Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Solid-state evidence constrains thermodynamic interpretation of public solubility data

**Minhao Liu**

minhliu@ttu.edu

Texas Tech University

**Pu Sun**

San Diego State University

**Research Article**

**Keywords:** Solubility, Solid-state thermodynamics, Polymorphism, Crystallization, Data provenance, Chemical informatics

**Posted Date:** November 20th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11534350/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Solid-state evidence constrains thermodynamic interpretation of public solubility data

**Minhao Liu**[1*], **Pu Sun**[2]

[1] Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas, USA

[2] Department of Electrical and Computer Engineering, San Diego State University, San Diego, California, USA

*Corresponding author: Minhao Liu; email: minhliu@ttu.edu

Pu Sun: psun9932@sdsu.edu

## Abstract

Public solubility records are increasingly reused to rank solvents and infer solid-form stability, although these decisions require different evidence. We combine solid-state thermodynamics with a retrospective audit of public data constructions and published phase measurements. Equilibrium activity ratios constrain the sign of a polymorph solubility difference and its transition temperature; concentration-ratio magnitudes require additional nonideality assumptions. BigSolDB 2.2 contains 14,813 solid-labelled measurements, but exact same-source multi-form matching yields only seven solute systems. The three systems supporting same-temperature cross-solvent comparisons show no sign reversals across 183 matched form pairs. In sulfamerazine, four minimum-concentration phase assignments disagree with terminal slurry powder X-ray diffraction (PXRD) across eight numerical representations. Tolbutamide provides a second assay-specific comparison in which declared equilibration does not resolve sampling limitations. Replaying SC3 identifies 87 of 4,172 temperature intervals whose label responses oppose continuing-source responses, while structural evidence resolves molecular attribution affecting 576 records. Conversely, all 15 eligible external solvent panels retain their highest-concentration candidate. These results establish concrete limits on thermodynamic reuse without estimating a database-wide failure rate. The audit links each retained comparison to its molecular identity, source support, uncertainty and measured endpoint.

Keywords: Solubility; Solid-state thermodynamics; Polymorphism; Crystallization; Data provenance; Chemical informatics

## Introduction

Solubility measurements connect solution chemistry to the free energy of a solid. They guide solvent selection, formulation and crystallization, and increasingly supply training labels for data-driven concentration prediction [1, 2]. However, a concentration measured during dissolution and a concentration in equilibrium with a specified crystalline phase need not answer the same chemical question. Thermodynamic interpretation requires both a well-defined solution quantity and a specified material state [3]. A useful solvent ordering therefore does not automatically establish which polymorph is stable.

Public resources make this distinction consequential. BigSolDB and MixtureSolDB compile measurements across sources, temperatures and media [4, 5]. BigSolDB 2.2 adds solid-form descriptions, characterization and protocol metadata, explicitly intended to support thermodynamically homogeneous subsets [6]. SC3 supplies a reproducible cleaning and source-fit consensus construction, nested quality tiers and source dispersion [7]. These developments make large-scale comparison possible, but the evidential requirements for a phase inference remain more specific than the presence of a concentration and a form label.

The relevant chemical principles are established. Voigt showed that the required accuracy depends on use and that experimental descriptions and independent caloric evidence can constrain solubility selection and phase stability [8]. Zhao and colleagues examined post-assay residual-solid information when studying analytical variability and amorphous-associated bias in intrinsic-solubility prediction [9]. The Agreement–Entropy Map (AEM) diagnoses differences between numerical mappings on matched domains [10]. Relative-solubility approaches exploit contributions shared by measurements [2]. Our contribution is consequently not a new thermodynamic law or a general claim that chemical metadata matter. It is a traceable application of these requirements to particular public-data operations and separately measured phase endpoints.

We ask when a reported concentration comparison admits an equilibrium solid-form interpretation. First, we distinguish exact activity-based constraints from approximate concentration-ratio tests and quantify the available same-source form-labelled coverage. We then examine sulfamerazine and tolbutamide using phase readouts acquired in separate published experiments. Finally, we trace temperature-response and molecular-attribution problems to their data-construction operations and retain solvent-ordering controls. These retrospective analyses identify what must be checked before a numerical contrast is reused as a thermodynamic statement. Their scope is the audited records and source families, rather than a prevalence estimate for public solubility data (Table 1). The resulting checks are intended for chemical data curation and for deciding whether a reported concentration can support a specified thermodynamic use.

Table 1 Evidence requirements for thermodynamic reuse of solubility measurements. The required evidence depends on the comparison target; a reported-concentration answer does not by itself establish phase stability.

| Decision target | Physical scope | Evidence required | Supported conclusion |
|---|---|---|---|
| Reported concentration | Specified molecule and quantity | Identity, solvent composition, temperature, method and unit basis | Reproduce the source-defined concentration; equilibrium status is separate. |
| Solvent ordering | Fixed conditions and candidate set | Compatible quantities; form and endpoint evidence if a specified solid is requested | Order within the supported set; no universal cross-polymorph guarantee. |
| Polymorph stability | Same-composition bulk solids at fixed T and p | Retained phases, equilibrium or constrained saturation, solution chemistry and uncertainty | Lower saturation concentration identifies lower solid chemical potential under these conditions. |
| Transition temperature | Equal solid chemical potentials | Overlapping equilibrium branches and phase identity across the temperature range | Common transition temperature; clear-point crossings alone are insufficient. |
| Solid–solid enthalpy | Activity-ratio derivative versus 1/T | Fixed phases and activity correction or justified ratio approximation | Concentration-derived slope remains conditional on the activity term. |
| Consensus temperature response | Known contributor support | Source membership and fitted domains at both temperatures | Separate changes in source composition before interpreting a slope physically. |

## Results

### Thermodynamic constraints define the eligible comparison

For two pure, non-solvated polymorphs A and B of the same chemical composition at fixed temperature T and pressure p, saturation with each retained form equates its molar solid chemical potential with that of the dissolved solute. Activities must share the same solution standard state within each solvent. Subtracting the two saturation relations gives [11, 12]:

$$RT\ln\left(\frac{a_A^{\mathrm{sat}}}{a_B^{\mathrm{sat}}}\right) = \mu_A^{\mathrm{solid}} - \mu_B^{\mathrm{solid}} = \Delta\mu_{AB} \tag{1}$$

The solid chemical-potential difference is solvent-independent under these conditions. Within a stable homogeneous liquid branch at the same solvent composition, increasing solute concentration increases its chemical potential. The sign of the within-solvent saturation-concentration difference must therefore be consistent across solvents at a fixed temperature, provided the solid phases and solution chemistry remain those specified. This is a sign constraint on the two forms within each solvent; it is not a theorem that solvent rankings are invariant between forms. Solvent-specific nonideality can still affect cross-solvent concentration comparisons.

At an equilibrium solid–solid transition temperature, the chemical-potential difference vanishes. The two saturation branches then meet at a common transition temperature across eligible solvents. A crossing inferred from dissolution clear points or another finite-time endpoint need not represent this equilibrium transition. Likewise, failure to observe a crossing within a limited measurement range does not falsify enantiotropy.

Writing the saturation activity as mole fraction times activity coefficient makes the limitation of a concentration-ratio test explicit:

$$\ln\left(\frac{x_A^{\mathrm{sat}}}{x_B^{\mathrm{sat}}}\right) = \frac{\Delta\mu_{AB}}{RT} - \ln\left(\frac{\gamma_A^{\mathrm{sat}}}{\gamma_B^{\mathrm{sat}}}\right) \tag{2}$$

The activity coefficients refer to different saturation concentrations in the same solvent. They need not cancel, even though the dissolved chemical species is identical [13]. Equal or sufficiently similar coefficients justify the familiar concentration-ratio approximation; concentration data alone do not establish its accuracy. In particular, a large ratio discrepancy is not by itself a thermodynamic inconsistency. Differentiation at fixed pressure gives a further conditional test:

$$\frac{\partial\ln\left(x_A^{\mathrm{sat}}/x_B^{\mathrm{sat}}\right)}{\partial(1/T)} = \frac{\Delta h_{AB}}{R} - \frac{\partial\ln\left(\gamma_A^{\mathrm{sat}}/\gamma_B^{\mathrm{sat}}\right)}{\partial(1/T)} \tag{3}$$

Here ΔhAB = hA − hB is the molar enthalpy difference of the two solids. A concentration-ratio slope can represent ΔhAB/R only when the activity-coefficient derivative is negligible or accounted for; cross-solvent agreement of apparent slopes alone does not establish this condition. The derivatives follow the saturation branches at fixed solvent formulation, including composition dependence of the coefficients. A concentration-gap stratification can describe sensitivity but cannot remove the unknown nonideality term.

These constraints apply only after the chemical and physical comparison has been established. Hydrate–anhydrate or solvate–unsolvated pairs exchange a component with the medium and require its chemical potential; they are outside this invariant test [14]. They are not expected to violate it in every dataset. Amorphous materials, different salts, cocrystals, unresolved mixtures and changed chemical species were not interpreted as eligible polymorph contrasts. Particle size, surface effects or strain would also require extending the bulk-solid model. When an eligible numerical contrast conflicts with the constraint, the appropriate conclusion is inconsistency with the stated equilibrium interpretation. Incomplete equilibration, transformation, unrecognized solvation, identity mismatch or quantity errors are alternative explanations, not diagnoses supplied by the sign alone.

