RETRO-FORGE: A MULTI-STEP PAIRWISE RETROSYNTHESIS FRAMEWORK FOR SOLID-STATE MATERIALS SYNTHESIS

Anonymous authors
Paper under double-blind review

ABSTRACT

AI-driven materials discovery has made remarkable strides in generating stable, unique, and novel structures at unprecedented scale, yet a critical gap remains between generation and realization: without valid synthesis routes, proposed materials stay in silico. Existing approaches to precursor prediction (PP)—the first and most consequential step of materials synthesis planning (MSP)—formulate it as a single-step problem, overlooking the established domain knowledge that solid-state reactions proceed pairwise. We introduce Retro-Forge, a multi-step pairwise retrosynthesis framework that for the first time casts PP as a sequence of learnable pairwise reactions. A single-step pairwise reactant prediction model is trained on a curated pairwise reaction dataset, expanded through synthetic data augmentation to address data scarcity. This model is composed recursively within a tree search to produce complete synthesis routes, and further optimized for this task via multi-step tree-RL. We construct a clean benchmark of 857 novel targets verified absent from all training data, on which Retro-Forge achieves state-of-the-art performance. Beyond precursor prediction accuracy, we independently corroborate that the large majority of its predicted pairwise steps are stoichiometrically balanced and thermodynamically favorable under DFT-derived reaction energies—demonstrating that the multi-step pairwise formulation is not only learnable but also chemically valid.

1 INTRODUCTION

The application of artificial intelligence to inorganic materials discovery has grown rapidly. Generative and predictive models now propose candidate materials at unprecedented scale (Merchant et al., 2023; Zeni et al., 2025), yet bringing them into the laboratory remains the central bottleneck: without valid synthesis routes, even the most promising predicted materials cannot be realized experimentally (Szymanski et al., 2023b). Materials synthesis planning (MSP)—determining how to produce a target material from available starting materials—is intrinsically difficult, admitting a one-to-many mapping in which a single target may be reachable through multiple routes involving different precursors, intermediates, and processing conditions. Among the subtasks of MSP, precursor prediction (PP)—selecting which starting materials react to form a given target—is the first and most consequential decision: errors here propagate through every subsequent step of the synthesis plan.

Previous approaches to PP fall along two broadly distinct lines. Early work adopted retrieval-based and similarity-driven strategies over known reactions (He et al., 2023), thermodynamic optimization over computed formation energies (Aykol et al., 2021), and ranking-based approaches that embed targets and precursor sets in a shared latent space (Prein et al., 2025a). More recently, large language models fine-tuned on synthesis corpora have achieved strong performance on precursor prediction (Prein et al., 2025b; Noh et al., 2026; Song et al., 2025; Kim et al., 2024a). What these approaches share is a common formulation: precursor prediction is treated as a single-shot mapping from a target composition to a complete precursor set, bypassing the sequential structure of how solid-state reactions actually proceed.

In solid-state inorganic synthesis, experiments have shown that multi-component transformations proceed through sequences of *pairwise* reactions between adjacent reactant phases (Malkowski et al., 2021): exactly two solid phases react at their mutual interface, driven by diffusion-limited

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

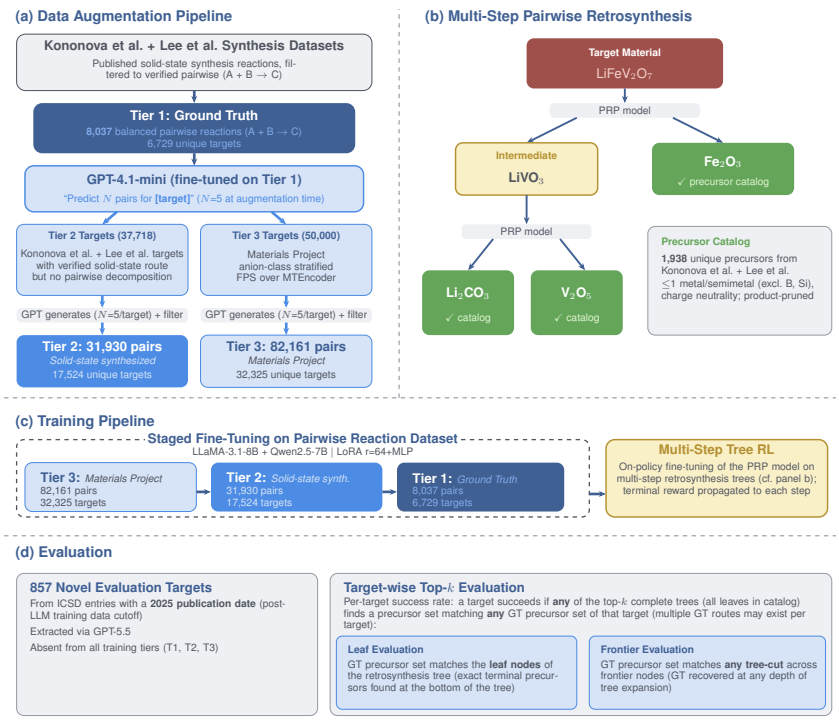

Figure 1: Overview of Retro-Forge. **(a)** Data augmentation: a GPT-4.1-mini model fine-tuned on verified Tier 1 pairwise reactions (8,037) generates candidate reactions for solid-state synthesized targets (Tier 2; 31,930 after filtering) and Materials Project targets (Tier 3; 82,161). **(b)** Multi-step pairwise retrosynthesis: the PRP model recursively decomposes intermediates (amber) until every leaf is a purchasable precursor in the catalog (green). **(c)** Training: staged fine-tuning (T3→T2→T1), then multi-step tree-RL on complete synthesis trees. **(d)** Evaluation on 857 novel ICSD-2025 targets absent from all training tiers, scored by leaf and frontier top-k (Section 4.4).

kinetics, and the product of one pairwise step may serve as a reactant in the next. For example, the well-known material $YBa_2Cu_3O_{6+x}$ (YBCO) is synthesized through an initial pairwise reaction of BaO_2 and CuO forming $Ba_2Cu_3O_6$, which later reacts with Y_2O_3 to form the target material (Miura et al., 2021). This principle has been translated into algorithmic frameworks that rank candidate pairwise reactions by thermodynamic driving force and have achieved experimental discovery of novel inorganic phases in autonomous laboratory settings (Szymanski et al., 2023a;b). However, these approaches rely on rule-based heuristics computed from *ab initio* formation energies, which are computationally expensive and unavailable for the majority of candidate or intermediate materials. It remains an open question how to cast the pairwise reaction as a *learnable* unit trained directly on large-scale synthesis corpora—turning isolated recipes into recombinable building blocks, in the same spirit as organic retrosynthesis extends known reactions through learnable bond disconnections (Corey & Wipke, 1969; Segler et al., 2018).

Addressing PP also requires confronting two practical obstacles. First, the datasets underpinning existing models—most notably Kononova et al. (2019) and Lee et al. (2025)—are systematically biased toward publisher-accessible literature, overrepresenting well-studied material families while leaving large regions of chemical space undercharacterized (Prein et al., 2025b; Sun & David, 2025), and carry substantial noise from automated extraction pipelines (Chung et al., 2025; Prein et al., 2025b). This is particularly consequential for PP, where the targets of interest are newly proposed materials that are least likely to resemble the well-documented families the corpora cover. Second, generalization to genuinely novel targets remains difficult to establish. Prior work recognizes this and probes it through split design, evaluating models on temporally held-out reactions (Noh et al., 2024) or on targets whose constituent element systems are absent from training (Prein et al., 2025a), with performance degrading substantially under these harder settings. Such splits, however, are still drawn from the same text-mined corpus, so held-out targets are novel only relative to a training cut rather

than to the documented literature as a whole. For LLM-based approaches the difficulty compounds: a temporal split within the corpus offers no protection when the held-out targets still predate the model's pretraining cutoff, making genuine generalization indistinguishable from knowledge recall (Magar & Schwartz, 2022; Sainz et al., 2023). Establishing generalization therefore calls for evaluation on targets that are both absent from the training corpus and postdate the pretraining cutoff of the models involved.

In this paper, we introduce **Retro-Forge**, a multi-step pairwise retrosynthesis framework for solid-state inorganic synthesis. We construct a chemically curated and validated dataset of solid-state pairwise reactions by rigorously filtering the Kononova and Lee corpora (Kononova et al., 2019; Lee et al., 2025) to retain only verified, chemically valid pairwise reactions. We expand coverage through LLM-based data augmentation, in which a fine-tuned model generates candidate pairwise reactions for a diverse pool of target compositions. A single-step pairwise reactant prediction (PRP) model trained via staged fine-tuning on this pairwise reaction dataset is then deployed within a multi-step tree search that recursively decomposes a target into pairs of reactants until a complete synthesis route is formed. The PRP model is further optimized within this format through multi-step tree-RL: the policy rolls out complete retrosynthesis trees and is rewarded based on the generated route, with that outcome credited to every decomposition step along the route. Critically, we evaluate Retro-Forge and existing PP models on 857 novel targets that are both absent from the training corpus and postdate the pretraining cutoff of every model considered, achieving state-of-the-art performance.

To the best of our knowledge, Retro-Forge is the first work to formulate precursor prediction for solid-state synthesis as a sequence of learnable pairwise reactions, and to benchmark precursor prediction models on targets that are genuinely novel with respect to both the training corpus and the pretraining data of the underlying models.

Our contributions are as follows:

- We propose Retro-Forge, a novel multi-step pairwise retrosynthesis framework that formulates PP as a sequence of learnable pairwise reactions, solved via a tree search over recursive two-reactant decompositions, enabling composition of pairwise reactions into complete synthesis routes.
- We propose a data augmentation pipeline in which an LLM fine-tuned on a small ground-truth dataset generates synthetic pairwise reactions from diverse target pools, demonstrating that augmentation substantially improves both single-step and multi-step prediction.
- We implement a tree-RL fine-tuning pipeline for multi-step retrosynthesis to reduce the error propagation of a single-step PRP model, enhancing multi-step precursor prediction while largely preserving its single-step performance.
- We construct a benchmark of 857 novel synthesis targets that are absent from the training corpus and postdate the pretraining cutoff of every model considered, and benchmark existing precursor-prediction methods on it, showing that Retro-Forge achieves state-of-the-art performance.

2 Related Works

Multi-step retrosynthesis in organic chemistry. Retrosynthesis planning in organic chemistry, formalized by Corey and Wipke (1969) as the disconnection of a target molecule into simpler precursors via known reaction templates, has been extensively studied with machine learning. Modern approaches include template-based methods that retrieve and rank known reaction templates (Segler et al., 2018; Chen et al., 2020) and template-free methods that treat retrosynthesis as a sequence-to-sequence problem over molecular SMILES strings (Liu et al., 2017). Multi-step retrosynthesis extends single-step prediction with tree search algorithms—most notably Monte Carlo tree search and best-first search—to find complete routes from purchasable starting materials (Segler et al., 2018; Chen et al., 2020). Despite the maturity of multi-step retrosynthesis in organic chemistry, analogous frameworks for inorganic solid-state synthesis remain absent.

Precursor prediction. PP is a cornerstone of synthesis planning, tasked with identifying starting materials for target compounds. In the inorganic domain, this process has traditionally relied on density functional theory (DFT) to screen precursors based on thermodynamic stability (Engel & Dreizler, 2011; Aykol et al., 2021), though such calculations often fail to account for practical experimental preferences such as kinetic accessibility (Szymanski & Bartel, 2024). Subsequent work encoded inorganic compositions as numerical vectors (Ward et al., 2016; Wang et al., 2021), enabling retrieval-based frameworks that infer precursors from synthesis precedents of similar targets (He

et al., 2023; Noh et al., 2024), and elementwise representations that rank candidate precursor sets conditioned on target composition (Kim et al., 2024b; Prein et al., 2025a). More recently, LLMs have been applied to the task by leveraging their broad chemical knowledge: Prein et al. (2025b) use them to generate synthetic reaction data that augments limited datasets and improves downstream expert models, while MSP-LLM (Noh et al., 2026) organizes synthesis planning into a chemically consistent decision chain by predicting the material group as an intermediate step, conditioning autoregressive decoding on hierarchical precursor types. While effective in improving precursor selection, all of these approaches still treat reactions as single-step events without explicitly modeling intermediate transformations—a framing particularly inadequate for solid-state synthesis, where stepwise reaction pathways are required to capture the thermodynamic landscapes underlying distinct phase transformations.

Reinforcement learning for LLM reasoning. Beyond fine-tuning on fixed demonstrations, recent work has shown that reinforcement learning with verifiable, outcome-based rewards substantially improves LLM reasoning without requiring a learned reward model (Guo et al., 2025). Group Relative Policy Optimization (GRPO) (Shao et al., 2024), a variant of proximal policy optimization (Schulman et al., 2017) that forgoes a learned critic in favor of a group-normalized baseline over sampled completions, has become a standard recipe for this class of methods. As RL pipelines increasingly combine several rewards capturing distinct purposes, GDPO (Liu et al., 2026) observes that applying GRPO's normalization to combined multi-reward rollouts collapses distinct reward combinations into identical advantage values, degrading the resolution of the training signal, and shows that decoupling the normalization of each individual reward preserves their relative differences and improves both convergence and training stability.

3 PRELIMINARIES

Precursor prediction (PP). The goal of PP is to predict a set of N precursor materials $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$ required to synthesize a target material T, given only its chemical formula. No additional structural or contextual information is provided as input. For example, given the target $\mathrm{LiFeV_2O_7}$, the task is to predict a precursor set such as $\{\mathrm{Li_2CO_3},\ \mathrm{V_2O_5},\ \mathrm{Fe_2O_3}\}$. Since a single target material may be synthesized via multiple valid precursor combinations, evaluation is performed target-wise: a prediction is considered correct if the predicted set matches any one of the ground-truth precursor sets recorded for that target.

Pairwise reaction. A pairwise reaction is a binary solid-state reaction in which exactly two solid reactants combine to form a single product. We denote a pairwise reaction as $\mathbf{p} = (p_1, p_2) \rightarrow t$, where p_1 and p_2 are the reactants and t is the product. For example, $(\mathrm{Li_2CO_3},\ \mathrm{V_2O_5}) \rightarrow \mathrm{LiVO_3}$ and $(\mathrm{LiVO_3},\ \mathrm{Fe_2O_3}) \rightarrow \mathrm{LiFeV_2O_7}$ are both pairwise reactions, where the left-hand side denotes the reactant pair $\mathbf{p}$ and the right-hand side denotes the product t.

