisiting the polymorph landscape with crystal structure prediction and form 4 structure solution. *Commun Chem.* 2025;8(1):404. doi:10.1038/s42004-025-01814-6
- [4] Wang C, Ma CY, Hong RS, Turner TD, Rosbottom I, Sheikh AY, et al. Influence of solvent selection on the crystallizability and polymorphic selectivity associated with the formation of the “disappeared” Form I polymorph of ritonavir. *Mol Pharm.* 2024;21(7):3525-39. doi:10.1021/acs.molpharmaceut.4c00234
- [5] Firaha D, Liu YM, van de Streek J, Sasikumar K, Dietrich H, Helfferich J, et al. Predicting crystal form stability under real-world conditions. *Nature.* 2023;623(7986):324-8. doi:10.1038/s41586-023-06587-3
- [6] Desgranges C, Delhommelle J. Deciphering the complexities of crystalline state(s) with molecular simulations. *Commun Chem.* 2025;8(1):281. doi:10.1038/s42004-025-01667-z
- [7] Price LS, Paloni M, Salvalaglio M, Price SL. One size fits all? Development of the CPOSS209 data set of experimental and hypothetical polymorphs for testing computational modeling methods. *Cryst Growth Des.* 2025;25(9):3186-209. doi:10.1021/acs.cgd.5c00255
- [8] Hong RS, Mattei A, Tuckerman ME, Sheikh AY. 3D conformation and crystal interaction insights into drug development challenges for HCV drug analogues via molecular simulations. *Commun Chem.* 2025;8(1):229. doi:10.1038/s42004-025-01618-8
- [9] Preston JA, Parisi E, Murray BS, Tyler AII, Simone E. Elucidating the polymorphism of xanthone: A crystallization and characterization study. *Cryst Growth Des.* 2024;24(8):3256-68. doi:10.1021/acs.cgd.3c01506
- [10] Chattopadhyay A, Hill AR, Wright SE, Beran GJO, Cruz-Cabeza AJ. Lattice energy partitions in crystals of flexible molecules and the 40% limit. *J Am Chem Soc.* 2025;147(43):39192-203. doi:10.1021/jacs.5c09997
- [11] Zhou D, Bier I, Santra B, Jacobson LD, Wu C, Garaizar Suarez A, et al. A robust crystal structure prediction method to support small molecule drug development with large scale validation and blind study. *Nat Commun.* 2025;16(1):2210. doi:10.1038/s41467-025-57479-1
- [12] Nayal KS, O'Connor D, Zubatyuk R, Anstine DM, Yang Y, Tom R, et al. Efficient molecular crystal structure prediction and stability assessment with AIMNet2 neural network potentials. *Cryst Growth Des.* 2025;25(21):9092-106. doi:10.1021/acs.cgd.5c01001
- [13] Beran GJO. Frontiers of molecular crystal structure prediction for pharmaceuticals and functional organic materials. *Chem Sci.* 2023;14(46):13290-312. doi:10.1039/D3SC03903J
- [14] Shimazaki Y, Ueda H, Kikuchi T, Nishiyama Y, Fukami T. Discovery of a new polymorph, κ -form of indomethacin. *Cryst Growth Des.* 2026;26(11):4113-23. doi:10.1021/acs.cgd.5c01534
- [15] Jeziorna A, Paluch P, Zajac J, Dolot R, Dudek MK. Crystallization of elusive polymorphs of meloxicam informed by crystal structure prediction. *Cryst Growth Des.* 2023;23(8):5998-6010. doi:10.1021/acs.cgd.3c00559
- [16] Mortazavi M, Hoja J, Aerts L, Quéré L, van de Streek J, Neumann MA, et al. Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine. *Commun Chem.* 2019;2(1):70. doi:10.1038/s42004-019-0171-y
- [17] Yerragunta M, Veliz AC, Pulluri R, Del Campo M, Guan X, Le Magueres P, et al. Solvent-dependent relative stability of crystal forms. *Cryst Growth Des.* 2026;26(9):3233-49. doi:10.1021/acs.cgd.5c01461
- [18] Yang M, Dybeck EC, Sun G, Peng C, Samas B, Burger VM, et al. Prediction of the relative free energies of drug polymorphs above zero Kelvin. *Cryst Growth Des.* 2020;20(8):5211-24. doi:10.1021/acs.cgd.0c00422