The released BigSolDB 2.2 tables join all 125,752 measurements to their metadata and contain 14,813 measurements with nonempty solid labels (11.8%). Exact matching by source, solute, solvent and native temperature gives 211 two-form groups, comprising 422 records from seven solute systems. Only three of these systems provide matched form pairs across multiple solvents at the same temperature (Fig. 1). These counts describe available metadata and numerical support; they do not certify equilibrium phase identities. The small, selected subset motivates source-resolved examples and an explicit account of unresolved evidence instead of a database-wide polymorph-error rate.

Rivaroxaban, imidacloprid and ticagrelor supplied the cross-solvent comparisons at identical reported temperatures [15, 16, 17]. These three systems supplied 183 paired solvent cells in 28 temperature panels, with source-reported concentration uncertainties and different levels of solid-state control. With fixed source-label orientation, $g = \log10(xB/xA)$, none of the panels contained both positive and negative form-solubility differences. No uncertainty-qualified reversal emerged under zero-covariance or worst-case pair-covariance sensitivity calculations. A conservative subset anchored to directly documented imidacloprid residue X-ray powder diffraction (XRPD) at 293.15 and 323.15 K retained 14 paired cells in two panels and also showed no reversal. This negative diagnostic result preserves the observed sign ordering in the eligible subset; it does not validate every labelled phase, establish equilibrium in every assay, or estimate the database-wide frequency of inconsistency. The 28 temperature panels represent three selected chemical systems, not 28 independent replicates.

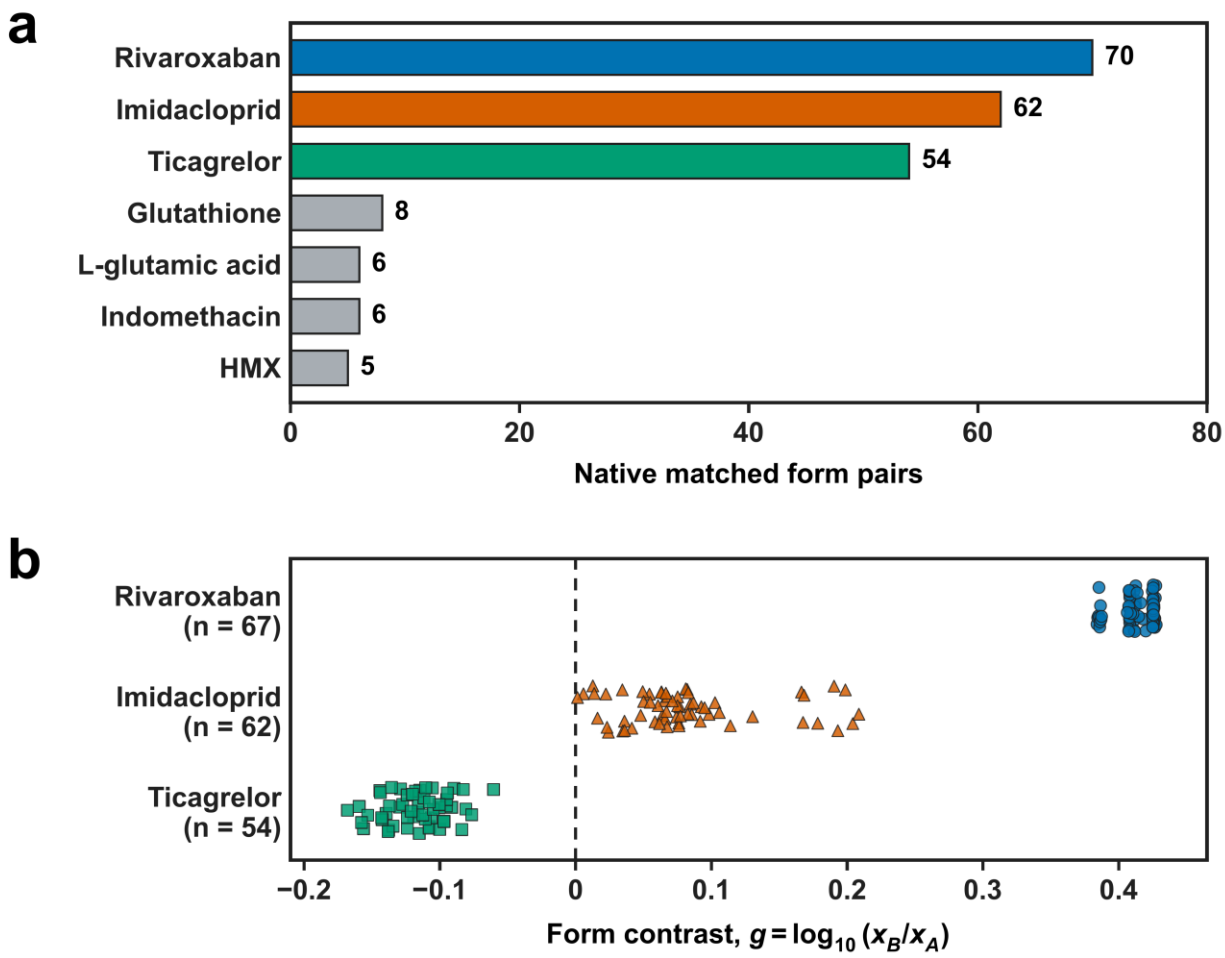



Fig. 1 Available form-labelled comparisons and their concentration contrasts. a Exact native-temperature matching of BigSolDB 2.2 yields 211 form pairs from seven solutes. Coloured bars identify the three systems supporting cross-solvent comparisons at a common temperature; grey bars identify single-solvent systems. Counts include three rivaroxaban pairs outside the cross-solvent subset. b All 183 cross-solvent paired cells in 28 temperature panels: 67 rivaroxaban, 62 imidacloprid and 54 ticagrelor pairs. The contrast is g = log10(xB/xA), with source-native A/B = I/II for rivaroxaban and imidacloprid and II/V for ticagrelor. Vertical jitter separates points; the dashed line indicates equal concentrations. No panel contains opposite point-estimate signs across solvents. Points are correlated concentration contrasts, not independent solutes. Source-reported uncertainties and covariance sensitivity analyses are retained in the source data. The metadata and published controls do not certify equilibrium or retained phase identity at every observation; the directly documented residue-XRPD subset is reported separately

**Separate slurry PXRD limits phase interpretation of dynamic clear points**

The BEST-CSP sulfamerazine study provides concentration curves and a distinct terminal-phase endpoint [18]. Forms I, II and V were measured by dynamic clear-point dissolution in acetonitrile/water, 80:20 v/v. The complete raw curve set contains 55 observations. Separate ternary slurries were examined by residue powder X-ray diffraction (PXRD) in Innsbruck and at University College London. We retained all 34 recorded ternary time points and selected the longest-time single-phase endpoint for each eligible vial or temperature condition. This yielded 13 terminal records at seven temperatures: ten duplicate Innsbruck vials and three UCL conditions within one collaborative study.

Using the minimum interpolated concentration of only I and II gave eight point disagreements with the terminal phase. Including V distinguished an incomplete candidate universe from the assay-endpoint problem: the complete I/II/V proxy still selected V in four terminal records at 45, 46 and 48 °C, whereas PXRD identified II (Fig. 2b). The same four records disagreed across all eight declared combinations of concentration basis and numerical representation (Fig. 2c). Other disagreements at 40 or 60 °C depended on representation and were retained as sensitivity results.

The source explicitly states that clear-point dissolution is non-equilibrium and places the slurry transition between 48 and 50 °C [18]. Parallel slurry experiments using the same heating program provided sparse PXRD checks at approximately 10, 30, 40, 50 and 65 °C that supported retention of the initial phases. These checks were assay anchors rather than residue characterization of every clear-point measurement, so the contrast does not require falsely assigned starting solids. It instead concerns the transfer of a dissolution-boundary measurement to a long-slurry phase judgment. The small phase gaps and unavailable concentration uncertainties prevent an uncertainty-resolved error claim. Nevertheless, the independent assay evidence gives a concrete reason to withhold a stable-phase interpretation from these dynamic curves. The original study discovered form V and established the endpoint distinction; our analysis makes that distinction into a fully specified comparison task.