Pairwise reactant prediction (PRP). Given a composition t, PRP is the reverse prediction problem of a pairwise reaction: predict a reactant pair $\mathbf{p} = (p_1, p_2)$ such that $p_1 + p_2 \rightarrow t$. Here t may be the final synthesis target T or an intermediate formed during the synthesis route, and likewise p_1 and p_2 may be purchasable starting materials or intermediates produced in earlier steps. For example, PRP of $\mathrm{LiFeV_2O_7}$ as target t may result in $(\mathrm{LiVO_3},\ \mathrm{Fe_2O_3})$ as the reactant pair $\mathbf{p}$.

Pairwise retrosynthesis. Pairwise retrosynthesis is a backward decomposition process that sequentially solves PRPs to identify the precursor set $\mathcal{P}$ for a given target T. Starting from T, each material is split into a reactant pair $\mathbf{p} = (p_1, p_2)$, where each p_i either serves as a final precursor or acts as an intermediate that is recursively decomposed in a subsequent step. The process continues until all materials have been reduced to final precursors, yielding $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$. For example, given the target $\mathrm{LiFeV_2O_7}$, the first PRP step yields the pair $(\mathrm{LiVO_3},\ \mathrm{Fe_2O_3})$. In the next step, $\mathrm{LiVO_3}$ is further decomposed into $(\mathrm{Li_2CO_3},\ \mathrm{V_2O_5})$ while $\mathrm{Fe_2O_3}$ is retained as a final precursor. The resulting precursor set is $\mathcal{P} = \{\mathrm{Li_2CO_3},\ \mathrm{V_2O_5},\ \mathrm{Fe_2O_3}\}$.

4 METHODS

Retro-Forge (Figure 1) comprises three components: (i) a curated pairwise reaction dataset augmented with LLM-generated synthetic data to address data scarcity (Section 4.1), (ii) a pairwise reactant prediction model based on LLMs further optimized with reinforcement learning (Sections 4.2 and 4.3),

and (iii) a multi-step tree search, reranking, and an optional fusion pipeline that composes PRPs into complete synthesis routes (Section 4.3). The following sections describe each component in detail.

4.1 Pairwise Reaction Dataset (PRD) Construction

We construct a ground-truth pairwise reaction dataset from two inorganic solid-state synthesis corpora (Kononova et al., 2019; Lee et al., 2025). Each source is preprocessed to resolve compositional variables and filtered for chemical validity and extraction consistency. The two sources are merged under a common schema. Ambient species (O_2, N_2, CO_2, H_2O) are separated from solid reactants, and only reactions with exactly two solid reactants are retained. After deduplication by canonical element composition, 8,037 pairwise reactions across 6,729 unique targets were obtained. Additional preprocessing details are provided in Appendix A.2.

To address the data scarcity of pairwise reactions, we expand training coverage through a three-tier augmentation pipeline, where the filtered reactions mentioned above form Tier 1—the most rigorously verified and chemically valid pairwise reactions. A GPT-4.1-mini (OpenAI, 2025) model fine-tuned on Tier 1 data generates candidate pairwise reactions for two additional target pools of increasing scale but decreasing verification quality. For Tier 2, we use targets drawn from the original Kononova and Lee corpora (Kononova et al., 2019; Lee et al., 2025), which are materials with documented solid-state synthesis records but without verified pairwise reaction decompositions. For Tier 3, novel targets are drawn from the Materials Project (Jain et al., 2013) by farthest-point sampling over MTEncoder (Prein et al., 2023) composition embeddings, stratified across 17 anion/material-group classes with an equal per-class quota (Appendix A.2). This synthetic data maximizes compositional diversity across chemical space but yields reactions that are purely model-generated and, therefore, noisier. Generated pairs are filtered for chemical validity, element consistency, and stoichiometric balancability between the pair and the target. This yields 17,524 and 32,325 targets, comprising 31,930 and 82,161 pairwise reactions, for Tier 2 and Tier 3 respectively, giving a Pairwise Reaction Dataset (PRD) of approximately 122,000 pairwise reactions across all three tiers.

4.2 Pairwise Reactant Prediction (PRP) Task

For the PRP task, we adopt large language models as the backbone for single-step pairwise reactant prediction, using two open-source model families: Qwen2.5-7B-Instruct (Qwen Team, 2025) and LLaMA-3.1-8B-Instruct (Grattafiori et al., 2024). Both models are fine-tuned on the PRD using a staged fine-tuning strategy that proceeds from large-scale noisy data to high-quality verified data: Tier 3 $\rightarrow$ Tier 2 $\rightarrow$ Tier 1. This ordering exposes the model first to broad chemical space coverage before specializing on verified pairwise reactions. Both models are fine-tuned using QLoRA (Dettmers et al., 2023) (rank 64, targeting both attention and MLP projections), continually tuning the same LoRA adapter across all three stages. Training details and hyperparameters are provided in Appendix B.1.

4.3 Precursor Prediction (PP) Task

For the PP task, we deploy the PRP models within a multi-step tree search that recursively applies pairwise reactant prediction until all leaf nodes correspond to purchasable precursors in the precursor catalog.

Precursor catalog. We define the *precursor catalog* as the set of materials considered purchasable terminal nodes in the tree search. The catalog is built from the raw Kononova and Lee records (Kononova et al., 2019; Lee et al., 2025), filtered by the same chemical-validity criteria as the reaction pipeline (Section 4.1) and restricted to materials containing at most one metal or semimetal, with B and Si excluded from the semimetal count. We then prune *product-dominant* entries—materials appearing strictly more often as a synthesis *product* than as a *precursor* in the raw corpus—since these let search shortcut-terminate at intermediates one would synthesize rather than purchase, leaving 1,938 precursors (Appendix A.3).

Tree search. We formulate multi-step retrosynthesis as a best-first search over partial synthesis routes, following the AND-OR search framework used in organic retrosynthesis (Segler et al., 2018; Chen et al., 2020). Each route r is an AND-OR tree: a material that admits multiple candidate decompositions constitutes an OR node, while each specific decomposition $\mathbf{p} = (p_1, p_2)$ constitutes an AND node whose two children must both be resolved. Leaf materials in a route are classified

as either *terminal nodes* (present in the catalog $\mathcal{C}$) or *frontier nodes* (not present in $\mathcal{C}$). A route is *complete* when all its leaves are terminal, and *partial* when at least one frontier node remains. Search proceeds best-first over cumulative sequence log-probability, repeatedly expanding the highest-scoring partial route by one frontier node until the expansion budget is exhausted, accumulating a pool of complete routes. Since distinct routes can terminate at the same leaf precursor set, we then marginalize over the derivations: complete routes are grouped by leaf set, and each distinct set is scored by the log-sum-exp of the log-probabilities of its derivations. Search thus returns a ranked list of distinct precursor sets, each retaining its constituent derivations. The full procedure is given in Appendix B.3.

Tree-RL fine-tuning. Each PRP model is further fine-tuned with multi-step tree-RL. Rather than optimizing a single decomposition step, the policy rolls out complete retrosynthesis trees and is rewarded on the terminal outcome, with the reward shared by every decomposition step along the route. The reward combines three terms: a **match** term for whether the route's leaf precursor set equals a ground-truth precursor set, a **balance** term for the fraction of stoichiometrically balanced steps, and a **diversity** term that discourages mode collapse across the routes sampled for a target. Following GDPO (Liu et al., 2026), each term is normalized within its sampled group before aggregation, and a KL penalty to the pre-RL model regularizes training. Reward weights, the training objective, and the weight ablation are given in Appendices B.2 and D.

Reranking and fusion. The candidate precursor sets returned by search are reranked by a gradient-boosted tree classifier (Friedman, 2001) trained per PRP model to predict whether a set matches a ground-truth recipe (Appendix B.3). Optionally, when more than one PRP model is available, their reranked candidate-set lists can be merged via reciprocal rank fusion (RRF, $K = 60$), with each candidate set scored by the sum of $1/(K + \text{rank})$ across the models' rankings. Finally, within each candidate set, we output the *most stoichiometrically balanced* derivation as the predicted route.

4.4 Evaluation Setup

Target-wise top-k metric. Given a target T with ground-truth precursor sets $\{\mathcal{P}_1^*, \mathcal{P}_2^*, \ldots\}$, a prediction is considered correct at rank k if any one of the ground-truth sets is matched by any of the top-k predicted precursor sets $\hat{\mathcal{P}}_1, \ldots, \hat{\mathcal{P}}_k$. We refer to this base metric, computed on the leaf set of the predicted route, as **leaf**; it is the only variant directly comparable to single-step methods that predict a flat precursor set with no synthesis tree, and is used for every baseline comparison. We additionally report **frontier** top-k, which credits a route if any valid tree-cut—taking some intermediate nodes as purchasable rather than decomposing them further—matches a ground-truth precursor recipe. Using the example from Figure 1(b): a two-step route for $\mathrm{LiFeV_2O_7}$ yields leaf set $\{\mathrm{Li_2CO_3}, \mathrm{V_2O_5}, \mathrm{Fe_2O_3}\}$, but the frontier cut that treats $\mathrm{LiVO_3}$ as purchasable gives $\{\mathrm{LiVO_3}, \mathrm{Fe_2O_3}\}$; the route is credited under frontier if either cut matches. Because the cut is evaluated on the single most stoichiometrically balanced representative route, ranking order is well defined. Both metrics are reported at $k \in \{1, 3, 5, 10\}$, with matching performed at the stoichiometric level. We rank candidates either by the policy's own sequence log-probability (**raw**) or after set aggregation and reranking (**agg+rerank**, Section 4.3).

Multi-step evaluation targets. We construct an out-of-distribution evaluation benchmark from Inorganic Crystal Structure Database (ICSD) entries published in 2025, postdating the knowledge cutoffs of all models used for prediction and augmentation. Candidate targets must contain at least one metal or semimetal (B and Si excluded). We extracted 2,527 candidate solid-state syntheses from 1,278 source articles via an LLM pipeline (prompt validated on 65 records), then curated with ICSD-consistency, flux-aware, and chemical-validity filters, yielding **857 unique targets**. Full construction details and prompt templates are in Appendices A.4 and H.

A secondary, in-distribution benchmark derived from a held-out split of the Kononova–Lee corpus is described and evaluated in Appendix E.

5 Experiments

We evaluate Retro-Forge on both the PRP and PP tasks, comparing against MSP-LLM and four additional precursor-prediction baselines on the PP task, ablate the two training components—data augmentation and tree-RL—on both tasks, and independently assess the stoichiometric and

Table 1: Standalone PRP task on the Tier-1 held-out test set (686 unique targets, target-wise top-k, %). **Bold** = best per column.

System	@1	@3	@5	@10
LLaMA, pre-RL	**82.9**	88.6	**92.0**	93.7
LLaMA, tree-RL	80.2	87.0	89.4	91.8
Qwen, pre-RL	81.2	**90.1**	**92.0**	**94.5**
Qwen, tree-RL	80.5	88.5	90.4	92.9

Table 2: Precursor prediction on 857 novel targets, top-k (%). Retro-Forge cells show leaf / frontier; baselines are leaf only. **Bold** = best leaf per column.

Method	@1	@3	@5	@10
Retro-Forge (agg+rerank)				
LLaMA (single)	57.8 / 58.2	63.1 / 63.8	**65.9** / 67.0	67.1 / 68.3
Qwen (single)	56.9 / 57.4	**64.6** / 65.8	**65.9** / 67.2	67.0 / 68.5
Fused (LLaMA+Qwen)	**58.7** / 59.2	63.7 / 64.3	65.6 / 66.3	**67.6** / 68.8
Baselines				
MSP-LLM (Noh et al., 2026)	52.9	62.3	63.8	66.9
ElemwiseRetro (Kim et al., 2024b)	51.8	55.7	56.5	56.8
CSLLM (Song et al., 2025)[†]	51.8	—	—	—
SynthGPT (Kim et al., 2024a)	47.8	54.7	55.9	56.7
Retrieval-Retro (Noh et al., 2024)	36.9	38.3	38.9	40.3

[†]CSLLM outputs a single precursor set; top-k for $k>1$ is not applicable.

thermodynamic validity of predicted routes. Details of baseline reproduction are given in Appendix C; full training, search, and reranking configurations are given in Appendices B.1–B.3.

5.1 Pairwise Reactant Prediction (PRP) Task

We evaluate the PRP model in isolation on the Tier-1 held-out test set (806 reactions, 686 unique targets), measuring target-wise top-k: a target succeeds if any of the top-k predicted precursor pairs matches any ground-truth precursor pair recorded for that target.

Table 1 shows that single-step accuracy is largely retained after tree-RL: tree-RL models remain within 0.7–2.7 points of their pre-RL counterparts at every k, even though the reward is defined only on the leaf set of a rolled-out tree. We analyze this trade-off, and the contribution of each augmentation tier, in Section 5.4.

5.2 Precursor Prediction (PP) Task

Each backbone is deployed within the full inference pipeline of Section 4.3 on the 857 novel evaluation targets of Section 4.4.

Table 2 compares Retro-Forge against existing models on the 857-target benchmark; baseline reproduction is detailed in Appendix C. Retro-Forge leads every baseline at every k even with a single model backbone. The lead widens under the frontier metric, which also credits routes that pass through a ground-truth precursor set but decompose it further into simpler precursors.

5.3 Route Chemistry Validity

Top-k measures whether Retro-Forge recovers a documented precursor set, but not whether an undocumented predicted route is itself chemically sound. We evaluate route quality along two independent axes beyond exact match: stoichiometric balance and thermodynamic feasibility. Table 3 summarizes both on the 857-target benchmark.

Stoichiometric balance. A pairwise step is *stoichiometrically balanced* if its two reactants exactly metal-balance the step-target (non-negative least-squares (NNLS) residual $< 10^{-6}$; O, C, H, N treated

Table 3: Route chemistry validity on 857 novel targets (%). **Stoich. balance**: NNLS check. **Thermo**: ARROWS3 $\Delta G_{\text{rxn}} < 0$ at 900°C, over ΔG-computable steps only ($\sim$85%).