- [19] Francia NF, Price LS, Nyman J, Price SL, Salvalaglio M. Systematic finite-temperature reduction of crystal energy landscapes. *Cryst Growth Des.* 2020;20(10):6847-62. doi:10.1021/acs.cgd.0c00918
- [20] Zhu Z, Zhang Y, Wang Z, Tang W, Wang J, Gong J. Artificial intelligence assisted pharmaceutical crystallization. *Cryst Growth Des.* 2024;24(10):4245-70. doi:10.1021/acs.cgd.3c01408

- [21] Sun G, Liu X, Abramov YA, Nilsson Lill SO, Chang C, Burger VM, et al. Current state-of-the-art in-house and cloud-based applications of virtual polymorph screening of pharmaceutical compounds: A challenging case of AZD1305. *Cryst Growth Des.* 2021;21(4):1972-83. doi:10.1021/acs.cgd.0c01266
- [22] Galanakis N, Tuckerman ME. Rapid prediction of molecular crystal structures using simple topological and physical descriptors. *Nat Commun.* 2024;15(1):9757. doi:10.1038/s41467-024-53596-5
- [23] Hoja J, Ko HY, Neumann MA, Car R, DiStasio RA Jr, Tkatchenko A. Reliable and practical computational description of molecular crystal polymorphs. *Sci Adv.* 2019;5(1):eaau3338. doi:10.1126/sciadv.aau3338
- [24] Nikhar R, Szalewicz K. Reliable crystal structure predictions from first principles. *Nat Commun.* 2022;13(1):3095. doi:10.1038/s41467-022-30692-y
- [25] Butler PWV, Day GM. Reducing overprediction of molecular crystal structures via threshold clustering. *Proc Natl Acad Sci U S A.* 2023;120(23):e2300516120. doi:10.1073/pnas.2300516120
- [26] Della Pia F, Shi BX, Kapil V, Zen A, Alfè D, Michaelides A. Accurate and efficient machine learning interatomic potentials for finite temperature modelling of molecular crystals. *Chem Sci.* 2025;16(25):11419-33. doi:10.1039/D5SC01325A
- [27] Beran GJO. From polymorphs to cocrystals and salts: Successfully predicting axitinib's challenging crystal forms. *Chem Sci.* 2025;16(48):23026-37. doi:10.1039/D5SC06271C
- [28] Zheng P, Abramov YA, Sun CC, Isayev O. Exploring celecoxib polymorph landscape using AIMNet2 machine learning interatomic potential. *Chem Sci.* 2026;17(30):14691-9. doi:10.1039/D5SC09784C
- [29] Bofill L, de Sande D, Barbas R, Prohens R. Hydrogen bond polarization overcomes unfavorable packing in the most stable high Z' polymorph of pterostilbene. *Cryst Growth Des.* 2019;19(5):2552-6. doi:10.1021/acs.cgd.9b00374
- [30] Matlinska MA, Wasylishen RE, Bernard GM, Terskikh VV, Brinkmann A, Michaelis VK. Capturing elusive polymorphs of curcumin: A structural characterization and computational study. *Cryst Growth Des.* 2018;18(9):5556-63. doi:10.1021/acs.cgd.8b00859
- [31] Liu Y, Marinova V, Davey RJ, Gabriele B, Salvalaglio M, Cruz-Cabeza AJ. Conformational self-poisoning in crystal growth. *JACS Au.* 2025;5(4):1781-90. doi:10.1021/jacsau.5c00043
- [32] Muthusamy AR, Chauhan D, Jain AK, Singh A. Navigating solid-form screening using in silico methods validated with experimental data for a drug-like molecule. *Cryst Growth Des.* 2025;25(9):2970-85. doi:10.1021/acs.cgd.4c01740
- [33] Newman JA, Iuzzolino L, Tan M, Orth P, Bruhn J, Lee AY. From powders to single crystals: A crystallographer's toolbox for small-molecule structure determination. *Mol Pharm.* 2022;19(7):2133-41. doi:10.1021/acs.molpharmaceut.2c00020
- [34] Abraham NS, Shirts MR. Statistical mechanical approximations to more efficiently determine polymorph free energy differences for small organic molecules. *J Chem Theory Comput.* 2020;16(10):6503-12. doi:10.1021/acs.jctc.0c00570
- [35] Kovács DP, Moore JH, Browning NJ, Batatia I, Horton JT, Pu Y, et al. MACE-OFF: Short-range transferable machine learning force fields for organic molecules. *J Am Chem Soc.* 2025;147(21):17598-611. doi:10.1021/jacs.4c07099