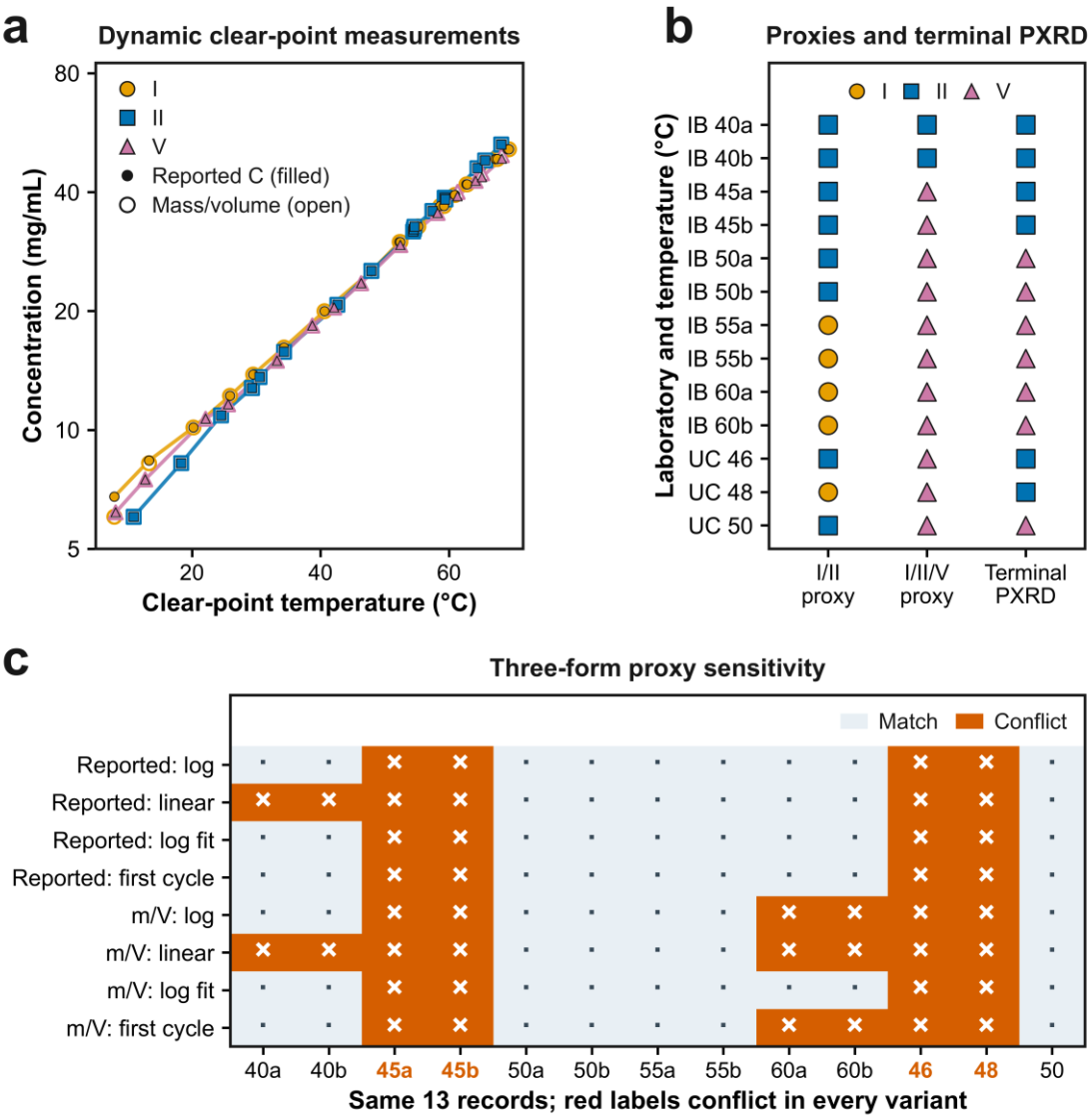



Fig. 2 Terminal-phase evidence limits phase judgments from clear-point concentrations. a All 55 sulfamerazine observations for forms I (24), II (17) and V (14) in acetonitrile/water (80:20 v/v). Filled symbols indicate reported concentrations and open symbols mass/volume interpretations of the same observations. Lines guide the eye; the concentration axis is logarithmic. b Two-form and three-form minimum-concentration proxies versus terminal slurry PXRD for 13 longest-time records, comprising ten duplicate Innsbruck vials (IB, a/b) and three UCL conditions (UC). The proxies disagree with PXRD in eight and four records, respectively. c Eight concentration-basis/representation variants; four records disagree in every variant, at IB 45 °C in duplicate and UC 46 and 48 °C. Log/linear denote piecewise interpolation, log fit denotes a global ln(concentration) fit, and first cycle excludes recrystallized form-I cycles. Sparse parallel-slurry PXRD checks anchor phase retention during heating; terminal judgments derive from long ternary slurries. No concentration uncertainty is inferred. These records concern one molecule and one collaborative study [18]

### Sampling conditions qualify the tolbutamide comparison

Tolbutamide supplied a source test absent from the audited library and SC3 tier populations [19]. The published study compares isothermal gravimetry (IT) and a polythermal visual clear-point method (PT), together with separate 48 h slurry-conversion Fourier-transform infrared (FTIR) observations in n-propanol. Its 2016 precursor belongs to the same laboratory family and provides related calorimetry and form-II solubility [11]. It is supporting context, not a second independent laboratory validation.

The reference pool contains five native slurry temperatures, 268.15–308.15 K. IT provides two-form concentration pairs at three of these temperatures: 278.15, 288.15 and 298.15 K. The minimum reported IT concentration selects I at the first two and II at the third, whereas the source reports conversion of I to II within 48 h at each reference temperature. Two of the three native-exact concentration minima therefore disagree with the reported 48 h endpoint (Fig. 3). This is an assay-specific point comparison in one system, not an estimated general failure rate.

PT has no condition at which both phases and the reference share exactly the same native temperature. In a separately declared within-domain log-linear interpolation supplement, PT can answer at 288.15 and 298.15 K and agrees with the reported endpoint at both. IT disagrees at one of the same two conditions. FI's PT range begins at 278.45 K, so the 278.15 K reference was not extrapolated. The agreement of the two interpolated PT points does not establish that PT is generally more reliable than IT.

The assay description shows why a declared equilibrium protocol is insufficient for interpreting these concentration minima. FI samples were withdrawn from clear liquid; FII suspensions remained cloudy after one hour of settling and were sampled using the same stated 0.2 μm filtration procedure. The authors already discussed particle separation and possible filtration-related effects [19]. If particles passed separation and contributed to the dried residue, they would bias the apparent FII concentration upward, which is the direction that could make FI appear less soluble. Actual particle carryover was not measured for every aliquot; this is a conditional explanation, not an established cause. Table 2 of the source supplies relative standard uncertainties and IT standard errors, but its stated solvent-mass basis conflicts with the printed equation's denominator. Both documented common-monotonic basis interpretations preserve the within-assay point order; absolute concentration remains unresolved.

The FTIR reference is a source-reported result from another experiment, with no retrieved per-temperature raw spectra, quantitative detection limit or explicit wet/dry recovery chain. Its two-vial initial allocation is also insufficiently clear to certify an equal-binary competitive slurry. We therefore evaluate the reported 48 h conversion endpoint, not an independently determined equilibrium boundary. Candidate selection and rules preceded numerical extraction, but qualitative outcome information was available before prediction locking. The test is explicitly an unblinded retrospective application of frozen rules, rather than a prospective validation.