	Stoich. balance		Thermo (900°C)	
System	**Per-step**	**Fully bal.**	**Steps fav.**	**Routes fully fav.**
LLaMA (raw top-1)	96.1	91.0	**72.9**	**53.0**
Qwen (raw top-1)	95.9	90.4	69.8	52.4
Fused (agg+rerank)	**98.8**	**97.4**	72.2	52.7

Table 4: Augmentation curriculum $\times$ tree-RL, LLaMA (%). **PRP**: Tier-1 held-out top-1 and the rate of stoichiometrically balanced top-1 pairs. **PP**: 857 novel targets; leaf top-k under raw and agg+rerank (abbreviated agg), and stoichiometric balance of the route pool.

		PRP		PP				
Stage	**Curriculum**	**@1**	**Bal.**	**raw @1**	**agg @1**	**agg @10**	**Step Bal.**	**Route Bal.**
pre-RL	Tier 1 only	79.6	94.8	37.5	54.8	63.2	63.3	13.0
	Tier 2→1	81.5	97.2	43.1	53.8	65.6	73.3	28.3
	Full	**82.9**	98.8	46.7	53.9	66.2	80.1	38.7
tree-RL	Tier 1 only	75.9	96.8	**56.2**	57.5	66.5	76.0	45.9
	Tier 2→1	79.4	99.0	55.4	**58.6**	66.4	83.4	57.8
	Full	80.2	**99.7**	54.3	57.2	**67.1**	**92.7**	**76.0**

as free byproducts). We report the per-step balanced rate and the fraction of routes in which *every* step is balanced. Raw top-1 routes of each backbone already exceed 95% per-step and 90% fully balanced. The full pipeline (aggregation, reranking, fusion, and balance-based route selection) raises the fused top-1 to 98.8% per-step and 97.4% fully balanced.

Thermodynamic feasibility. A stoichiometrically balanced step may still be thermodynamically uphill. Following the ARROWS3 formulation (Szymanski et al., 2023a), we compute each step's per-atom reaction Gibbs energy from temperature-corrected DFT formation energies and call the step favorable if $\Delta G_{\text{rxn}} < 0$ at 900°C. Over the $\sim$85% of steps with a computable ΔG (Appendix F), 70–73% of steps and 52–53% of routes are favorable, and this is sustained after reranking and fusion.

5.4 Effect of Data Augmentation and Tree-RL

We isolate the two training components with a factorial ablation on the LLaMA backbone: each of three augmentation curricula (Tier 1 only; Tier 2→1; full Tier 3→2→1) is evaluated before and after tree-RL trained with the identical recipe (Table 4). For PP, stoichiometric balance (Section 5.3) is measured on the **route pool**: all complete candidate routes the search returns for a target ($\sim$120–150), before aggregation and reranking. We use the pool rather than the selected top-1 because reranking and balance-based route selection push top-1 balance toward its ceiling for every arm (top-1 values are in Appendix D.2).

Data augmentation. Reading down each block, augmentation improves the single-step predictor monotonically at both stages: PRP top-1 rises by 3.3 points pre-RL and 4.3 points after tree-RL, and the rate of balanced top-1 pairs rises alongside it. On the multi-step task, augmentation raises the pre-RL raw top-1 from 37.5 to 46.7, and it improves stoichiometric balance at both stages, with the fully balanced share of the route pool rising from 13.0 to 38.7% pre-RL and from 45.9 to 76.0% after tree-RL. Once tree-RL and reranking are applied, top-k differences between curricula fall within 1–2 points, the single-seed noise floor, while the balance gains remain large and monotonic. Augmentation therefore improves multi-step performance chiefly by making the generated routes chemically sound, which is why we adopt the full curriculum for the final model.

Tree-RL. Reading across blocks, the small single-step cost seen in Section 5.1 holds at every curriculum (2.1–3.7 points of PRP top-1), while PP raw top-1 rises by 7.6–18.7 points and the fully balanced share of the route pool roughly doubles. This is the pattern expected when training on the complete route reduces error propagation: individual pairwise predictions change little, while the fraction of routes whose steps compose correctly into a full, balanced synthesis rises sharply. The two

Table 5: Calcination and sintering temperature prediction (MAE, K). **ICSD-2025**: the 181-reaction temperature-labeled subset of the 857-target benchmark.

	Kononova		ICSD-2025	
Method	**Calc.**	**Sint.**	**Calc.**	**Sint.**
SyntMTE (Prein et al., 2025b)	160.0	138.0	151.9	139.4
Ours (route-augmented)	**133.4**	**104.9**	**123.8**	**95.6**

components are complementary. The full curriculum with tree-RL attains the highest pool balance (92.7% step, 76.0% route) and the best agg+rerank top-10, and it is the configuration we use in the final model. Full-k tables for both tasks and both backbones are given in Appendix D.2.

5.5 Calcination and Sintering Temperatures

Once precursors are chosen, calcination and sintering temperatures are the most consequential processing parameters in solid-state synthesis, governing which phases form, how completely the reaction proceeds, and the density and microstructure of the product. As a preliminary case study, we test whether the intermediates in predicted routes help predict these temperatures. A gradient-boosted regressor (Ke et al., 2017) predicts both from the target's embedding under a Masked Precursor Completion (MPC) encoder (He et al., 2023), the temperatures of recipes retrieved in that embedding space from a 33,371-record Kononova–Lee recipe–temperature corpus, and predicted melting points. We retrain MPC with 28,301 pairwise steps extracted from top-1 Retro-Forge routes of the corpus targets (Appendix G). We evaluate on the 918-reaction Kononova test split and on the 181 reactions of our 857-target benchmark that report calcination or sintering temperatures (Appendix A.4). Table 5 shows that our predictor lowers MAE relative to SyntMTE (Prein et al., 2025b) by 27–44 K on both sets.

6 Conclusion

Summary. We presented Retro-Forge, the first framework to formulate precursor prediction for solid-state synthesis as a sequence of learnable pairwise reactions composed by tree search, evaluated on targets novel to both the training corpus and the pretraining data of the underlying models. Two findings indicate that the pairwise formulation is not merely a convenient decomposition but a faithful one: augmentation that improves single-step prediction carries through to multi-step planning, most visibly in the chemical validity of the generated routes, and tree-RL on complete routes improves route accuracy and validity while leaving single-step accuracy largely intact. Both follow from the fact that a correct route is by construction a composition of correct pairwise reactions, so pairwise gains accumulate across the route; the resulting system surpasses every existing precursor-prediction model at every rank on our 857-target benchmark, with routes that are stoichiometrically balanced and thermodynamically favorable under DFT-derived reaction energies.

Limitations and future work. This validation is computational: stoichiometric balance and thermodynamic favorability establish that a route is chemically sound, but kinetics, phase competition, and processing conditions are determined in the laboratory. The most direct next step is experimental validation, for which our thermodynamic screening offers a principled basis for prioritizing candidates. Beyond validation, our formulation makes experimental feedback directly usable as training signal: the policy is already optimized with an outcome-based reward on complete routes, so a synthesis attempt—successful or otherwise—is the same kind of signal as a corpus match, and enters training the same way. A second limitation concerns the scope of what the framework predicts. We show that its routes also inform calcination and sintering temperature prediction (Section 5.5), but a complete synthesis plan additionally specifies the sequence of operations that realizes a route; predicting these operations alongside the precursors and temperatures would bring the framework to a complete materials synthesis planning system.

By casting synthesis planning as an outcome-optimized policy rather than a mapping fit to recorded recipes, Retro-Forge can continue to improve through such feedback where no recipes exist—precisely where AI-driven materials discovery now proposes its candidates, and where experimentally actionable synthesis plans are most needed and least available.

AI USE STATEMENT

In this work, we used generative AI tools to generate synthetic datasets, to clean and reformat dataset entries, to design and give feedback on research methodology and experiments, to implement methods, and to interpret results. Synthetic data generation is a core component of our method: a fine-tuned GPT-4.1-mini model produces the candidate pairwise decompositions that constitute Tiers 2 and 3 of our training corpus (Section 4.1). All generated reactions were filtered programmatically for chemical validity, element consistency with the target, charge neutrality, and stoichiometric balance before use, and the filtering criteria and pass rates are reported in Appendix A.2; no model-generated reaction enters the training corpus unverified. We additionally used a large language model for structured extraction of synthesis information from published papers when constructing the evaluation benchmark (Section 4.4), with extracted entries subject to the same validity filters. Our use of AI tools to implement methods took the form of code writing, with the resulting code reviewed by the authors.

We did not use generative AI tools to propose or refine hypotheses. We did not use them to develop theoretical models or conceptual frameworks, to formulate mathematical claims or assist in the writing of proofs, or for translation or qualitative and thematic data analysis; these tasks are not applicable to this work.

Additionally, we used generative AI tools for tasks under recommended disclosure: writing and editing software code, creating the figures in this paper, drafting and editing portions of the manuscript text, suggesting document structure, identifying and summarizing related literature, and formatting bibliographic references.

We have reviewed all AI-assisted work. All AI-generated training data passed the automated chemical-validity pipeline described above; all AI-written code was reviewed by the authors and its outputs verified against the reported results; all literature identified with AI assistance was verified against the primary sources; and all interpretations, manuscript text, and figures were reviewed and edited by the authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work does not involve human subjects, personal data, or crowdsourced annotation. Our training corpus derives from two publicly released text-mined synthesis datasets (Kononova et al., 2019; Lee et al., 2025) and the Materials Project (Jain et al., 2013), each used in accordance with its published license.

Constructing our evaluation benchmark required extracting synthesis information from primary literature, which raises questions of publisher copyright that we addressed as follows. All articles were obtained through institutional subscription access and handled in accordance with our institutions' policies on licensed content. Only the portions of each article describing the synthesis procedure were supplied to a language model for structured extraction, and no article text was redistributed at any point. The resulting benchmark records factual experimental information—target compositions, precursor identities, and reaction conditions—together with a citation to the source article, rather than any portion of the articles' text; this factual data is what we report and release.

REPRODUCIBILITY STATEMENT

We release the full pipeline needed to reproduce every result in this paper.

Code. All code—pairwise-reaction dataset construction and chemical-validity filters, staged fine-tuning, tree-RL, multi-step search, reranking, fusion, evaluation, and baseline reproduction—is available at `https://anonymous.4open.science/r/_Retro-Forge` (MIT license), with step-by-step reproduction instructions in the repository README.

Model weights. Both final backbones are released as LoRA adapters (the staged-fine-tuning adapter plus the tree-RL adapter at step 1,800), split into two repositories each carrying its base model's license: `https://anonymous-hf.com/a/v5mfz5viox5b/` (LLaMA-3.1-8B) and `https://anonymous-hf.com/a/rswiurlxcn1z/` (Qwen2.5-7B). Applied to the stock base models

they reproduce the reported system without redistributing base weights; the two per-backbone set-rerankers are released with the code.

Data. The 857-target evaluation benchmark—each entry annotated with its source DOI—and the 1,938-entry precursor catalog are released with the code, so inference and evaluation run without any access to raw data. The training corpus (the Tier-1 pairwise reactions and the Tier-2/3 augmented tiers) is derived from the publicly released Kononova and Lee corpora (Kononova et al., 2019; Lee et al., 2025) and the Materials Project (Jain et al., 2013); rather than redistribute it, we release the construction pipeline that rebuilds it locally from those original sources.

Method details. Dataset construction and all chemical-validity filters, including the charge-neutrality algorithm, are in Appendices A.2 and A.1; staged fine-tuning and tree-RL configurations in Appendices B.1 and B.2; multi-step search, reranking, and fusion in Appendix B.3; baseline reproduction in Appendix C; full ablation tables in Appendix D; the thermodynamic-validity methodology of Section 5.3 in Appendix F; and all prompt templates in Appendix H.

REFERENCES

Muratahan Aykol, Joseph H. Montoya, and Jens Hummelshøj. Rational solid-state synthesis routes for inorganic materials. *Journal of the American Chemical Society*, 143(24):9244–9259, 2021. doi: 10.1021/jacs.1c04888.

Binghong Chen, Chengtao Li, Hanjun Dai, and Le Song. Retro*: Learning retrosynthetic planning with neural guided A* search. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 1608–1616, 2020. doi: 10.48550/arXiv.2006.15820.

Mouyang Cheng, Weiliang Luo, Hao Tang, Bowen Yu, Yongqiang Cheng, Weiwei Xie, Ju Li, Heather J. Kulik, and Mingda Li. Enhancing materials discovery with valence constrained design in generative modeling. *Nature Computational Science*, 2026. doi: 10.1038/s43588-026-01037-2.

Vincent Chung, Aron Walsh, and David J. Payne. Solid-state synthesizability predictions using positive-unlabeled learning from human-curated literature data. *Digital Discovery*, 4:2439–2453, 2025. doi: 10.1039/d5dd00065c.

E. J. Corey and W. T. Wipke. Computer-assisted design of complex organic syntheses. *Science*, 166: 178–192, 1969. doi: 10.1126/science.166.3902.178.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. QLoRA: Efficient finetuning of quantized LLMs. In *Advances in Neural Information Processing Systems*, volume 36, pp. 10088–10115, 2023. doi: 10.48550/arXiv.2305.14314.

E. Engel and R.M. Dreizler. *Density Functional Theory: An Advanced Course*. Theoretical and Mathematical Physics. Springer Berlin Heidelberg, 2011. ISBN 9783642140907.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Vaughan, et al. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. doi: 10.48550/arXiv.2407.21783.

Daya Guo, Dejian Yang, Haowei Zhang, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645:633–638, 2025. doi: 10.1038/s41586-025-09422-z.

Tanjin He, Haoyan Huo, Christopher J. Bartel, Zheren Wang, Kevin Cruse, and Gerbrand Ceder. Precursor recommendation for inorganic synthesis by machine learning materials similarity from scientific literature. *Science Advances*, 9(23):eadg8180, 2023. doi: 10.1126/sciadv.adg8180.

Qi-Jun Hong. A melting temperature database and a neural network model for melting temperature prediction. *arXiv preprint arXiv:2110.10748*, 2021. doi: 10.48550/arXiv.2110.10748.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013. doi: 10.1063/1.4812323.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3149–3157, 2017.

Seongmin Kim, Yousung Jung, and Joshua Schrier. Large language models for inorganic synthesis predictions. *Journal of the American Chemical Society*, 146(29):19654–19659, 2024a. doi: 10.1021/jacs.4c05840.

Seongmin Kim, Juhwan Noh, Geun Ho Gu, Shuan Chen, and Yousung Jung. Predicting synthesis recipes of inorganic crystal materials using elementwise template formulation. *Chemical Science*, 15(3):1039–1045, 2024b. doi: 10.1039/D3SC03538G.

Olga Kononova, Haoyan Huo, Tanjin He, Ziqin Rong, Tiago Botari, Wenhao Sun, Vahe Tshitoyan, and Gerbrand Ceder. Text-mined dataset of inorganic materials synthesis recipes. *Scientific Data*, 6:203, 2019. doi: 10.1038/s41597-019-0224-1.

A. B. Kuznetsov, A. Y. Jamous, V. A. Svetlichnyi, M. I. Rakhmanova, N. E. Bogdanov, A. E. Kokh, V. S. Shevchenko, and K. A. Kokh. From phase diagram to functional properties: $TbGa_3(BO_3)_4$ crystal growth and its stoichiometry-dependent optical behavior. *Journal of Alloys and Compounds*, 1040:183562, 2025. doi: 10.1016/j.jallcom.2025.183562.

Sanghoon Lee, Kevin Cruse, Viktoriia Baibakova, Gerbrand Ceder, and Anubhav Jain. Text-mined dataset of solid-state syntheses with impurity phases using large language model. *Scientific Data*, 12:1969, 2025. doi: 10.1038/s41597-025-06222-y.

Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, and Vijay Pande. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science*, 3(10):1103–1113, 2017. doi: 10.1021/acscentsci.7b00303.

Shih-Yang Liu, Xin Dong, Ximing Lu, Shizhe Diao, Peter Belcak, Mingjie Liu, Min-Hung Chen, Hongxu Yin, Yu-Chiang Frank Wang, Kwang-Ting Cheng, Yejin Choi, Jan Kautz, and Pavlo Molchanov. GDPO: Group reward-decoupled normalization policy optimization for multi-reward RL optimization. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. doi: 10.48550/arXiv.2601.05242.

Inbal Magar and Roy Schwartz. Data contamination: From memorization to exploitation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 157–165, 2022. doi: 10.18653/v1/2022.acl-short.18.

Thomas F. Malkowski, Robert L. Sacci, Rebecca D. McAuliffe, Shree Ram Acharya, Valentino R. Cooper, Nancy J. Dudney, and Gabriel M. Veith. Role of pairwise reactions on the synthesis of $Li_{0.3}La_{0.57}TiO_3$ and the resulting structure–property correlations. *Inorganic Chemistry*, 60(19): 14831–14843, 2021. doi: 10.1021/acs.inorgchem.1c02136.

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624:80–85, 2023. doi: 10.1038/s41586-023-06735-9.

Akira Miura, Christopher J. Bartel, Yosuke Goto, Yoshikazu Mizuguchi, Chikako Moriyoshi, Yoshihiro Kuroiwa, Yongming Wang, Toshie Yaguchi, Manabu Shirai, Masanori Nagao, Nataly Carolina Rosero-Navarro, Kiyoharu Tadanaga, Gerbrand Ceder, and Wenhao Sun. Observing and modeling the sequential pairwise reactions that drive solid-state ceramic synthesis. *Advanced Materials*, 33 (24):2100312, 2021. doi: 10.1002/adma.202100312.

Heewoong Noh, Namkyeong Lee, Gyoung S. Na, and Chanyoung Park. Retrieval-Retro: Retrieval-based inorganic retrosynthesis with expert knowledge. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.48550/arXiv.2410.21341.

Heewoong Noh, Gyoung S. Na, Namkyeong Lee, and Chanyoung Park. MSP-LLM: A unified large language model framework for complete material synthesis planning. *arXiv preprint arXiv:2602.07543*, 2026. doi: 10.48550/arXiv.2602.07543.

OpenAI. GPT-4.1. `https://openai.com/index/gpt-4-1/`, 2025. Accessed: 2025.

OpenAI. GPT-5.5. https://openai.com/index/introducing-gpt-5-5/, 2026. Accessed: 2026.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Thorben Prein, Elton Pan, Tom Doerr, Elsa Olivetti, and Jennifer L.M. Rupp. MTEncoder: A multi-task pretrained transformer encoder for materials representation learning. In *AI for Accelerated Materials Design - NeurIPS 2023 Workshop*, 2023. URL https://openreview.net/forum?id=wug7i3O7y1.

Thorben Prein, Elton Pan, Sami Haddouti, Marco Lorenz, Janik Jehkul, Tymoteusz Wilk, Cansu Moran, Menelaos Panagiotis Fotiadis, Artur P. Toshev, Elsa Olivetti, and Jennifer L. M. Rupp. Retro-Rank-In: A ranking-based approach for inorganic materials synthesis planning. *arXiv preprint arXiv:2502.04289*, 2025a. doi: 10.48550/arXiv.2502.04289.

Thorben Prein, Elton Pan, Janik Jehkul, Steffen Weinmann, Elsa A. Olivetti, and Jennifer L. M. Rupp. Language models enable data-augmented synthesis planning for inorganic materials. *ACS Applied Materials & Interfaces*, 17(51):69221–69233, 2025b. doi: 10.1021/acsami.5c11229.

Qwen Team. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2025. doi: 10.48550/arXiv.2412.15115.

Banik Rai, Sandip Kumar Kuila, Rana Saha, Sankalpa Hazra, Chandan De, Jyotirmoy Sau, Venkatraman Gopalan, Partha Pratim Jana, Stuart S. P. Parkin, and Nitesh Kumar. Peculiar magnetic and magneto-transport properties in a noncentrosymmetric self-intercalated van der Waals ferromagnet Cr_5Te_8. *Chemistry of Materials*, 37(2):746–755, 2025. doi: 10.1021/acs.chemmater.4c02996.

Oscar Sainz, Jon Campos, Iker García-Ferrero, Julen Etxaniz, Oier Lopez de Lacalle, and Eneko Agirre. NLP evaluation in trouble: On the need to measure LLM data contamination for each benchmark. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 10776–10787, 2023. doi: 10.18653/v1/2023.findings-emnlp.722.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347.

Marwin H. S. Segler, Mike Preuss, and Mark P. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 555:604–610, 2018. doi: 10.1038/nature25978.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. doi: 10.48550/arXiv.2402.03300.

Zhilong Song, Shuaihua Lu, Minggang Ju, Qionghua Zhou, and Jinlan Wang. Accurate prediction of synthesizability and precursors of 3D crystal structures via large language models. *Nature Communications*, 16:6530, 2025. doi: 10.1038/s41467-025-61778-y.

Wenhao Sun and Nicholas David. A critical reflection on attempts to machine-learn materials synthesis insights from text-mined literature recipes. *Faraday Discussions*, 256:614–638, 2025. doi: 10.1039/D4FD00112E.

Nathan J. Szymanski and Christopher J. Bartel. Computationally guided synthesis of battery materials. *ACS Energy Letters*, 9(6):2902–2911, 2024. doi: 10.1021/acsenergylett.4c00821.

Nathan J. Szymanski, Pragnay Nevatia, Christopher J. Bartel, Yan Zeng, and Gerbrand Ceder. Autonomous and dynamic precursor selection for solid-state materials synthesis. *Nature Communications*, 14:6956, 2023a. doi: 10.1038/s41467-023-42329-9.

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson, Yan Zeng, and Gerbrand Ceder. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624:86–91, 2023b. doi: 10.1038/s41586-023-06734-w.

Anthony Yu-Tung Wang, Steven K Kauwe, Ryan J Murdock, and Taylor D Sparks. Compositionally restricted attention-based network for materials property predictions. *npj Computational Materials*, 7:77, 2021. doi: 10.1038/s41524-021-00545-1.

Logan Ward, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2:16028, 2016. doi: 10.1038/npjcompumats.2016.28.

Si-Da Xue and Qi-Jun Hong. Materials properties prediction (MAPP): Empowering the prediction of material properties solely based on chemical formulas. *Materials*, 17(17):4176, 2024. doi: 10.3390/ma17174176.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, Ryota Tomioka, and Tian Xie. A generative model for inorganic materials design. *Nature*, 639:624–632, 2025. doi: 10.1038/s41586-025-08628-5.

A DATASET

A.1 CHEMICAL VALIDITY FILTERS

The following formula-level validity filters are applied at every stage of the pipeline—during pairwise reaction dataset construction, augmented data generation, and precursor catalog construction—to remove entries with chemically invalid or computationally intractable formulas.

Notational filters. Formulas are rejected for: non-physical subscripts (negative or zero); disallowed characters or empty parentheses left by failed variable substitution; unresolved or unresolvable symbols, which fall into three categories—mathematical variables (x, y, δ) remaining after substitution failures, generic element placeholders referring to classes of elements (e.g., "Ln" for lanthanides), and parsing artifacts where a variable has been concatenated with an element symbol (e.g., "Wx", "Ok"). Abbreviated chemical groups such as isopropoxide ("i-OPr") are expanded to their full molecular formulas before any further filtering.

Charge neutrality. All retained formulas must additionally pass the charge-neutrality check. The filter labels each formula with oxidation states using a breadth-first search (BFS) algorithm adapted from CrysVCD (Cheng et al., 2026), and accepts a compound only if a charge-neutral (or near-neutral) assignment exists; compounds that cannot be labeled are discarded as chemically invalid.

The algorithm operates over a vocabulary of oxidation states spanning all 118 elements (393 (element, oxidation-state) tokens, including a neutral 0 state for the 105 elements that admit an elemental/metallic form), covering main-group elements, transition metals, and the full lanthanide and actinide series. Each token corresponds to a specific ionic species (e.g., Fe^{2+}, Fe^{3+}, O^{2-}), except the 0 token, which denotes the neutral (elemental) form and is used only by the alloy pass below. A compound's oxidation-state assignment maps each atom to one of its element's allowed tokens. Given a formula with element counts $\{(e_i, c_i)\}$, the algorithm attempts to assign oxidation states that make the absolute charge residual satisfy $|Q| \leq 0.1$, where $Q = \sum_i c_i \cdot v_i$, via the three-stage search below. This threshold is relaxed from a strict $|Q| = 0$ requirement specifically to accommodate the non-integer element counts introduced by retaining fractional and doped stoichiometries, while still rejecting compounds with substantial charge imbalance under every assignment as chemically invalid.

1. **Pass 0 (alloy / elemental):** A formula is treated as an alloy when it is single-element, or when *every* constituent element admits the neutral 0 oxidation state (i.e. each element can occur in metallic/elemental form). Such formulas are labeled directly with oxidation state 0 for every atom and accepted immediately, bypassing the two ionic passes. This also handles elemental molecular solids such as I_2, S_8, N_2. Only this pass uses the neutral 0 tokens; Passes 1 and 2 restrict each element to its *non-zero* oxidation states.
2. **Pass 1 (simple):** Assigns one (non-zero) oxidation state per element. The algorithm performs BFS over elements, trying each allowed non-zero oxidation state v for element e_i and accumulating a running charge $Q = \sum_i c_i \cdot v_i$. Branches are pruned when, at the last element, its oxidation state has the same sign as the running charge (and thus cannot neutralize it). The formula is accepted if any leaf achieves the neutrality criterion.
3. **Pass 2 (mixed-valence):** If Pass 1 fails, the algorithm expands each individual atom rather than each element, allowing different oxidation states for the same element (e.g., Fe_3O_4 requires both Fe^{2+} and Fe^{3+}). A consistency constraint forbids assigning both positive and negative oxidation states to the same element. For fractional stoichiometries, each element count is split into unit-weight atoms plus a single fractional-weight slot (e.g., $Li_{0.2}Mn_{1.8}O_4 \rightarrow \mathrm{Mn}(1.0), \mathrm{Mn}(0.8), \mathrm{Li}(0.2), \mathrm{O}(1.0) \times 4$), with each slot contributing weight $\times$ valence to Q. The search frontier is capped at 10,000 candidate compositions to bound computational cost.

Stoichiometric balancability criterion. The filters above are per-compound properties applied to individual formula strings. We additionally define a *reaction-level* criterion that governs whether a proposed pairwise reaction $(p_1,\ p_2) \rightarrow t$ is stoichiometrically self-consistent.

Given element count vectors $\mathbf{a}_{p_1},\ \mathbf{a}_{p_2} \in \mathbb{R}^E$ for the two reactants and $\mathbf{b}_t \in \mathbb{R}^E$ for the target (over the full E-dimensional element space), we ask whether non-negative mixing coefficients $c_1, c_2 \geq 0$ exist such that $c_1\mathbf{a}_{p_1} + c_2\mathbf{a}_{p_2} = \mathbf{b}_t$ holds up to free byproducts.

Free byproducts. Oxygen, carbon, hydrogen, and nitrogen (O, C, H, N) participate in solid-state reactions as gaseous phases (O_2, CO_2, H_2O, NH_3) and their surplus or deficit is not penalized. These

four elements are excluded from both sides before balancing; all remaining elements—the *metal-only* subset $E' \subset E$—must balance exactly.

NNLS formulation. Let $A \in \mathbb{R}^{E' \times 2}$ be the matrix whose columns are the metal-only element vectors of p_1 and p_2, and let $\mathbf{b} \in \mathbb{R}^{E'}$ be the metal-only element vector of t. We solve

$$\mathbf{c}^* = \arg\min_{\mathbf{c} \geq \mathbf{0}} \|A\mathbf{c} - \mathbf{b}\|_2,$$

and accept the reaction if the relative residual satisfies

$$\frac{\|A\mathbf{c}^* - \mathbf{b}\|_2}{\|\mathbf{b}\|_2 + 10^{-12}} < 10^{-6}.$$

A.2 Pairwise Reaction Dataset Construction

Our pairwise reaction dataset is constructed from two inorganic synthesis datasets (Kononova et al., 2019; Lee et al., 2025). The two sources are preprocessed independently to accommodate their distinct structures, then unified under a common schema before pair extraction.

Source-specific preprocessing. For the dataset from Kononova et al., we resolve element and amount variables using the "element_substitution", "amounts_vars", and "targets_string" fields provided in the original metadata. For the dataset from Lee et al., element substitutions are parsed from inline annotations in the reaction strings. Entries with variables that cannot be resolved are discarded. In addition, this dataset stores multiple reactions per source publication in a single entry, which we flatten into individual entries.