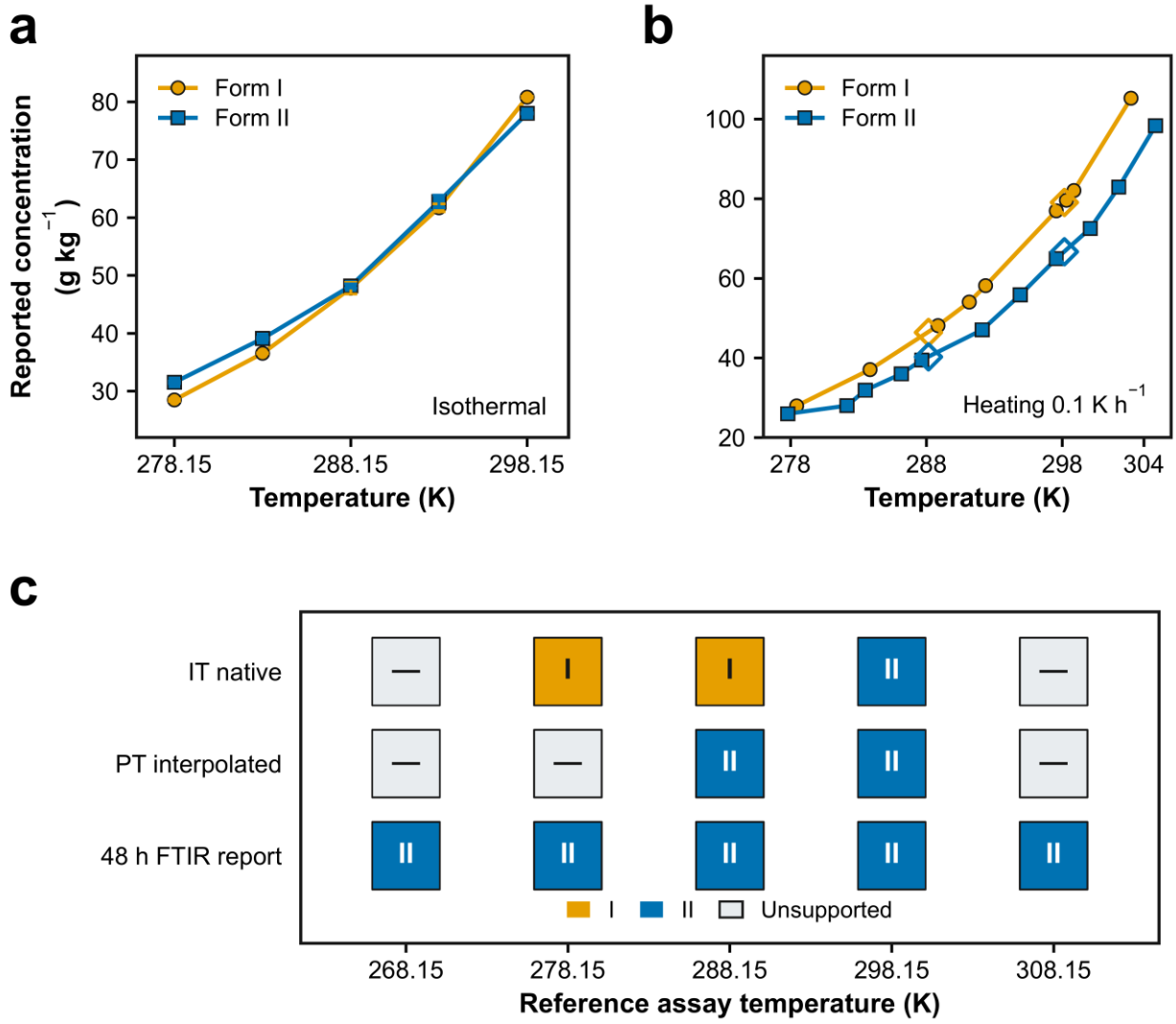



Fig. 3 Sampling and assay endpoints qualify the tolbutamide comparison. a Ten native isothermal (IT) observations in n-propanol; bars are reported standard errors over four samples. b Twenty polythermal (PT) observations: nine for form I and eleven for form II, during heating at 0.1 K per hour. Lines connect native values; open diamonds mark within-domain log-linear interpolation at 288.15 and 298.15 K. c Minimum-concentration proxies versus the source-reported phase after 48 h in separate FTIR slurry experiments. IT supports three native-exact comparisons and disagrees at two. PT supports no native-exact comparison; both supported interpolated comparisons agree. Grey cells are unsupported conditions. The IT form-II suspension remained cloudy before the stated 0.2 µm filtration; post-filtration cloudiness was not established. The conflicting solution/solvent mass bases preserve within-method point order under the two documented monotonic interpretations. Relative standard concentration uncertainties are 0.02 (IT) and 0.01 (PT), with unknown pair covariance. Per-temperature raw FTIR spectra, detection limits and wet/dry handling were unavailable. This retrospective, unblinded comparison in one laboratory family does not establish a precise equilibrium boundary [19]

**Source membership changes the apparent temperature response**

We reproduced all 6,331 labels in the released SC3 Bronze population of 469 solute–solvent pairs, including its nested Silver and Gold populations, with a maximum absolute difference of $8.9 \times 10^{-16}$ in the reported logarithmic quantity. The comparison therefore concerns the released construction rather than a substitute averaging rule. For each adjacent-temperature interval, we retained the original fitted source curves and compared the actual label difference with the difference obtained from sources contributing at both endpoints.

The Gold population contained 4,172 adjacent intervals across 335 pairs. The contributing set changed in 525 intervals, affecting 275 pairs. In 87 intervals, the actual label difference had the opposite sign to the continuing-source difference. Of these, 62 spanned less than 1 K, 11 spanned 1 to less than 5 K, and 14 spanned at least 5 K (Fig. 4a). Only 14 of the 87 intervals had defined source dispersion at both endpoints. Missing dispersion was retained, and the correlated intervals were not treated as independent experimental systems.

Terephthalic acid in N,N-dimethylacetamide (DMAC) provides a directly traceable example in the Silver population. At 363.15 K, two fitted source groups contribute to the label; at 363.20 K, the 2021 group has reached its support boundary and only the 2008 group remains [20, 21]. The released concentration label decreases by 26.7% over this 0.05 K interval, although the continuing source fit increases (Fig. 4b,c). No source was evaluated beyond its fitted range. The difference is explained by the change in aggregation support, without assigning it to a physical phase transition.

Recovered-solid evidence imposes a separate limitation on interpreting the curve. The 2021 study collected solids at 323.15 and 363.15 K and dried them for 24 h at the respective temperature before PXRD [20]. Its DMAC and N-methyl-2-pyrrolidone (NMP) patterns differed between these two sampled conditions, and the authors interpreted the low-temperature patterns as solvent-associated forms. Their original fitting treatment respected solvent-dependent branches. These anchors do not establish an in-situ phase identity at every intermediate temperature. A smooth empirical concentration fit consequently does not, by itself, certify one fixed solid-state thermodynamic process. Source-support changes and physical-state scope remain separate checks; neither the 87 sign differences nor this example establishes thermodynamic anomalies or a newly measured dissolution enthalpy.

The closely spaced 363.15/363.20 K labels magnify the consequence of changing contributors, but a temperature-rounding explanation alone is incomplete: 14 of the opposite-sign Gold intervals span at least 5 K. A thermodynamic slope should be calculated on a fixed supported source set, with any phase change assessed separately. A continuing-source contrast removes the membership term on its matched interval; it does not create an equilibrium measurement or solve an unresolved solid-state assignment. The numerical repair therefore preserves native temperatures and exposes contributor membership rather than merging nearby temperatures without justification.

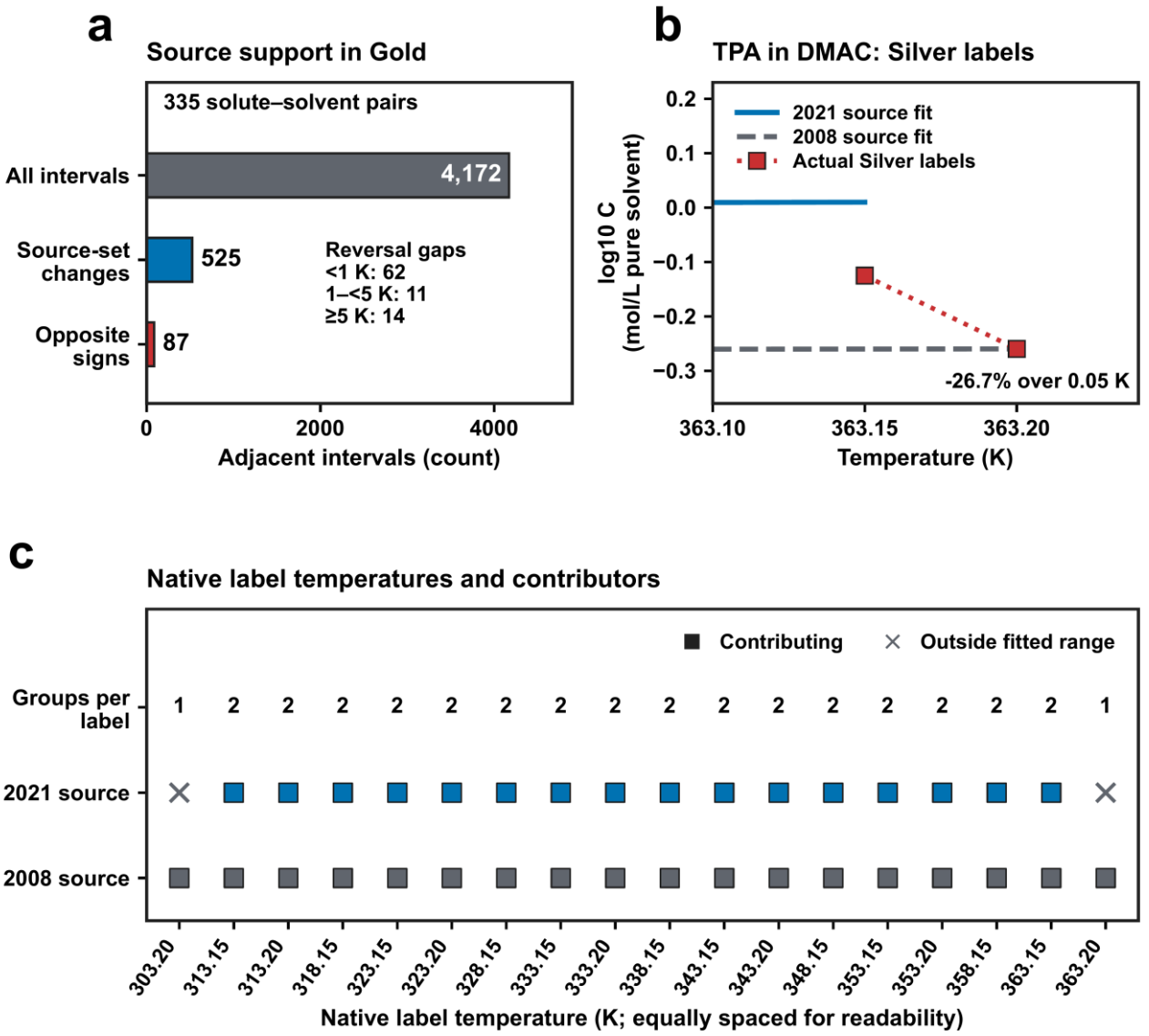



Fig. 4 Changing source support alters a constructed temperature response. a The SC3 Gold population contains 4,172 adjacent-temperature intervals across 335 solute–solvent pairs, including 525 changes in contributor membership and 87 label differences opposite in sign to the continuing-source response. Of the 87 intervals, 62 span less than 1 K, 11 span 1 to less than 5 K, and 14 span at least 5 K. b Terephthalic acid (TPA) in N,N-dimethylacetamide (DMAC), from the Silver population. Two source groups contribute at 363.15 K, whereas only the 2008 group contributes at 363.20 K. The label decreases by 26.7% while the continuing fit increases. Curves retain their fitted domains; the 2021 fit is not extrapolated. The ordinate is log10 concentration in mol per litre of pure solvent. c All 18 native label temperatures, equally spaced for display, with contributor counts. This is an aggregation effect, not evidence of a phase transition or a measured dissolution enthalpy. Undefined dispersion at 363.20 K remains missing [7, 20, 21]