Chemical validity filters. All formula strings are passed through the pipeline-wide validity filters of Appendix A.1. Additionally, self-reactions whose reactant and product sets are identical are removed. Ambient species (O_2, N_2, CO_2, H_2O) are separated and removed from the solid reactants and products, as they do not serve as primary solid-phase reactants in solid-state synthesis. Elemental coverage of the target by its precursors is recorded as a flag and used as a tiebreak during deduplication, but does not remove Tier 1 reactions at this stage.

Extraction-error filter. We additionally drop reactions whose formula *string* parses to a different composition than the composition *stored* in the source record, under a tolerance of $\mathrm{atol} = 0.005 + \mathrm{rtol} = 0.02 \cdot |b|$ on the stored value b. This removes 27 of 8,064 candidate Tier 1 reactions, all confirmed genuine extraction errors (phase/purity tags, single-element duplication artifacts, ratio errors).

Merging and pair extraction. The two preprocessed datasets are converted to a common schema and merged into a single reaction pool. From this pool, we retain reactions with exactly two solid reactants, as required by our pairwise formulation. We then deduplicate by constructing a canonical reaction key from the element-composition representation of both sides, which collapses notational variants of the same chemistry into a single entry. When multiple entries share the same key, a single representative is retained.

After all processing steps, the resulting Tier 1 corpus comprises 8,064 pairwise reactions over 6,740 unique targets; removing the 27 string/composition extraction errors of the preceding filter yields the ground-truth 8,037 pairwise reactions over 6,729 targets.

Tiers 2 and 3: augmented data generation. The augmentation model is a fine-tuned `gpt-4.1-mini-2025-04-14`, trained via the OpenAI fine-tuning API on a multi-plan reformulation of Tier 1: each training example asks the model to predict N pairwise decompositions for a target in one call, where N is that target's true number of distinct recorded routes in Tier 1 (1 plan: 89.2% of targets; 2 plans: 7.2%; 3 or more: 3.6%, with a maximum of 27 recorded routes for $LiMn_2O_4$), and the target output is N newline-separated pairs. At augmentation-generation time, the fine-tuned model is instead prompted for a fixed $N = 5$ candidate plans per Tier 2/3 target. Tier 2 targets are 37,718 products from the Kononova and Lee corpora that lie outside Tier 1: materials with documented solid-state synthesis records but no verified pairwise decomposition. Tier 3 targets are novel Materials Project compositions selected for chemical diversity by anion-class–stratified farthest-point sampling (FPS) over MTEncoder (Prein et al., 2023) composition embeddings, yielding 50,000 targets; the class taxonomy and per-class allocation are given in Appendix A.2. These targets are verified disjoint from the Tier 1 dataset under canonical stoichiometry-key matching. Generated

pairs are filtered for chemical validity as in Appendix A.1 to 31,930 pairs for 17,524 targets (Tier 2) and 82,161 pairs for 32,325 targets (Tier 3).

Tier-3 group-balance breakdown. The stratifier defines 17 material-group classes, of which 16 are ever populated (the `other` catch-all never fires). Table 6 tracks each group from the Materials Project candidate pool (after the Tier-1/2 firewall, a per-element stoichiometric-amount bound of ≤ 100, and charge-neutrality) through the 50,000 stratified-FPS selection, to the targets that survive into the final staged fine-tuning training set after augmentation and the stoichiometric balancability filter. Under-populated groups (hydroxide, nitrate, carbonate, sulfate, silicate, hydride, phosphate) are taken in full; every large group is capped at the redistributed equal quota ($\sim$4,943–4,944), providing a coverage floor for minority chemistries that global FPS could not supply (oxide alone constitutes 25% of the pool).

Table 6: Tier-3 material-group stratification. **MP pool**: Materials Project compositions after Tier-1/2 firewall, stoichiometric-amount bound, and charge-neutrality filter (154,949 total). **Selected**: chosen by anion-class–stratified FPS with equal-quota redistribution (50,000 total). **Surviving**: targets retained after augmentation and the stoichiometric balancability filter (32,325 total).

Material group	MP pool	Selected	Surviving
oxide	38,739	4,944	3,580
hydroxide	117	117	72
carbonate	286	286	184
nitrate	146	146	105
sulfate	547	547	284
phosphate	2,383	2,383	614
silicate	565	565	200
sulfide	6,748	4,943	2,937
halide	9,815	4,943	3,213
nitride	5,668	4,943	4,276
se/te	7,139	4,943	2,885
pnictide (P/As/Sb)	9,875	4,943	2,872
hydride	1,466	1,466	1,168
C/B/Si/Ge	15,841	4,943	3,453
mixed-anion	18,172	4,944	2,977
intermetallic	37,442	4,944	3,505
Total	**154,949**	**50,000**	**32,325**

A.3 Precursor Catalog Construction

The precursor catalog defines the set of materials at which multi-step search may terminate (Section 4.3). It is constructed from the raw Kononova and Lee records (Kononova et al., 2019; Lee et al., 2025), re-parsing each material formula from its string representation with a fractional-aware parser whenever the cached element dictionary is empty or incomplete. This recovers entries that would otherwise be dropped: papers whose reaction equations fail to parse lose all of their precursors, entries flagged as acronyms arrive with empty element dictionaries, and partially parsed formulas silently lose elements (e.g., $H_2CO_3 + (NH_4)_2CO_3$ resolving to only $\{N{:}2, H{:}8\}$, dropping C and O).

Candidates are then filtered by the same chemical-validity criteria used for the reaction pipeline (Appendix A.2): hydrate stripping, multi-component collapse, a per-element stoichiometric-amount bound (coefficients ≤ 100), and charge neutrality under the relaxed threshold. Entries are deduplicated by stoichiometry key at four-decimal precision, retaining the most frequently occurring formula string as the representative. Two further filters determine what search is permitted to treat as purchasable:

Metal filter. Entries are restricted to materials containing at most one metal or semimetal, excluding boron and silicon from the semimetal count. Multi-metal compounds are, in solid-state practice, materials one synthesizes rather than procures, and admitting them as terminal nodes allows search to stop short of a genuine precursor set.

Product-dominance pruning. We drop entries appearing strictly more often as a synthesis product than as a precursor across the raw corpus. These materials (e.g., Zn_2SiO_4, $BaSiO_3$, Bi_2Se_3) let search

shortcut-terminate at intermediates that would themselves require synthesis. The strict inequality is deliberate: a criterion that excluded any material ever appearing as a product would wrongly remove genuinely purchasable reagents such as TiO_2, $CaCO_3$, and ZnO, which are both widely sold and frequently synthesized.

A.4 Evaluation Dataset Construction

To enable contamination-free evaluation, we construct up-to-date out-of-distribution test sets for both precursor prediction and calcination/sintering temperature prediction using recent ICSD entries and the filtering procedures described below.

Candidate selection. We source candidate targets from the Inorganic Crystal Structure Database (ICSD) and restrict the pool to entries with publication year 2025. The 2025 publication-year cutoff was chosen to support a contamination-free evaluation setting, based on the knowledge cutoffs and release dates of the pretrained models used for prediction and data augmentation. Candidate targets are required to contain at least one metal or semimetal, with B and Si excluded from the semimetal definition. This filtering yields 4,489 ICSD entries corresponding to 1,278 unique source articles.

Synthesis extraction and validation. The collected source articles are paired with the corresponding ICSD metadata and processed using GPT-5.5 (`gpt-5.5-2026-04-23`) (OpenAI, 2026) with seed 42 for structured synthesis extraction. The model determines whether the reported procedure is solid-state, then extracts the target composition, precursor set, explicitly identified flux compounds, and ordered synthesis operations. Flux growth is treated as solid-state, whereas primarily solution- or vapor-phase routes are excluded. Verbatim material strings are retained separately from parsed chemical formulas, and reported units are preserved without conversion. Before full-corpus extraction, we manually validate the procedure on 65 records. For the target and precursor fields used in precursor-prediction evaluation, we observe two target mismatches and 15 precursor extraction errors. All 15 precursor errors arise from flux-related compounds being treated as reactants. These error modes are addressed during downstream curation rather than by manually correcting individual records. Applying the extraction procedure to the full corpus produces 4,474 entry-level records, of which 2,527 are labeled as solid-state. The extraction model is used only to construct evaluation labels and is not used for Retro-Forge training or inference. The complete extraction prompt and structured output format are provided in Appendix H.

Curation. We first retain only records with `is_solid_state = true` and `found_in_paper = true`. Records with empty precursor or operation lists are excluded. We also remove records whose verbatim operation tokens indicate wet-chemical or other non-solid-state routes. Targets containing unresolved compositional variables or explicit charge annotations are removed. The extracted target is then checked against the curated ICSD sum formula. Formula comparison expands bracketed groups, hydrate notation, and fractional stoichiometries. We require the two formulas to contain the same elements and to agree in every elemental coefficient after rounding to two decimal places. For example, an ICSD entry with composition $Cr_{4.885}Te_8$ was extracted as the nominal composition Cr_5Te_8 and is removed by the stoichiometry-consistency filter (Rai et al., 2025).

We also enforce bidirectional elemental consistency between the target and retained precursors, removing target-disjoint precursors and records with incomplete elemental coverage of the target. Records containing a precursor identical to the target or malformed formula strings with natural-language tokens are also excluded. Flux compounds are retained as precursors only if they supply a target element not provided by any other precursor; otherwise, they are removed. For example, the initially extracted precursor set for $TbGa_3(BO_3)_4$ contains Bi_2O_3, MoO_3, B_2O_3, Ga_2O_3, and Tb_4O_7 (Kuznetsov et al., 2025). Curation removes Bi_2O_3 and MoO_3 while retaining B_2O_3 because it supplies B to the target. Similarly, Co_3O_4 is removed from the extracted precursor set of $Bi_{25.4}Fe_{0.6}O_{40}$, for which Co is absent from the target composition. After each precursor-level removal, records containing fewer than two remaining precursors are discarded.

Formula normalization and contamination filtering. Formula strings are normalized after chemical filtering. Deuterium is mapped to hydrogen in composition fields, residual $\pm\delta$ and phase-prefix notation are removed, integer fractions are converted to decimal stoichiometries, and whitespace is standardized. Hydration annotations are also removed from precursor formulas where appropriate. We further require the evaluation targets to be compositionally novel with respect to earlier literature and the complete Retro-Forge training corpus. Therefore, targets whose ICSD sum formula is already registered in an ICSD entry from 2024 or earlier are excluded. We additionally remove

target compositions that overlap any data used for model training, including the pairwise-reaction and augmented training tiers. After the overall process, 857 unique targets remain and form the out-of-distribution benchmark used throughout the paper.

Temperature assignment. For calcination and sintering temperature experiments (Section 5.5), we further reference temperatures from the extracted heating operations. We identify explicitly reported calcination and sintering steps by matching the paper-verbatim `operations.token` fields against `calcin*` and `sinter*` keywords, respectively, taking the highest temperature when multiple operations share the same label. We retain records with at least one unambiguous calcination or sintering setpoint and remove training-recipe overlaps. This yields 181 reactions, of which 133 contain a calcination label and 113 contain a sintering label.

B Implementation Details

B.1 Staged Fine-Tuning Details

Both backbones (Qwen2.5-7B-Instruct (Qwen Team, 2025), LLaMA-3.1-8B-Instruct (Grattafiori et al., 2024)) are fine-tuned with 4-bit QLoRA (Dettmers et al., 2023) (NF4 quantization, double quantization, bf16 compute) in a continual curriculum that reuses a single LoRA adapter across Tier $3 \to 2 \to 1$. LoRA rank $r = 64$, $\alpha = 128$, targeting both attention (`q,k,v,o`) and MLP (`gate,up,down`) projections. Per-tier schedule: Tier 3, 1 epoch at learning rate 2×10^{-4}; Tier 2, 2 epochs at 1×10^{-4}; Tier 1, 3 epochs at 5×10^{-5}. Per-device batch size 2 with gradient accumulation 8 (effective batch 16), maximum sequence length 512, cosine learning-rate schedule, warmup ratio 0.05, weight decay 0.01, bf16 mixed precision, seed 42.

B.2 Tree-RL Training Details

Training data. Tree-RL is supervised by complete synthesis recipes rather than pairwise reactions: the 40,951 training targets of the Kononova–Lee split of Appendix E, each paired with all of its recorded precursor sets, any of which counts as a match. Because these recipes are not restricted to two precursors, they supervise entire multi-step routes.

Optimization. The final PRP model is further optimized with multi-step tree-RL, on top of the pre-RL checkpoint of Appendix B.1. Training is on-policy: at each step, the current policy rolls out $G = 8$ complete retrosynthesis trees per target (temperature 1.0, top-p 0.95, maximum depth 3), expanding each frontier node to a catalog-terminal leaf. The reward combines three components computed on the resulting leaf set L: **match** (weight 1.0, 1 if L exactly equals a ground-truth precursor set, else 0), **balance** (weight 0.3, the fraction of the route's steps that are stoichiometrically balanced under the NNLS criterion of Appendix A.1), and **diversity** (weight 0.5, the inverse frequency of L among the G sampled routes for that target). Each component is z-normalized within its sampled group and then weighted-summed into a single advantage A (Section 4.3); the terminal reward is shared by every step of the route, and the training loss is $-(A \cdot \sum_{\text{step}} \log \pi) + \beta \cdot \mathrm{KL}(\pi \| \pi_{\text{pre-RL}})$, with the KL reference given by the frozen pre-RL backbone. Optimization uses $\beta = 0.04$, learning rate 1×10^{-6}, 4,000 updates, checkpointing every 300 steps, on 4×A100 with data-parallel training; the final model uses the checkpoint at step 1,800. LoRA continues to target attention and MLP projections at $r = 64$. A bf16 logits sanitizer and a non-finite-gradient guard were required to keep Qwen's on-policy sampling numerically stable. All reinforcement-learning prompts are firewalled against every evaluation and test set (Section 4.4). The reward-weight ablation that selected match/balance/diversity $= 1.0/0.3/0.5$ is reported in Appendix D.