**Molecular attribution must precede a thermodynamic comparison**

We examined 20 previously fixed mixed-solvent overlap windows using original materials descriptions, structures, tables and Supplementary Information (SI). Nineteen were joins between different reported molecular objects. The two sources responsible for these windows contributed 576 library rows: 324 measurements of 2,4,5-trimethoxybenzaldehyde were attributed to its 3,4,5 constitutional isomer, and 252 measurements of N-methyl trinitropyrazole were attributed to 2,4,6-trinitrotoluene (TNT) [22, 23, 24]. Original names, material identifiers and structural diagrams agreed within each paper and differed between the joined objects (Fig. 5a). These are source-correlated attribution errors. The adjudication certifies the object declared in the publication, rather than independently verifying the contents of the experimental sample bottle.

The consequences were quantified with a fixed equal-source arithmetic-mean counterfactual. For a query targeting the reported 3,4,5-trimethoxybenzaldehyde source, pooling the incorrectly joined isomer curves lowered mole-fraction concentration values by 12.4–40.7% across all 144 common-domain query cells. Two cells absent from the extracted library target-source array were explicitly interpolated (Fig. 5b); reconstructing the comparison from the complete published SI retained the level-shift range and six point-choice changes. Six of 72 acetone-versus-N,N-dimethylformamide (DMF) choices changed. Their largest relative target-source gap, 7.82%, was smaller than the source's reported 12% relative standard uncertainty, but missing cross-solvent covariance prevents a calibrated contrast interval. The level shifts are reproducible, whereas a robust solvent-choice benefit is not established. This arithmetic mean is an audit counterfactual, not a label released by MixtureSolDB, SC3 or either source paper.

The existing AEM implementation also returned nonzero mapping diagnostics for these arrays. Using the locked author core with the declared transformation 1/T versus ln(x), the 18 trimethoxybenzaldehyde windows gave structural coordinates of 2.864–4.439 and residual-dispersion coordinates of 2.856–4.446 nats [10]. These are descriptive results of our numerical stress test under that transformation, not phase assignments or a calibrated molecular-identity classifier. Primary chemical evidence identifies the two different isomers being compared; the example therefore does not support a claim that numerical diagnostics are generally blind to attribution errors.

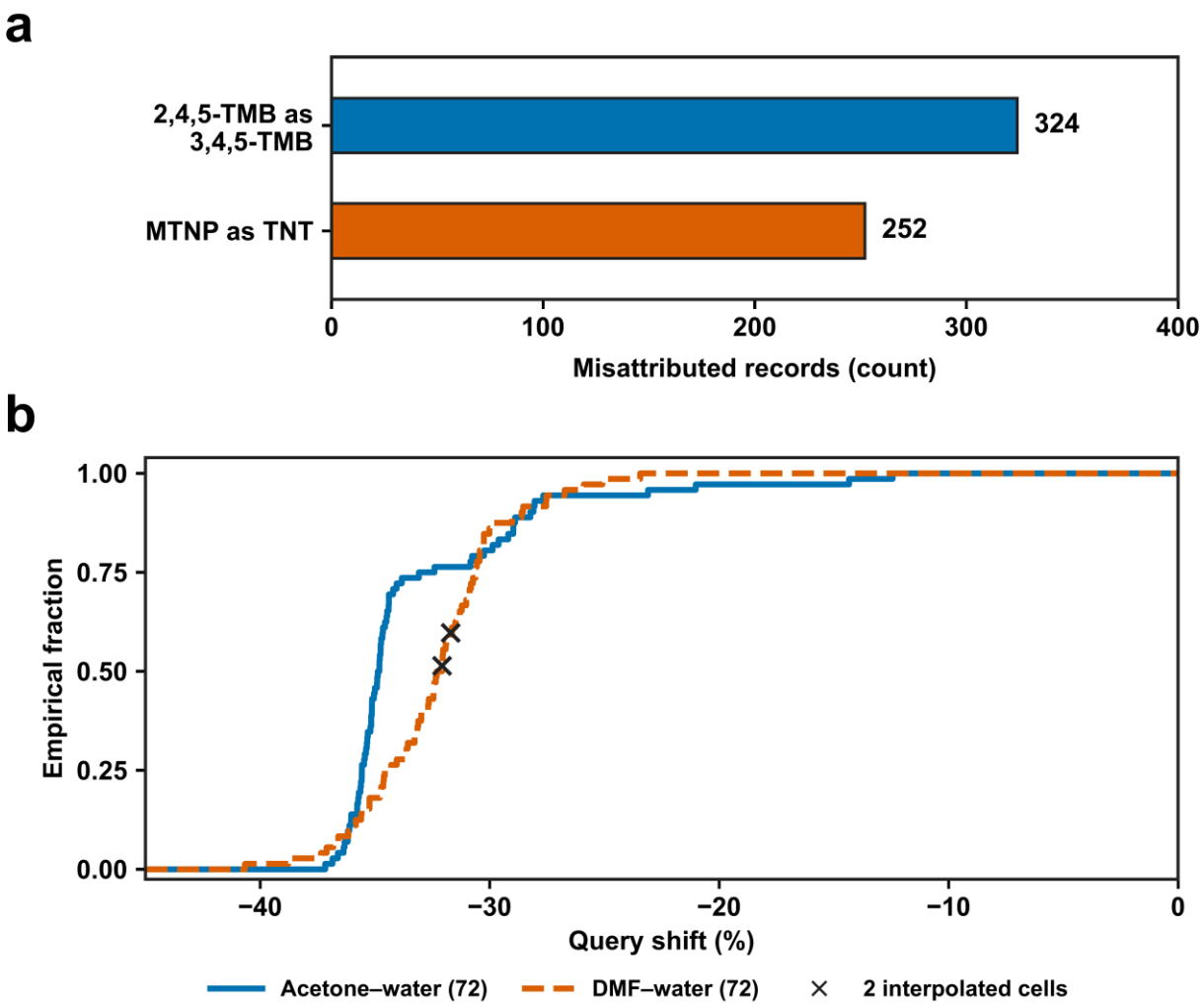



Fig. 5 Molecular attribution changes the object of a concentration query. a Two source-level identity corrections affect 576 records: 324 measurements reporting 2,4,5-trimethoxybenzaldehyde (TMB) were assigned to the 3,4,5 isomer, and 252 measurements of N-methyl trinitropyrazole (MTNP) to TNT. Original values are preserved. b All 144 concentration shifts under a fixed equal-source arithmetic mean, with 72 query cells each for acetone/water and DMF/water. The target is the reported 3,4,5-TMB source. Shift = 100[(x245+x345)/(2x345)−1], where x is mole-fraction solubility. Crosses mark two values absent from the extracted target-source array and explicitly interpolated. This pooling counterfactual was not released by MixtureSolDB, SC3 or the original authors. The shifts do not establish an uncertainty-resolved solvent-choice benefit [22–24]

### Restricted solvent-ordering controls retain useful answers

The thermodynamic restrictions did not imply a general loss of solvent-ordering utility. In the fixed external SolProp comparison, all 15 eligible panels retained the highest-concentration candidate and 41 of 42 pairwise directions agreed with the reference; these panels represent four solutes, and source-native concentration uncertainty was unavailable [25, 26, 27, 28, 29]. Conditioning the released SoluBench choices at their stored panel temperatures preserved all 3,618 evaluable pairwise answers and all 856 evaluable best-solvent answers; 81 and 22 questions, respectively, remained unsupported [30]. These existing controls are summarized in Supplementary Section S6. They support preservation within the audited candidate sets, not a universal claim that solvent rankings survive changes of polymorph.

## Discussion

Solid-state thermodynamics provides a practical test of the interpretation attached to a solubility record. Its strongest constraints concern chemical potentials, the sign of a within-solvent form contrast and an equilibrium transition temperature. The concentration-ratio magnitude and its temperature slope require additional assumptions. Making this distinction prevents two opposite errors: treating nonideal concentration ratios as violations of thermodynamics, and treating smooth or representation-stable concentration curves as proof of phase stability.

The two phase comparisons show why independent assay information matters. In sulfamerazine, four disagreements persist after retaining all three candidate forms and changing the numerical representation eight ways. The published distinction between dynamic dissolution and terminal slurry behavior supplies a physical explanation for withholding a stable-phase interpretation; the numerical exercise does not discover a new polymorphic transition. Tolbutamide provides a complementary limitation: a nominally isothermal equilibrium procedure can still leave an unresolved sampling issue. Its persistently cloudy FII suspension before filtration, recorded separation procedure and disputed concentration denominator must be considered together. Neither example supplies a calibrated error probability, and neither supports ranking whole classes of solubility methods.

The data-construction analyses identify concrete operations that precede any thermodynamic comparison. Changing contributor membership introduces a term that can reverse an adjacent-label temperature response even when each continuing fit rises. The SC3 replay isolates that term without presuming the baseline is defective for concentration prediction. Molecular attribution acts at an earlier level: an isomer or different energetic compound cannot be pooled as the same molecular object merely because the resulting curves are smooth. The 576 affected rows arise from two source-level assignments and do not estimate the frequency of independent extraction errors. Their chemical adjudication supports a derived identity correction, while the published original records remain preserved.

Useful solvent-ordering results coexist with these limitations. All eligible external panel maxima and all temperature-conditioned SoluBench answers were retained. These controls constrain the paper's conclusion: the audited evidence supports some concentration comparisons while limiting specific phase interpretations. It does not justify the categorical statement that public data can rank solvents reliably but cannot inform polymorph decisions. For suitable equilibrium measurements with verified phase identity, the thermodynamic comparison is well founded. The practical task is to identify whether a particular record supplies that evidence.

The distinction also clarifies what should be recorded in future data releases. A minimal reusable record should separate the starting solid designation from characterization of the solid associated with the concentration assay, retain solvent composition and quantity basis, state the temperature and operational endpoint, document equilibration and separation, and preserve the reported uncertainty definition. Measurements and derived consensus labels should remain distinguishable, with contributing source identifiers at each condition. Missing entries should remain explicit. These fields allow chemical eligibility to be checked before evaluating a numerical ratio; they do not automatically certify a record as equilibrium evidence (Table 1).

The present results have a deliberately limited statistical scope. The same-source multi-solvent subset is small and selected by availability of labels and exact-condition support. Repeated temperatures, solvents and vials are not independent solutes. Existing source inspection was targeted rather than a randomized audit, so its adjudicated counts cannot be reported as an automatic-pairing precision estimate with a Wilson interval. The phase cases were already exposed during development, and the tolbutamide analysis is explicitly retrospective and unblinded. We do not designate the new thermodynamic organization or its follow-up sign screen as a preregistered experiment. Missing concentration uncertainty and incomplete endpoint-phase evidence limit uncertainty-resolved inference.

Likewise, a shrinking form-concentration gap in more soluble media would be an association rather than direct evidence of solvent-mediated transformation. Direct sulfamerazine experiments show that increased solubility need not produce faster nucleation when solvent interactions differ [31]. Nucleation, growth and the operational observation time can alter that relationship. The present data do not identify those kinetic contributions, so no transformation-rate mechanism is inferred from concentration compression. An adequately controlled analysis would require source-specific timing and phase-retention evidence in addition to concentration values.

## Conclusions

This audit links thermodynamic requirements to the molecular identities, source support and physical endpoints of public solubility records. The selected cross-solvent form comparisons retain their observed sign ordering, while the phase and data-construction examples identify specific limits on equilibrium interpretation. Solvent-ordering controls preserve useful concentration comparisons within their supported candidate sets. Reuse should therefore be tied to the intended chemical decision and to the evidence available for that record. These retrospective findings do not establish a database-wide failure rate or replace phase-resolved measurements.

## Methods

### Study design and provenance

This work is a retrospective secondary analysis of public solubility resources, published physical experiments and released computational constructions. The thermodynamic reframing and additional sign-screen calculations were specified on 11 October 2026 after the earlier source and numerical analyses. They are not a prospective preregistration. Original source arrays, units, observation identifiers and prior frozen outputs were preserved; all new results were written to a versioned revision directory. No new physical measurements or trained machine learning models were added.

The metadata census uses exact original source, solute, solvent and temperature keys. A source-native form label is a candidate identity, not a globally interchangeable crystallographic designation. Cross-paper FI/FII labels were not normalized into a shared physical form without structural evidence. Distinct salts, hydrates, solvates, amorphous records and unresolved mixtures were not promoted to eligible neat-polymorph contrasts. Source papers, solutes, fitted source groups, intervals, queries and physical endpoints retain separate counts.

### Conditional sign screen and uncertainty

The retrospective sign diagnostic retained exact source, solute and solvent simplified molecular-input line-entry system (SMILES) identifiers and reported temperature matches, without interpolation or temperature rounding. Forms A/B were I/II for rivaroxaban and imidacloprid, and II/V for ticagrelor; the diagnostic used g = log10(xB/xA). Reversing this orientation changes every sign but leaves reversal detection unchanged. Same-temperature panels required at least two common solvents. Published-control and directly documented residue-XRPD-anchor subsets were reported separately. We evaluated ρ = 0 and the maximum-variance bound ρ = −1; positive-correlation sensitivity outputs were also retained. An uncertainty-qualified flag required one solvent with g > 2u(g) and another with g < −2u(g). These are descriptive scale checks, not simultaneous 95% tests across panels. Unknown temperature-error contributions, cross-solvent covariance and systematic assay effects were not inferred. No population confidence interval was estimated from the three selected solutes. All 211 paired values were independently reconstructed from the unchanged raw database.

For the source-labelled diagnostic, g = log10(xB/xA), with form A and form B fixed in the source ledger. This is the reverse of the ratio used to derive equations (1)–(3); reversing all signs leaves the cross-solvent reversal criterion unchanged. First-order propagation of relative standard uncertainties retains the measurement-error correlation ρ [32]:

$$u(g) = \frac{\sqrt{u_{r,A}{}^2 + u_{r,B}{}^2 - 2\rho u_{r,A} u_{r,B}}}{\ln(10)} \tag{4}$$

Setting ρ to zero is conservative only if the true covariance is nonnegative. Same-source provenance alone does not establish that condition. Uncertainty definitions were retained from the original papers; replicate SDs or SEs were not silently substituted for total standard uncertainty. Missing uncertainties were not set to zero. Expanded bands are conditional sensitivity scales, not calibrated simultaneous confidence intervals across all tested solvents or temperatures.

The thermodynamic conditions, rather than the numerical sign, determine physical eligibility. A nominal-label reversal prompts source review and cannot identify transformation, equilibration failure or mislabelling by itself. Non-significant or unresolved contrasts do not establish equivalence. The three selected systems were described in full without a database-wide prevalence estimate or a cluster-bootstrap interval.

### Replay of the released consensus and source-support decomposition

The SC3 release was treated as the actual computational baseline [7]. Its cleaned checkpoint, source-fit groups, fitted-domain support, quality tiers and published label arrays were read directly. Source fits were evaluated only within their retained domains. Per-row means and contributor counts were reconstructed and checked against all 6,331 Bronze labels, retaining the nested Silver and Gold membership.

Let $G_0$ and $G_1$ be the groups contributing at temperatures $T_0$ and $T_1$, and let $f_g(T)$ denote the fixed reported-log-concentration fit of group g. The constructed label is defined by equation (5):

$$m(T) = \frac{1}{|G(T)|} \sum_{g \in G(T)} f_g(T) \quad (5)$$

For $H = G_0 \cap G_1 \neq \emptyset$, we defined the continuing-source response by equation (6):

$$\Delta_{\text{cont}} = \frac{1}{|H|} \sum_{g \in H} \{ f_g(T_1) - f_g(T_0) \} \quad (6)$$

The composition contribution is $\Delta_{comp} = m(T_1) - m(T_0) - \Delta_{cont}$. Equivalently, it is the change between the endpoint offsets of the actual contributor mean relative to the continuing-group mean. When $G_0 = G_1$, $\Delta_{comp}$ is zero apart from floating-point tolerance. When H is empty, the control is undefined and the interval is retained without a sign comparison. Opposite responses require opposing nonzero signs. The operation changes only source support, leaving temperatures and source fits fixed. The continuing-source mean is a processing control, not independently measured truth. For at least two contributing fit evaluations, the released σ is their sample standard deviation (ddof = 1), floored at 0.012 log S; it is undefined when fewer than two groups contribute [7]. Source dispersion remains the released construction's dispersion and is not interpreted as a calibrated interval for a phase transition or temperature derivative.

**Molecular identity and fixed pooling counterfactuals**

The mixed-solvent candidates were fixed before numerical pooling results were inspected. Original solute names, material/CAS declarations and structural figures were compared with the assigned full molecular graph. Graph matching preserved stereochemistry, charge, isotopes and disconnected fragments; it did not reduce salts to parent fragments or use similarity matching. Original table and SI values were checked for temperature, scale and concentration basis. An agreement between publication identifiers and diagrams adjudicates the reported experimental object, while bottle-level composition remains outside the audit.

For trimethoxybenzaldehyde, the 18 fixed composition windows were retained across acetone/water and DMF/water. The counterfactual pooled concentration was $(x_{245} + x_{345})/2$, with x the reported solute mole fraction and equal source weights. The target-reference shift was $100[(x_{245}+x_{345})/(2x_{345})-1]$. Queries stayed within both native temperature domains. Two native values omitted from the extracted library array were separately marked as interpolated; the complete published SI was evaluated as a sensitivity analysis, and full native-data sensitivity was retained. Solvent-choice comparisons used the same temperature, organic-solvent composition and candidate pair. This operation was not substituted for any released official consensus.