B.3 Search, Reranking, and Fusion Details

Search. The search maintains a priority queue $\mathcal{Q}$ of partial routes ranked by cumulative log-probability $S(r)$. At each step, the highest-scoring partial route r is popped and a frontier node t is selected. The PRP model generates B candidates, $\mathrm{PRP}(t, B)$, each returning a reactant pair $\mathbf{p}_i = (p_{1,i}, p_{2,i})$ with sequence log-probability s_i. Each valid candidate—one in which neither $p_{1,i}$ nor $p_{2,i}$ appears among the ancestors of t in r (cycle check)—produces a new partial route r' by replacing t with its two children, with updated score $S(r') = S(r) + s_i$. Each child is marked as a terminal node if it belongs to $\mathcal{C}$, or as a frontier node otherwise. A route with no expandable frontier node is added to the completed set $\mathcal{R}$ and the search continues; the loop ends once the queue is

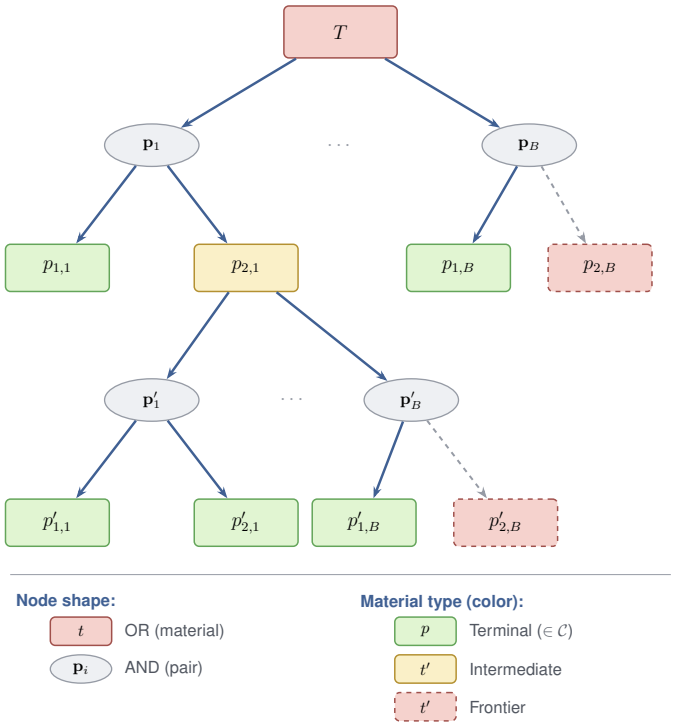

Figure 2: AND-OR tree structure of pairwise retrosynthesis. Rectangles are OR nodes (materials admitting B candidate pairs); ellipses are AND nodes (pairs whose two children must both be resolved). Color indicates material status: green = terminal ($\in \mathcal{C}$), amber = intermediate (recursively expanded), coral dashed = frontier node (awaiting expansion).

exhausted or the expansion budget E is spent. Completed routes are then grouped by leaf set $L(r)$, each distinct set receiving the log-sum-exp of the log-probabilities of its constituent derivations, and the resulting sets are returned in descending order of this marginal score. Figure 2 illustrates the AND-OR tree structure.

Set-marginalized aggregation. Rather than maintaining routes naively, we aggregate them into distinct leaf (precursor) sets, as the same precursor set can be reached through various routes (derivations). We assign each distinct set a marginal log-probability equal to the log-sum-exp of the log-probabilities of all of its constituent derivations, so that a frequently rediscovered set is scored more highly than any single derivation of it.

Reranker. We train one gradient-boosted tree classifier per PRP model (`HistGradientBoostingClassifier` from scikit-learn (Pedregosa et al., 2011), a histogram-based implementation following LightGBM (Ke et al., 2017); 300 maximum iterations, learning rate 0.05, 31 maximum leaf nodes, ℓ_2 regularization 1.0, balanced class weights, early stopping with a 10% internal validation split, seed 42) on candidate routes generated by the search (beam 8, depth 4, expansions 400, top-n 80) for the targets of the Kononova–Lee training split (Appendix E). Each candidate set is labeled 1 if it stoichiometrically matches a ground-truth recipe and 0 otherwise. The 25 input features all describe a candidate set, in three groups: (i) aggregates over the set's derivations (marginal log-probability with its within-target rank, and $\log(1 + \text{number of contributing derivations})$); (ii) properties intrinsic to the set (stoichiometric balance score $\in \{0, 1, 2\}$, number of precursors, and a 13-way one-hot encoding of the target's anion family); and (iii) features of the set's highest-log-probability derivation (its log-probability with within-target rank and z-score, a conventionality score with rank, tree depth, and whether the route is complete). The reranker weights are fixed after training and are not refit at inference time.

Anion-family feature. The anion-family one-hot uses the reranker's own coarse taxonomy of 13 families, separate from the 17-class Tier-3 stratification of Appendix A.2: oxide, phosphate, sulfate, carbonate, silicate, borate, oxyhalide, oxynitride, chalcogenide, nitride, halide, pnictide,

Algorithm 1 Set-Marginalized Best-First Search over AND-OR Tree

Input: Target T, catalog $\mathcal{C}$, branching factor B,
max expansions E, max depth D
Output: Top-k distinct precursor sets by marginal score
Initialize $r_0 \leftarrow T$; $S(r_0) \leftarrow 0$; $\mathcal{Q} \leftarrow \{r_0\}$; $\mathcal{R} \leftarrow \emptyset$; $n \leftarrow 0$
while $\mathcal{Q} \neq \emptyset$ **and** $n < E$ **do**
  $r \leftarrow \mathcal{Q}$.PopMax()
  $t \leftarrow$ frontier node of r s.t. $t \notin \mathcal{C}$ and depth$(t) < D$
  if no such t exists **then**
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{r\}$; **continue**
  end if
  $\{\mathbf{p}_i, s_i\}_{i=1}^{B} \leftarrow$ PRP(t, B)
  for $i = 1$ **to** B **do**
    if $p_{1,i}, p_{2,i} \notin$ Anc(t, r) **then**
      Expand t into $p_{1,i}, p_{2,i}$ in $r \rightarrow r'$
      $S(r') \leftarrow S(r) + s_i$
      $\mathcal{Q}$.Push$(r', S(r'))$
    end if
  end for
  $n \leftarrow n + 1$
end while
// marginalize derivations by leaf set
for each distinct leaf set L in $\{L(r) : r \in \mathcal{R}\}$ **do**
  $\mathcal{D}(L) \leftarrow \{r \in \mathcal{R} : L(r) = L\}$
  $\tilde{S}(L) \leftarrow \log \sum_{r \in \mathcal{D}(L)} \exp S(r)$
end for
return top-k sets L by $\tilde{S}(L)$, each with $\mathcal{D}(L)$

and intermetallic. Each target is assigned by a priority rule. If it contains oxygen, it is classified as an oxoanion family (phosphate, sulfate, carbonate, silicate, or borate) when P, S, C, Si, or B is present, and otherwise as an oxyhalide, oxynitride, or oxide. Oxygen-free targets are classified as chalcogenide, nitride, halide, or pnictide, in that order of priority, and default to intermetallic.

Stoichiometric balance score. We define $\mathrm{bal}(\text{precursors}, \text{target}) \in \{0, 1, 2\}$: 0 if the precursors' metal set does not match the target's metal set; otherwise, 2 if the step is stoichiometrically balanced under the NNLS criterion of Appendix A.1 (relative residual $< 10^{-6}$, O/C/H/N treated as free byproducts), and 1 otherwise. We define $\mathrm{bal_frac}$ as the fraction of a route's steps that are exactly two-precursor and achieve $\mathrm{bal} = 2$. Within a fused candidate set, the output top-1 route is $\arg\max$ over $(\mathrm{bal_frac}, \text{total log-probability})$ among that set's derivations. This balance-based selection is a distinct check from the dataset-level charge-neutrality/valence filter of Appendix A.1, which separately governs which compounds are admissible as targets or precursors during corpus construction.

Fusion. The two backbones' reranked candidate-set lists are combined via reciprocal rank fusion (RRF) with $K = 60$: each candidate set's fused score is $\sum 1/(60 + \mathrm{rank})$ across the two backbones' (zero-indexed) rankings, and the merged pool is re-sorted by this score.

C Baseline Reproduction

MSP-LLM (Noh et al., 2026) We train the LLaMA-3.1-8B (Grattafiori et al., 2024) base model under the published LoRA configuration. We use the authors' publicly released training and validation splits without modification for data parity. Our retrained model reaches 71.32% top-1 accuracy on the released test set, 0.6 points above the 70.75% reported for LLaMA-3.1-8B in the paper, confirming a faithful reproduction.

ElemwiseRetro (Kim et al., 2024b) We retrain the authors' template classifier with their unmodified code and default configuration (ElemNet element embeddings) on the same dataset, using the authors' random split. Top-k candidate sets are the highest joint-score template combinations returned by

the authors' `find_precursors_set`. Our retrained models reproduce the reported accuracies on both splits (top-1 80.4% vs. 78.6% reported for RandSplit, 80.1% vs. 80.4% for TimeSplit; top-5 within 0.8 points on both).

CSLLM (Song et al., 2025) We use the authors' released LoRA adapter on its LLaMA-3-8B base (4-bit NF4), with the authors' evaluation prompt and greedy decoding (temperature 0). The model emits one precursor list per target, so only top-1 is reported.

SynthGPT (Kim et al., 2024a) We re-fine-tune the same base model (`gpt-3.5-turbo-0125`) on the authors' released 5-fold fine-tuning files via the OpenAI API with default hyperparameters. At inference we use the authors' system prompt and "provide k synthesis plans" query at temperature 0, and strip O_2/H_2O from predictions as in the authors' scoring script.

Our port reproduces the authors' statistical baseline to within 0.3 points at both k=1 and k=5, confirming that the data pipeline and scoring are faithful. Our re-fine-tuned GPT-3.5 is within 0.7 points of the reported top-1 and 3.8 points below at top-5; we attribute the residual to re-running the fine-tuning rather than reusing the authors' account-bound models.

Retrieval-Retro (Noh et al., 2024) Using the authors' released year-split knowledge base, we retrain both the MPC retriever and the Retrieval-Retro predictor (K=3 retrieved references per retriever) with their unmodified code; the neural reaction energy (NRE) predictor is the authors' pretrained checkpoint. Our retrained model reaches 60.5% / 65.9% / 67.7% / 69.2% top-1/3/5/10 on the authors' year-split test set, within 0.7 points of the reported 61.2% / 65.9% / 67.2% / 69.5% at every k, confirming a faithful reproduction.

D Ablations

Unless noted otherwise, ablations in this appendix use the LLaMA backbone and the full search budget of Appendix D.6 (B/D/E/top-n = 5/4/400/200) on the 857-target benchmark. As throughout this work, all models are trained with a single seed; treat deltas of roughly 1–2 points as within the expected run-to-run noise floor. Ablation runs use reward supervision and a shared reranker built before the Kononova–Lee corpus was re-filtered and split (Appendix E), whereas the final models in Table 2 use the rebuilt supervision and a reranker trained on their own search outputs. The headline LLaMA top-1 differs by less than single-seed noise between the two settings (57.2 vs. 57.8), so conclusions transfer.

D.1 Data Pipeline Levers

Stoichiometric balancability filter. Table 7 compares training the pre-RL PRP model on augmented pairs that pass the stoichiometric balancability criterion (Section 4.1, Appendix A.1) against a coverage-only filter that requires the pair to cover the target's elements without checking stoichiometry. The balancability filter adds 1–2 points of raw leaf top-1 and roughly doubles the per-step stoichiometric balance of generated routes ($\sim$30% $\rightarrow$ 55%), with the largest gains on families where metal/anion balance is most discriminative (e.g. pnictides).

Table 7: Stoichiometric balancability filter vs. coverage-only filter, single pre-RL model, raw leaf top-k (%).

Backbone	Filter	@1	@3	@5	@10
Qwen	coverage	44.3	50.2	51.9	55.7
Qwen	**balancability**	45.4	50.8	53.6	55.8
LLaMA	coverage	43.4	49.6	54.0	**57.4**
LLaMA	**balancability**	**45.6**	**51.2**	**54.3**	56.5

D.2 Augmentation Curriculum and Tree-RL (Full Tables)

This section expands the factorial ablation of Section 5.4 to every k, to both backbones for the PRP task, and to the stoichiometric balance of the reranker-selected top-1 route for the PP task.

PRP task. Table 8 reports stoichiometry-level, target-wise top-k on the Tier-1 held-out set for each curriculum stage, before and after tree-RL, together with the rate of stoichiometrically balanced top-1 pairs (a single step, so step- and route-level balance coincide). Both trends in the main text hold for Qwen: augmentation lifts top-1 by 2.8 points pre-RL and 6.3 points after tree-RL, and the balanced top-1 rate rises monotonically at both stages. The single-step cost of tree-RL also shrinks as augmentation grows (Qwen $-4.2 \rightarrow -1.6 \rightarrow -0.7$ @1), suggesting that a richer pre-RL base absorbs the shift of tree-RL toward route-level reward with less loss of per-step precision.

Table 8: Augmentation curriculum and tree-RL on the PRP task, Tier-1 held-out set (686 targets), stoichiometry-level target-wise top-k (%) and rate of stoichiometrically balanced top-1 pairs (**Bal.**, %).

		LLaMA					Qwen				
Stage	**Curriculum**	**@1**	**@3**	**@5**	**@10**	**Bal.**	**@1**	**@3**	**@5**	**@10**	**Bal.**
pre-RL	Tier 1 only	79.6	87.8	91.4	93.1	94.8	78.4	86.7	89.9	92.3	96.8
	Tier 2→1	81.5	**89.5**	91.8	**94.0**	97.2	80.9	**90.2**	**92.0**	94.0	98.1
	Full	**82.9**	88.6	**92.0**	93.7	98.8	**81.2**	90.1	**92.0**	**94.5**	98.5
tree-RL	Tier 1 only	75.9	84.1	86.9	89.9	96.8	74.2	82.9	86.2	88.8	97.8
	Tier 2→1	79.4	87.2	89.5	91.8	99.0	79.3	88.6	90.7	92.6	99.0
	Full	80.2	87.0	89.4	91.8	**99.7**	80.5	88.5	90.4	92.9	**99.4**

PP task. Table 9 reports leaf top-k at every k for the LLaMA arms of Table 4, together with the stoichiometric balance of the reranker-selected top-1 route. Top-1 balance follows the same monotonic trend as the pool balance reported in the main text but is compressed toward its ceiling: tree-RL with the full curriculum reaches 97.3% per-step and 93.6% fully balanced top-1 routes, against 92.8% / 85.8% for Tier-1 only. Because the reranker's features include stoichiometric balance and the output route is selected for it, the top-1 view understates how differently the policies behave across the whole search, which is why the main text reports the pool.