The FOX-7 comparison [33, 34] retained its fixed NMP composition and all native records. Level and shape comparisons used within-domain interpolation only and left the fixed-temperature solvent-addition and heating endpoints separately labeled. The author AEM core was applied without modifying its min-points rule or searching an agreement threshold [10]. Its outputs summarize compatibility under the declared numerical transformation; no coordinate was converted into a chemical-identity or phase label.

### Sulfamerazine concentration proxy and terminal-phase reference

The sulfamerazine comparison used the original reported form-I, II and V clear-point arrays in acetonitrile/water, 80:20 v/v, together with independently performed ternary-slurry residue measurements [18]. The numerical primary representation was piecewise linear interpolation of log concentration, restricted to the shared native domain. The phase proxy selected the candidate with the lowest interpolated concentration. I/II-only and complete I/II/V candidate universes were evaluated separately.

The primary physical reference retained the longest reported time for each eligible vial or condition and required a single assigned terminal phase within the numerical domain. The 34 ternary time points yielded ten duplicate Innsbruck vials and three UCL conditions, totaling 13 terminal records. Earlier mixtures and shorter times remained in the reference ledger; duplicated SI figures were not additional experiments. PXRD assignments and temperature/time labels were inspected in the published main figures and SI. Dynamic-phase-retention checks were recorded as sparse assay anchors rather than propagated as per-row in-situ characterization.

Eight numerical sensitivity variants combined reported versus mass/volume-derived concentration with log-concentration interpolation, linear-concentration interpolation, global log-concentration fitting and first-cycle-only form-I data. Source inconsistencies remained in the originals. Agreement across representations was described as representation stability, with no concentration uncertainty or statistical confidence inferred from the alternatives.

### Tolbutamide reference scope and retrospective evaluation

Tolbutamide 2020 was selected using title/access metadata after its candidate and field contract had been frozen [19]. Molecular and source overlap were checked against BigSolDB 2.2, SC3 canonical and cleaned data, fitted groups, tier pairs and the original development-source rosters. The 2016 predecessor was separately recorded for prior exposure and shared authorship [11]. Qualitative endpoint information reached the analysis before prediction locking; all results therefore retain retrospective, unblinded status.

The original Table 2 supplied ten IT and twenty PT observations, with original temperatures, IT standard errors and reported standard uncertainties. The independent-reference table contains the five temperatures at which the paper reports I-to-II conversion after 48 h and FTIR sampling. It was not supplied to prediction rules. Missing raw spectra, detection quantitation, recovery handling and initial vial allocation remained explicit limitations.

Exact native-temperature matching was the primary analysis. IT supported three queries; PT supported none. The separately labeled supplement used within-domain piecewise linear interpolation in log of the reported concentration and supported two PT queries. The basis sensitivity applied the two documented common monotonic mass-basis interpretations after this declared representation. It verified within-assay order invariance without resolving the original denominator discrepancy or certifying cross-method absolute concentrations. Relative standard uncertainty, four-sample SE and unknown inter-form covariance were stored separately; no SE-to-total-error conversion or automatic confidence interval was applied.

### External concentration-order and benchmark-temperature controls

External SolProp candidates were selected before computing the reference contrasts [25]. Original DOI labels had to be absent from the released SC3 canonical table; full molecular graphs and concentration conventions were matched, and admissible source fits had to represent the actual tier-pair population within range. Ambiguous phase-specific rows and unsupported cells remained excluded or unresolved with reasons.

References were constructed from the original quantities, not blindly copied from the stored converted column. Mole fraction was converted as x/(1−x) times solvent density divided by solvent molar mass. Reported grams per 100 grams of solvent were converted using the stated mass-ratio convention. The resulting concentration is mol per litre of pure solvent, rather than a certified molarity of the saturated solution. Fixed release densities were retained. P is the mean of available source fits for each solvent; C is the mean contrast of groups supporting both solvents at the query temperature. C is undefined without common groups. At released tier temperatures, P was checked against the published construction; elsewhere it is explicitly an in-range reconstruction. All eligible contrasts, candidate panels and top choices were retained. Original uncertainty absent from the release remained absent.

SoluBench was audited at its fixed released repository snapshot [30]. Every Task-1 and Task-2 component was matched by DOI, solute, solvent and stored LogS to the cited original BigSolDB 2.0 array, with numerical tolerance $10^{-10}$. Native temperature matching was unique. The candidate set and stored panel T were fixed. Conditioning used native values or within-domain linear interpolation in the original stored log10(mol/L); incompatible duplicates and out-of-domain queries abstained. Model APIs were not called and published model performance was not re-ranked. Generator metadata, model-question wording and reference-answer stability were evaluated separately.

### Statistical analyses

Analyses retain the finite audited population and its dependence structure. No pooled p-value or chemical-error rate is formed across molecules, source arrays or assay endpoints. Source dispersion, reported standard uncertainty, standard error and categorical phase evidence are distinct. The 87 SC3 intervals, 576 attribution rows and repeated slurry records are not independent biological or chemical replicates. Existing software verification checks computational logic and numerical replay; it does not supply independent experimental replication.

Each main figure has machine-readable source data, retained input hashes and a construction script or a documented exact reuse of an existing figure. Supplementary records preserve unresolved comparisons and the scope of original source inspection. Raw inputs and earlier frozen outputs were retained without modification. Cross-source form labels were not treated as structurally equivalent without supporting evidence, and source-held-out errors were not interpreted as performance on unseen solutes.

### AI assistance

Large language model tools, including OpenAI ChatGPT and Codex, assisted literature retrieval, extraction-script and analysis-code drafting, and manuscript revision. Numerical claims were checked against executable outputs and retained primary-source records. The thermodynamic assumptions, source inspection scope, retrospective design and unresolved uncertainty are stated explicitly. The authors retain responsibility for the scientific interpretation and final manuscript.

## Declarations


### Funding

No funding was received for this work.


### Competing interests

The authors declare no competing interests.

### Author contributions

Minhao Liu conceived the study, developed the methodology and analysis code, curated and analysed the data, prepared the figures, and drafted the manuscript. Pu Sun reviewed the analysis code, provided technical consultation, and contributed to manuscript revision. Both authors contributed to reviewing and editing the manuscript.

### Ethics approval and consent to participate

Not applicable. This study analysed published chemical measurements and public datasets and involved no human participants, human biological material or animal experiments.

### Consent for publication

Not applicable.

### Data availability

The data analysed in this study were obtained from published articles and the public resources identified in the reference list and Supplementary Information, Section S8. The derived data supporting Figures 1–5 and Supplementary Fig. S1, the molecular-attribution corrections, the sign-screen outputs and the retained solvent-ordering controls are provided in Source_Data.zip. The archive includes source identifiers, an index of figure data and input/output checksums. Original database files and published source materials remain available from their cited providers under their respective access and licensing terms. No new experimental data were collected.

### Code availability

The revision-specific endpoint-analysis code and an exported-data verification script are included in Source_Data.zip. Recomputing the endpoint audit from the original records also requires the BigSolDB 2.2 measurement and solid-form tables identified in Supplementary Information, Section S8. The earlier audit implementation is available at https://github.com/chineseschrodingercat/solubility-comparison-evidence/releases/tag/v1.0.0. That release predates the present thermodynamic revision; it is not an archive of the current manuscript or its complete revised analyses.

**Supplementary information**

Supporting_Information.docx contains supplementary methods, source-evidence qualifications, sensitivity analyses, Fig. S1 and supplementary references. Source_Data.zip contains the derived figure data, endpoint-analysis code, exported-data verifier and provenance checksums. These files accompany the article.

## References


1. Attia, L., Burns, J. W., Doyle, P. S. & Green, W. H. Data-driven organic solubility prediction at the limit of aleatoric uncertainty. *Nature Communications* **16**, 7497 (2025). https://doi.org/10.1038/s41467-025-62717-7
2. Wollschläger, J. & Montanari, F. Solvmate – a hybrid physical/ML approach to solvent recommendation leveraging a rank-based problem framework. *Digital Discovery* **3**, 1749–1760 (2024). https://doi.org/10.1039/D4DD00138A
3. Joint Committee for Guides in Metrology. *International vocabulary of metrology—Basic and general concepts and associated terms (VIM).* 3rd edition, 2008 version with minor corrections, JCGM 200:2012 (2012). Clause 2.3, notes 1 and 3. https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf
4. Krasnov, L., Malikov, D., Kiseleva, M., Tatarin, S., Sosnin, S. & Bezzubov, S. BigSolDB 2.0, dataset of solubility values for organic compounds in different solvents at various temperatures. *Scientific Data* **12**, 1236 (2025). https://doi.org/10.1038/s41597-025-05559-8
5. Malikov, D., Krasnov, L., Kiseleva, M., Meshcheriakova, E., Kuznetsov, F., Elistratov, V., Vasiyarov, M., Tatarin, S. & Bezzubov, S. Dataset of solubility values for organic compounds in binary mixtures of solvents at various temperatures. *Scientific Data* (2026). https://doi.org/10.1038/s41597-026-07047-z
6. BigSolDB 2.2: A Comprehensive Dataset of Solubility Values for Organic Compounds in Organic Solvents and Water at Various Temperatures. Version 2.2, *Zenodo*, published 7 September 2026. Dataset. https://doi.org/10.5281/zenodo.22648301
7. Ramani, V., Arora, H. A., Kuchhal, D., Tatarin, S., Krasnov, L., Ranu, S. & Karmakar, T. SC3: The Multi-Solvent Solubility Challenge and Benchmark. *arXiv*, 2606.07656, version 1, 3 June 2026. Preprint. https://arxiv.org/abs/2606.07656
8. Voigt, W. Evaluation of Equilibrium Solubilities: Statistics Versus Selected Experiments. *Journal of Solution Chemistry* (2026). https://doi.org/10.1007/s10953-026-01579-z
9. Zhao, J., Hermans, E., Sepassi, K., Tistaert, C., Bergström, C. A. S., Ahmad, M. & Larsson, P. Effect of Data Quality and Data Quantity on the Estimation of Intrinsic Solubility: Analysis Based on a Single-Source Data Set. *Molecular Pharmaceutics* **21**, 5261–5271 (2024). https://doi.org/10.1021/acs.molpharmaceut.4c00685
10. Haddadi, H., Kloskowski, A. & Mironowicz, P. Physics-informed structural diagnostics of model–data agreement beyond scalar metrics. *Scientific Reports* **16**, 18622 (2026). https://doi.org/10.1038/s41598-026-49445-8. Author code: https://github.com/HedayatHaddadi/AEM
11. Svärd, M., Valavi, M., Khamar, D., Kuhs, M. & Rasmuson, Å. C. Thermodynamic Stability Analysis of Tolbutamide Polymorphs and Solubility in Organic Solvents. *Journal of Pharmaceutical Sciences* **105**, 1901–1906 (2016). https://doi.org/10.1016/j.xphs.2016.03.021

12. Cheuk, D., Svärd, M. & Rasmuson, Å. C. Thermodynamics of the Enantiotropic Pharmaceutical Compound Benzocaine and Solubility in Pure Organic Solvents. *Journal of Pharmaceutical Sciences* **109**, 3370–3377 (2020). https://doi.org/10.1016/j.xphs.2020.07.022

13. Valavi, M., Svärd, M. & Rasmuson, Å. C. Improving Estimates of the Crystallization Driving Force: Investigation into the Dependence on Temperature and Composition of Activity Coefficients in Solution. *Crystal Growth & Design* **16**, 6951–6960 (2016). https://doi.org/10.1021/acs.cgd.6b01137

14. Qu, H., Louhi-Kultanen, M. & Kallas, J. Solubility and stability of anhydrate/hydrate in solvent mixtures. *International Journal of Pharmaceutics* **321**, 101–107 (2006). https://doi.org/10.1016/j.ijpharm.2006.05.013

15. Zhai, J., Chen, Z., Liu, X., Zhang, L. & Lu, J. Solubility measurement, model evaluation and thermodynamic analysis of rivaroxaban polymorphs in organic solvents. The Journal of Chemical Thermodynamics 104, 218–229 (2017). https://doi.org/10.1016/j.jct.2016.09.036

16. Li, W., Wang, L., Shi, P. & Gao, Z. Determination and Correlation of Solubility of Forms I and II of Imidacloprid in Seven Pure Solvents from 293.15 to 333.15 K. Journal of Chemical & Engineering Data 65, 5421–5427 (2020). https://doi.org/10.1021/acs.jced.0c00558

17. Yang, S., Wu, Y., Zhang, H., Sun, J. & Liu, Y. Solubility Measurement of Crystal Form II and Form V Ticagrelor in Several Pure Organic Solvents. Journal of Chemical & Engineering Data 66, 2022–2033 (2021). https://doi.org/10.1021/acs.jced.0c01073

18. Wood, W. P., et al. BEST-CSP Benchmark Study of Polymorphs I and II of Sulfamerazine and the Perils of Polytype Polymorphs. *Crystal Growth & Design* **26**, 476–493 (2026). https://doi.org/10.1021/acs.cgd.5c01406

19. Zeng, L., Rasmuson, Å. C. & Svärd, M. Solubility of Two Polymorphs of Tolbutamide in n-Propanol: Comparison of Methods. *Journal of Pharmaceutical Sciences* **109**, 3021–3026 (2020). https://doi.org/10.1016/j.xphs.2020.06.022

20. Celante, D., Diehl, L. O., Brondani, L. N., Bizzi, C. A. & de Castilhos, F. Measurement and Correlation of the Solubility of Terephthalic Acid in Six Pure Solvents and Five Binary Mixtures. *Journal of Chemical & Engineering Data* **66**, 3512–3519 (2021). https://doi.org/10.1021/acs.jced.1c00373

21. Guo, X., Cheng, Y.-W., Wang, L.-J. & Li, X. Solubility of Terephthalic Acid in Aqueous N-Methyl Pyrrolidone and N,N-Dimethyl Acetamide Solvents at (303.2 to 363.2) K. *Journal of Chemical & Engineering Data* **53**, 1421–1423 (2008). https://doi.org/10.1021/je700635r

22. Yu, S., Cheng, Y., Feng, W., Xing, W., Li, H. & Xue, F. Solid-liquid equilibrium solubility measurement, model evaluation and Hansen solubility parameter of 2,4,5-trimethoxybenzaldehyde in four aqueous binary solvent mixtures. *Journal of Molecular Liquids* **339**, 116750 (2021). https://doi.org/10.1016/j.molliq.2021.116750

23. Xue, F., Wang, Z., Ma, Y., Feng, W., Xing, W., Li, H. & Yu, S. The Solubility Profile of 3,4,5-Trimethoxybenzaldehyde in Aqueous Binary Solvent Mixtures at Several Temperatures. *Journal of Chemical & Engineering Data* **66**, 3911–3918 (2021). https://doi.org/10.1021/acs.jced.1c00578

24. Guo, H.-J., et al. Solubility and thermodynamic properties of a kind of explosives in four binary solvents. *Journal of Molecular Liquids* **247**, 313–327 (2017). https://doi.org/10.1016/j.molliq.2017.09.097

25. Vermeire, F. H., Chung, Y. & Green, W. H. Predicting Solubility Limits of Organic Solutes for a Wide Range of Solvents and Temperatures. *Journal of the American Chemical Society* **144**, 10785–10797 (2022). https://doi.org/10.1021/jacs.2c01768. SolProp v1.2 experimental data: https://zenodo.org/records/5970538

26. Suren, S., Sunsandee, N., Stolcova, M., Hronec, M., Leepipatpiboon, N., Pancharoen, U. & Kheawhom, S. Measurement on the solubility of adipic acid in various solvents at high temperature and its thermodynamics parameters. *Fluid Phase Equilibria* **360**, 332–337 (2013). https://doi.org/10.1016/j.fluid.2013.10.003

27. Jia, S., Zhang, K., Wan, X., Gao, Z., Gong, J. & Rohani, S. Effects of Temperature and Solvent Properties on the Liquid–Solid Phase Equilibrium of γ-Pyrazinamide. *Journal of Chemical & Engineering Data* **65**, 3667–3678 (2020). https://doi.org/10.1021/acs.jced.0c00286

28. Stephens, T. W., Loera, M., Calderas, M., Diaz, R., Montney, N., Acree, W. E. & Abraham, M. H. Determination of Abraham model solute descriptors for benzoin based on measured solubility ratios. *Physics and Chemistry of Liquids* **50**, 254–265 (2012). https://doi.org/10.1080/00319104.2011.637628

29. Schmidt, A., et al. Determination of Abraham model solute descriptors for isophthalic acid from experimental solubility data in organic solvents at 298 K. *Physics and Chemistry of Liquids* **54**, 747–757 (2016). https://doi.org/10.1080/00319104.2016.1149178

30. SoluBench. Released task-generation, question, result and evaluation repository accompanying *Can LLMs Solve Solubility Tasks? The SoluBench Benchmark for Pure and Mixed Solvent Systems*. Computational resource; audited Git snapshot 00ad03f584a1764a8aeaf30aec66b3e584aa45c6, accessed 9 October 2026. https://github.com/levakrasnovs/SoluBench/tree/00ad03f584a1764a8aeaf30aec66b3e584aa45c6

31. Gu, C.-H., Young, V., Jr. & Grant, D. J. W. Polymorph screening: Influence of solvents on the rate of solvent-mediated polymorphic transformation. *Journal of Pharmaceutical Sciences* **90**, 1878–1890 (2001). https://doi.org/10.1002/jps.1137

32. Joint Committee for Guides in Metrology. *Evaluation of measurement data—Guide to the expression of uncertainty in measurement*. JCGM 100:2008 (2008). https://doi.org/10.59161/JCGM100-2008E

33. Zhao, X., Xu, J., Zhang, L., Chen, L. & Wang, J. Solubility and thermodynamics properties of FOX-7 in binary (N-methyl pyrrolidone + acetone, ethanol, water) mixed solvents. *Journal of Energetic Materials* **41**, 632–655 (2023; first published online 20 December 2021). https://doi.org/10.1080/07370652.2021.2017075

34. Liu, Q., et al. Solubility determination and prediction for FOX-7 in three binary solvents at different temperatures. *Journal of Energetic Materials* **42**, 68–83 (2024; first published online 21 February 2022). https://doi.org/10.1080/07370652.2022.2028034

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformation.docx