Table 9: Augmentation curriculum and tree-RL on the PP task, LLaMA, 857 novel targets. Leaf top-k (%) without a reranker (raw) and with the full pipeline (agg+rerank), and stoichiometric balance of the reranker-selected top-1 route (**step**: per-step balanced rate; **route**: fraction of routes where every step is balanced). Each tree-RL arm is trained with the identical recipe on top of the corresponding pre-RL base; all arms share one fixed reranker.

	raw				agg+rerank				Top-1 stoich. balance
Curriculum	**@1**	**@3**	**@5**	**@10**	**@1**	**@3**	**@5**	**@10**	**step / route**
pre-RL									
Tier 1 only	37.5	49.5	54.5	58.0	54.8	60.2	61.3	63.2	65.2 / 33.4
Tier 2→1	43.1	52.2	56.2	58.6	53.8	62.2	64.3	65.6	77.7 / 54.0
Full (T3→2→1)	46.7	54.6	56.7	58.7	53.9	61.4	63.8	66.2	83.8 / 64.6
tree-RL									
Tier 1 only	**56.2**	**59.9**	**61.3**	62.3	57.5	62.9	64.3	66.5	92.8 / 85.8
Tier 2→1	55.4	58.8	61.0	62.1	**58.6**	**63.7**	65.5	66.4	96.4 / 92.1
Full (T3→2→1)	54.3	59.7	60.9	**63.0**	57.2	62.7	**65.7**	**67.1**	**97.3 / 93.6**

D.3 Search Aggregation and Reranking

Set-marginalization. On a fixed pool of 200 candidate routes per target, aggregating routes into distinct leaf sets (with marginal log-probability given by log-sum-exp over derivations, Section 4.3) before reranking reaches 53.8 leaf top-1, compared to 49.5 for the same reranker applied to naively deduplicated routes (+4.3@1, +7.4@10)—confirming that the aggregation step is essential rather than cosmetic.

Reranker and fusion. Table 10 isolates the contribution of the learned reranker and of two-backbone fusion on the final tree-RL PRP models. The reranker alone lifts each backbone's leaf top-1 by 2.9–3.5 points over its raw policy, and fusing the two rerankers' outputs via RRF adds a further 0.9 points over the better single backbone (LLaMA) and 1.8 points over Qwen.

Table 10: Effect of reranking and fusion on the final tree-RL PRP model, leaf top-1 (%), 857 novel targets.

System	raw (no reranker)	agg+rerank
LLaMA (single)	**54.3**	57.8
Qwen (single)	54.0	56.9
Fused (RRF)	—	**58.7**

D.4 Tree-RL Reward-Weight Ablation

The final reward weights (match/balance/diversity $= 1.0/0.3/0.5$, Section 4.3) were selected by varying each shaping term with the match weight fixed at 1.0, on the LLaMA backbone, agg+rerank, best checkpoint. Table 11 also reports the fraction of raw top-1 routes that are fully stoichiometrically balanced, since this axis trades directly against top-k.

Table 11: Tree-RL reward-weight ablation (match/balance/diversity), LLaMA, agg+rerank leaf top-k (%) and fully balanced raw top-1 route rate.

Weights (match/bal/div)	@1	@3	@5	@10	raw top-1 balanced %
1.0 / 0.0 / 0.0 (pure match)	**58.6**	64.8	**67.6**	**68.1**	72.6
1.0 / 0.0 / 0.5	57.6	**65.1**	66.6	67.8	76.3
1.0 / 0.3 / 0.5 (final)	57.2	62.7	65.7	67.1	91.0
1.0 / 0.5 / 0.5	57.1	62.1	63.9	66.0	**94.4**
1.0 / 0.1 / 0.1	57.1	63.6	65.7	66.9	91.5
1.0 / 0.3 / 0.25	56.8	62.4	65.7	66.7	91.9
1.0 / 0.5 / 0.7	55.7	62.0	63.7	65.9	94.6
1.0 / 0.3 / 1.0	55.2	61.6	63.7	65.5	92.1

The balance weight trades top-k against stoichiometric balance, and the trade is favorable. The two balance-zero configurations reach only 72–76% fully balanced raw routes, while every configuration with a nonzero balance weight reaches 91–95%; whether the term is on at all matters considerably more than its magnitude (0.1 to 0.5). The accuracy cost is small: the final $1.0/0.3/0.5$ gives up at most 1.4 points of top-1 relative to the pure-match configuration. Given comparable accuracy across these settings, we prefer the one that yields stoichiometrically balanced routes independent of the reranker, and use $1.0/0.3/0.5$ on that basis. The diversity term shows an inverted-U: 0.5 is optimal (0.25 under-spreads the candidate pool; 1.0 over-spreads and erodes the match signal).

D.5 Reward Aggregation: GRPO vs. GDPO

Table 12 isolates the effect of decoupled per-component reward normalization (GDPO, Section 4.3) against two baselines that differ only in how the three reward signals are aggregated. **Plain GRPO** sums match and balance into a single scalar and group-normalizes it (diversity off). **GRPO+diversity** adds the within-group novelty term and sums all three before group-normalizing—the same three signals as GDPO, but collapsed into one normalized scalar. **GDPO (final)** z-normalizes each component *independently* within the rollout group, then weighted-sums ($1.0/0.3/0.5$), preserving each component's gradient even when another saturates group-wide. All three variants share identical rollout, KL penalty, and schedule; only the aggregation differs (LLaMA single-backbone, agg+rerank leaf, 857 novel targets, checkpoint at step 1,800).

Table 12: Effect of reward aggregation scheme on the final tree-RL model, leaf top-k (%), 857 novel targets. **Bold** = best per column.

Aggregation	@1	@3	@5	@10
Plain GRPO (scalar-sum → normalize)	56.4	62.5	64.2	65.3
GRPO + diversity	55.9	**62.8**	**65.7**	67.0
GDPO (final)	**57.2**	62.7	**65.7**	**67.1**

GDPO leads at @1 ($+0.8$ over plain GRPO) and at @10 ($+1.8$). The GRPO+diversity row isolates the effect of decoupled normalization specifically: it carries the same three signals as GDPO but sums them before normalizing, and still trails. The mechanism is the one identified by Liu et al. (2026): when the match component saturates group-wide (all G sampled routes for a target either all match or all miss), a single normalized scalar carries no gradient from the remaining balance and diversity terms—they are washed out. GDPO's per-component normalization preserves the gradient in those terms regardless of match saturation, supplying a denser training signal throughout training.

D.6 Search Budget Efficiency

Every result reported in this paper uses the **full search budget**: branching factor $B = 5$ (candidates generated per expansion), maximum tree depth $D = 4$, maximum expansions $E = 400$ (total expansion steps before the search terminates), and top-$n = 200$ (candidate precursor sets ranked and returned after set-marginalization). Table 13 shows the maximum capacity of this full budget and demonstrates that a substantially smaller configuration achieves equivalent performance.

Table 13: Search-budget ablation, leaf top-k (%), 857 novel targets. Parameters: $B/D/E$/top-n = branching factor / max depth / max expansions / candidate sets returned. The full budget (5/4/400/200) is used for all reported results in this paper.

$B/D/E$/**top-n**	**@1**	**@3**	**@5**	**@10**
5/4/400/200 (full)	**57.8**	63.1	**65.9**	67.1
5/4/150/60	**57.8**	**63.5**	65.7	**67.4**
5/4/80/30	57.6	62.2	64.5	66.2
5/4/40/15	56.0	60.9	62.5	62.9
3/4/40/15	56.2	61.0	62.1	62.5

The 5/4/150/60 configuration—approximately $2.7\times$ fewer expansions than the full budget—matches it at every k (identical @1, within ± 0.4 elsewhere), showing that the PRP model finds high-quality routes efficiently and the extra capacity of the full budget does not meaningfully improve results. Reducing the budget below $E = 80$ erodes recall by 1–5 points at @3–10. We retain the full budget as the reported configuration for full reproducibility; 5/4/150/60 is a validated no-loss configuration for faster inference.

D.7 Few-Shot Base-Model (No Fine-Tuning) Baseline

To confirm that task-specific fine-tuning is necessary rather than an artifact of our pairwise-plus-tree-search formulation, we evaluate the base, non-fine-tuned LLaMA-3.1-8B-Instruct and Qwen2.5-7B-Instruct driven by K fixed in-context demonstrations (Tier 1 training pairs injected as user/assistant turns before the query, using the PRP prompt template of Appendix H) and plugged into the same multi-step search, with shot counts $\{20, 40\}$ matching MSP-LLM's own few-shot configuration (Noh et al., 2026).

The best few-shot base-model configuration (LLaMA, 40-shot, agg+rerank) reaches 31.5 leaf top-1—roughly half of the fine-tuned PRP model (57.8)—and 0-shot performance is a near-total floor for both backbones without any in-context demonstration. Shot count scales monotonically ($0 \ll 20 \ll 40$) but never closes the gap to fine-tuning, consistent with MSP-LLM's own finding that fine-tuning outperforms few-shot prompting for this task; we observe a larger gap for our open 7–8B backbones than MSP-LLM reports for much larger API models, as expected for in-context learning at this model scale. We conclude that task-specific fine-tuning is necessary: the pairwise-plus-tree-search formulation alone does not rescue a non-fine-tuned model.

E Kononova–Lee Held-Out Benchmark

We additionally report our model's performance on a held-out test set of the Kononova and Lee corpora following other precursor-prediction works. We process the merged Kononova–Lee corpus with the same preprocessing and chemical-validity filters used for the ground-truth pairwise reaction dataset (Appendix A.2), without the two-reactant restriction, reducing 106,376 raw reactions to 99,190 reactions over 45,501 targets. A target-wise 90/10 split (seed 42) yields 40,951 training and

Table 14: Few-shot base models (no fine-tuning) vs. the fine-tuned, tree-RL PRP model, leaf top-k (%), 857 novel targets.

Configuration	@1	@3	@5	@10
raw (no reranker)				
LLaMA 0-shot	0.2	0.2	0.2	0.2
LLaMA 20-shot	6.8	13.4	19.6	24.2
LLaMA 40-shot	21.5	31.0	34.9	40.8
Qwen 0-shot	5.3	7.4	8.9	9.5
Qwen 20-shot	10.6	18.9	22.9	26.7
Qwen 40-shot	18.6	30.2	35.7	42.4
Retro-Forge (raw)				
LLaMA (pre-RL)	46.7	54.6	56.7	58.7
LLaMA (tree-RL)	**54.3**	**59.8**	**61.0**	**63.1**
Qwen (tree-RL)	54.0	59.4	60.8	63.0
agg+rerank				
LLaMA 20-shot	19.3	26.4	28.1	29.8
LLaMA 40-shot	31.5	43.1	46.8	48.8
Qwen 20-shot	14.2	29.8	32.9	34.5
Qwen 40-shot	27.1	42.0	45.6	47.7
Retro-Forge (agg+rerank)				
LLaMA (tree-RL)	**57.8**	63.1	**65.9**	**67.1**
Qwen (tree-RL)	56.9	**64.6**	**65.9**	67.0

4,550 held-out targets. The training split supplies the tree-RL rollout targets (Appendix B.2) and the reranker's training data (Appendix B.3); models are evaluated on the held-out targets.

Results. Table 15 reports the fused system on the 4,550-target held-out split. As expected for an in-distribution evaluation, absolute performance is substantially higher than on the 857-target benchmark across every k.

Table 15: Precursor prediction on the 4,550-target Kononova–Lee held-out set (fused, agg+rerank, top-k, %). In-distribution; reported for comparison with prior work's typical evaluation regime, not as a contamination-free result.

Metric	@1	@3	@5	@10
leaf	70.4	80.2	82.2	84.0
frontier	71.0	81.4	83.7	85.7

F Thermodynamic Route Validity

Stoichiometric balance (Section 5.3) is necessary but not sufficient for a predicted pairwise step to be a real reaction. As a second validity axis, we additionally test whether each step is *thermodynamically downhill*, following the ARROWS3 formulation (Szymanski et al., 2023a): for each forward step (precursors $\rightarrow$ step-target, with O_2/CO_2/H_2O/NH_3 permitted as gaseous byproducts under open-system oxygen uptake), we compute the per-atom reaction Gibbs energy ΔG_{rxn} and call the step favorable if $\Delta G_{\text{rxn}} < 0$.

Energetics. We use 0 K DFT formation energies from the Materials Project, temperature-corrected via the Bartel $G(T)$ machine-learned descriptor, with a convex-hull fallback for compositions absent from the database and Materials-Project gas chemical-potential corrections $\Delta\mu(T, p)$. We additionally cross-validate the hull-fallback energies against Alexandria (PBE) formation energies for integer-stoichiometry phases missing from the Materials Project; the two sources agree closely (median offset ≈ 0 eV/atom: -0.010 for oxides, $+0.001$ for non-oxides), supporting their use as a consistency check on the fallback.

Coverage. Two distinct coverage notions matter here. *DFT-exact coverage*: only 21–24% of step-targets have a real database formation energy (Materials Project + Alexandria); the majority are

fractional or doped compositions that rely on the convex-hull fallback. ΔG*-computable coverage*: steps with either a real DFT entry *or* a working convex-hull fallback—approximately 85% for the 857 set and 93% for the Kononova–Lee held-out set. The reliable-DFT subset tracks the ΔG-computable headline closely (reliable-DFT steps favorable: 70.7% for 857, 65.9% for Kononova–Lee), confirming the fallback is not materially skewing the estimate. All figures are calculated presuming reaction at 900°C in air.

Table 16: Route chemistry validity at 900°C (Materials Project + Alexandria). **Stoich. balance**: NNLS check. **Thermo**: ARROWS3 $\Delta G < 0$, ΔG-computable steps only ($\sim$85% for 857, $\sim$93% for Kononova–Lee held-out). The pipeline raises stoichiometric balance but leaves thermodynamic favorability flat.

		Stoich. balance (%)		Thermo (%)	
Set	**System**	**Per-step**	**Fully bal.**	**Steps fav.**	**Routes fully fav.**
857 novel	LLaMA (raw top-1)	96.1	91.0	**72.9**	**53.0**
	Qwen (raw top-1)	95.9	90.4	69.8	52.4
	Fused (agg+rerank)	**98.8**	**97.4**	72.2	52.7
Kononova–Lee held-out	LLaMA (raw top-1)	98.1	95.7	71.6	50.8
	Qwen (raw top-1)	97.8	95.0	71.0	50.4
	Fused (agg+rerank)	**99.6**	**99.3**	**72.0**	**51.2**

Adding Alexandria to the Materials-Project-only basis raises DFT-exact coverage by 1.4× (Kononova–Lee) to 2.5× (857) while shifting favorability by only $+0.7$ to $+3.9$ points, confirming the fallback is mildly pessimistic and the headline is robust. Across both sets, the pipeline raises stoichiometric balance substantially (raw 90–96% $\rightarrow$ agg+rerank 97–99% fully balanced) while leaving thermodynamic favorability nearly unchanged (70–73% steps / 50–53% routes), confirming that thermodynamic feasibility is not optimized by the balance reward or reranker.

Temperature dependence. On a Materials-Project-only hull basis, favorability rises mildly and monotonically with assumed firing temperature (Kononova–Lee steps: 69.5% / 71.1% / 72.9% at 700 / 900 / 1100°C; 857: 66.6% / 68.3% / 68.7%), so the headline conclusion does not hinge on the particular temperature assumed. Evaluated instead at each Kononova–Lee target's *real*, literature-recorded synthesis temperature (the maximum firing plateau from the cleaned Kononova–Lee procedures), favorability is highest of all: 73.8% of steps and 53.4% of routes fully favorable ($n = 2{,}725$)—the most realistic available estimate.

Predicted onset vs. actual firing temperature. For each step we scan for the lowest temperature at which $\Delta G_{\text{rxn}} < 0$ (the thermodynamic onset); a route's onset is the maximum over its steps (the rate-limiting step). Among Kononova–Lee routes with a recorded synthesis temperature, 71.2% have a predicted onset at or below the temperature actually used, with a median headroom of 500°C—consistent with real solid-state syntheses being run well above their thermodynamic threshold, as expected for kinetically limited reactions. The onset does not predict the actual firing temperature used (Spearman ≈ 0.03): it is a thermodynamic floor, not a synthesis-temperature model. 22–29% of routes are thermodynamic dead ends under these energetics (a step never becomes favorable up to 2000°C), concentrated in the more exotic, hull-fallback-dependent compositions.

Takeaway. Retro-Forge's final (fused) top-1 routes are 97–99% fully stoichiometrically balanced (Table 16); 70–73% of pairwise steps are additionally thermodynamically favorable at realistic firing temperatures; and roughly 71% of routes are thermodynamically feasible at or below their recorded firing temperature, with $\sim$500°C of kinetic headroom. Thermodynamic feasibility is the strictest filter, and a solid majority of predicted pairwise steps pass it; $\sim$10–14% of steps could not be balanced even with the permitted gaseous byproducts and are excluded from this analysis.

G Calcination and Sintering Temperature Prediction

This section details the preliminary case study of Section 5.5.

MPC encoder. We use the Masked Precursor Completion (MPC) encoder of He et al. (2023), a graph neural network pretrained to predict a masked precursor from the target and the remaining precursors

of a recipe. MPC produces a 32-dimensional embedding per chemical formula, so compositions that are synthesized from similar precursors lie close together in the embedding space.

Recipe–temperature corpus. Retrieval and the route-derived augmentation both draw on a recipe–temperature corpus built from the Kononova and Lee datasets (Kononova et al., 2019; Lee et al., 2025). Each source is preprocessed to resolve compositional variables and reconciled under a shared schema; from every reaction we extract the target, its set of solid precursors, and its heating operations. Atmospheric species (O_2, H_2O, N_2, CO_2) are removed from precursor lists, and every target and precursor must admit a charge-neutral oxidation-state assignment under the algorithm of Appendix A.1. Heating temperatures are converted to Kelvin; values outside a plausible [250, 4000) K range are repaired when they arise from a trailing-zero or concatenation artifact of text mining, and discarded otherwise. Calcination and sintering temperatures are then assigned by the keyword rules of Appendix A.4, with a fallback that assigns the hottest unlabeled heating operation to sintering when a label is missing, and records with calcination below 573 K or sintering below 773 K are dropped. Finally, to prevent leakage, we remove every record that matches a test instance in recipe (target and precursor set) and in both temperatures within ± 5 K. The resulting corpus contains 33,371 recipe–temperature records.

Route-derived step pairs. Taking the top-1 Retro-Forge route for the target of each corpus record, we extract a pair at every node and keep those whose reactants all lie in the literature precursor vocabulary. This yields 28,301 step pairs that add 7,761 intermediate compositions as new MPC training targets without enlarging the precursor vocabulary. We retrain MPC on the corpus recipes together with these step pairs; this is the only point at which route information enters the predictor.

Predictor. The predictor is a gradient-boosted regressor (Ke et al., 2017) over three feature groups: (i) the target's MPC embedding; (ii) temperature statistics of the 50 nearest corpus recipes retrieved in the MPC embedding space; and (iii) melting-point features of the precursors and target predicted by MAPP (Hong, 2021; Xue & Hong, 2024).

Evaluation. We evaluate on two test sets. The Kononova test set is a held-out split of 918 Kononova reactions, matching the data of SyntMTE's year-split training stage. The ICSD-2025 test set is the subset of our 857-target benchmark (Appendix A.4) whose extracted procedures report an unambiguous calcination or sintering temperature: 181 reactions, of which 133 carry a calcination label and 113 a sintering label (Appendix A.4). These 2025 reactions postdate the training data of both methods. SyntMTE (Prein et al., 2025b) is fine-tuned from its released checkpoint on its year-split training set. Relative to SyntMTE, our predictor lowers MAE by 26.6 and 33.1 K on Kononova calcination and sintering, and by 28.1 and 43.8 K on ICSD-2025 (Table 5).

Ablation of Retro-Forge routes. To isolate the contribution of the routes themselves, Table 17 holds the regressor, retrieval corpus, and feature set fixed and varies only the encoder's training data. Augmenting the 33k literature pairs with the route-derived step pairs reduces out-of-distribution MAE on ICSD-2025 from 126.9 to 123.8 K for calcination and from 99.3 to 95.6 K for sintering. In-distribution performance on the 918 Kononova reactions remains unchanged within 0.5 K. These results suggest that supervision from route-derived intermediates shapes the composition embedding in a way that improves temperature prediction for novel targets, a promising direction for extending route prediction toward synthesis-condition prediction.

Table 17: Calcination and sintering temperature prediction (MAE, K; mean over 3 encoder seeds) as a function of MPC encoder training data. The regressor, retrieval corpus, and feature set are identical across rows.

	Kononova		**ICSD-2025**	
Encoder training data	**Calc.**	**Sint.**	**Calc.**	**Sint.**
Literature pairs (33k)	**132.9**	105.4	126.9	99.3
+ Retro-Forge step pairs (33k+28k)	133.4	**104.9**	**123.8**	**95.6**

H PROMPT TEMPLATES

In this section, we present the prompt templates used for the multi-plan data augmentation model (GPT-4.1-mini) and for the deployed single-step PRP models (Qwen and LLaMA). Both PRP back-

bones share the same system prompt, user prompt, and output format (`p1 + p2`). The augmentation model instead uses a multi-plan reformulation (Section 4.1) that asks for N decompositions in a single call. We additionally provide the GPT-5.5 prompt used for structured synthesis extraction during evaluation dataset construction. The corresponding extraction and curation procedure is described in Appendix A.4.

Table 18: Shared prompt for both PRP models (Qwen and LLaMA).

System Prompt: You are an expert in solid-state inorganic synthesis. Assume all solid-state inorganic synthesis happens pairwise between intermediates or precursors.Given a target material formula, predict exactly two precursor or intermediate materials that reacts pairwise via a solid-state reaction to synthesize the target. Output only the two material formulas separated by `' + '`.

User: Predict two solid-state precursors or intermediates for {`target_formula`}

Output format:

```
precursor1 + precursor2
```

Table 19: Multi-plan prompt for GPT-4.1-mini (data augmentation model only; never used at inference). `N` is the target's true recorded plan count during fine-tuning, and a fixed 5 at augmentation-generation time (Section 4.1).

System Prompt: You are an expert in solid-state inorganic synthesis. Assume all solid-state inorganic synthesis happens pairwise between intermediates or precursors. Given a target material formula and a count N, predict N pairs of precursor or intermediate materials that each react pairwise via a solid-state reaction to synthesize the target. Output each pair on its own line as `'A + B'`.

User: Predict {`N`} {`pair|pairs`} of solid-state precursors or intermediates for {`target_formula`}

("pair" if $N = 1$, "pairs" otherwise.)

Output format:

```
precursor1a + precursor1b
precursor2a + precursor2b
...
```

Table 20: Prompt for GPT-5.5 used for structured synthesis extraction from the ICSD evaluation corpus. The extraction model is used only for evaluation-dataset construction and is not used for Retro-Forge training or inference.

System Prompt: You are an inorganic materials chemistry expert. Extract the solid-state synthesis for one ICSD entry from a paper into the given schema.

Rules:
1. Determine whether the synthesis is a solid-state synthesis. Flux growth counts as solid-state. If solution-phase or vapor-phase is the primary route, set `is_solid_state=false`, `solid_state_option=null`, `flux_compounds=[]`, and return `target=null`, `precursors=[]`, `operations=[]`. Stop here. Otherwise, set `solid_state_option` and `flux_compounds`:

- If a flux is used (compound explicitly named as "flux" in the paper, e.g., `ZnCl2`, `NaCl`, `B2O3`, `PbO`, `CsBr`), write `solid_state_option` as `"flux with <formula>"` using the paper's exact compound (e.g., `"flux with ZnCl2"`, `"flux with CsBr"`). If multiple flux compounds are used together, comma-separate them inside the same string (e.g., `"flux with NaCl, KCl"`). Also list each flux compound as written in `flux_compounds` (e.g., `["ZnCl2"]`, `["NaCl", "KCl"]`).
- Otherwise: `solid_state_option="normal"`, `flux_compounds=[]`.

2. This call targets one ICSD entry. A procedure may cover multiple entries (e.g., solid solution x values); extract as it applies to this one.
3. Use only what's explicitly in the paper. No external knowledge, intuition, or defaults like "air".
4. If procedure is SI-only or not described for this entry, set `found_in_paper=false` with a one-line reason, plus `target=null`, `precursors=[]`, `operations=[]`. Don't guess.
5. `material_string` = paper's verbatim text; `material_formula` = parsed formula. Don't normalize between them.
6. Preserve units verbatim (no °C↔K, no h↔min). Ranges → `min_value`/`max_value` with `value=null`; single values → `value` with `min/max=null`.
7. Conditions per operation type:

- `HEATING/DRYING`: `heating_temperature`, `heating_time`, `heating_atmosphere`
- `COOLING`: `heating_temperature` (end temperature, e.g., "cooled to 900 °C"), `heating_time` (duration if given), `heating_atmosphere`
- `SHAPING`: `shaping_pressure`
- `MIXING/QUENCHING`: all null (info in `type`/`token`)

8. Operations numbered from 1 in narrative order. Repeats (regrinding, refiring) each get their own number.
Types and typical tokens (paper verbatim):

- `MIXING`: "mixed", "ground", "ball-milled", "milled", "weighted", "loaded", "stirred"
- `HEATING`: "heated", "calcined", "sintered", "fired", "annealed", "reacted", "treated"
- `DRYING`: "dried", "baked", "evaporated"
- `SHAPING`: "pressed", "pelletized", "compacted", "uniaxially pressed", "cold isostatic pressed"
- `COOLING`: "cooled", "slow-cooled", "ramped down", "furnace-cooled" (slow/controlled cooling, including controlled ramp rates such as "cooled at 1.4 °C/h")
- `QUENCHING`: "quenched", "ice-water quenched", "air-quenched", "water-quenched" (rapid cooling)
- `NOT_OPERATION`: when the classifier identifies a token as a non-synthesis action

Distinguish `COOLING` (slow, controlled, with explicit rate or duration) from `QUENCHING` (rapid, often without rate). Cooling rate values (e.g., 1.4 °C/h) cannot be stored in the current schema; capture the end temperature in `heating_temperature` and the duration in `heating_time` when given. Multiple cooling segments (e.g., 1100→900 °C then 900→30 °C) get separate `COOLING` operations with their own order.

User:
`Paper:`

{paper_text}

ICSD entry:
{icsd_metadata}

Output format:
The response is constrained by a strict JSON schema with the following structure:

```
{
  "is_solid_state": boolean,
  "solid_state_option": string | null,
  "found_in_paper": boolean,
  "not_found_reason": string | null,
  "flux_compounds": [string, ...],

  "target": material | null,

  "precursors": [
    material, ...
  ],

  "operations": [
    {
      "type": "MIXING"
            | "HEATING"
            | "DRYING"
            | "SHAPING"
            | "QUENCHING"
            | "COOLING"
            | "NOT_OPERATION",
      "token": string,
      "order": integer,
      "conditions": {
        "heating_temperature": quantity | null,
        "heating_time": quantity | null,
        "heating_atmosphere": string | null,
        "shaping_pressure": quantity | null
      }
    },
    ...
  ]
}

material = {
  "material_string": string,
  "material_formula": string
}

quantity = {
  "value": number | null,
  "min_value": number | null,
  "max_value": number | null,
  "unit": string
}
```

API configuration:
model = "gpt-5.5"
response_model = "gpt-5.5-2026-04-23"
seed = 42
structured outputs = strict

Table 21: Prompt for the few-shot, no-fine-tuning baseline (base LLaMA-3.1-8B / Qwen2.5-7B, $N \in \{0, 20, 40\}$ shots). Demonstrations are ground-truth Tier-1 training pairs, sampled with a fixed seed; the system and query turns are identical to the single-step (PRP) prompt.

System Prompt: You are an expert in solid-state inorganic synthesis. Assume all solid-state inorganic synthesis happens pairwise between intermediates or precursors. Given a target material formula, predict exactly two precursor or intermediate materials that reacts pairwise via a solid-state reaction to synthesize the target. Output only the two material formulas separated by `' + '`. **Demonstrations** ($\times N$, $N \in \{20, 40\}$; omitted when $N = 0$):
User: Predict two solid-state precursors or intermediates for {`demo_target`$_i$}
Assistant: {`precursor1`$_i$} `+` {`precursor2`$_i$} **User:** Predict two solid-state precursors or intermediates for {`target_formula`} **Output format:**

```
precursor1 + precursor2
```